ISSUES AND STRATEGIES IN EVALUATING SYSTEMIC REFORM

PAPERS PREPARED FOR THE U.S. DEPARTMENT OF EDUCATION OFFICE OF THE UNDER SECRETARY PLANNING AND EVALUATION SERVICE

David Cohen, University of Michigan
Michael Fullan, University of Toronto
Harry Hatry, The Urban Institute
Larry Suter, National Science Foundation
Kenji Hakuta and Diane August, Stanford University
Vinetta Jones, The College Board
Richard Elmore, Harvard University
Glynn Ligon, Evaluation Publishing Software Inc.
Brenda Turnbull, Policy Studies Associates

SEPTEMBER 1994

Evaluating the Inclusion of L.E.P. Students in Systemic Reform[7]

Diane August[8] and Kenji Hakuta[9]
Stanford University

At the national, state, and local levels there is a new push to set high academic standards and create a related system of student assessments. On March 31, 1994, President Clinton signed into law the Goals 2000: Educate America Act, an Act that codifies in law the national education goals and provides resources to states and communities to develop and implement systemic education reforms aimed at helping all students reach challenging academic and occupational standards. Moreover, the Elementary and Secondary Education Act is currently being reauthorized and is expected to become law before the close of the year. Many provisions of the new bill are coordinated with Goals 2000; In fact, the administration views ESEA as providing "the resources" needed to implement Goals 2000, which they view as setting the framework for reform.

In addition to the new laws, there are many efforts under way to develop content and performance standards in different academic areas and to create assessments that are aligned with these standards. Content standards are being developed or have been developed by professional organizations of teachers and specialists in English, mathematics, science, history, geography, foreign languages, citizenship/civics, the arts and other subjects. The *New Standards Project* is developing and field-testing innovative assessments tied to some of the new content standards.

States have also been very involved in some aspects of systemic reform. At least 45 states have created or are preparing new curriculum frameworks, while at least 26 states and the District of Columbia will be dealing with educational standards in 1994.

New York City, under the guidance of Schools Chancellor, Ramon Cortines, has undertaken the development of a curriculum framework for all the city's public schools. According to the Chancellor, standards are needed to address vast differences in the material taught to certain grades in each of the city's schools and community school districts.

---

[8] Consultant, School of Education, Stanford University.

[9] Professor, School of Education, Stanford University.

Systemic Reform and L.E.P. Students

Systemic reform holds promise for improving instruction and learning for all students, including L.E.P. students. But such an outcome is not a foregone conclusion. Thus far the reform movement has generally sidestepped the particular conditions, needs, and strengths of L.E.P. children. Difficult questions remain to be answered, for example:

- Entrusting states with greater responsibility can encourage coherent leadership, but what will ensure that L.E.P. students are not left behind and that they receive appropriate services and instruction?

- The movement toward high standards for all students is welcome, but what will guarantee that L.E.P. students have full access to a challenging curriculum, rather than programs which focus single-mindedly on English acquisition?

- No matter how well L.E.P. students' needs are acknowledged on paper, how much can schools really do to address these challenges when there is a growing shortage of bilingual and ESL teachers, and when many L.E.P. students attend schools with severely limited resources?

- Systemic reform depends in large part on relaxing rules and regulations in exchange for holding schools accountable based on student outcomes. Without appropriate assessments for L.E.P. students, how can schools that enroll these students be held accountable?

Overall Approach to Evaluation

To fully assess how systemic reform efforts incorporate L.E.P. students, we conceptualize the evaluation of activities at three levels: national, state, and local. This partitioning of the evaluation territory is not totally arbitrary in that the legislation specifies different activities at each of these levels, and different activities require different approaches to evaluation. However, given that we are evaluating systemic reform, while the focus may be on one level at a time, the evaluation should consider how each level relates to the others.

This paper addresses the implementation of systemic reform in the context of Goals 2000. Clearly, it is important to recognize that the Improving America's Schools Act (IASA) will also contribute in a major way to systemic reform. For example, IASA will require some form of greater coordination of resources at the local level. However, since this bill has not become law, we will not fully discuss at this time how to evaluate its impact on systemic reform.

Arenas

At each level, we propose the examination of different activities defined to a large extent by the legislation. For example, at the national level, the major arenas for evaluation in *Goals 2000* include the composition of various panels, the criteria for certification of standards, various activities funded through grants, i.e., the development of OTL standards, and the evaluation of the technical quality of the work of the Goals Panel and NESIC, and student progress toward meeting the National Goals. Similarly, distinct arenas exist at the other levels. Each carries with it a set of benchmarks and measurement methodologies.

Benchmarks

By benchmarks, we mean standards, desirable activities or outcomes against which the existing conditions are assessed. For any given arena, the benchmarks may differ depending on the assumptions and values of those that establish them. This will make establishing benchmarks a formidable task. For example, many advocates for L.E.P. students believe that the NESIC panel should include individuals with expertise in the assessment of L.E.P. students. Others may argue, however, that general knowledge of assessment is sufficient. While some experts believe that it is important to have separate English-as- a-Second-Language (ESL) content standards for L.E.P. students, others feel that a single set of standards for English language arts would be sufficient.

Various groups are working to define benchmarks for L.E.P. students in the context of systemic reform. For example, there have been two meetings on L.E.P. students and systemic reform, sponsored by the Office of Bilingual Education and Minority Languages Affairs (OBEMLA) in March and June, 1994. Participants included practitioners and policymakers concerned with and knowledgeable about the education of L.E.P. students. The group has made specific recommendations on how to incorporate L.E.P. students into systemic reform, embodied by Goals 2000. These recommendations are in draft form, pending review by the group and by OBEMLA. In addition, OBEMLA has sponsored hearings around the country on this topic. Professional groups such as Teachers of English to Speakers of Other Languages (TESOL) are also working to develop benchmarks. For the most part, these benchmarks define adequate resources (access to technology) and appropriate processes or interventions (i.e. articulation and coordination between the bilingual education program and the "regular" school program) rather than student performance.

This is no coincidence. As Elmore points out, no state or federal policy actually specifies what a good distribution of student performance will look like in a system where "all students are expected to meet the same high standards". Incorporating L.E.P. students into any such system is a daunting task. For one, most L.E.P. students, by definition, are limited in what they can learn from a school program that is conducted exclusively or primarily in English. In addition, L.E.P. students are a very heterogeneous group. Some L.E.P. students arrive in U.S. schools with no prior schooling; others arrive with substantial skills and knowledge. Many questions remain to be answered: How much time should L.E.P.

students be given to reach grade appropriate performance standards in English? Should this depend on the resources available to them? their prior schooling? Should they be taught to these standards and assessed in their native languages until they are proficient in English? Even if we define acceptable performance in terms of improvement, it is still problematic given the tremendous variation in how long it takes L.E.P. students to become English proficient.

Moreover, to complicate things further, even for English-proficient students, few valid and reliable instruments exist for assessing student achievement aligned with new conceptions of knowledge and skills embodied by the content standards, although development efforts are underway. For L.E.P. students, the problem is even more difficult. Current assessment instruments in English are inappropriate because they actually assess both content concepts and language ability, particularly reading comprehension and writing. The interconnection of language and content makes it difficult to isolate one feature from the other. As a result, it is difficult to know whether a student is unable to demonstrate knowledge because of a language barrier or whether the student does not know the content material being tested. Often these assessments, then, simply become measures of L.E.P student language proficiency rather than measures of content knowledge, as they are intended to be. Valid methods for assessing L.E.P. students' knowledge of content matter in English are yet to be developed. Further, reliable tests in languages other than English that measure this knowledge and skills have been virtually non-existent.

Measurement

The measurement methodology will change for the different benchmarks. For example, assessing the adequacy of the composition of various panels will require different techniques than assessing how well services for L.E.P. students are coordinated at the local level. Assessing changes in student performance will require yet other methodologies.

Collecting the information necessary to determine whether and to what extent the benchmarks have been met will be extremely challenging. This will be especially true for the assessment of local reform efforts. For example, observing and analyzing the kind of instruction and learning envisioned by the reform movement will be very difficult. For L.E.P. students the task is even more challenging given the limited knowledge base, let alone assessment measures, in this area.

Systemic reform calls for planning, implementation of reform practices, and improvements in student performance. Although implementation may be preferable to intent (i.e., state action is better than state planning), intent is a necessary and important first step as it serves as a starting point for examining the extent to which goals and objectives have been accomplished. Even though the desirable ultimate outcome of reform might be improvements in student performance, evaluating implementation is important for a number of reasons. First, it will probably take considerable time before changes in systems and schools result in changes in student performance. Second, the validity of outcome indicators

is uncertain for L.E.P. students. Third, without a very clear understanding of what was implemented, we have no chance of understanding why certain outcomes occurred.

Given the complexity of systemic reform (i.e, it is defined and implemented differently in different locations; reform strategies have been implemented at different times) it is unlikely that we will be able to attribute given student performance outcomes to particular inputs and processes beyond the school level. This is not to say that we cannot study and report on the impact of reform strategies on educational practice (both in terms of what is offered and apprehended). Cohen identifies three interesting areas in which to assess stability and change -- policy environments, guidance for instruction, and instruction itself.

Putting it Back Together

The spirit of systemic reform calls for an integrated system, such that the optimum results in student performance are attained when the levels and arenas are coordinated and work together. As such, evaluating each of the arenas independently does not serve justice to the theory.

We recommend concentrating resources for evaluation at the local level. Because of the complexity of systemic reform, we feel the best place to understand it is as close as possible to where learning occurs. However, it will be very important to evaluate programs or projects in the context of schools and schools in relationship to other administrative levels and institutions. As Elmore states, "the challenge of systemic reform is for the whole policy and administrative apparatus to learn to think in a new way about school improvement -- a way that stresses systemic relations, rather than isolated solutions. Evaluations that select samples of schools or classrooms without regard for the environment in which they sit, or demonstrations that single out single schools for attention as pilot sites, or projects that provide assistance to certain schools in a system and not others -- all of these approaches are suspect, because they remove the school as a unit of analysis from its environment and treat it as an isolated case of some phenomenon."

Additional Caveats

For completeness, we must mention three other issues very briefly. First, because of the complexity of systemic reform as well as our limited understanding in some arenas of how to incorporate L.E.P. students (e.g., how to assess them for accountability purposes), we recommend that the benchmarks that are established as well as the methodologies used for assessment be continually subjected to empirical scrutiny and revision.

Second, we would like to suggest that whenever possible, the evaluation of how L.E.P. children are incorporated into systemic reform be part of an overall evaluation effort. That is, we recommend against doing separate studies on L.E.P. student inclusion in systemic reform because this would reinforce the existing tendency to treat L.E.P. student issues as a "specialization" that should be of concern only to the specialists and advocates for these

students.

Third, we have used the framework of Goals 2000 and the evaluation needs of the Department of Education to guide our suggestions. But other approaches are certainly possible. Reform can be defined -- and is indeed frequently envisioned -- *outside the* framework of the legislation (i.e, as networking, "reculturing" and restructuring). Evaluation needs in such cases would be construed differently. For example, school districts will use evaluations to improve teaching and learning.

Evaluation of National, State, and Local Activities

In this section, we describe in more detail how one might approach the evaluation of the inclusion of L.E.P. students in national, state, and local systemic reform activities. For each level, we describe the various arenas, benchmarks, and methodologies that comprise the evaluation. In reading the remainder of this paper, note that the arenas are identified as examples only -- they should not be considered in any way to be exhaustive. Full enumeration of the arenas, among other things, must await the completion of the reauthorization process for ESEA (IASA), particularly as they relate to Titles I, II, and VII.

Evaluation of National Activities

At the national level we would recommend reviewing, against benchmarks the following arenas:

Arena 1: Whether the concerns of L.E.P. students have been included in generating systemic reform policies and activities.

Possible Benchmark: Persons knowledgeable and concerned about the education of L.E.P. students are involved in the various panels.

Arena 2: Whether the NESIC criteria developed to certify standards and assessments satisfactorily address the needs of L.E.P. students.

Possible Benchmark: The criteria to certify voluntary national and state content and performance standards include specific information regarding how such standards apply to L.E.P. students. Content standards, for example, reflect the best available knowledge about how L.E.P. students learn and about how the content can be most effectively taught to them.[10]

----

[10] In reviewing early drafts of the national content standards, we found that despite explicit principles that the standards apply to all students -- i.e., that they should be reflective of a multicultural society, should build on students' first languages and home culture, and that all students should have the opportunity to learn -- there is very little specific information or guidance regarding how this will come to be.

Possible Benchmark: NESIC, in certifying exemplary national and state opportunity to learn standards, ensures that such standards explicitly address the needs of L.E.P. students.

Possible Benchmark: the criteria to certify state assessments explicitly considers the needs of L.E.P. students, such that States can demonstrate such assessments are valid and reliable measures of what L.E.P. students know and can do.

Arena 3: Whether the grants awarded for the development of national opportunity to learn standards and to defray the cost of developing, field testing, and evaluating systems of assessments take into consideration L.E.P. student issues.

Possible Benchmark: The activities funded by these grants address the needs of L.E.P. students and involve persons knowledgeable about the education of L.E.P. students. For example, in the development of assessments, innovative approaches to incorporating L.E.P. students into assessment systems are explored. Such approaches might entail altering the procedures used to administer the assessment (e.g., giving instructions in students' native languages, allowing students to respond in their native languages), modifying the assessment itself so it is more comprehensible to L.E.P students, and exploring computer-assisted assessments that are tailored to the language needs and content knowledge of L.E.P. students.

Arena 4: Whether the studies that look at technical quality of the work performed by the Goals Panel and NESIC and student performance toward meeting national goals adequately incorporate L.E.P. students.

Possible Benchmark: A grant will be made to the National Academy of Sciences or the National Academy of Education to evaluate the technical quality of the work of the Goals Panel and NESIC and the process for the development and use of criteria for certification of standards and assessment used by the Goals Panel and NESIC. This evaluation process includes an assessment of the extent to which the provision to include "all students" is operationalized and monitored by NESIC and the Goals Panel. Moreover, persons knowledgeable about L.E.P. students are involved in conducting this evaluation.

Possible Benchmark: The Goals Panel, in reporting on progress that the Nation and states are making toward achieving the national education goals and the progress states are making in implementing opportunity-to-learn standards and strategies, reports specifically on how these efforts impact L.E.P. students. Moreover, when reporting on promising actions being taken at the national, State, and local levels to achieve the national goals, the Goals Panel describes how these actions have affected L.E.P. students.

Methodology

At the national level, the methodology involves comparing activities in various arenas against certain benchmarks, developed by persons knowledgeable about the education of L.E.P. students. The evaluation is qualitative in that it entails choosing appropriate benchmarks and then assessing how well the activities in each arena have been implemented.

We do not propose a separate analysis on a national level of L.E.P. student performance because Goals 2000 already requires that the Goals Panel report on the progress that the Nation and states are making toward meeting the national education goals. We recommend that the data used to assess such progress be disaggregated for L.E.P. students.

It should be pointed out that it will be very difficult to compare student performance across states given differences in assessments, in definitions of data elements, and data collection methods. For example, districts have substantial leeway in both how they define limited English proficiency and whether L.E.P. students are included or excluded in large scale assessments.

Evaluation of State Activities

As at the national level, we recommend reviewing various arenas against benchmarks of effective practice.

> *Arena 1:* Whether the state plan addresses the unique needs of L.E.P. students in the state.

> *Possible Benchmark:* States in their plans provide assurance that they have statewide criteria for the identification and reclassification of students from other thanEnglish backgounds. The reclassification criteria is such that L.E.P. children once reclassified are function in all-English classrooms.

> *Possible Benchmark:* States describe in their plans how they will increase the pool of well-trained bilingual and ESL teachers, through such activities as creating a credentialling process for bilingual or ESL teachers, recruiting bilingual undergraduates and graduates into the teaching profession, enabling bilingual paraprofessionals to become certified teachers, and assisting LEAs in recruiting them.

> *Arena 2:* Whether the needs of L.E.P. students have been included in generating systemic reform policies and activities.

> *Possible Benchmark:* The panel established to develop the state plan fully involves persons knowledgeable about and involved in the education of L.E.P. students, including L.E.P. secondary students and parents of L.E.P. students.

Possible Benchmark: People with expertise and interest in the education of L.E.P. students are given the opportunity to participate in the process of developing a state plan.

Possible Benchmark: Information related to the State Plan and its implementation is made available in languages substantially represented in the state and when necessary, discussions are conducted in non-English languages in order to give as many parents of L.E.P. students and community members an opportunity to participate

Arena 3: Whether state standards are appropriate for L.E.P. students

Possible Benchmark: The state has developed common content and performance standards for L.E.P. students that are the same as those established for all other students, and any additional standards that the community that represents these students considers important.

Possible Benchmark: States establish a multi-faceted approach to enhancing opportunities to learn with provisions to ensure that the unique educational needs of L.E.P. students are met. This approach includes both the enforcement of a core set of standards as well as the use of "indirect" strategies to build the capacity of schools and school districts.

Core standards are focussed on assuring equal access to learning embodied in the new content and performance standards. These core standards are legally required and externally regulated by states and the federal government. Examples of core minimal standards that all schools should meet, include, for example, appropriately certified staff and student access to core coursework. State education agencies employ a wide variety of indirect strategies to improve schooling. And in these efforts, they mobilize and cooperate with other institutions to enhance their capacity. One strategy is providing incentives to school districts to go beyond the core standards (i.e., additional state funds for schools who run summer programs to help L.E.P. students meet performance standards). A second strategy is helping projects evaluate themselves against benchmarks of excellence, through program quality reviews. California, for example, has a Program Quality Review System that relies upon peer review. Benchmarks would include school-wide and classroom factors that are known to improve the overall education of all children, including L.E.P. students. A third strategy is working with colleges, universities and state licensing agencies to increase the number and quality of school personnel prepared to work with L.E.P. students. A fourth approach is working with the legislature and other stakeholders to decrease funding inequities among school districts. This would greatly benefit L.E.P. students, the majority of whom are concentrated in high-poverty districts.

Arena 4: Whether the state has addressed the needs of L.E.P. students in the design and implementation of any assessment systems that may be developed--both

assessments of student performance and opportunity to learn.

Possible Benchmark: In states with substantial numbers of L.E.P. students in given language groups, the state plan has included a process for developing or borrowing (from other states or entities such as large school districts with substantial L.E.P. students) content area assessments in the native languages represented by these groups. This process might also involve cooperative efforts among two or more states. The process includes persons knowledgeable about the assessment of L.E.P. students and systems serving them.

Arena 5: Whether the state has developed a system or systems of school and LEA accountability that fully incorporate L.E.P. students.

Possible Benchmark: States have determined what constitutes adequate progress and L.E.P. students are required to make the same progress as non-L.E.P. students. In making this determination, states have considered the results of the required assessments as well as other measures of school success, such as grade retention and dropout rates. In cases where L.E.P. students have failed to make adequate progress, the state has taken corrective action, including but not limited to ensuring the implementation of opportunity-to-learn standards.

Arena 6: Whether States, in using their allotment of Goals 2000 and ESEA funds for state activities to implement the State improvement plan, attend to the special needs of L.E.P students.

Possible Benchmark: There is financial equity in the distribution of funds for state activities that focus on the special needs of L.E.P. students.

Arena 7: Whether states, in making competitive, peer-reviewed grants to LEAs or consortia of LEAs, IHEs, private nonprofit organizations, or combinations of these entities for preservice teacher education and professional development, ensure that these funds are used to improve preservice and professional development activities such that instruction for L.E.P. students improves.

Possible Benchmark: States develop a multi-layered strategy for preservice and staff development that seeks to train (1) specialists with expertise in the delivery of high level content to L.E.P. students, either through bilingual instruction or high- quality sheltered instruction; and (2) all teachers who are likely to teach L.E.P. students to have a solid grounding in effective instructional strategies.

Arena 8: Whether states that have grants to develop systemic statewide plans to increase the use of state-of-the-art technologies address the needs of L.E.P. students.

Possible Benchmark: L.E.P. students have equitable access to high quality

August & Hakuta - 10

technology and special efforts are made to use technology to address the special educational and assessment needs of L.E.P. students.

Methodology

At the state level, we recommend three strategies for assessing how well states have implemented systemic reform. They include a fifty state survey, case studies, and the analysis of student performance data.

Survey. A fifty state survey would collect baseline information on states' reform activities at the initial stages of implementation of Goals 2000 and the IAEA. This information could then be compared with activities that are in place at various time intervals after the legislation has taken effect. We note that the Council of Chief State School Officers (CCSSO) is currently collecting baseline information on how states are incorporating L.E.P. students into their reform efforts. To collect this information, the CCSSO developed a supplement to their systemic reform survey. The survey on L.E.P. students has been piloted in five states, revised on the basis this pilot test and will be sent out to all fifty states in early July. Because the data will be entered into a computerized data base, it can be continually updated. The Department of Education is partially supporting these efforts.

Case Studies. To collect more detailed information on how systemic reform is implemented at the state level, we recommend a case study approach that would examine in more detail the arenas described above. It would be best to describe the benchmarks as precisely as possible and specify how data should be collected and analyzed to determine whether and how they have been met. For example, study questions could be developed to correspond to the various arenas. Respondents could be identified for each study question and interview protocols could be designed to take into consideration the role of the respondent and the information requirements of the question. In addition, documents germane to the arenas under investigation could be reviewed.

States should be selected according to the purpose of the study. To study obstacles and successes in implementation, we recommend that six to ten states be selected to capture variation in the percentage of L.E.P. students in the population, in SEA activity levels related to the implementation of systemic reform, and in geographical region. To study "model" state activity on behalf of L.E.P. students, those states that are doing the best job should be studied, regardless of geographic region or percentage of L.E.P. students in the state. Study questions should focus on fleshing out model practices and how the state was able to put them in place.

Analysis of Performance Outcomes for L.E.P. Students.

There are potentially two sources of data on L.E.P. student achievement state-wide. One source of data are state assessments, linked to content standards, that exist or will be developed by states participating in systemic reform. Outcome data For L.E.P. students who

take these assessments can be collected, analyzed, and reported out. However, as previously mentioned, there are problems collecting performance data on L.E.P. students because of the limited number of assessments in L.E.P. students' native languages and L.E.P. students' inability to take assessments in English. There have been some efforts to make adjustments in the procedures used to administer state-wide assessments and in the assessments themselves to enable more L.E.P. students to take them. But these efforts are limited and in the development stage only.

Another source of information on L.E.P. student performance is the National Assessment of Educational Progress (NAEP) Trial State Assessment. However, given the limited participation of L.E.P. students in the past, there needs to be systematic experimentation to look at a range of inclusion strategies for making the assessments accessible for L.E.P. students. Inclusion strategies might entail: altering the procedures used to administer the assessments (e.g., giving instructions in students' native languages, allowing students to respond in their native languages, using think-aloud techniques; modifying the assessment itself so it is more comprehensible to L.E.P. students (i.e., decreasing its English language demands); and employing computer-assisted assessments that are tailored to the language needs and content knowledge of L.E.P. students.

In examining student performance, one recommendation is to determine percentage changes in outcomes from one year to the next rather than relying solely on absolute levels of performance. This makes sense given the great diversity in assessment procedures used for L.E.P. students. Another recommendation is to look beyond test scores at other indicators such as drop-out rates, expulsions and suspensions, and college-completion rates. Given the difficulty of collecting adequate data on L.E.P. student performance, the use of other indicators is highly recommended.

Evaluation of Local Activities

As with the other levels, at the local level we also recommend evaluating various arenas against a set of benchmarks.

> *Arena 1:* Whether persons knowledgeable about L.E.P. students have been involved in generating systemic reform policies and activities
>
> *Possible Benchmark:* In their applications to SEAs for funds, in LEAs where there are L.E.P. students, school staff and community members that represent L.E.P. students are participating in any discussions about additional local standards for curriculum and instruction.
>
> *Arena 2:* Whether the educational needs and contributions of L.E.P. students are considered in the LEA plans.
>
> *Possible Benchmark:* In school districts with L.E.P. student enrollment, the LEA

plans address: the recruitment and training of teachers and aides to provide effective instruction to L.E.P. students; the acquisition and use of instructional materials -- in all languages substantially present in the school district -- equivalent to those provided in the English language curriculum; the most effective means for providing L.E.P. students with access to high quality curricula, instructional materials, extra- curricular support systems and technology; and the development and use of assessment instruments appropriate to measure the academic progress of L.E.P. students.

Arena 3: Whether schools with limited English proficient students receive LEA funds to support school improvement initiatives toward providing all students in the school the opportunity to meet high academic standards.

Possible Benchmark: LEAs ensure that all schools in the district are aware of their right to apply for funds to support school improvement initiatives.

Arena 4: Whether there is a coherent program in which federal, state, and local funds are coordinated to enable L.E.P. students to meet high standards.

Possible Benchmark: There is articulation and coordination between the bilingual program and the "regular" school program so instruction is complementary.

Arena 5: Whether L.E.P. students have access to curriculum aligned with rigorous professional and community content standards.

Possible Benchmark: L.E.P. high school students have access to the coursework that will enable them to graduate from high school with skills needed for gainful employment or college.

Other Arenas: Other authors have suggested additional arenas that capture aspects of "systemic change". For Elmore they include the same standards of academic performance for all students, school- level accountability connected to performance standards, and continuous improvement. For Cohen they include the extent to which policy and governance support intellectually more ambitious instruction and become more coherent as well as the extent to which teaching, educational materials, and assessments are more intellectually ambitious and coherent.

Possible Benchmarks: In the above arenas, L.E.P. students and school personnel who work with L.E.P. students are explicitly considered and included. For example, in examining whether teaching is more intellectually ambitious, it is more intellectually ambitious for all students, including L.E.P. students. In assessing continuous improvement, teachers who work with L.E.P. students have access to learning and support required to develop new skills and competencies.

Methodology

At the local level, we recommend case studies of schools to assess the implementation of systemic reform. Besides investigating what has happened in various arenas, student performance data from large-scale assessments could be analyzed if available for all students. In addition, information from school assessments, including alternative assessments, could also be used to determine how well L.E.P. students are succeeding in school.

Two approaches to selecting schools come to mind. To get an overall picture of how systemic reform plays out at the local level, sites might be selected purposefully to reflect the diversity in the proportion of students who are L.E.P. in the district and schools, the proportion of students who are from the same language background, the size of the district, the kind of programs that exist (targeted assistance versus school wide, for example), and the poverty level.

Another strategy would be to find schools that are successful at educating L.E.P. students and examine why this is the case -- are there interventions associated with systemic reform that appear to be helping the schools serve these children well. What are they? What did it take to implement them? As others have pointed out, it would also be worth figuring out how to "go to scale" with successful strategies. This would involve more than replicating effective practices, but finding mechanisms to ensure this happens "systemwide".