



IMPROVING PERFORMANCE: A FIVE-STEP PROCESS

Version 2
August 2002

This guide is a product of the United States Department of Education, Office of Vocational and Adult Education. It may be used only for educational purposes. Commercial uses of this guidebook are prohibited.

TABLE OF CONTENTS

Introduction.....	1
Step 1: Document Performance Results	4
Step 2: Identify Root Causes	14
Step 3: Select Best Solutions	20
Step 4: Pilot Test and Evaluate Solutions	24
Step 5: Implement Solutions.....	31
Appendix A: Benchmarking	33
Appendix B: Calculating and Displaying Data.....	34
Appendix C: Data Quality Criteria	41
References.....	42

INTRODUCTION

Improving Performance: A Five-Step Process is a guide designed to help state education agencies, schools and colleges use data to improve performance on the Perkins III core indicators. It describes a generic five-step process that state education agencies, schools and colleges can use in its existing form or modify for incorporation into their existing improvement processes. This guide is one of the resources developed by the United States Department of Education, Office of Vocational and Adult Education (OVAE), for state education agencies, schools and colleges as part of Perkins III accountability efforts, in particular, the Program Quality Initiative (PQI).

Perkins III Accountability: Guiding Principles

The Perkins III accountability effort is based on two guiding principles of the U.S. Department of Education: accountability and doing what works. The Accountability principle means state education agencies, schools and colleges will be held accountable for results. Institutions that improve performance will be rewarded. Performance information will be regularly reported and widely disseminated. The other principle, Focus on What Works, means federal dollars will be spent on effective, evidence based programs and practices with proven results. Funds will be targeted to improve schools and colleges and to enhance teacher quality.

The principle of accountability promotes the development and reporting of valid and reliable performance information. The principle of doing what works, or evidence based education, promotes the use of empirically based methods to identify and implement educational solutions with proven results. It encourages state education agencies and schools and colleges to use scientific research and systematic evaluations to develop and test improvement strategies that can drive continuous improvement in school and college performance. It also encourages the educational community to use systematic methods to analyze institutional and program data to improve performance.

Data Quality Initiative (DQI)

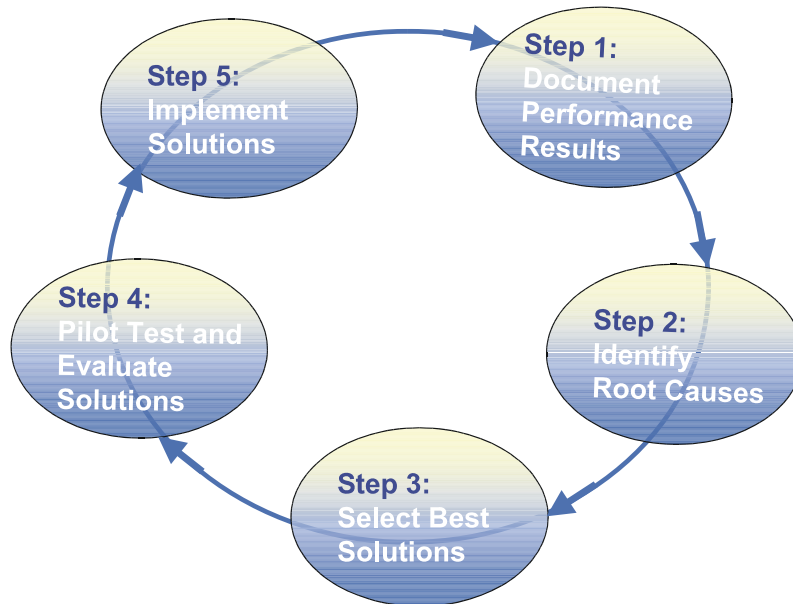
Since 1999, the Perkins III accountability effort has focused on improving state accountability systems. OVAE worked with states to develop the Perkins III core indicator framework that established performance measures, measurement approaches, and reporting requirements for state accountability systems. Over the past two years, OVAE has been collaborating with states on the Data Quality Initiative (DQI), an effort to focus on improving the quality of data collected and reported by states on the Perkins III core indicators. A result of this initiative has been the development of data quality criteria and scoring rubrics to assist states in evaluating and improving the quality of their own data.

Program Quality Initiative (PQI)

The current focus of improvement efforts is the Program Quality Initiative, designed to promote doing what works at the state and local levels. It builds on the Data Quality Initiative by shifting from collecting valid and reliable data to using this data to guide performance improvement. This initiative promotes the use of scientific methods and centers on how state education agencies, schools and colleges can use research and evaluation findings to develop performance improvement strategies. It also encourages the educational community to use its performance data and use scientifically based methods to determine the causes of performance gaps, and identify and evaluate strategies to improve performance on the Perkins III Core Indicators.

The 5-Step Improvement Process

The PQI uses a generic five-step improvement process that is based on practical yet rigorous methods and tools to guide state and local improvement efforts.



Step 1: Document Performance Results. The first step in the process is to describe state and school/college performance on the core indicators by comparing performance levels between schools/colleges, student populations, and programs over time. This step uses summary statistics and basic graphs and charts to document performance and identify improvement priorities.

Step 2: Identify Root Causes. The second step is to analyze performance data and use additional information and methods to determine the most important and most direct causes of performance gaps that can be addressed by improvement strategies and specific solutions. This step encourages states to use multiple methods to identify and evaluate potential causes and select a few critical root causes as the focus of improvement efforts.

Step 3: Select Best Solutions. The third step is to identify and evaluate potential solutions to performance problems, including both improvement strategies and program models, by reviewing and evaluating the underlying logic of these solutions and the empirical evidence of their effectiveness in achieving performance results.

Step 4: Pilot Test and Evaluate Solutions. The fourth step is to conduct pilot testing and evaluation of solutions. This step presents practical yet rigorous methods and tools for evaluating solutions before full implementation at the state or institutional levels.

Step 5: Implement Solutions. The fifth step is to implement fully tested solutions based on implementation plans that measure the implementation of the solution and evaluate the success of the solution in reaching the expected performance results. This step also addresses how to use evaluation results to plan the next steps in state and local improvement efforts.

Guide Overview

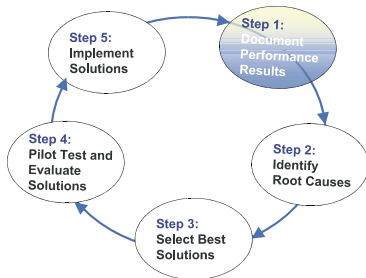
This guide is organized around these five steps of the improvement process with a separate section for each step of the process. Each section follows the same format:

- ❑ Why Important? Each section begins with a discussion of why the step is important and sets the stage for the next step.
- ❑ Methods, Tools, and Procedures. Each section describes the methods, tools and procedures that state education agencies, schools and colleges can use to complete the step. Each section also provides examples.
- ❑ Reminder Checklist. Each section concludes with a short checklist to make sure that the user has fully understood the requirements of the step and what has to be completed to move to the next step.

This guide also contains appendix materials and references to assist state education agencies, schools and colleges in learning more about specific methods, tools, and procedures and explore additional resources on continuous improvement.

- ❑ Appendix A: Benchmarking
- ❑ Appendix B: Calculating and Displaying Data
- ❑ Appendix C: Data Quality Criteria
- ❑ References

STEP 1: DOCUMENT PERFORMANCE RESULTS



To begin improving educational programs and increasing student performance, first assess current and past student performance by school and college, by program area, and by population groups. This section will assist you in documenting and understanding performance results, and help to identify areas for improvement.

Why Document Performance?

The performance accountability requirements of Perkins III are motivating states to collect and report more data than ever before. While the mass of numbers and data elements collected in your state may appear overwhelming, state accountability data contains substantial information that, if analyzed and managed, can help educators and stakeholders to understand what drives students' success. This data offers an important opportunity to gain insight into the quality of education provided by the public system. Further, understanding what patterns are present in student performance helps focus on appropriate improvement priorities.

Understand the problem completely before you seek solutions

Documenting and analyzing performance from a number of perspectives is critical to gaining a fuller understanding of where performance gaps lie, under what conditions exceptional performance is attained, and where improvements need to occur. This first step in the program improvement process offers:

- ❑ Recommendations for types of documentation that will aid in understanding state performance results;
- ❑ Suggestions of descriptive statistics and data displays to help describe student performance;
- ❑ Tips on how to evaluate the effect of data quality limitations in interpreting performance data; and
- ❑ Criteria for using performance documentation to establish program improvement priorities.

Please note that this section provides a minimum expectation for documenting performance results for state and local entities. States are encouraged to perform more sophisticated analyses of their data using multivariate statistics and more complex displays, if they possess the capability.

What to Document

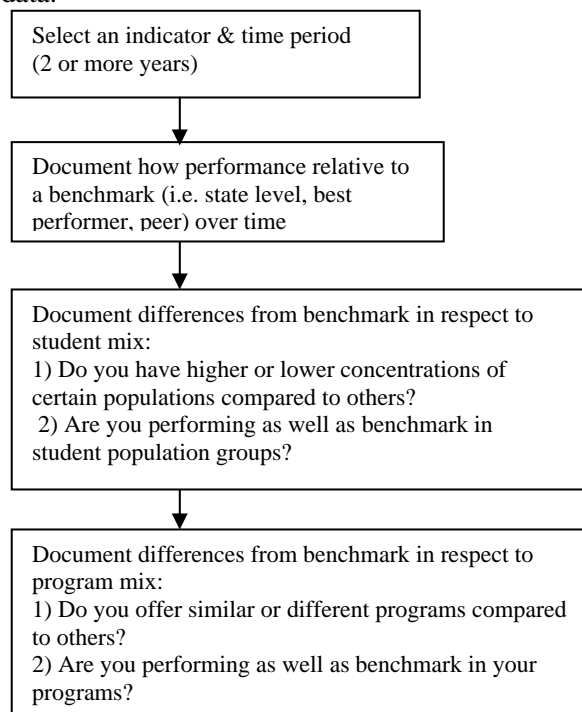
States are required to report student outcome data to the Consolidated Annual Report and for the Report to Congress, but to derive the greatest benefit from the data you have collected you need to go further than simply reporting your performance. Once you have documented student outcomes for the core and sub-indicators, assess outcomes for groups within each indicator and disaggregate the data a number of ways. Make comparisons within groups—comparing subgroups will allow you to further examine differences within the group and to hypothesize reasons for, and solutions to, those differences. Make use of benchmarks—they provide a point of reference to assess performance relative to similar groups. Strive to pinpoint trends by documenting how groups and subgroups are performing over time. The figure below presents three groups by which performance should be documented for each indicator along with suggested comparisons over time.

Appendix A provides further information on benchmarking.

Recommended Areas of Performance Documentation

Unit of Analysis	Comparisons	Trends
School/College Districts	<ul style="list-style-type: none"> State performance levels In-State Comparisons <ul style="list-style-type: none"> - Best performer in state - Selected Peer Benchmark External state benchmarks 	<ul style="list-style-type: none"> At least 2 years Preferred—3 to 5 years
Demographic Groups and Special Populations		
Programs—General Categories/Program Areas		

Comparisons such as the ones suggested can uncover performance differences between high and low performers and reveal whether these gaps become smaller or larger over time. Cutting the data a number of different ways will assist in identifying areas to target improvement efforts. There is not necessarily a linear or standard approach for describing data, but the process depicted below is one way to unpack your data.



Example

You decide to document postsecondary placement for the last 3 years for ABC College compared to the top performing institution in your state. Over this time period, you find that the top institution places 7-10 percentage points more students than does ABC. The institutions contain similar proportions of demographic and special population groups, however African Americans and Hispanics are placed at much lower levels at ABC than those groups in the top institution. In comparing programs, you find that the top performing college offers programs in IT and Financial Services, while ABC lacks these programs. Across similar programs, ABC places a somewhat lower percentage of its students than does the benchmark institution but performs worse in the Manufacturing program.

There are a couple of caveats to consider when conducting your analysis: 1) While analyzing gaps is ideal for identifying underperformers in the state, keep in mind that improvement efforts should not focus exclusively on lower performers; there is room for improvement among all educational institutions. 2) Comparing low and high performers may not be as useful in identifying problems if there exists small variability in performance within the state. States may want to use other states' performance outcomes as benchmarks for comparison purposes, if they use similar measurement approaches and have similar demographic characteristics.

How to Document: Methods and Tools

A useful way to document performance is to provide outcomes and comparisons in tabular format. Tables are a nice way to capture a lot of information, however they may not always be the most effective way to portray the meaning of your data. Graphical displays, accompanied by summary statistics, may provide a more expressive representation of patterns in your performance data and convey a clearer picture of performance outcomes and gaps. Further, graphics and summary statistics are good supplements to tables. Graphical displays include histograms, pie charts, line graphs, and bar charts. Summary statistics include the average, median, range, percentile ranking, and standard deviation. The figures that follow provide a quick reference on some summary statistics and examples of several graphical tools.

Summary Statistics

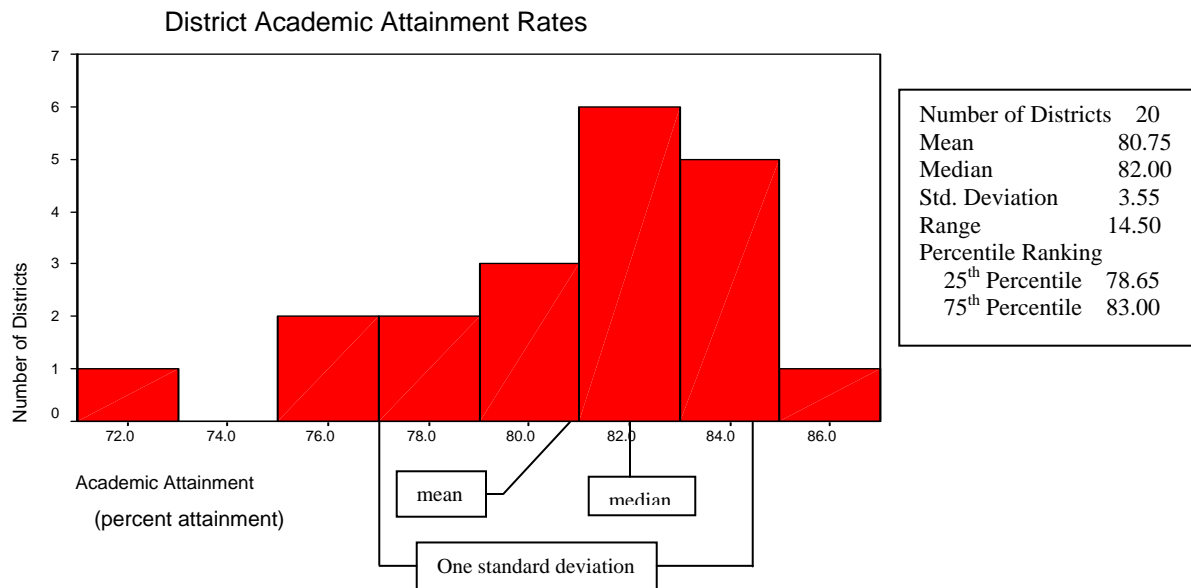
Refer to Appendix B for the calculation and applications of these statistics

Summary Statistic	Description
Mean or Average	The mean or average is the typical or representative value in a set of data. It is best used for data sets with few outliers.
Median	Represents the midpoint, or middle value in a distribution of numbers. It is useful if your data contains outliers that distort the average.
Standard Deviation	The standard deviation provides a measure of how widely the actual values are set apart from the average. It is useful for identifying the dispersion of your data. In a normal distribution, 68% of the cases fall within one standard deviation from the mean and 95% fall within two standard deviations.
Range	The range provides a measure of variability in a data set. It represents the difference between the highest and lowest numbers in a set of data and its value is entirely determined by the two extreme values.
Percentile Ranking	Provides an idea of position in a set of numbers. It is useful for indicating what percent of values fall below a particular value under consideration.

Graphical Displays

The examples below will demonstrate how you can uncover gaps in performance data using a variety of graphical techniques. The examples document performance for secondary academic attainment for a hypothetical state composed of 20 school districts.

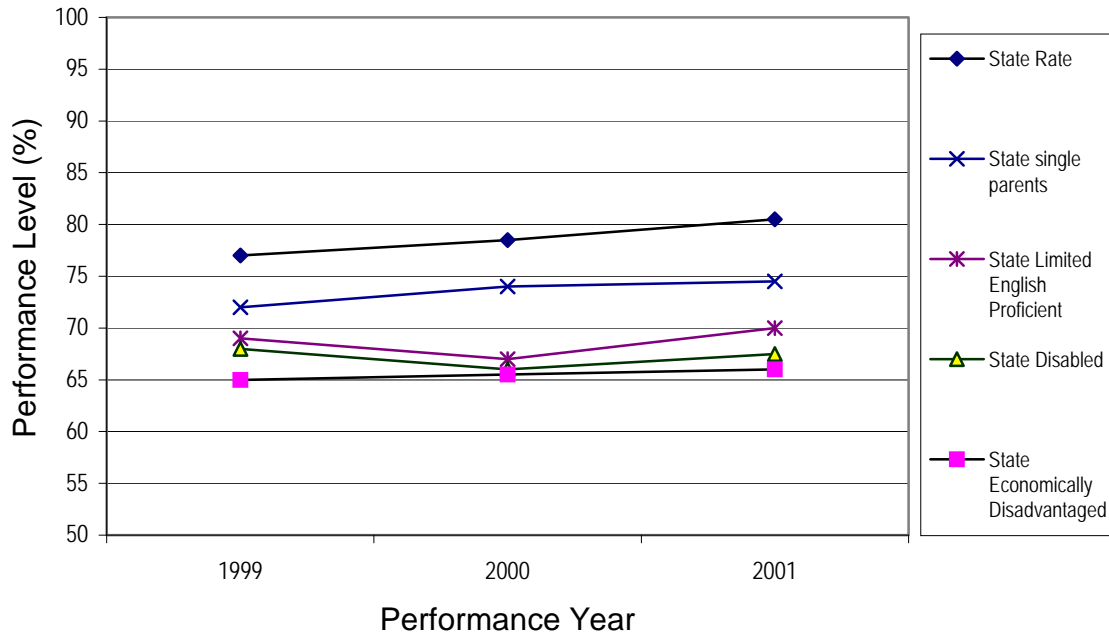
A *histogram* provides a visual indication of the data's distribution and is an easy reference for identifying summary statistics.



Interpretation: When examined along with summary statistics, the histogram allows you to see that a few school districts in the state are well below the average district's attainment rate, and that more than half the districts perform above the average. Is there something different about the higher and lower performing districts? What do the high-performing districts have in common?

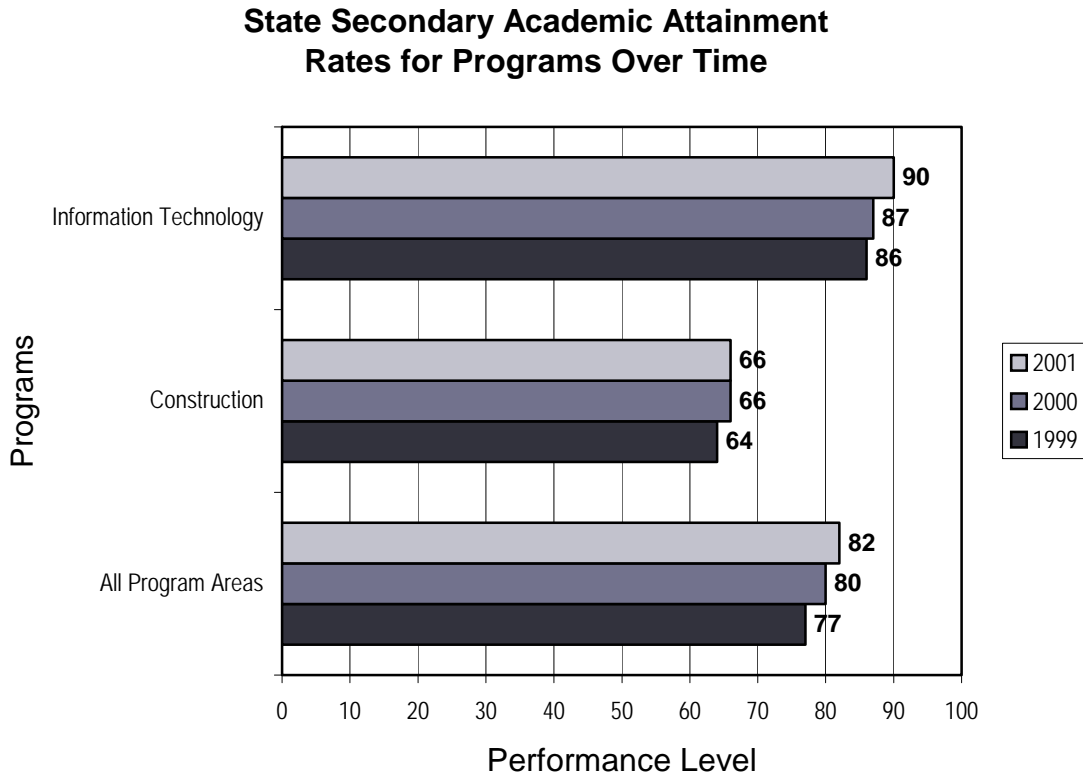
A *line graph* depicts comparative trends in performance among various sets of groups and benchmarks.

State Secondary Academic Attainment Rates for Special Populations



Interpretation: The line graph shows that all special population groups in the state perform below the overall state performance level over time. It also shows that the gap in performance between the state level and the economically disadvantaged and disabled groups is largest and seems to have grown from 1999 to 2001. Have there been noticeable trends within the state's special populations? This suggests that the state may want to focus improvement efforts on these populations.

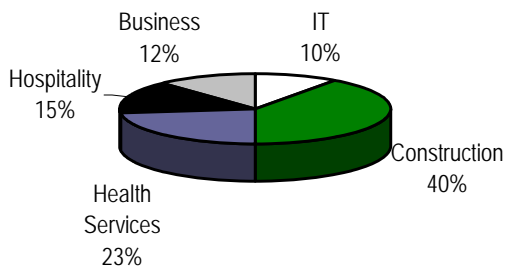
A *bar chart* enables comparison of sizes, quantities, amounts, proportions, etc., among groups of related items.



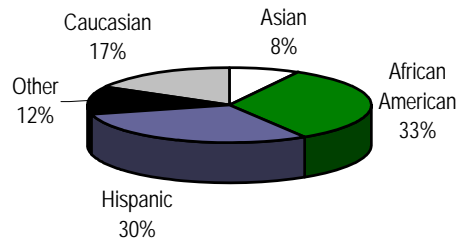
Interpretation: The bar chart depicted above compares the average of all program areas to the state’s lowest (Construction) and highest (IT) performing programs from 1999-2001. Over time, the attainment rate in the construction program has stagnated, while the overall state rate and the IT program rate have risen at similar incremental levels. Are there characteristics of the construction program that make it more challenging for students? This suggests that the state may want to focus improvement efforts on this cluster.

A *pie chart* illustrates the relative sizes of components that make up a whole.

Percent of Economically Disadvantaged Enrollments by Program



Percent of Construction Cluster Enrollments by Race/Ethnicity



Interpretation: The first pie chart shows that the economically disadvantaged students in the state are disproportionately enrolled in construction programs. The second pie chart indicates that Hispanics and African Americans account for a higher percentage of the state's construction cluster enrollments. What are the characteristics of the construction program that might explain the disproportionate enrollment?

How do I know I can trust my data?

All data are limited in some respects; the extent of the limitations determines its usability. It is important that you consider the quality of your data when performing your analyses to ensure that what you say is happening is a true reflection of student performance and not simply a problem of limited data quality. You should assess whether missing data or data quality problems create substantial bias in your numbers and consider if these problems are extensive enough to hinder useful performance analysis. For example, you need to identify whether the data problems exist for all groups, whether they are concentrated among particular schools, populations or programs, and whether you think the existing data is a reasonable representation of data that you may not have.

The Office of Vocational and Adult Education has established an objective set of criteria to determine data quality. These criteria include alignment, scope, reliability, timing, and coverage (refer to Appendix C for further information). States should do two things when looking at data quality:

- 1) identify major limitations, and
- 2) assess the implication for interpreting results.

To illustrate the effect of data quality, let's look at an example of inadequate coverage. Suppose that you mail surveys to student completers to measure placement but have accurate contact information for just two-thirds of them. Of those that received the survey, half of the students do not return it. Calculating a placement rate based only on the returned surveys will likely misrepresent the actual placement of students. In this case, you might be systematically ignoring the outcomes of specific groups who may have cultural, language, or other barriers that impede completion of the survey. Or perhaps the inaccurate contact information is concentrated among students who moved to college. Regardless, you need to assess the extent of the data quality problem to determine whether the numbers being reported are accurate or misleading.

Despite the previous discussion, please remember that data quality limitations should not keep you from using the data. Simply recognize that there may be flaws in the documented data and be cognizant of how you characterize and present the data. If you find gaps in performance, decide if data quality problems are responsible for the gaps or whether they are a true reflection of performance. In some cases, you may decide not to use the data, while in others you may find the data useful.

The data quality improvement process should occur alongside program improvement efforts, and each of these processes should influence the other. The more you analyze and use your data, the more you'll be able to recognize high quality data. Also, working with the data will assist you in discovering areas on which to focus data quality improvement efforts. And, the better the data quality is, the stronger the argument to use data for decision-making. Refer to OVAE's Peer Evaluation Resource Guide available at www.edcountability.net for a reference on data quality criteria and tools for assessing the quality of performance data.

Criteria for Establishing Improvement Priorities

After documenting and performing initial analyses of performance data, list where gaps exist for each core indicator. Quantify the nature of the gaps. Identify improvement priorities using the following criteria:

Criteria	Tips
Size of Gaps	Identify where performance differences are the largest for each core indicator. Is there an indicator where performance gaps are larger than others? For example, are attainment rates high but completion rates low?
Trends in Performance Gaps	Over time, are gaps getting larger or smaller? Are changes over time explained by unusual events or do gaps reflect an ongoing performance result?
Concentration of Gaps	Identify whether a particular program, population, or school/college disproportionately impacts state performance for each core indicator, either positively or negatively. Are there gaps that cut across measures? For example, does a particular population lag on all indicators or is a particular program in your state responsible for the gap across measures? Can you identify performance pertaining to certain core indicators that is unique to particular areas or populations? For example, are secondary completion rates high for a particular school district while that district lags on the other indicators?
State Improvement Priorities	Determine whether performance areas that you have been targeting for improvement in the past continue to be problematic or have shown improvement. Also take into consideration what key stakeholders in the state have identified as priorities.

Is there a gap on your list that, if improved, is likely to raise student performance more than others? Rank your performance gaps from likelihood to most raise performance to least likely to raise performance.

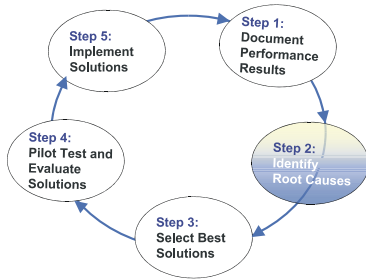
Moving to the Next Step

Before you move to the next step, make sure you have laid the foundation necessary for a successful transition to identifying root causes. Have you done the following:

- Documented your performance for each indicator?
- Determined reasonable performance benchmarks for schools/colleges, population groups, and program areas?
- Used graphical tools and statistics to document at least two years of performance for schools/college districts and compared data among other districts and to benchmarks?
- Used graphical tools and statistics to document at least two years of performance for each demographic and special population group and compared data among other groups and to benchmarks?
- Documented your student population mix and compared to benchmarks?
- Documented at least two years of performance for every CTE program in the state and compared that performance to other programs and benchmarks?
- Documented your program mix and compared to benchmarks?
- Assessed the quality of your data and factored in its likely effect on the interpretation of your performance results?
- Determined your performance improvement priorities?

If you've accomplished all the above, you have successfully identified performance areas that require program improvements. You are ready to move on to Step 2 of the improvement process: Identify Root Causes.

STEP 2: IDENTIFY ROOT CAUSES



Once you have described performance results and identified improvement priorities, the next step is to identify the most critical direct causes of performance problems—what are called root causes. Your analysis in Step 1 uncovered which students are attaining the desired outcomes and which students are not, but it did not tell you why. Step 2 is designed to address the why questions—why do these performance problems and student differences exist? What

are the major root causes that determine performance and explain student differences in performance? Which of these causes should we address first in our improvement efforts?

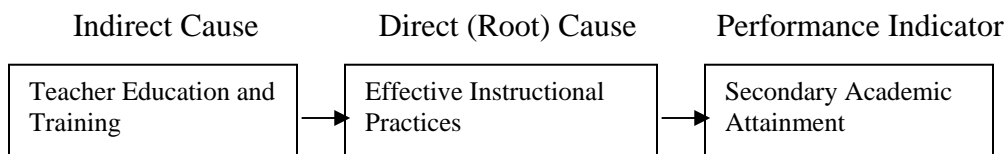
The search for root causes should be done through a systematic process that first identifies all potential causes, both within your control and outside your control, then determines whether causes within your control are direct root causes or indirect causes, and finally selects the most critical root causes to target for improvement efforts. This section will assist you in identifying and evaluating the root causes of performance to help guide your search for solutions.

Why Search for Root Causes?

Most of us want the quick fix. Responding to day-to-day problems consumes time, and often we don't feel we have the time to examine systematically what is really going on—what works, what doesn't work, and why not. Many times, we settle for conventional wisdom or accept convenient answers to performance problems rather than taking the time to question whether we really understand what is happening. We jump for the “silver bullet” answer heard at a conference or meeting, without trying to understand whether it really addresses problems at our own schools and colleges or can explain our current performance gaps described in Step 1. We want to adopt a new idea or innovative practice without asking whether it will really address the true underlying causes of the problem and have an impact on performance. Quite often, this means that we invest a lot of time and effort in improvement activities that do not achieve the expected results and, worst of all, we don't know why.

Don't Settle for Conventional Wisdom and Symptoms—Never Stop Asking Why

Program improvement is, in part, a search for answers to a very basic question: what causes poor performance? Root causes are those conditions or factors that directly cause or permit a performance gap to occur. They are direct, not indirect, causes. For example, effective instructional practices are a direct cause of student academic achievement because they have a direct impact on academic achievement. In contrast, teacher training is an “indirect” cause because it has an effect on student achievement only to the extent that the training results in improved instructional practices in the classroom, which, in turn, affect academic achievement.



Indirect causes can be either within the control of schools and colleges or outside the control of schools and colleges. For example, teacher education and training is an indirect cause of academic attainment, but schools and colleges can take actions to increase teacher education and training. However, another indirect cause may be overall school resources. Although federal and state governments and local taxing districts can increase school and college funding, this may be considered an indirect cause outside the control of school and college staff for the purposes of improvement planning.

How to Identify Root Causes

Determining root causes is a search for the most direct and highest impact causes of performance gaps on core indicators that are within the control of schools and colleges. This search should employ a systematic evidence-based process, one that allows for the formulation and testing of theories or hypotheses about the underlying cause-effect relationships for each core performance indicator. It should draw on current research and evaluation, and use multiple methods and data sources to test specific hypotheses. Rarely are performance problems caused by a single factor; rather, they are caused by a combination of root causes and indirect causes some of which are beyond the immediate control of schools and colleges.

There are many different approaches to identifying root causes, but most approaches involve three basic phases: (1) identify potential causes, (2) analyze and evaluate potential causes, and (3) select a critical few root causes.

Phase 1: Identify Potential Causes

Start the process by first developing a comprehensive list of all possible causes. Use multiple methods to develop lists so that you can take advantage of multiple sources of information and multiple perspectives from students, faculty, and other internal and external stakeholders.

Methods	Advantages and Contributions
Reviewing Research Literature	Reviewing the literature provides insights from researchers on major causes of performance gaps from a wide variety of school and college settings.
Reviewing Program/Institutional Evaluations and Effectiveness Reviews	Assessing evaluation data may provide you the perspective of external evaluators, peers, and stakeholders on what they think are potential root causes of performance gaps at your school, college, or a peer institution.
Analyzing Student Data	Analyzing your own student data offers a unique opportunity to identify and evaluate potential causes by attempting to explain why some students achieve desired outcomes and others do not, even those students with similar backgrounds and characteristics and those in the same programs. This could be done for all students or a sample of students.
Conducting Focus Groups	Conducting meetings with students and professional staff provides a mechanism to gain the perspective of customers and stakeholders on what they feel are the major causes. This subjective information can then be used to develop and test hypotheses using objective methods. This method provides a basis for customers and stakeholders to take ownership of your conclusions.

Brainstorming	Providing people a way to freely express their opinions and generate new ideas is important in identifying all potential causes, including those that may be counterintuitive or outside conventional wisdom. This method provides a systematic approach for generating new ideas that can be more fully tested and evaluated by other methods. It also provides the basis for ownership of your conclusions and the search for solutions.
Peer Benchmarking	Networking with peers provides an opportunity to exchange information and perspectives on root causes and may save valuable time. Focus your peer exchange with similar peers and best performers, so that you can make sure to identify the major causes that best performers believe are the most critical causes that must be addressed and controlled to achieve performance excellence.

Phase 2: Analyze and Evaluate Potential Causes

After identifying a potential set of causes, group these causes into two major categories: (1) causes within your control that could potentially be addressed by improvement strategies, and (2) causes beyond your control that to consider when developing improvement strategies. For example, for secondary academic attainment, you may start a listing of your causes in a table like the one below.

Causes Within Control	Causes Outside Control
<ul style="list-style-type: none"> ● Student Motivation and Engagement ● Effective Instructional Practices ● Teacher Training/Education ● School Expectations/Incentives ● Perceived Career Relevance 	<ul style="list-style-type: none"> ● Student Transfer/Mobility Levels ● Family Income ● Parents' Education ● School Resources

Then, use many of the same methods used to identify potential causes to fully test and evaluate them. Although all potential causes should be tested and evaluated, spend most of your time and resources on those that you have determined to be within your control. Evaluate those causes by addressing the following criteria:

Theory

Is there a clear and compelling theory or rationale for the cause?

Evidence

Is there strong and compelling evidence that this is a major cause of performance problems? Is there evidence to suggest that it is a major cause at your school or college or at peer institutions?

Root or Indirect Cause

Is this cause a direct cause of performance gaps or is it an indirect cause that is has an impact only through another related cause?

Need

Does this cause represent a major problem? Do conditions exist that make it an obvious factor to be addressed? Has this cause already been successfully addressed at your school or college?

Impact

Are there opportunities and resources to address this cause and make major impacts on performance gaps?

Stakeholder Support

Do major stakeholders who must develop and implement solutions support the cause widely?

Of the methods listed above, the two that are most helpful for evaluating root causes are analyzing student data and conducting focus groups. For example, if during exploratory focus groups and brainstorming “established student career objectives” was identified as a cause affecting secondary placement rates, you could then evaluate and validate the extent of a cause-effect relationship. This can be accomplished by analyzing your current data to see if those students who had clear career objectives after completing career exploration and planning activities were more likely to transition to employment and/or further education. In absence of this data, you could survey students prior to high school completion to see if those receiving career services had clear objectives and whether those students have greater likelihood of transitioning to employment or further education the following school year. Since this would take time, an alternative would be to conduct surveys or focus groups with students, faculty, and staff to review and confirm your hypotheses.

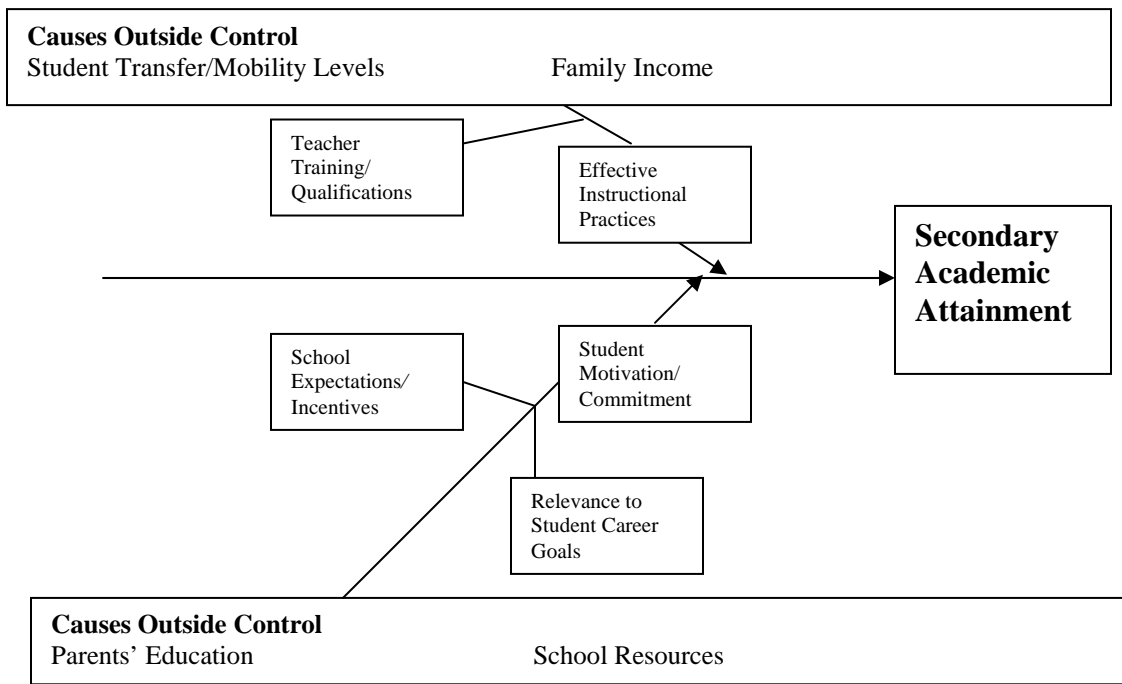
This will result in an overall evaluation of the potential causes within your control similar to the example provided below.

Causes Within Control	Theory	Evidence	Root/Indirect	Need	Impact	Support
Effective Instructional Practices	Strong	Strong	Root	High	High	High
Teacher Training/Education	Strong	Strong	Indirect	High	High	High
Student Motivation/Engagement	Strong	Strong	Root	High	High	High
School Expectations/Incentives	Strong	Moderate	Indirect	High	High	High
Perceived Career Relevance	Strong	Weak	Indirect	High	High	Moderate

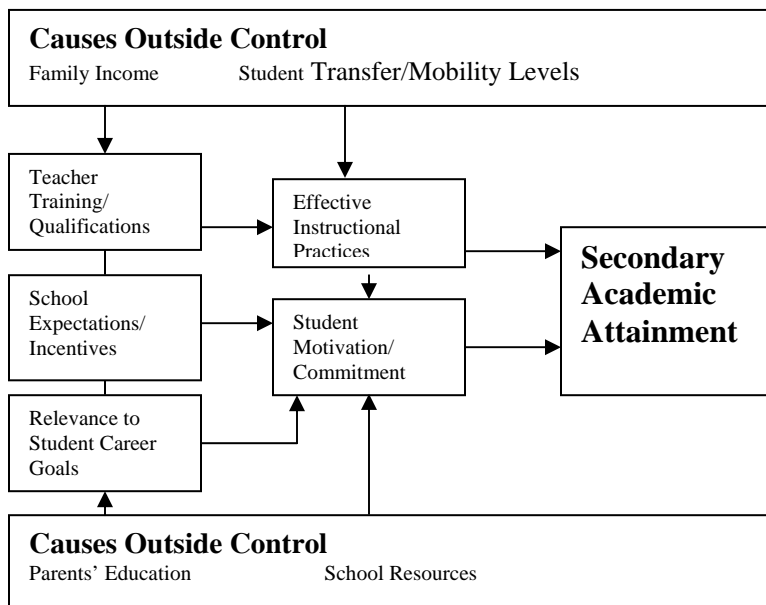
Phase 3: Organize Your Theory and Select the Most Critical Root Causes

After completing your evaluation, attempt to formulate and gain consensus on a comprehensive theory of performance for a core indicator.

It may be useful to use visual models to communicate your theory of performance on a core indicator. Many different visual models are useful, but the rule in developing these visual models is to “keep it simple.” One approach is to construct a modified fishbone diagram like the one shown on the next page. In this example, the model demonstrates the causal relationships among various causes outside of and within your control on secondary academic attainment.



Another is to use a cause-effect model like the one below.



After organizing and gaining consensus on your overall theory of performance, select the most critical root causes on which to focus in developing solutions. You can rarely address all of the root causes at one time—set priorities and address the most critical causes first.

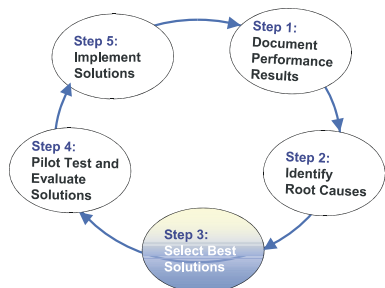
Use the results from your testing and evaluation to select those root causes that have the strongest theory and evidence to support them, address the most critical needs, and provide the best opportunity to have high impacts on performance.

Moving to the Next Step

Before you move to the next step and start the search for solutions, make sure you have laid the necessary foundation. Have you done the following:

- ❑ Used multiple approaches to identify potential causes based on data and perspectives of stakeholders?
- ❑ Grouped potential causes into those within your control and those outside your control?
- ❑ Used multiple approaches to analyze and evaluate potential causes and assess them according to objective criteria?
- ❑ Organized your theory of the causes of performance into a visual model or figure that shows: (1) root causes, (2) indirect causes, and (3) causes outside your control?
- ❑ Identified the most critical root causes on which to focus efforts in searching for solutions?

STEP 3: SELECT BEST SOLUTIONS



Once you have identified the most critical root causes to address in your improvement effort, the next step is to identify and select the solutions that seem most promising for testing and evaluation. This section will assist you in reviewing and selecting potential solutions for testing in Step 4.

Why Take the Time to Search for and Evaluate Alternative Solutions?

It pays to take the time to systematically review and select all of the potential solutions that have the best chance to be successful, including your own “home grown” solutions. Testing and implementing solutions that don’t work can be time consuming and expensive, and can undercut staff morale. Also, most schools and colleges have limited resources and need to make sure they are working on solutions that have the largest impact. Finally, build consensus and stakeholder support to implement solutions. It is important to reach agreement among staff and other stakeholders that you have properly considered all possible solutions supported by different people so that you can secure full commitment and support for testing and evaluating one or more of these potential solutions.

Don't Be Too Quick to Adopt Best Practices Before Getting the Facts Straight

How do you identify the best solutions? First, identify or develop a full range of potential solutions. Selecting a full range of choices stretches your thinking and helps develop more creative solutions. Next, select the most promising of these potential solutions. Any systematic analysis of alternative solutions has two parts:

- 1) reviewing the underlying logic or rationale of the solution—is it based on sound theory of root causes (Step 2) and how does the solution address these causes, and
- 2) reviewing the empirical evidence—has the solution worked, that is, has it produced results under similar or comparable circumstances to yours and is the evidence strong and compelling.

How to Develop Solutions: Improvement Strategies and Models

In developing solutions, begin by identifying the potential improvement strategies that can impact the root and indirect causes in your cause-effect model from Step 2. Next, identify specific school or college model practices that are based on these strategies. This can be illustrated by the example below for the Secondary Academic Attainment indicator.

Root Cause	Improvement Strategy	Model (Model Practices)
Time on Task	School Class Scheduling	Block Scheduling Model 1 Block Scheduling Model 2

Identifying Potential Strategies and Models: Three Methods

For best results, use multiple methods to identify potential improvement strategies and models. It is important to identify a broad range of potential strategies and models. Seeking both conventional and seemingly radical strategies and models yields the best results because it stretches your thinking. Below are three methods that you should utilize.

Review What Others Propose

Always consider what your profession, peers, and researchers have found to be leading improvement strategies and models that have shown results in other schools and colleges. There are many different sources for this information including:

- ❑ National Research and Dissemination Centers for Career and Technical Education (www.nccte.org)
- ❑ National Research and Development Centers and Regional Educational Laboratories (www.ed.gov)
- ❑ Peer Collaborative Resource Network (www.edcountability.net)
- ❑ ERIC Publications and Databases
- ❑ Professional Journals and Magazines
- ❑ Professional Association Publications

Benchmark Peers and Leading Performers

It is also beneficial to take the time to learn from other institutions and programs that consistently show the top performance among institutions and programs like yours. You need to fully understand what improvement strategies and models they are using and what is the underlying cause-effect model on which they are based. These strategies and models may or may not be widely known in the professional and research literature. They may also not be fully described at the level of detail to be transferred to your school or college. See Appendix A for tips on benchmarking strategies and models.

Develop Your Own Solutions

Regardless of what others propose and your benchmark partners do, develop, integrate, and/or customize strategies and models to meet your unique circumstances. In most cases, strive to implement strategies and models that can address more than one critical root cause and can work effectively under conditions or constraints (causes outside your control) you have identified in Step 2 of this guide. Try to integrate these strategies and models with existing policies, programs, and practices that are already in place and have been determined to be effective, building on and improving previous effective strategies and models while taking into account new ideas from peers and best performers.

Narrowing the Choices: Assessing and Comparing Alternative Strategies and Models

Since not all solutions you have identified will work for you, narrow the choices by assessing and analyzing their rationale or underlying logic and the empirical evidence that supports them. Use the following criteria in your assessment.

Sound Theory and Logic

Why Can It Be Expected to Work for You? The best solutions are the solutions based on clearly understood and sound theory that explains how the improvement strategy and model works and why—what some have called the “logic” of improvement strategies and models. Best strategies and models provide a compelling logic that makes sense to the people who will be expected to test, evaluate and eventually implement and support the solution. They describe the major causes the strategy and model

address and how the solution addresses these causes. For example, a school or college may consider alternative teacher training and mentoring models by asking questions on whether they address all necessary “effective instructional practices” and how they make sure that the model results in effective instructional practices and student achievement. In other words, what is the logic of each model? Why do their sponsors think they work and under what conditions?

Strong Evidence

Has it Been Shown to Work? In Schools or Colleges Like Yours? The best solutions also are supported by strong evidence indicating they have worked under conditions similar to yours, especially in schools and colleges that have similar “causes outside your control” identified in Step 2. For example, if your school has large concentrations of poor families and high interschool mobility, do these improvement strategies and models work in schools with similar conditions? The evidence should be based on sound evaluations using scientific methods.

Improvement strategies and models with sound theory and compelling evidence rarely come ready made for your situation. You generally have to choose between imperfect alternatives that have some limitations in theory and evidence. Therefore, it is necessary to fully compare and contrast the tradeoffs among alternative improvement strategies and models before choosing which solution or combination of solutions to test and evaluate yourself. Use the following criteria and rate each solution like you rated the causes in Step 2 of this guide.

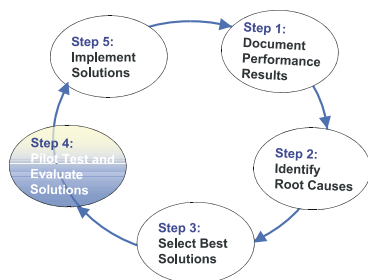
Selection Criteria	Rationale
Sound Theory	The strategy and model should make sense to all major stakeholders and you should be satisfied that it clearly addresses your critical root causes.
Strong Evidence	Unless developing your solution from scratch, the strategy or model or elements of the strategy and model should have worked somewhere; ideally somewhere that has the same conditions as yours (i.e., causes outside your control). The evidence should be credible enough to consider it a solution that works.
Costs/Time of Further Testing	Some strategies and models are too costly and time-consuming to thoroughly test and evaluate. Consider the costs and time of testing when assessing alternative solutions.
Resources and Support	The costs of implementing and maintaining strategies and models should be a major consideration. Consider the resources and capabilities necessary to support the solution and the extent of external resources and support to help sustain the solution at your school or college.
Stakeholder Support	The process of identifying and selecting the best strategies and models should be designed to build consensus and ownership among major stakeholders. Always consider the level of stakeholder support when choosing the best solution.

Moving to the Next Step

Before you move to the next step, make sure you have identified a full range of potential improvement strategies and models and have chosen the best solutions based on a systematic review. Have you done the following:

- Identified potential strategies and models by reviewing research and professional literature?
- Identified potential strategies and models by conducting peer and best performer benchmarking?
- Developed new, integrated, or customized strategies and models based on an internal design process?
- Assessed and rated alternative strategies and models based on the five selection criteria?
- Selected the best strategies and models for further testing and evaluation?

STEP 4: PILOT TEST AND EVALUATE SOLUTIONS



Now that you have identified a set of promising solutions—improvement strategies and models—on which to base initial improvement efforts, create an evaluation approach that will allow you to assess how well the improvement strategies and models are working. This section will assist you in selecting practical evaluation designs and analysis tools that you can use to gauge the success of your improvement efforts.

Why Evaluate?

Even though findings from the literature or your own common sense may suggest you have found a winning formula, it is important to pilot test improvement strategies and models to see if they produce desired outcomes. Monitoring student and program performances can provide a clear indication of how well your improvement strategy is operating and, in the event gains are not at an acceptable level, whether there is a need to rethink some or all of its components. Pilot testing is essential to your improvement process because it can help you to refine the strategy and may save considerable resources if the strategy does not work. Additionally, testing and evaluating improvement strategies provides an evidenced based method for justifying your educational improvement efforts and goes beyond theorizing about what works—an evaluation will supply you proof of whether the strategy works.

Make sure your improvement strategy works somewhere before you attempt to apply it everywhere...

A well-designed evaluation is one that enables you to assess the overall impact of improvement efforts, as well as your success in addressing the root causes that underlie improvement goals. Since student demographics, teacher characteristics, and district and state conditions all exert some influence on your proposed activities, structure your evaluation to take into account factors over which you have little control. If properly designed, evaluations can help to test and evaluate your overall theory of root causes of performance and how your improvement strategy and model addressed these causes and improves performance.

If you're successful, you'll not only see increases in student performance, but also be able to use your evaluation results to leverage additional support for your programs. Positive results can convince state policymakers, educators, parents, students, and other stakeholders of the value of your efforts, making it easier to obtain program funding, to motivate others to join the effort, and to convince students to participate in classes. Demonstrating success can also reinvigorate those already involved, convincing them of the value of collecting high quality, reliable data to document program outcomes.

How to Test Solutions: Designing an Evaluation Strategy

Mounting an evaluation can seem a daunting task, particularly given that you will simultaneously be working to introduce and run a program improvement effort. To lessen the burden, identify in advance a research methodology that you can use to collect the necessary information.

Choose a Study Design

Because changes in student performance may occur for a variety of reasons, design your evaluation to control for factors (e.g., other root or indirect causes and causes outside your control) that may confound your interpretation of results. Ideally, you will randomly assign students among pilot sites, in part because students in a given location may be more likely to be influenced by similar, outside factors that may affect their performance.

Since random assignment may not always be possible, the next best option is to structure the evaluation to allow assessment of changes in student or program performance over time; if possible, comparing those in pilot-sites against those not participating in the intervention. Listed below are three study designs you might consider when structuring your evaluation.

Option A: Random Assignment with Control Groups

Randomly assigning students into experimental and control groups offers the strongest basis for testing your improvement strategy, primarily because it can allow you to rule out external factors that might affect student outcomes. To employ this approach, begin by randomly assigning students into treatment and non-treatment groups. Then, over time, compare outcomes for each group to assess whether the program improvement model is responsible for any observed improvements.

Since random assignment of students can ensure some comparability across experimental and control groups, any difference in outcomes between the two groups can be directly attributed to your program effects. Unfortunately, scheduling conflicts, ethical considerations, and the need to collect data on all students make this approach difficult to employ.

Option B: Comparisons with Similar Populations

All things being equal, students enrolled in treatment districts might be expected to have different outcomes than those not participating in the project. To assess the overall effect of your improvement strategy, find some way of comparing student outcomes by taking into account the influence of other factors on performance.

One means of addressing other factors is to identify a comparison population of students in sites that are not yet receiving services. To ensure valid comparisons, select a group of students that, though not randomly assigned, has similar characteristics as those participating in the treatment group. If you are successful in matching the treatment and non-treatment groups, you will be more likely to prove that the improvement program caused any observed changes.

Option C: Comparing Individuals Against Themselves

Perhaps the simplest approach is to measure how performance outcomes change over time among students participating in the intervention. Based on the belief that, without any action things will proceed as before, this approach assumes that the intervention itself caused any observed changes. Typically, gains are measured using pre- and post-tests of student performance or other measures that compare outcomes prior to, and following participation.

The advantage of this approach is that it can simplify evaluation efforts and reduce the need to devote extensive time and resources to tracking student outcomes. The drawback is that, since there is no way of controlling for outside factors, it may be more difficult for you to claim that your improvement strategy alone accounted for any observed changes.

Select Pilot Sites

While it is tempting to roll out your program improvement strategy statewide, it is advisable to select a subset of sites in which to pilot-test ideas. Since different approaches may work better in different contexts or with different student populations, identify a small, representative group of schools, districts, or postsecondary institutions in which to test your approach. If your intent is to eventually expand your program statewide, try to select sites that mirror your state's demographic conditions. Some factors to consider include:

Student Characteristics

- Race-ethnicity
- Socio-economic status
- English language proficiency
- Special needs status

Site Characteristics

- Size of student population
- Geographical location
- Performance on state indicators
- Postsecondary participation rate
- Intensity of vocational coursework offerings
- Organization of instructional delivery

One means of identifying pilot sites is to rank local education agencies within your state on a range of dimensions, selecting a cross-section that allows you to observe how well the improvement strategy works in a variety of contexts. Randomly select a manageable number of sites that allows you to both test your approach and provide you with sufficient feedback to assess whether you're making desired progress.

Select Outcome Measures

It can often take a long time to observe improved performance on the Core Indicators associated with your improvement strategies. To help track performance changes, develop both short- and long-term measures to provide some indication of the success of your improvement efforts.

Short-term measures focus on intermediate results that must be achieved to eventually obtain results on the Core Indicators. These measures focus on how well the improvement strategy is being implemented and whether you are successfully addressing indirect or root causes. For example, in measuring implementation, you may survey students and teachers or examine student records to assess if the improvement strategy is implemented consistently and they are doing what you expected. In measuring intermediate results, develop measures to determine changes in root and indirect causes addressed by the improvement strategy. Plan to use this data to make mid-course corrections as your implementation proceeds.

Long-term measures are designed to provide direct evidence of your success in improving performance on the Core Indicators. These measures are typically collected at or near the end of your experimental project and may be collected during a post-intervention follow-up period depending on the Core Indicator.

Identify a number of differing measures that will provide *quantitative* and *qualitative* information. Quantitative information is information that can be reported in numeric form, such as the percentage of vocational concentrators who earn a score at or above the “proficient” level on a state academic exam. Qualitative data—characterizing what participants think or feel—may also be used to support the numeric results you obtain.

Identify Data Sources

After selecting short and longer-term outcome measures, it is important to identify data sources and collection instruments that will allow assessment of whether the improvement strategy is affecting student or program outcomes. Examples of data sources are listed below.

Student transcript records	Interviews or focus group discussions
State standardized test scores	Classroom visits or observations
Guidance records	Local assessments
State or locally administered surveys	

While multiple measures can help ensure that you obtain a comprehensive set of information on program outcomes, the selection of data sources and collection instruments should reflect the purposes of the study. Recognize that there are advantages and disadvantages associated with different data collection strategies.

Evaluating Data Collection Sources		
Method	Advantages	Disadvantages
Student Transcript Records Standardized State Tests Guidance Records	<ul style="list-style-type: none"> • Accurate • Inexpensive • Up-to-date information 	<ul style="list-style-type: none"> • May require programming systems • Reliability may vary across sites • May provide limited measure of student performance
Student Surveys	<ul style="list-style-type: none"> • Low cost • Opportunity to tailor questions to address specific issues 	<ul style="list-style-type: none"> • Response rates may be low • Potential bias in reporting • Students may not be truthful • Limited to direct, closed-ended questions
Interviews or Focus Groups Classroom Visits	<ul style="list-style-type: none"> • Can provide relatively more in-depth information • Reliability can be higher • Range of questions may be asked 	<ul style="list-style-type: none"> • Expensive to conduct • Staff require training • Number of participants may be limited
Local Assessments	<ul style="list-style-type: none"> • Low cost • Can tailor to address local needs • Targeted data on reform efforts 	<ul style="list-style-type: none"> • Reliability may vary across sites • Measures may not be consistent

Choose data sources and collection instruments that are *valid*, meaning that they directly assess what you are trying to measure, that are *reliable*, meaning that they will provide consistent results over time, and

that are *cost-effective*, meaning that they will not bankrupt the project budget. Also consider the best period for *timing* data collection to ensure timely access to accurate information.

Train Pilot Site Staff

Since the outcomes of the improvement effort will hinge on the work of participating administrators and school staff, communicate the purposes and activities related to the improvement effort to ensure that pilot sites are faithful to the planned intervention.

Communicating the objectives of the improvement strategy may include:

- ❑ Sharing research literature about the outcomes associated with various activities
- ❑ Conducting training workshops to clarify how agency staff can contribute
- ❑ Developing technical assistance materials to support educators in implementing solutions
- ❑ Training project staff on the procedures that should be used to collect and report data

Given that you will be relying on school faculty and administrators to collect data, specify clear, easy-to-follow instructions on data collection procedures to ensure that data are consistent across sites.

Analyzing Initial Results

Whenever possible, employ basic descriptive or summary statistics to assess outcomes. The statistics, presented in Step 1 of this guidebook, include the use of measures of central tendency (e.g., means, medians); measures of dispersion (e.g., range, standard deviation), position (e.g., ranking); and simple statistical tests, such as comparisons of mean using the student t-test.

To assess whether gains are realized equally across all groups, plan to disaggregate data for various subgroups of students, for example controlling for race-ethnicity or special population status. If outcomes differ, consider whether variations are due to the manner in which your solutions have been implemented or to characteristics of the subgroup itself.

As you near the end of your pilot project timeline, determine whether you are ready to move to full implementation of your program improvement model. In the event that the program improvement strategy does not work, consider whether you:

- ❑ Correctly identified the root causes of the performance problem
- ❑ Selected the best solutions—the improvement strategy and model—to address the problem
- ❑ Correctly implemented improvement strategies
- ❑ Need to give yourself more time before assessing results

If you are pleased with the results and believe others will be convinced of its merit, prepare to move to full implementation.

A CAUTIONARY NOTE: BUDGET TIME FOR EVALUATION

Rolling out a new educational reform model can be a daunting task, even if initial project activities are confined to a limited number of sites. To ensure that project implementation proceeds smoothly, plan to set aside time each week to step back and critically assess the results of your efforts. In particular, take a look at the available data on project implementation and early results and think how you can use your existing information to improve program effectiveness. Consider sharing this information with project participants and solicit their interpretation of the gains they have made.

Moving to the Next Step

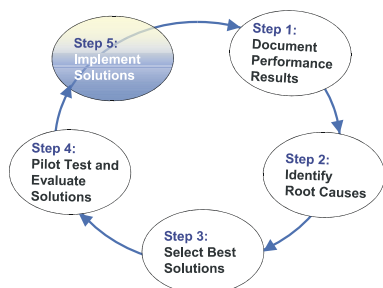
Before you roll out your improvement strategy and model statewide or suggest it to other local agencies, ask yourself a number of questions about your effort. Have you:

- Conducted a rigorous evaluation of program outcomes that will stand up to scrutiny from outside sources?
- Properly identified all of the strategies and models linked to root causes that contribute to program improvements? Do some work better than others? If so, should you consider dropping some strategies or models in favor of others?
- Considered how well your pilot-site results translate to other sites? Was there anything special or unique about your pilot sites that predisposed them to success?
- Secured a base of support among pilot-site participants? Can you find ways of enlisting experienced staff in rolling out your program?
- Identified sufficient resources to ensure that your improvement strategy is correctly implemented in pilot sites? How will you assess that new sites are true to your program goals?
- Evaluated whether sites are achieving success? Can all sites report on the data elements you have identified?

TIPS FOR DESIGNING A PROGRAM EVALUATION

Choose a Study Design	<u>Issues for Consideration</u>
<ul style="list-style-type: none"> ✓ Determine whether comparisons will be made against a randomized or non-randomized control group or for individual student's learning gains ✓ If using a comparison group, match students using a variety of demographic, social and economic characteristics 	<ul style="list-style-type: none"> ▪ Is it important to demonstrate that your solutions are solely responsible for observed gains? If so, use a control population for comparative purposes. ▪ What criteria will you use to identify comparison groups? Consider using sites rejected for the first round because they were too similar to those selected. ▪ How will changes in student or program outcomes be assessed over time? If against others, be sure to identify a representative control group. If against themselves, consider using some type of standardized pre- and post-test.
<p style="text-align: center;">Select Pilot Sites</p> <ul style="list-style-type: none"> ✓ Identify sites representative of your state's demographic and conditions ✓ Determine whether the analysis will focus on the district or school level 	<ul style="list-style-type: none"> ▪ What types of student or district characteristics might affect outcomes? ▪ How representative is your selected group? Will your findings at your pilot sites transfer to other agencies? ▪ How receptive will staff at pilot sites be toward your reform? ▪ Will you collect sufficient information to allow you to disaggregate results for different sub-populations of students?
<p style="text-align: center;">Select Outcome Measures</p> <ul style="list-style-type: none"> ✓ Plan for both short- and long-term measurement ✓ Seek to collect quantitative data 	<ul style="list-style-type: none"> ▪ Will you be able to identify immediate outcomes that can inform your effort? ▪ Will the data sources you are using provide comprehensive, quantitative information on student outcomes? Can qualitative data be converted into quantitative results?
<p style="text-align: center;">Identify Data Sources</p> <ul style="list-style-type: none"> ✓ Link data to root causes and solutions ✓ Identify multiple measures ✓ Routinize data collection 	<ul style="list-style-type: none"> ▪ How will you assess the validity of your data collection instrument? ▪ Can you be sure that participants will collect reliable data across sites? Consider developing data entry procedures documenting the process and timing for collecting information.
<p style="text-align: center;">Train Pilot Site Staff</p> <ul style="list-style-type: none"> ✓ Communicate objectives to participants ✓ Develop supporting materials to assist pilot site staff in collecting information ✓ Visit sites to assess implementation 	<ul style="list-style-type: none"> ▪ Do participants understand the purpose of the effort? Hold workshops to communicate program improvement objectives to participants. ▪ Develop technical assistance tools and materials that clearly describe the purposes of your intervention ▪ How sure are you that pilot site staff are following your solutions? Consider conducting site visits to participant agencies to observe project activities

STEP 5: IMPLEMENT SOLUTIONS



If you have come this far, you have likely identified a set of solutions—improvement strategies and models—that increase student performance in pilot sites. As you prepare to expand the pool of participants, plan to simultaneously create a monitoring process that will allow you to obtain ongoing feedback on the improvement strategy. This section will help you to develop implementation plans to monitor outcomes across the full range of implementation

sites.

Moving from Pilot Testing to Statewide Implementation

Successfully piloting improvement strategy does not necessarily guarantee that the improvement strategy will continue to work as the number of participants in the effort expands. Replicating small successes on a larger scale can be complicated by the addition of new factors and administrative challenges that can reduce program effectiveness so approach this effort with renewed focus.

Don't say you've solved the problem until you've fully implemented solutions and achieved results!

Full implementation will require flexibility and a willingness to revise strategies as new sites join the effort. Expect to modify solutions and evaluation approaches to accommodate unique site characteristics or unexpected situations that arise in mid-stream.

Just as with the pilot, monitor site performance on a number of dimensions to ensure you're achieving the intended results. Plan to evaluate both the desired outcomes as well as the process you are using to roll out the implementation across agencies. To encourage this process, seek to make on-going assessment a part of the organizational culture, and consider establishing short-term and long-term milestones to share with project participants.

Monitoring Ongoing Implementation

Expanding your improvement efforts requires the development of a comprehensive strategy for monitoring the implementation and evaluation approach, one that will enable you to monitor how well the implementation is progressing. Plan to enlist the support of pilot site staff to help explain the importance of monitoring to new participants and to serve as field-based mentors, for example by having educators from new sites visit pilot sites to speak directly with experienced staff.

Steps to consider taking for monitoring on-going implementation may include:

- ❑ Preparing a “Process Evaluation Plan” to help assess the manner in which you are implementing new improvement strategies. The plan may include assigning activities, a timeline for completion and a quality check. Have responsible parties keep a log of unanticipated obstacles and ways to improve activities.
- ❑ Establishing short-term outcome measures (e.g., changes in attendance or student grades) that can be used to assess on-going efforts.

- ❑ Identifying process indicators that capture how well improvement activities are being implemented.
- ❑ Calling or visiting school and district staff to observe whether they are correctly applying proposed solutions.
- ❑ Requiring local faculty and staff to attend technical assistance workshops to discuss their observations and challenges.
- ❑ Conducting focus groups with students and faculty to assess changes in beliefs or practices.

Sustaining Improvement Efforts

As its name implies, continuous improvement is a never-ending process that requires that you constantly review and critique the outcomes of improvement efforts. To sustain your effort, schedule time—much as you did in Step 4—to revisit your strategy. Consult the following checklist for additional ideas to help structure your inquiry:

- ❑ Do the root causes you've identified in Step 2 hold true in new implementation sites?
- ❑ How well do the solutions you've selected address the obstacles to high performance in new schools or districts? Are some more applicable than others?
- ❑ Are new sites properly implementing improvement strategies? Does staff appear to understand what is being asked of them?
- ❑ Is the data you are collecting in sites accurately capturing the progress being made?
- ❑ Are there any mid-term course corrections to make that can improve performance outcomes?

APPENDIX A: BENCHMARKING

What Is Benchmarking?

According to the American Productivity and Quality Center, benchmarking is the “process of identifying, understanding, and adapting outstanding practices from organizations anywhere in the world to help your organization improve its performance.”

The process of benchmarking is a systematic data collection and analysis process for determining:

- ❑ What Must Be Improved?—What must we do better to reach the performance levels of our peers and recognized leaders?
- ❑ How Much?—How much must we improve over the coming years to reach these performance levels?
- ❑ How Can It Be Improved?—How have recognized performance leaders achieved these results? What are the improvement strategies and models they have used to get there?

This process is systematic because the organization goes through a formal, step-by-step process for identifying and determining improvement priorities, improvement goals, and improvement strategies and models. The process is largely one of data collection and analysis because every decision in the process must be based on empirical evidence collected and analyzed during the process.

Although educators have always shared ideas and practices to help each other improve performance, many times they have not used a systematic, fact-based process to determine if ideas and practices have in fact worked. Benchmarking provides an opportunity for educators to improve how they learn from others to get results.

Types of Benchmarking in the Five-Step Process

The five-step improvement process in this guide uses two types of benchmarking:

- ❑ Performance Benchmarking—is the determination of what must be improved by how much to achieve performance levels of peers and recognized leaders on the Core Indicators (Step 1).
- ❑ Strategy and Model (Best Practices) Benchmarking—is the determination of what peers and leaders have determined to be the root causes (Step 2) and improvement strategies and models (Step 3) that have shown results or they expect to show results.

The reference section lists additional resources on benchmarking that states and colleges and schools can use to develop their own benchmarking process.

APPENDIX B: CALCULATING AND DISPLAYING DATA

This appendix is intended to provide a reference on summarizing, describing and displaying data to state-level staff that work with Perkins III performance data. It presents standard statistical concepts and terms and seeks to explain these with examples appropriate to Perkins data. This resource is targeted at individuals who need to understand and draw conclusions from their data, but who may not be acquainted with data analysis.

Summarizing Data

The first step in understanding data is summarizing it. The use of a handful of statistics can help to identify patterns and describe what is happening with your student population. There are a number of well-established methods for describing and summarizing data that will assist you to understand masses of information. Many of these techniques may be familiar while others may be less familiar. In the following section, definitions and examples of summary statistics are presented that you will be most likely to use in describing your own data. The table below summarizes the statistical tools that will be described in more detail.

Measure	Summary Statistics
Central Tendency	Mean Median
Dispersion	Range Standard Deviation
Position	Simple Ranking Percentile Ranking

Measures of Central Tendency

Mean

The mean, commonly called the average, is the most widely used measure of central tendency. The mean gives a sense of a representative or typical value and falls in the center of a distribution of values. To compute a mean, sum all the values in a set and divide by the number of values in that set. A mean is sensitive to a change in any value.

Example

Find the average placement rate of vocational secondary completers per school district in Table 1.

District	Placed Completers	All Completers	Placement Rate
District A	3,098	5,500	.56
District B	2,256	5,786	.39
District C	4,327	2,845	.72
District D	1,772	4,907	.36
District E	3,573	4,309	.83
District F	2,399	3,116	.77
District G	2,988	3,675	.81
District H	5,265	6,645	.79
Total	25,678	36,783	

The average placement rate per district $= (.56 + .39 + .72 + .36 + .83 + .77 + .81 + .79) / 8 = .65 \times 100 = 65\%$

The typical school district in your state places 65% of its vocational completers. This example gives equal weight all schools in the state in the calculation of the average.

Keep in mind the unit of analysis when looking at numbers. In the example above you found the average placement rate per secondary school district, which gave equal weight to schools regardless of the number of completers. However, if you wanted to know how your *state* performed in placing secondary vocational completers, you perform a different calculation.

State placement rate = Sum of all district's placed completers / Sum of all district's completers $\times 100$

$(25,678) / (36,783) = .70 \times 100 = 70\%$

Seventy percent of the state's secondary vocational completers were placed. Another interpretation would be that if you picked a vocational completer at random in your state, they would have a 70% likelihood of being employed, in higher education or in the military. This example is the outcome that is reported in the CAR. It allows you to see the average of the total and is based on the number of completers across schools.

Median

Another common measure of central tendency is the median. This represents the midpoint in a set of numbers—half of the units fall below the median and half above. This statistic differs from the average in that extreme values—whether they are very low or very high—do not affect its value. In a perfectly symmetric distribution of data, the mean and median are equal. To compute the median, sort the values of all cases and select the middle value. With an odd number of values, the median is the middle one. With an even number of values, add the two middle values together then divide by 2.

Example

You suspect that a few low performing school districts in your state are greatly affecting the secondary placement rate for the state. You would like to compare the average district performance to the median to see if this is true.

To find the median, you can order the district completion rates from lowest to highest:

.36, .39, .56, .72, .77, .79, .81, .83

The median performance is $.72 + .77 / 2 = .745 \times 100 = 74.5\%$

In this example, the average may give you a misleading indicator of district level performance. When using the average you conclude that districts have a 65% secondary placement rate. However, when you look at the median district completion rate of 75%, you are able to see that half of the districts are actually performing much better than the average completion rate, while a few districts are struggling in this performance area. Using both the mean and median can help you determine how much your mean is being affected by extreme values.

Measures of Dispersion

In describing your performance data, it is also often helpful to not only identify the middle values, but also to find how your performance data *varies* or is *dispersed* around the average. Three states may report very similar performance outcomes, however their individual set of numbers may vary dramatically. For example, State A may report an average school outcome of 50% on Postsecondary Academic Attainment (1P1) for Limited English Proficient students, with individual institutions reporting outcomes listed in the table. State B and State C also report an average outcome of 50% but their postsecondary institutions outcomes are distributed much differently.

LEP PERFORMANCE FOR 1P1 BY INSTITUTION

Table 2 Postsecondary LEP Performance		
State A Institutional Performance	State B's Institutional Performance	State C's Institutional Performance
40%	17%	15%
45%	24%	22%
50%	25%	30%
50%	75%	45%
55%	79%	90%
60%	82%	95%

State A											
0				40	45	50	55	60		100	
State B											
0	17	24	25					75	79	82	100
State C											
0	15	22		30		45			90	95	100

While the average institutional performance for LEP students on 1P1 is similar among the three states, the distribution of outcomes differs considerably. State A's institutions tend to clump around 50% while State B's are very spread out and State C's are skewed to the left with some outliers at higher performance levels. Measures such as range and standard deviation are statistics that can quantify and describe the dispersion of your performance data.

Range

The range is, quite simply, the difference between the smallest and largest values. It can give you an idea of the tails in your data set and an impression of how widely distributed your values are. It is entirely determined by the extreme values, though, and is not influenced by middle values. To calculate the range, subtract the smallest value in your data set from the largest value.

Example

Using the data for State A above, the range calculated is

$60 - 40 = 20$ percentage points

For State B, the range is

$82 - 17 = 65$ percentage points

For State C, the range is

$95 - 15 = 80$ percentage points

Even though both states reported a 50% performance level on postsecondary academic attainment for LEP students, you can see that the range of the data differs substantially. Using this statistic gives you a sense of whether students, schools, and institutions, are attaining similar levels or if they differ.

Standard Deviation

The standard deviation is a bit more complicated than the range, both conceptually and in its calculation. It is, however, the most useful statistic for measuring the spread of your data. You might want to think of it as the “average difference from the mean” in a distribution of numbers. To calculate it, do the following:

1. Find the difference between each data value and the average
2. Square all the differences and add them together
3. Divide the sum the squared differences by the number of data points (assuming you are working with a population and not a sample)
4. Take the square root of the quotient

Below is an example of the calculation of standard deviation using the data for State A.

Average= 50%

Differences: $(50 - 50) = 0$ $(40-50) = -10$ $(60 - 50) = 10$ $(50-50) = 0$ $(45-50) = -5$ $(55-50) = 5$

Sum of Squared Differences: $0 + 100 + 100 + 0 + 25 + 25 = 250$

Divide by N: $.250/6 = 41.67$

Square Root of $41.67 = 6.46$

The standard deviation from 50% is 6.46% for State A. In State B, the standard deviation from 50% is 28.5%. Without plotting the data, you can tell from the numerical value of the standard deviation that state B’s institutions’ performance in 1P1 for LEP students is more widely distributed than institutions in State A.

Why is this a useful statistic? Because in a single number it provides you with a representative value for a population’s variability, and it’s not heavily influenced by the tails of the data set. Also, in a normal distribution, 68% of the cases fall within one standard deviation from the mean and 95% fall within two standard deviations. This might be useful criteria to use when determining where performance gaps lie—for example, you may decide that the norm for performance should fall within one standard deviation from the mean (although in cases of very small variability, this may not be a good standard).

Measures of Position

Besides knowing the midpoint and dispersion of your performance data, it is often useful to know how a particular value compares to other values; that is, the *position* of a particular data value. Simple ranking and percentile ranking are two ways for identifying the position of a value.

Simple Ranking

To do a simple ranking, arrange the data elements in some order and determine where in that order a particular value falls. For example, you might want to compare a particular school district to all other districts in the state on secondary non-traditional participation. By ordering the performance outcomes, you find that it has the 3rd highest participation among the districts.

Percentile Ranking

A percentile ranking can provide an indication of the percentage of values falling below the value under consideration. It is commonly used in standardized testing to indicate how a particular student score ranks compared to all the scores in the state. For example, a score of 50 does not indicate performance relative to others; however, if knowing that a score of 50 places you in the 80th percentile, then you can ascertain that score of 50 places you in the top 20% of the tested group.

To calculate percentile ranking, order and rank the values. Select the rank under consideration and divide it by the total number in the population then subtract from 100.

Example

A school district in your state ranks 146th out of 625 districts nationally in non-traditional completion rates. To determine the district’s percentile ranking:

$$146 / 625 = .23$$
$$.23 * 100 = 23\%$$
$$100 - 23 = 77^{\text{th}} \text{ percentile}$$

This particular school district’s non-traditional completion rate is higher than 77% of all other districts in the country.

Communicating Performance Data

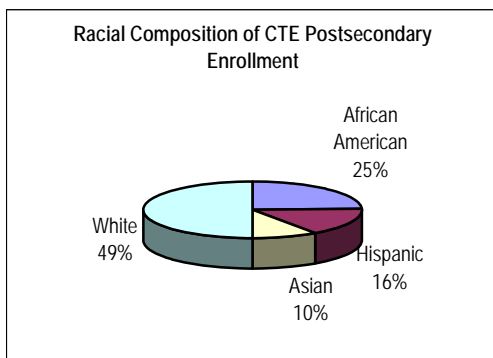
Beyond understanding your performance data, it is important to effectively communicate it to the public, parents, students, educational boards, and stakeholders. The first example below offers one way to communicate data and the second example provides an improved method.

Example 1: Text Only

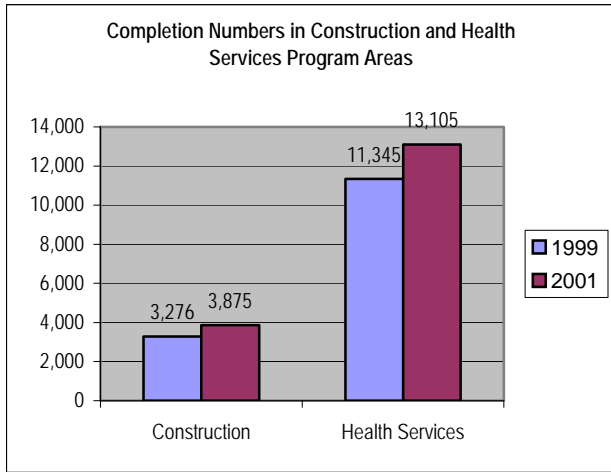
Sixty-four percent of all students enrolled in public postsecondary education attend 2-year colleges or other CTE programs. Of this number, one-fourth are African Americans, 16% are Hispanics, one-tenth are Asian, and the rest are White. In the past several years, completion rates at state CTE institutions have graduated less than 42% of enrollees, so reforms have been focused on improving completion rates. The last two years have seen gains in completion rates increase to 46% last year and 50% this year. In particular, since implementing reforms in 1999, construction programs awarded 600 more certificates this year compared to 1999, and 1760 more certificates have been awarded in health services programs.

Example 2: Text and graphics

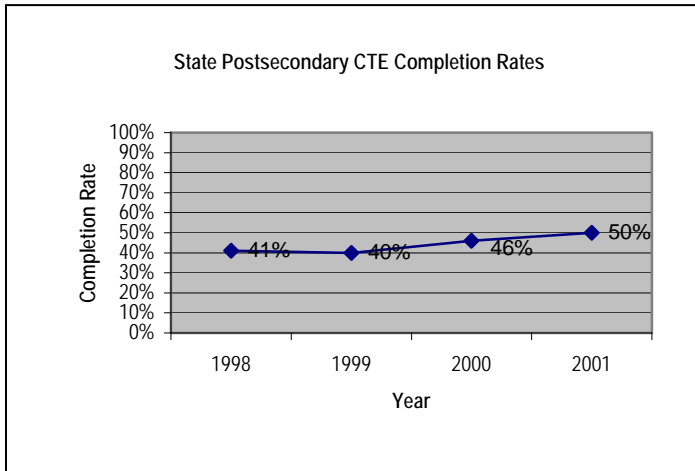
A lot of information and numbers are contained in the paragraph above, confusing the message that is being communicated. Visual tools like the graphs that follow, along with text relate the same information but more effectively communicate the various statistics that are presented.



Whites comprise nearly half of all postsecondary CTE enrollments, while African Americans make up a quarter of enrollments. Hispanics and Asians account for 16% and 10% of enrollments respectively.

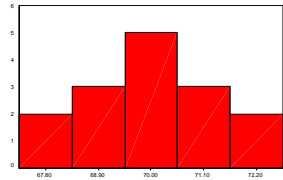
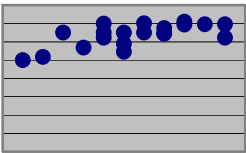
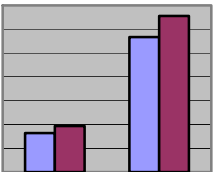
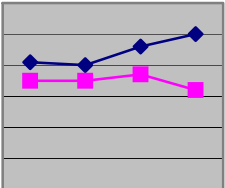



Since implementing improvement efforts aimed at increasing postsecondary completion, completion rates have improved from 41% in 1998 to 50% in 2001.



Significant growth has occurred in the completion numbers for both construction programs and health services programs. The number of students completing programs in these two areas has increased substantially since 1999.

There are a number of graphical tools you can use to communicate your data.

Graphical Tool	Use	Tips
<p>Histogram</p> 	<p>Reveals the distribution of data and tells you the likelihood of various values; helpful in revealing central tendency and dispersion.</p>	<p>Examine the shape of a histogram—does your data follow a symmetric pattern or is there a “tail” extending on one side? How spread out are the data? Look for whether values cluster around a central value, whether small and large values are equally likely, and whether outliers exist.</p>
<p>Scatterplot</p> 	<p>Displays the relationship between two variables with numeric values. For example, you might want to look at the relationship of attainment and completion rates.</p>	<p>Look for a pattern in the scatterplot—are points randomly scattered or do the points seem to be related? Look for whether the relationship is positively or negatively related, or whether no relationship is evident. One variable may be a predictor of performance in the other.</p>
<p>Bar Chart</p> 	<p>Useful for comparing sizes, quantities, amounts, proportions of similar variables. Can be used to show trends.</p>	<p>In a bar chart, you are comparing the relative length of the bars, since the bars represent the number of occurrences of a value in a particular category. Bars can be displayed both vertically and horizontally, depending on what you want to communicate. Be sure to use the appropriate scale on the axis—you can give the impression of extremely small differences or very large differences depending on the scale you select.</p>
<p>Line Graph</p> 	<p>Ideal for demonstrating trends and variation over time. Line graphs are useful for comparing performance among similar categories and change in a particular value.</p>	<p>Make sure that the time period being tracked is sufficient to be meaningful. When using the line graph to display outcomes for comparative variables, look for opening or closing of gaps over time and whether changes occur at similar rates. Look for “spikes” or “valleys” in the line and research whether something unusual happened at that point in time.</p>
<p>Pie Chart</p> 	<p>Illustrates the relative sizes of parts comprising a whole. Useful for showing the distribution of components.</p>	<p>Pie charts can be used to show both numbers and percentages. To emphasize a particular slice in the pie, you may want to make it appear to “separate” from the rest of the pie.</p>

The graphical displays provided above are some of the more common techniques for communicating data, but certainly there exist other meaningful displays. Keep in mind that proper and complete labeling of graphs is essential to clearly communicating the data.

APPENDIX C: DATA QUALITY CRITERIA

The quality criteria provide a set of objective criteria that states can use to critique the measurement approaches they have proposed in their Perkins III accountability systems. These criteria are intended to ensure that states have chosen a data collection strategy that will yield data that are both valid and reliable. States can assess each core indicator in relation to five quality criteria:

Alignment

The extent to which a measure directly and fully measures intended student outcomes. Measurement is aligned when assessment and other data collection instruments are unbiased and have strong content validity; that is, when they accurately measure what they are supposed to evaluate.

Scope

The breadth of measurement with respect to state-identified performance outcomes. Measurement is of sufficient scope when it quantifies students' performance in all state-identified measurement areas.

Reliability

The degree to which measurement is conducted consistently using standardized or comparable data collection instruments and procedures. Measurement is reliable when repeated measurements yield similar results and when it is conducted consistently across student groups and educational agencies.

Timing

The relationship between performance measurement and student participation in vocational programs. Measurement is well timed when it is conducted at appropriate intervals; that is, when it is concurrent with or follows student participation in or completion of vocational programs.

Coverage

The base of students included in measurement. Measurement has sufficient coverage when it includes all or a representative group of targeted students within a state.

REFERENCES

- American Productivity and Quality Center, *Benchmarking Pure and Simple: A Quick Reference Guide to Benchmarking*, Houston, TX: American Productivity and Quality Center, 1997.
- Andersen, B., and Fagerhaug, T. *Root Cause Analysis: Simplified Tools and Techniques*. Milwaukee: American Society for Quality, 2000.
- Bemowski, K., and Stratton, B. (Editors) *101 Good Ideas: How to Improve Just about any Process*. Milwaukee: American Society for Quality, 1998.
- Campbell, D. T., and Stanley, J. C. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin Company, 1963.
- Clark, T. J. *Success through Quality: Support Guide for the Journey to Continuous Improvement*. Milwaukee: American Society for Quality, 1999.
- GOAL/QPC, *The Memory Jogger: A Pocket Guide of Tools for Continuous Improvement*, Melthuen, MA: GOAL/QPC, 1991.
- Hoerl, Roger and Ronald Snee, *Statistical Thinking: Improving Business Performance*, Pacific Grove, CA: Duxbury Publishing, 2002.
- Hutton, D. W. *From Baldrige to the Bottom Line: A Road Map for Organizational Change and Improvement*. Milwaukee: American Society for Quality, 2000.
- Kelly, Michael R., *Everyone's Problem-solving Handbook: Step-by-Step Solutions for Quality Improvement*, White Plains, NY: Quality Resources, 1992.
- Levesque, K., Rossi, K., Bradby, D., Teitelbaum, P. *At Your Fingertips: Using Everyday Data to Improve Schools*. Washington: MPR Associates, Inc., 1998.
- Norušis, M. J. *SPSS 8.0 Guide to Data Analysis*. Upper Saddle River: Prentice-Hall, Inc., 1998.
- Pande, Peter S., Neuman, Robert P., Cavanagh, Roland R., *The Six Sigma Way*, New York, NY: MacGraw-Hill, 2000.
- Sternstein, M. *Statistics*. Hauppauge: Barron's Educational Series, Inc., 1996.
- Yap, Kim, et al. *Evaluating Whole School Reform Efforts*. Northwest Regional Educational Laboratory, Portland, Oregon, August 2000.

WEBSITES

American Productivity and Quality Center, <http://www.apqc.org/>

American Society for Quality, <http://www.asq.org>

Baldrige National Quality Program, <http://www.apqc.org>

The Guide to Managing for Quality, <http://erc.msh.org/quality/index.cfm>

The National Dissemination Center for Career and Technical Education, <http://www.nccte.com>

Peer Collaborative Resource Network, <http://www.edcountability.net>

Smartdraw Diagram tools, <http://www.smartdraw.com/specials/decisionfishbone.asp?id=9142>