

Evaluation of Graduation 2010: Performance Indicators and Problems*

Mary H. O Phelan, Antony D. Norman, and Gayle W. Ecton

Western Kentucky University

Bowling Green, Kentucky

*Presented at the Eighth Annual Conference of the SQÉP, October 29, 1999, Quebec City, Quebec. This is a working document. Please do not cite without permission. For more information, please contact Mary H. O Phelan, Educational Leadership, Western Kentucky University, One Big Red Way, Bowling Green, KY 42101, or by telephone at 270-745-6330, or email at mary.ophelan@wku.edu

Evaluation of Graduation 2010: Performance Indicators and Problems

In this paper, we will present the background and description of the major components of Graduation 2010, most of the indicators of progress being used in the evaluation, and finally, at least some of the problems we are encountering using those indicators, and with the evaluation in general. Problems discussed will include those inherent in trying to reach any conclusions about a program that is changing continuously such that the program itself is not the same now as it was two years ago, or even a month ago. Discussion will also include the level of participation in data collection as well as in the implementation of the program, interaction with the media, and level of communication among people.

Other problems that seem less urgent, but nevertheless need to be addressed, are funding for the evaluation itself, and how to handle the massive amounts of data that will be forthcoming from so many indicators in so many schools at so many different implementation levels. Our team hopes that the discussion we begin in this paper will lead to productive ways to solve some of these problems.

Background and Essential Components

In early 1997, a steering committee of public educators and concerned citizens in Kentucky's Daviess County school system formed to examine new developments in how the brain learns and to determine what curricular changes could be made that would enhance the intellectual capacity of children in its schools. Naming its effort Graduation 2010 for the graduation year of the children who would enter kindergarten in the fall of 1997, the committee identified eight areas in which they thought educational changes would prove beneficial: music, the arts, foreign language, thinking skills, family involvement, health and emotional health,

reading and language development, and community involvement.

After a community forum where the committee's initial findings and ideas were presented, the committee established separate subcommittees composed of volunteers from the forum to further examine each area. The expressed task for each subcommittee was to gather available literature and to make recommendations about how best to apply the ideas gleaned. The steering committee asked the subcommittees to consider the following questions when developing each recommendation: Is it research based? Is it reasonable, practical, and feasible? What particular activities and experiences would occur at what age? Is this the most effective and efficient way to carry out this addition?

The eight committees submitted a total of 106 short-range and long-range recommendations for Graduation 2010. From these, the steering committee selected 16 goals for the 1997-98 school year, most of which centered around implementing the eight areas of innovation. For music, piano labs were purchased for all 12 elementary schools, and music teachers were trained to teach keyboarding. For the arts, each school identified an arts facilitator and coordinated nine arts experiences (three visual arts, three dance, and three theater) for each kindergartner. For foreign language, the district funded the development of a videotaped Spanish program and kindergarten teachers began teaching simple Spanish vocabulary. For thinking skills, teachers were trained in Talents Unlimited, a thinking skills program, and schools began to implement strategic games, such as chess, in the classroom. For family involvement, several strategies were initiated, including the development of new student orientations, homework committees composed of teachers and parents to establish reasonable and defensible guidelines for homework, and family involvement committees. For health and emotional health, each

school was to have access to a nurse at least one day a week and schools were to develop fitness programs for students. For reading and language development, the district implemented the Kentucky Reads grant targeting primary children who are at least one grade level behind in reading. For community involvement, a corporate sponsor, Owensboro Mercy Hospital, agreed to adopt the kindergartners who made up the 2010 graduation class.

A final goal of the steering committee was to obtain a research team that would study both the level of implementation of the above goals at each school and the overall effect of this program on student intellectual development and achievement.

Performance Indicators and Evaluation plan

The evaluation team became involved in Graduation 2010 when the Dean of the College of Education and Behavioral Sciences at Western Kentucky University asked us to consider working with the school district's steering committee, superintendent, and administrators, to develop a research evaluation plan. We met with steering committee members and administrators several times during the first year of the project. Most of those meetings were dedicated to clarifying the goals and procedures associated with the project.

During the course of these meetings, we became aware that we were going to face some difficulties because we had not been involved in the initial planning phases of the program, but had been called in after the fact, that is, after implementation had already begun. An example of that is that although our first tendency was to begin searching the literature for information on each of the eight strands adopted for implementation by the district, we soon realized that the evaluation would have to be based on what was actually used by the district. We asked the steering committee members to share the research articles and/or references they had used during

the planning phase. In reviewing those articles provided to us, we found that much of the information was anecdotal or opinion in nature, and few references were from scientific journals, although empirical evidence was cited in some of the literature. The result was that some strands had more solid research bases than others. In spite of the fact that many of the initiatives already taken had come from this somewhat limited literature base, only those references used by the district were included in the project description and background in the report on implementation progress at the end of the first year. To us, this was the first sign that our role was to be on the periphery, providing summative evidence for confirmation or rejection, rather than an integral part of the process as a source of information for the modification and/or improvement of the project.

There was much discussion about what the elements of the project included, what they were trying to accomplish, and what indicators of success could be used to evaluate the project. It took a full year of communication before we could come to agreement about what we would be looking for, what information we would need to gather, and how to go about gathering it. Originally, committee members wanted student scores as the bottom line, but gradually, we started hearing capacity to learn, which we interpreted to mean aptitude, as well as achievement. We started with the following indicators for achievement: Woodcock-Johnson, Talents Unlimited tests, SAT scores, Kentucky's State Assessment scores, CTBS/5 (TerraNova), ACT scores, Advanced Placement tests, and scores on the Primary Test of Cognitive Skills (P/TCS). The school district already uses most of these tests, and they agreed to provide the scores to us as they became available.

We wanted to do more than count the things that are easily counted (e.g., how many minutes per week of music instruction). We knew we needed some qualitative evaluation to find out how different stakeholders feel about the project, whether or not teachers, students and parents are buying into it, and even general atmosphere and level of cooperation and collegiality at individual schools. We agreed that data analyses would have to include qualitative descriptions of survey and interview responses, quantitative comparisons of test scores and frequency variables, and trend analyses for tested groups and individuals. Results would be reported at yearly intervals for the program as a whole, by school, and by program strand.

These discussions resulted in 10 research questions, each of which had several different indicators. A list of indicators by research question is presented in the second column of Table 1 below. We decided on a variety of information to help us assess the overall impact of the program. We asked for information such as number of students taking the SAT, enrolled in Advanced Placement (AP) courses and taking AP tests, or enrolled in other higher level courses. We asked for the number of students going to college or getting scholarships, dropping out of school, the number of pregnancies during high school, the number of students taking music lessons outside of school, choosing to study abroad, or getting involved in foreign language, art, or music clubs. We included surveys and interviews, as well as data sheets (containing some of the information listed above) to be completed by the schools, culture audits, and individual testing for a sample of students.

There are some limitations on the research design as well as the data gathering procedures worth noting here. First, the absence of any allocation of funds for the evaluation of the program will most likely result in less data gathering and less time commitment from us. Second,

Table 1. Research Plan for Graduation 2010

Research Question	Instrumentation	Time Frame
1. To what degree has each strand of the program be implemented?	Survey of principals School visitations # of teachers trained in Talents Unlimited School data sheets	Yearly
2. To what degree has the learning of children been increased?	P/T.S. Woodcock Johnson (individuals) Talents Unlimited Tests	Yearly
3. Has the program increased achievement?	KIRIS (CATS) CABS/5 (TerraNova) # of students in upper level courses Advanced Placement tests # of students taking SAT/ACT/SAGE SAT/ACT/SAGE tests Kentucky Reads results % of students reading at or above grade level School data sheets	Yearly
4. Has the school learning environment improved?	School culture audits Student/teacher/parent interviews/surveys Attitude surveys of music, arts, foreign language	Yearly
5. Has family involvement increased?	Parent/student/teacher surveys School data sheets # of PTO members # attending school meetings # of parent volunteers	Yearly
6. Has the health of students improved?	# of absences Demographics (pregnancies, etc.) Parent/teacher/student surveys	Yearly
7. To what extent are students well educated?	School data sheets # taking foreign language, AP classes, art, music, upper level courses # and % taking ACT/SAT ACT/SAT scores # and % of dropouts # and % in college prep track # and % going to college # receiving scholarships	Yearly
8. Are students benefitting from the arts?	CATS performance assessment School data sheets # taking private music lessons # enrolled in fine arts and music class # in arts clubs and activities	Yearly
9. Are students benefitting from foreign language?	School data sheets # enrolled in foreign language classes # participating in foreign language clubs and activities	Yearly
10. Has community involvement increased?	Teacher/student/administrator surveys Letters from sponsors Documentation of sponsor participation	Yearly

administrators do not want to overburden faculty and staff with data collection procedures by asking them to do more paperwork than they already have to do. Third, administrators felt it would be inappropriate for the research design to include a control group, either from within the school system or from a comparable district elsewhere. Finally, they did not want surveys to go out to the community at large, but only to staff and parents of the county school system, since there is an independent city public school system, separate from the county public schools where Graduation 2010 is being implemented, and a politically sensitive environment surrounding the two school systems.

Evaluation in Progress

During the first two years, the evaluation focus has been on implementation. Principals were given an open ended survey half way through the first year to report on progress in each of the eight strands of the program. At the end of the first year, principals were asked to rate to what extent each of the goals for that year had been accomplished in their schools. Eleven of the twelve elementary principals responded to the survey. A similar survey was given at the end of year two.

Toward the end of year two, we began collecting performance baseline data. We individually tested those students who were part of a sample of 150 kindergarten students, and for whom the schools had received parental permission for testing. We were able to test 115 out of the original 150 students selected.

At about the same time, we conducted school culture audits for all 12 elementary schools. Teachers at all elementary schools were asked to participate by rating the extent to which each of 13 characteristics was present in the school and then, the importance of that characteristic.

Schools were promised reports on the culture audits, and the evaluators visited each school and presented the results to the teachers and principals. At that time, teachers were interviewed about positive and negative changes occurring in the school, problems with teaching and learning, suggestions for improvement, and finally, their opinions about Graduation 2010.

To gather data from individual schools, we devised a data sheet to be completed by the schools during the year. Data sheets were adapted for, and distributed to all district schools, elementary, middle and high schools. So as not to overburden school personnel, the administration requested that we consolidate this data collection sheet to include all the information we could not otherwise obtain, and give it to the schools at the beginning of the year. They were to progressively complete the form during the year, as activities arose, and return the completed sheets to us at the end of the year. Members of the steering committee helped modify the data sheets to make them appropriate for elementary, middle and high school use.

Problems encountered

At the end of the second year of our participation, the following difficulties and problems have become evident: Each one will be addressed briefly.

Participation by stakeholders

We have not been altogether successful in collecting the data we set out to collect. Some schools returned their data sheets, and others didn't. We have been informed by district staff that since the program is voluntary, data collection is also voluntary. Even some schools judged to be high in the implementation of program initiatives have not been timely nor thorough with data collection.

Interviews with elementary teachers have indicated that teachers beyond the kindergarten and primary levels may not yet consider themselves integral parts of the project although grades four and five are already receiving music keyboarding, Spanish, and in some cases other strands. This is especially important for those strands requiring teacher training or professional development like Spanish or the chess curriculum, since teachers are asked to participate in training sessions in these areas offered by the district.

Subcommittees for project strands were originally formed with volunteers from the community. According to accounts of the original public forum held to get the program started, 166 out of 200 in attendance offered to serve on these committees. Once the project got started, however, the steering committee and some of its subcommittees decided to decrease their involvement in the ongoing project. At this writing, it is our understanding that some of these committees are still meeting and others are not.

Dealing with the media

The project has attracted national attention, including a segment on national television. More than one feature article has appeared in newspapers around the state and nation. One day a month has been set aside in the district for visitors from other places who want to come and observe the programs which make up Graduation 2010. Although most are enjoying so much attention, some statements in the media do not distinguish between the baseline data being collected and results of the program. The district's test score results are generally good, and it is impossible to control the impression given by reports about the project. The evaluators have been placed in a somewhat awkward position since media representatives have contacted us, asking about the validity of claims made in the media elsewhere.

Implementation

Implementation has been voluntary, and some schools have adopted program ideas with more enthusiasm than others. Likewise, some schools have implemented certain strands and not others, and what constitutes full implementation of a strand is not necessarily constant from school to school. For example, the chess curriculum, in some places, consists of having the chess board and pieces in the classroom, while in other schools, children are playing chess. These variations in the implementation have to be taken into account during the evaluation of the program, but they will make it difficult to attribute successes or failures to any particular set of procedures.

Dynamic aspect of the program

The project is continually changing and adapting to the demands of the environment. In the last year alone, the definition of the community involvement strand has been modified, steps have been taken to divide the health/emotional health strand into two separate strands, and initial plans are being made to add a technology strand. There have also been changes in leadership. As previously mentioned, some of the original committees elected not to continue meeting, so that some key people who helped create the program are no longer involved in the day to day operation of it. The expanding program has resulted in the hiring of a full time coordinator who, along with administrators, is in charge of many aspects of the program formerly performed by the original subcommittees. One result of this dynamic characteristic of the program is that there was no formal statement of goals for the second year or for the third year of the project as there was for the first year. With the passage of time, new goals are set and new policies are put in place which were not in the original plan. Some of the original ideas have been abandoned

because they were found to be unsuitable or impossible. Though adaptation is usually considered a positive thing, things are moving so fast that it is difficult to maintain anchors upon which to base the evaluation of results. Constant change affects our ability to collect long-term data and our ability to isolate what seems to be effective and what does not. Collecting so much information about so many different aspects of the program has become burdensome for all parties, and at a recent meeting, the mood was more about what information can we get rather than what information do we need.

Communication

One problem we keep running into, and one that overlaps with all other problems mentioned here, could be characterized as a communication problem. Communication with the evaluation team has been sporadic, and in some cases, we have not been informed about important changes occurring in the district. We found out about impending changes in the strands during a presentation on the program in which we were participating with a discussion of the evaluation of the program.

In the beginning, the eight strands had more or less equal importance, at least in our perception, but recently, there has been a change in language used to describe the strands, and they are now referred to as learning enhancing or barrier removing strands. For the evaluation team, this change in posture regarding the strands makes some strands more important than others, and shifts attention from problem areas (barriers) to the new ideas being implemented (enhancers). This creates problems for the evaluation since this shift in relative importance has not been taken into account in the research design. Administrators maintain that the idea of barriers and enhancers was part of original discussions in planning for the project, and

acknowledge that this may not have been communicated to the evaluation team.

Recent developments

After discussing some of these problems at a recent meeting with school administrators, it was agreed that the scope of the evaluation efforts will have to be limited since it will not be feasible to collect data on all indicators originally listed (see Table 1, page 7), and that the emphasis should shift from trying to evaluate all strands (both enhancers and barriers) to enhancer strands only. This shift will result in the study of those schools recognized as high implementers of the four innovative or enhancer strands: thinking skills, foreign language, music and the arts. While this will simplify the collection of data and evaluation process, it will limit the external validity of the findings since it will no longer be a county wide evaluation plan.

One of our concerns remains that we are not regarded as partners in the effort to make the project successful, but rather as having a more summative role of approval or disapproval. We need to find appropriate ways to communicate information, positive or negative, about the ongoing program, such that the information can contribute to constructive modifications and improvements of the program. There was no provision in the original plan, for example, for taking into consideration the results of any part of the evaluation and using that information to make modifications in the program for the future. We are discovering the advantage of educational reform initiatives that take into account research concerns before implementation.

Summary

We have described a Kentucky school district's reform project, some of the indicators of success for the project, and some of the problems we have encountered in using these indicators and in our evaluation efforts in general. We hope that this description of our experiences and

concerns will lead to productive discussion and contribute to solutions for problems that program evaluators face.