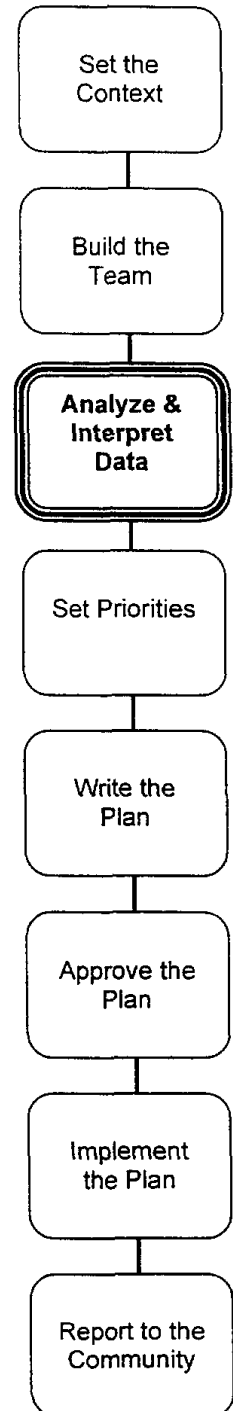


EQUITY AND EXCELLENCE

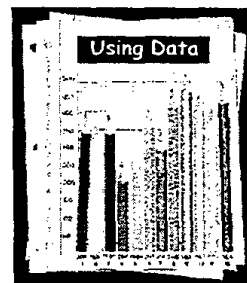
Analyze and Interpret Data

Improved student performance in relation to standards is the goal of action planning. To reach this goal we need to focus on *excellence* and *equity*. Excellence as a goal means that we need to move toward greater numbers of students achieving the standards or achieving the standards with honors. Equity means that as we move all students toward excellence, we also decrease the gaps among different groups of students.

This section provides you with guidance for analyzing data so that you can make a hypothesis about what the data is indicating in order to identify next steps for analysis or actions to be taken.



Using Data for Action Planning To Improve Student Performance



INTRODUCTION

“Data are to goals what signposts are to travelers; data are not endpoints, but data are essential to reaching them--the signposts on the road to school improvement.”

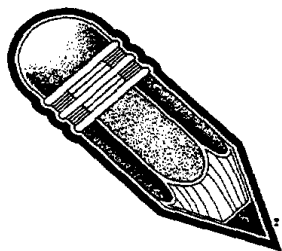
Results: the key to continuous school improvement, p. 30

Success will foster success as you focus on continuous improvement by finding trends, setting and reaching short- and long-term goals, and achieving incremental changes in student performance.

Using data to decide on specific actions to take that will result in students meeting and exceeding high standards is the purpose of **Action Planning**. It involves asking the questions:

- How are your students performing in relationship to standards?
- Why are your students performing at the level they are?
- What could you do differently in order to increase the number of students meeting or exceeding standards?
- How can you monitor progress?
- How are your students performing in relationship to standards?

This is the process of continuous improvement. Student performance is the beginning and the end in the process of continuous improvement. Only student data and data on those things that are directly related to



Notes

improving student performance on standards are to be examined and dealt with in action planning.

Schools are in different places with regards to action planning. Some schools will be examining data in this way for the first time or will be using only baseline data; other schools will be continuing an ongoing process. However, for all schools, there are some basic steps to follow:

- Collect and organize data
- Examine student performance data
- Identify strengths and weaknesses based on student performance data
- Look at equity as well as excellence
- Hypothesize causes
- Look deeper at other measures/indicators
- Formulate actions
- Set initial performance targets
- Set priorities

There are two purposes being served by this chapter of the *Action Planning Guide*: (1) understanding data; and, (2) using data. Since it is difficult to understand data without having a context, and difficult to use data without understanding the characteristics of some types of data, this chapter tries to share information about data within the context of using it. As you work your way through this chapter, it is recommended that you look at real score reports (examples are provided in Appendix I.) and try out the templates provided.

Remember that as you examine your data, your ultimate goal is to have ALL students meeting and exceeding high standards. As Michael Schmoker, author of *Results: The Key to Continuous School Improvement*, said, teams must examine data from every student in order to improve programs and achieve “broad and continuous benefits for greater numbers of students.”

Success will foster success as you focus on continuous improvement by finding trends, setting and reaching short- and long-term goals, and achieving incremental changes in your student performance. Your reward for using data is equity and excellence in student learning.

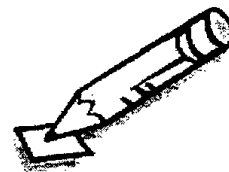
What kinds of data will be useful in action planning?

In order to do meaningful action planning, it is important to consider a variety of data. Data from standardized assessments, portfolio assessments, local assessments, and other indicators should be considered to gain a complete picture. No single piece of information tells the whole story, but analysis across different assessments can point out areas of needed improvement.

In action planning, student performance data from assessments are the first things to analyze. These data highlight potential areas of strength and weakness in a school's program. While assessment data can tell you where to look, these data may not be able to indicate what you need to do to improve the situation.

To determine what is needed to do to improve student learning, you must review other indicators such as curriculum alignment to standards, teaching practices, teacher preparation, dropout rates and attendance.

Key idea



No single piece of information tells the whole story, but analysis across different assessments can point out areas of needed improvement.

Assessments used to highlight strengths and weaknesses in reaching standards:

Administered to all students:

- ✓ New Standards Reference Exams in English language arts and mathematics (NSRE)
- ✓ The Vermont Science Assessment
- ✓ Vermont Developmental Reading Assessment
- ✓ Vermont History and Social Sciences Assessment (to be implemented in 1999)
- ✓ Locally administered norm-referenced assessments
- ✓ Local portfolio scores
- ✓ Other locally administered assessments

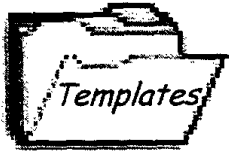
Administered to some groups of students:

- ✓ SAT (Scholastic Achievement Test)
- ✓ Advanced Placement Exams (AP)
- ✓ Diagnostic tests
- ✓ Other



Other indicators used to help plan actions:

- ✓ Curriculum alignment to standards
- ✓ Curriculum implementation
- ✓ Teacher practices
- ✓ Resource allocation
- ✓ Dropout rates
- ✓ Attendance
- ✓ Course taking patterns
- ✓ Course completion rates
- ✓ Advanced Placement course taking and success rates
- ✓ Other



Use template in appendix C.1 to organize assessment data.

The first step in using data

The first step is to organize data by what you want students to know and be able to do. To accomplish this, each team should list the assessments administered to all students by the standards and curricular areas addressed. A template in Appendix C.1 is provided for this task. (Note: If a local comprehensive assessment plan exists for your school, this task should already be completed.)

Why use multiple measures in action planning?

Using multiple assessments helps us to provide a full picture of student performance, one that shows students...

- 1) ... performing at a point in time and over time
- 2) ... recalling and using information
- 3) ... answering factual questions and problem solving through complex situations.

A Complete Data Picture Leads to a Strong Set of Actions

New Standards Reference Exam in Mathematics Results

Mathematics Portfolio Results

Strong Action Plan in Mathematics

Other information

- curriculum alignment to standards
- teaching practices
- assessment practices
- professional development support

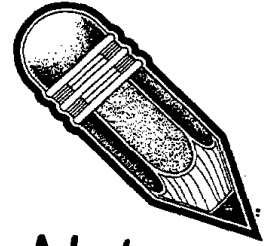
Examples of local measures

- norm-referenced test
- success on local criterion-referenced calculation test
- success on local uniform tasks

Analyze one measure at a time and then look across measures for consistent patterns and areas in which to focus the action plans.

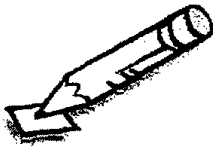
Important considerations when analyzing student performance data

- Whenever possible, use data from the same measures from year to year.
- Collect and analyze as much data as possible. Use multiple measures.
- Always consider the nature and the size of your group tested. (Discussion on sample size is found in Appendix C.18.)
- Be sure you know what is being measured.
- Large scale assessment data are about the school's program - not just about the grade at which the assessment was administered.
- Consider the characteristics of the group being tested.
- Compare performance from different groups of students (disaggregate). The goal is to reduce differences in performance between groups while increasing excellence for all.



Notes

Key idea



"Standards-based assessments provide scores that describe student performance against a set of high academic standards, not against other students."

New Standards Project Pamphlet

Before you begin...

... spend some time reviewing the reports in Appendix I.

School Standards Levels Summaries: Report the percentage of students who achieved at different performance levels in each of the clusters of standards. It summarizes instructional needs.

Student Standards Summary: Reports the performance level a student achieved in each cluster and provides information on instructional activities that might help individual students improve their performance.

School Demographic Summary: Reports performance by different demographic groups.

Instructional Needs Roster: This report is a summary of classroom instructional needs (only available for New Standards Reference Exams).

Standards-Based Assessments

What are standards-based assessments?

Standards-based assessments are any assessments that provide information about how students are performing in relation to the expectations in the standards. This includes assessments such as standardized standards-based tests, criterion-referenced tests, and portfolios. They also include classroom projects, tests, assignments, or quizzes that are aligned to the standards. For action planning, you need to look at standards-based assessments that are administered to all students at a grade level. This includes state and local measures.

What are Vermont's state-level standardized standards-based assessments?

Vermont administers state-level standards-based assessments in mathematics, English language arts, science, early reading and, starting in 1999, history and social science. These assessments provide information on how Vermont's students are performing in relationship to Vermont Standards.

Vermont State-level Standardized Assessments:

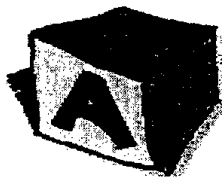
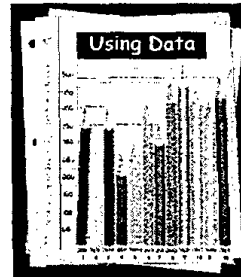
- New Standards Reference Exam in Mathematics, grades 4, 8, 10
- New Standards Reference Exam in English/Language Arts, grades 4, 8, 10
- Vermont Science Assessment, grades 6 and 11
- Vermont Developmental Reading Assessment, grade 2
- Vermont History and Social Science Assessment, grades 6, 9, 11 (starting in 1999)

The results from these standards-based assessments are reported in clusters of standards and indicate the percentage of students who have shown evidence of achievement at the following performance levels:

Achieved the Standard with Honors
Achieved the Standard
Nearly Achieved the Standard
Below Standard
Little Evidence of Achievement

What are the ways that standardized standards-based assessment results can be used in action planning?

- ✓ Use the distribution of scores across performance levels to determine how students are performing in relationship to Vermont's Standards.
- ✓ Use gaps between the school's performance and targeted performance to focus the plan.
- ✓ Use performance at both ends of the continuum:
 - 1) Decrease the number of students achieving Below the Standard or at Little Evidence.
 - 2) Move all students toward Achieving the Standard.
 - 3) Increase the number of students Achieving the Standard with Honors.
- ✓ Use data across time to identify improvements and trends in performance.
- ✓ If provided, use the Summary of Instructional Needs at the bottom of the School Standards Level Summary to identify the types of school-wide instructional activities that might improve performance.
- ✓ Use the School Demographic Summary to identify gaps in performance between different groups of students.
- ✓ Use scale scores provided from the New Standards Reference Exams in Mathematics and English/Language Arts as well as the Vermont Science Assessment to identify strengths and weaknesses in content areas/standards and to track progress over time.



TECHNICAL PRIMER

Important Note: Standards-based assessments are a type of criterion-referenced assessment. See page 3.15 for more information.

Vermont's state standards-based assessments report by clusters of standards.

New Standards Reference Exam
English Language Arts Clusters

- ✓ Reading: Basic Understanding
- ✓ Reading: Analysis and Interpretation
- ✓ Writing
- ✓ Conventions

New Standards Reference Exam
Mathematics Clusters

- ✓ Concepts
- ✓ Skills
- ✓ Problem Solving

Scale scores will be provided in *Arithmetic & Number, Geometry & Measurement, Probability & Statistics, and Algebra/functions.*

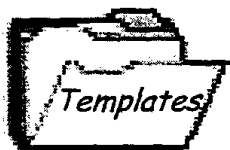
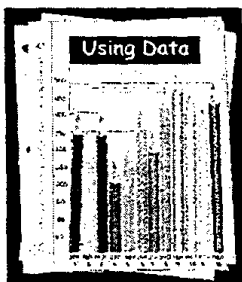
Vermont Science Assessment
Clusters

- ✓ Science

Scale scores will be provided in *Inquiry & Investigation, Systems, Living World, Physical Science, and Earth Science.*

Vermont Developmental Reading Assessment

- ✓ Accuracy and Comprehension



Use template in appendix C.2 to analyze the distribution of scores across each reporting cluster.

What information does the distribution of scores across the performance levels provide?

Vermont’s standards-based assessment results are directly tied to the Vermont Standards assessed. The results from standards-based tests are sensitive to the degree to which students experience curriculum and instruction aligned to the standards and the degree to which students are engaged with the program. These data, therefore, can provide rich information into issues of curriculum and instruction alignment to Vermont Standards.

For example: It is likely that curricula in schools where a large percentage of students score in either the Little Evidence or Below the Standard performance levels are not aligned with the standards. Performance at “Nearly” may indicate a program that is aligned here and there but not consistently articulated. Performance at “Achieved” indicates a program that is consistently aligned. Performance at “Honors” indicates a program that is consistently aligned and provides opportunities to learn at very high levels.

On the pages that follow, you will find three examples of distributions of scores with analyses of possible implications for action planning. In these examples, each analysis first identifies a key feature of the data to be explored; hypothesizes as to the cause; and then uses other data to confirm the hypothesis in preparation for identifying actions.

Distributions of Scores: Example One: N = 50
Percent of Students Who Achieve at Each Performance Level in Mathematical Concepts

	Little Evidence	Below the Standard	Nearly Achieved the Standard	Achieved the Standard	Achieved the Standard w/ Honors
School A	10%	60%	20%	10%	0%

70%

Key feature about this distribution that should be explored:
 70% of the students performed *Below the Standard*.

What might be causing this distribution? (Hypothesis)
 The curriculum and/or teaching practices that the students are receiving may not be aligned with the expectations in the standards assessed.

How can we find out if this is the cause?

- ☞ Determine if there is an articulated curriculum aligned with the standards. Be sure to look across classrooms within grades, across grades in a school, and across schools.
- ☞ Determine if that is the curriculum being taught.
- ☞ Find out if the teaching practices are consistent in delivering the curriculum.
- ☞ Determine if *all* the teachers have the content knowledge and skills necessary to deliver the curriculum.
- ☞ Determine if *all* teachers have the materials needed to deliver the curriculum.
- ☞ Analyze performance for different groups of students.

Equity & Excellence ▲ Action Planning Guide

Distributions of Scores: Example Two: N=150 Percent of Students Who Achieve at Each Performance Level in Mathematical Concepts					
	Little Evidence	Below the Standard	Nearly Achieved the Standard	Achieved the Standard	Achieved the Standard w/ Honors
School B	10%	45%	5%	30%	10%
	55%			40%	

<p>Key feature about this distribution that should be explored: A little less than half of the students Achieved the Standard or Achieved the Standard with Honors, and more than half are performing Below the Standard.</p> <p>What might be causing this distribution? (Hypothesis) This distribution suggests that groups of students in this school may be receiving two different curriculums.</p>	<p>How can we find out if this is the cause?</p> <ul style="list-style-type: none"> ☞ If there is an articulated curriculum aligned with the standards, determine if it is the curriculum being taught to all students or are some students receiving a different curriculum? ☞ Did students at different performance levels come from different feeder schools? ☞ Determine if the teaching practices are consistent across the school. ☞ Determine if <u>all</u> the teachers have the content knowledge and skills necessary to deliver the curriculum. ☞ Determine if <u>all</u> the teachers have the materials to support the curriculum. ☞ Determine how different groups of students performed.
--	--

Distributions of Scores: Example Three: N=50 Percent of Students Who Achieve at Each Performance Level in Mathematical Concepts					
	Little Evidence	Below the Standard	Nearly Achieved the Standard	Achieved the Standard	Achieved the Standard w/ Honors
School C		10%	50%	30%	10%

<p>Key feature about this distribution that should be explored: 50% of the students performed in the Nearly Achieved the Standard performance level.</p> <p>What might be causing this distribution? (Hypothesis) This distribution may mean there is an uneven implementation of the curriculum or an uneven implementation of teaching practices aligned with standards.</p>	<p>How can we find out if this is the cause?</p> <ul style="list-style-type: none"> ☞ Determine if curriculum aligned with the standards is being implemented across all classrooms within each grade, across grades in a school, and across schools. ☞ Determine if the teaching practices are consistent across the school or are appropriately implemented. ☞ Determine if all the teachers have the content knowledge and skills to deliver the curriculum. ☞ Determine if all the teachers have the materials to support the curriculum. ☞ Determine how different groups of students performed.
--	---

Definitions

Performance Target:

The goal that a school sets for performance. (e.g. Sixty percent of the students will Achieve the Standard or Achieve the Standard with Honors.) See page 5.2 for considerations when setting performance targets.

Performance Gap:

The difference between present performance and targeted performance.

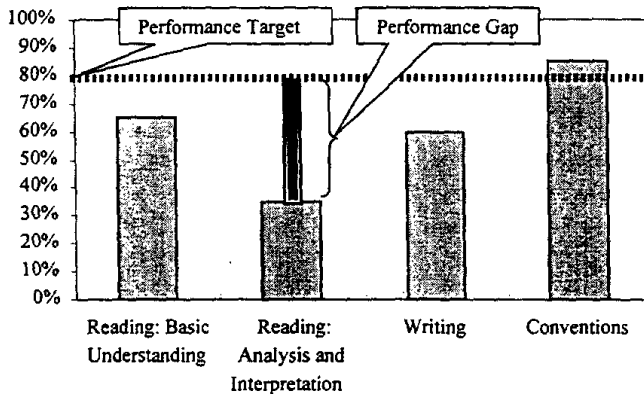
What information can the gap in performance provide?

The gap in performance can point to areas of need as well as point to areas in which the school has met or is coming close to meeting its goal.

The graph below indicates the percentage of students who *Achieved the Standard* or *Achieved the Standard with Honors* for a school. The line at 80% is the *initial* "performance target" that has been set by that school. The difference between the school's performance and its target is the "performance gap." Closing the gap should be the focus of action planning.

The ultimate goal is to have all students meeting or exceeding the standards.

Elementary English Language Arts:
Percent of Students Achieving the Standard or
Achieving the Standard with Honors N=50



Whenever you see the computer man icon (right), it signals that an internet-based graphing capacity is available. This site will allow you to graph the data for your school. (Available Summer, 1998)



Interpreting the Graph:

In this example, the greatest "performance gap" is with Reading Analysis and Interpretation. The school has met its goal in Conventions, and is closing "performance gaps" in the two other areas. The school should set new performance targets for those areas in which it has met its target and identify strategies for those students still not Achieving the Standard or Achieving the Standard with Honors.

(**Note:** Analyzing the performance of different groups of students will be very important in all areas but particularly where progress is being made but not for all students.) Work should be continued in these areas while building the program and teaching practices to strengthen students' abilities to analyze and interpret text.

What can be learned from tracking data over time (years)?

Tracking data over time will provide information about improvement in student performance.

Important Consideration: The biggest difficulty in looking at data across time is knowing if a change in numbers represents a true change in performance. The smaller the school the greater the challenge. First, in looking at programs, remember that the group of students in any one year is only a sample of all the students who have been and will be in that grade in the school. Since this group is a sample of all possible students, you must account for its size.

Secondly, it is important to understand that the smaller the size of any given group, the greater the chance that the data does not reflect the performance of all possible students in your school. Account for differences in numbers of students tested in two ways.

- All schools should represent the possible difference in performance with an “error bar.”
- Small school should also accumulate data over a number of years.

The graph below uses error bars to compare performance between years. An “error bar” indicates that the score could be as high as the top of the bar or as low as the bottom of the bar. An overlap of error bars indicates that there is no significant statistical difference between the two years.



Sample/group size is defined as the number of students included in the results. Sample/group size is usually represented by “n=” or “total number reported =”. This information can be found on the score reports and must be considered in your analysis.

Small schools must look at a distribution based on at least 20 students before they can be sure that judgments can be made about the program.

(More information on sample size can be found in Appendix C.18.)



Interpreting Improvement

In 1996, 30% of the students Achieved the Standard or Achieved the Standard with Honors in this school. In 1997, that proportion increased to 40%. Although there is a trend upwards, the overlap of the error bars for these two years indicates that there is not a statistically significant improvement in performance between 1996 and 1997. However, because the bars do *not* overlap between 1997 and 1998, it is likely that the performance has definitely improved between these two years. We can say that this school has shown significant improvement over the three-year period in this reporting area.

Accumulating Data Across Years

Small schools must also accumulate the scores across time. The table and explanation below provides an example of how data can be accumulated.

Reading for basic understanding	Little Evidence	Below the Standard	Nearly Achieved the Standard	Achieved the Standard	Achieved the Stand. w/Honors
1996 N=8			4	3	1
1997 N=10		2	2	4	2
1998 N=7		2	1	3	1
Total N=25		4	7	10	4
Distribution of Scores	0%	16%	28%	40%	16%
	16%		28%	56%	

Key feature that needs to be explored:

The data are accumulated to 25 students. It took three years to accumulate at least 20 students.

Questions you might ask about these data:

☛ Taking into account the characteristics of the students over all three years, what accounts for their performance?

☛ Are the students who score in the Below the Standard performance level receiving the same curriculum as the other students in the school? Do these students have similar characteristics (special education, LEP, etc.)?

☛ What are the characteristics of the students that score at each of the levels? Are they consistent throughout the three years?

How should the summary list of instructional needs be used?

The New Standards Reference Exam provides a Summary List of Instructional Needs on the bottom of the School Level Report. As analysis is completed and curriculum gap issues are being explored, this information should be used.

An Instructional Needs Roster by students is also provided. The instructional roster by students should be used *only* by school personnel in order to protect the privacy of individual students. (Family Educational Rights and Privacy Act - 34 CFR 99.35)

Instructional Needs Roster by student should be used by school personnel only.

Sample Summary List of Student Instructional Needs	
What was tested	Writing Students must demonstrate the ability to write effectively in a variety of formats for a variety of purposes, audiences, and contexts.
What students need The percentages to the right show the proportion of students in this group whose performance might be improved by each kind of activity.	[64%] creating cohesion between sentences and paragraphs [56%] organizing the essay to purpose, audience, and context [52%] including relevant information [44%] using specific details [44%] using appropriate vocabulary and sentence structure

How do scale scores relate to standards-based assessments?

Scale scores are created to allow us to look at achievement in the content areas of the standards-based tests. These scores also allow us to see how students perform over time. Mean scale scores are available for groups of students by school, supervisory union, and the state.

New Standards Reference Exam in Mathematics:

Scale scores are available for the mathematics content areas: Arithmetic & Number, Measurement & Geometry, Probability & Statistics, and Algebra/Functions. The Mathematics Content Area Scale runs from 0 to 100, and scores may be aggregated and averaged.

The performance on the total test is also reported in terms of a scaled score. The Total Mathematics Scale runs from 200 to 300. A Total Score of 150 is equivalent to Achieving the Overall Standard in Mathematics.

New Standards Reference Exam in English Language Arts:

The performances on total reading and total writing are reported in terms of scaled scores. The Total Reading and Total Writing Scales run from 200 to 300. A Total Score of 150 is equivalent to Achieving the Overall Standard in Reading or in Writing.

Vermont Science Assessment:

Scale scores are available for the science content areas, Inquiry, Systems, Space/Time/Matter, Living World, and Universe/Earth/Environment. In addition to the Mean Scaled Score, its standard deviation is also reported. This number is what is needed to draw the error bar.

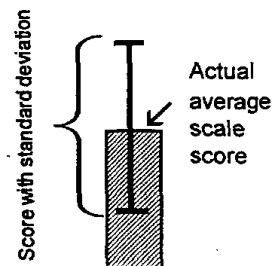
Scale scores are found on assessment reports. See appendix I.



TECHNICAL PRIMER

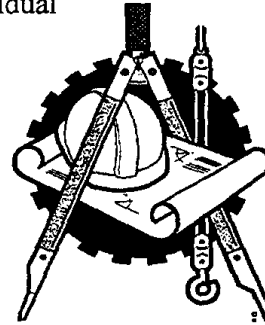
The scale scores provided with the standards-based assessments can be aggregated and averaged.

Reports for scale scores include the score's standard deviation. This number is what is needed to draw the error bar.



How can technical centers use the data from standards-based assessments?

- ✓ Instructors can analyze the “Instructional Needs Roster” for individual students to help build student programs.
- ✓ Centers can use the standards-based assessments as pre- and post-test program evaluation as a part of Federal Requirements.
- ✓ Centers can work with sending high schools to build action plans that address the needs of students in preparation for success in technical programs, and for ongoing success while in technical centers programs.
- ✓ Centers can aggregate performance data on post-tests to evaluate the effectiveness of their programs and make necessary adjustments.



Do's and Don'ts When Using Standards-Based Assessments Results

Do

- ◆ Track data across time whenever possible. Look for trends.
- ◆ Use the distribution of scores across performance levels in analysis.
- ◆ Disaggregate by different groups of students whenever possible.
- ◆ Use “instructional needs bullets” to help identify gaps in the curriculum or in instructional practices.
- ◆ Test hypothesis with other pieces of data or analysis, such as curriculum gap analysis, disaggregation of data, and surveys of instructional practices.

Don't

- ◆ Do not make high-stakes conclusions based upon a single assessment or with small sample sizes.
- ◆ Do not average percentages.
- ◆ Do not combine the percentages across reporting categories.
- ◆ Do not make decisions without collecting and analyzing other sources of data.
- ◆ Do not use these data as a tool for teacher evaluation. These data are about programs. The whole school is accountable for the performance of all its students.

Criterion-Referenced Assessments

What is a criterion-referenced test?

A **criteria-referenced test (CRT)** measures performance relative to specified content and skills.

How are criterion-referenced tests like standards-based tests?

Both assess performance in relation to specific content and skills. Standards-based assessments are one kind of criterion-referenced test. Other types of criterion-referenced tests measure content or skill objectives that may not be standards.

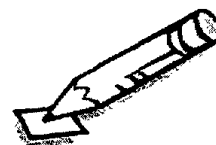
How are criterion-referenced tests reported?

- ✓ reported by concepts or skills. Individual student rosters are usually provided with the results.

Example *Skill Assessed: Student can detect errors in verb tense. (Jane Doe got 4 out of 5 items correct.)*
- ✓ reported by the percent of students who have mastered a skill or concept.

Example *Skill Assessed: Student can detect errors in verb tense. (Central High: 55% of students mastered the skill)*

Key ideas



Vermont's standards-based assessments are necessary to determine gaps in what standards-based programs are providing. However, they are administered only once a year or every other year and do not gather data from every grade level. Districts and schools should collect other student achievement data that addresses standards or specific areas in depth. Criterion-referenced tests are the type of assessment that can provide these data.

For action planning, the most useful criterion-referenced score is the percent of students who have mastered the concepts assessed!

Do's and Don'ts When Using Criterion-Referenced Tests

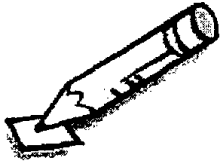
Do

- ◆ Track data across time whenever possible. Look for trends.
- ◆ Disaggregate by different groups of students whenever possible.
- ◆ Use the percent of mastery of a concept or skill.
- ◆ Use other results with these to help make decisions.
- ◆ Use for pre- and post-test situations.
- ◆ Test hypothesis with other pieces of data or analysis, such as curriculum gap analysis, disaggregation of data, surveys of instructional practices.

Don't

- ◆ Do not make high-stakes conclusions with small sample sizes.
- ◆ Do not average percentages.
- ◆ Do not combine the percentages across reporting categories.
- ◆ Do not make decisions without collecting and analyzing other sources of data.
- ◆ Do not use these data as a tool for teacher evaluation. These data are about programs. The whole school is accountable for the performance of all its students.

Key ideas



Uniform "point-in-time" assessments such as the standards-based and criterion-referenced tests show how students perform in a testing situation. To look at data that shows how students perform when working over time in a standards-based environment, look at data from portfolio assessments.

"We need on-going feedback, not just end-of-the-year assessments. These come too late and often are not keyed to the curriculum."
- Grant Wiggins

Portfolio Assessment

What is a Vermont assessment portfolio?

The Vermont Mathematics and Writing Portfolios are standards-based assessments. An assessment portfolio is a collection of work that is assessed based upon an explicit set of criteria aligned with Vermont Standards and defined with examples of student work. The portfolio should reflect the student's "best work."

In mathematics, students are provided with multiple experiences in complex problem solving and are then assessed on the following criteria in complex problem solving and communication:

- Approach and reasoning used
- Accuracy of the solution provided
- Connections or generalizations made
- Ways in which the students mathematically communicate their solutions using mathematical vocabulary, symbolism and representation.

In writing, students are provided with multiple experiences in writing for different purposes and audiences, and are assessed on:

- Purpose
- Voice/Tone
- Detail
- Organization
- Grammar, Usage and Mechanics

Criteria for the Mathematics and Writing Portfolios are found in Appendices C.19 through C.21.

How can portfolio assessment results be used?

In the classroom:

The most powerful use of portfolio assessment is in the classroom. Students and teachers should assess student work as they complete assignments. In this way students receive continuous feedback and know on which areas to focus in order to improve.

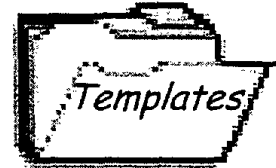
At the school level:

Portfolio results can only be used for action planning at the school

level if the school completes local scoring and has an agreement level of approximately 70%. This means at least two teachers have scored a large enough number of the school's portfolios and have assigned the same scores on 70% of the portfolios.

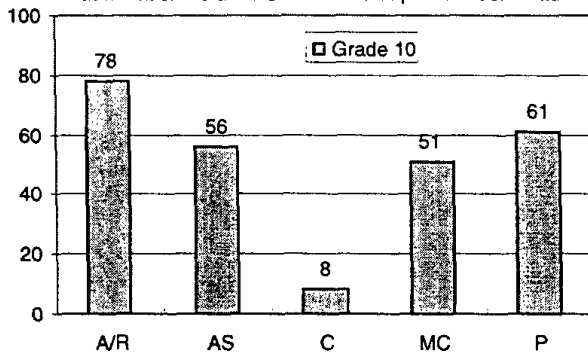
What are the ways portfolio assessment results can be used for action planning?

- ✓ Use the percent of students who scored three or higher on a criterion to identify strengths and weaknesses.
- ✓ Use the distribution of scores across the score points of each criterion to identify possible causes.
- ✓ Use portfolio results with other data from the NSRE and local assessments to confirm or complement other findings.



Use the Templates in Appendices C.3 and C.4 to help analyze results.

Percent of Students Who Scored Three or Higher On the Vermont Mathematics Portfolio Statewide Sample in 1996: Grade 10



Key feature that should be explored:

There is evidence that only 8% of the students are able to consistently apply and generalize their mathematics.

The data above indicate the percent of tenth grade students who scored a three or four on the Vermont 1996 mathematics portfolio sample. The criteria descriptions to the right of the graph provide meaning to these scores. These data help point to areas of strengths and weaknesses statewide but can be used in this same way locally. An overall strength in this sample is the ability to find effective approaches to solving complex problems. An area of weakness statewide is in the students' ability to apply and generalize their mathematics.

Criteria for Assessment

Descriptions represent a score of three on the scoring guide.

- Approach and Reasoning:** Students find effective strategies to solve problems and their reasoning supports the approach.
- Accuracy of the Solution:** Students solve problems accurately.
- Connections:** Students apply and generalize their mathematics.
- Mathematical Communication:** Students use accurate and appropriate mathematical language and representation (symbolic notation, charts, graphs, etc.).
- Presentation:** Students present their solution effectively.

Some additional questions that should be explored to support this finding:

- ☞ Do the results in New Standards Reference Exam Problem-Solving cluster support this finding?
- ☞ Does the curriculum include opportunities for students to solve complex problems that demand that they apply and generalize their mathematics? Is this opportunity uniformly implemented in the school?
- ☞ What can the distribution of the scores across each criterion tell us?

(See example on next page)

The second way to look at these data is by distribution across each criterion.

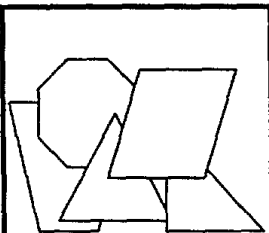
Distribution of Scores Across Criteria:

Mathematics Portfolio Sample

	1	2	3	4
Approach and Reasoning	17%	5%	70%	8%
Accuracy of Solution	5%	39%	56%	
Connections	30%	62%	8%	
Mathematical Communication	11%	38%	40%	11%
Presentation	2%	37%	61%	

Some key features about this distribution that should be explored:

- ✎ 17% of the students did not find an effective strategy to solve problems. This is very high. It could point to a curriculum inconsistency or a student motivation problem.
- ✎ 44% of the students did not consistently solve problems accurately.
- ✎ 92% of the students did not consistently make mathematical connections or generalizations.



✎ **Sample Eighth Grade Problem:**

The sum of the angles in polygons is unique. Determine the sum of the angles in a three-sided, four-sided, five sided, ten-sided, and n-sided polygon?

Number of Sides	Sum of Angles
3	180
4	360
5	540
n	(n-2) 180

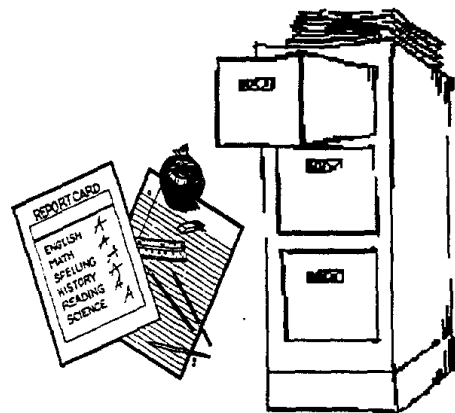
Do's and Don'ts When Using Portfolio Assessments

Do

Don't

- ◆ Use portfolio assessment for instructional decisions at the classroom level.
- ◆ Use local scoring portfolio results if scorer agreement is better than 70%.
- ◆ Use local portfolio scores with other sources of data.
- ◆ Analyze the percent of scores three or higher and the distribution of scores across the criterion.
- ◆ Use local portfolio scores to help identify strengths and weaknesses in the different aspects of writing and problem solving.

- ◆ Do not use portfolio scores as your sole source of data for action planning.
- ◆ Do not use local portfolio results without a scorer agreement rate of 70% or better.

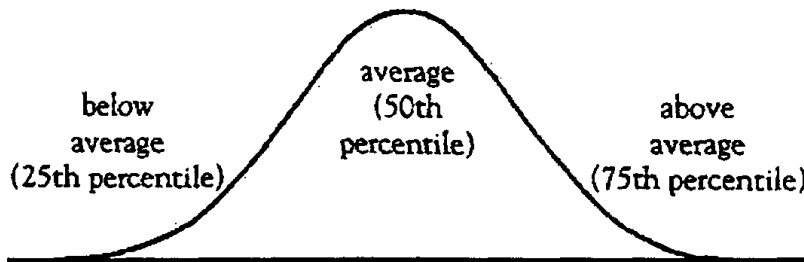


Norm-Referenced Assessments and Action Planning

What is a norm-referenced test and how is it different from a criterion-referenced test?

A **norm-referenced test (NRT)** compares the performance of an individual to that of a norm group. National norm-referenced tests select the norm group from a random sample stratified according to race, gender, socioeconomic status, etc. Scores of the norm group become the basis of comparison for those taking the NRT.

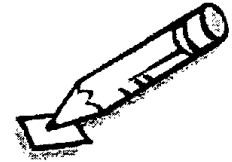
A norm-referenced test (NRT), therefore, compares the performance of one student to another or one group of students to another.



What kind of scores are available from norm-referenced tests?

Most test publishers provide criterion-referenced information through a skill or concept objective analysis. Many publishers also provide raw scores (the number of items correct for each subtest). In addition, publishers report several types of scores related to the norm groups. These include normal curve equivalents, scale scores, stanines, percentiles, and grade level equivalents. Examples of the uses and definitions of the types of scores follow.

Key ideas



Standards-based, criterion-referenced and portfolio assessment measures are developed to address standards. They can tell us how students are performing in relationship to the expectations in standards. To set this learning in a national context, use norm-referenced tests.

Both standards-based and norm-referenced data are important. Both sets of data allow us to portray Vermont's achievement of standards in a national context.

For an analysis of the alignment of norm-referenced tests to Vermont Standards in English Language Arts and Mathematics, see the *Norm Referenced Test Review* published by the Vermont Department of Education in November, 1997.

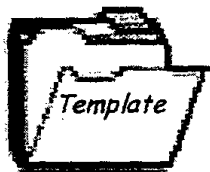


TECHNICAL
PRIMER

Definitions

Stanine: Stanines are nine-point scales that allow the conversion of percentile ranks into nine larger units.

Quartile: Quartiles divide scores into four equal groups. The lowest quartile consists of the lowest 25% of the scores of the norm group. The highest quartile represents the highest 25% of the scores of the norm group.



See template C.6

How can NRT focus in on high and low performing students?

The two types of reporting in norm-referenced results most useful for analyzing high and low performance are “stanines” and “quartiles.”

Stanines and quartiles can be used to quickly identify high and low performing students.

Stanines are based upon a nine-point scale with an expected distribution. The expected percent in the first and last three stanines is approximately 23% respectively. In Example A below, 32% of the students are in the first three stanines. This is a significantly large percentage of low performing students and should prompt the school to develop an action aimed specifically to improve reading comprehension skills for those students.

Quartiles are based upon percentile rankings. Based upon the norm group, it is expected that 25% of the students would perform in the lowest or highest percentile. Example B below provides an example of how this analysis can be used.

To look at the high and low performing group, it is NOT necessary to analyze both stanine and quartiles. Both provide similar data.

Example A:

Use of Stanine:

Reading Comprehension N = 50

Stanines

1 2 3 4 5 6 7 8 9

Expected distribution of scores based upon a normal curve

4%	7%	12%	17%	20%	17%	12%	7%	4%	
School A	7%	15%	10%	x	x	x	13%	8%	3%
32%						24%			

Key feature about this distribution that should be explored:

The distribution of the scores in the first three stanines is greater than expected.

What might be causing this distribution? (Hypothesis)

- ☛ Students in the bottom three stanines are not receiving the same curriculum as the other students.
- ☛ Students in these stanines have similar characteristics.

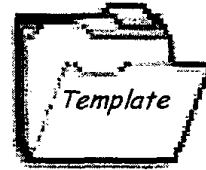
Example B:

Use of quartiles
Reading Comprehension N = 50

	First Quartile	Second Quartile	Third Quartile	Fourth Quartile
<i>Expected Distribution based upon norm group</i>	25%	25%	25%	25%
<i>Distribution of performance – School B</i>	35%	21%	20%	24%

Key feature about this distribution that should be explored:
The distribution of the scores in the first quartile is greater than expected.

Interpreting this data: (Hypothesis)
This distribution would point to the same questions as the distribution in Example A.



See template C.7

When using either stanines or quartiles for analysis, always compare the expected results with the actual results.

How can “normal curve equivalents” be used?

Normal Curve Equivalent scores (NCE) :

- can be aggregated and averaged
- can be compared from one subtest to another
- can be used to compare the performance of one group to another; male to female
- can show gains.

Compare one subtest to another:

A school may want to know if the same group of students is performing similarly in mathematics and English language arts. The NCE score is the only measurement from a norm-referenced test that will allow the comparison between two subtests. We can say that a reading NCE of 30 is 10 points lower than an NCE of 40. You may conclude from this discrepancy that the same groups of students are performing differently in the two subtests.

Aggregate and average:

You can determine the average NCE score for a grade. No other norm-referenced score can be treated arithmetically.

Compare different groups of students:

You can use the average NCE scores of one subgroup and compare them to another.

Demonstrate Gains:

Use NCE scores to show growth from one year to the next.



TECHNICAL PRIMER

Normal Curve Equivalent (NCE):
Identifies the students' position on a standard normal curve (score of 1 to 99).



TECHNICAL PRIMER

Using Norm-Referenced Test Results

Question: *What are areas of priority for action planning?*

Score(s) to use: Use stanines and then objective analysis.

Question: *How are our fourth graders performing in reading over the years?*

Score(s) to use: Use Normal Curve Equivalent Scores.

Question: *How are our fourth graders performing in reading v. math?*

Score(s) to use: Use Normal Curve Equivalent Scores.

Question: *How can we track this year's fourth graders from 1st grade through graduation?*

Score(s) to use: Use scale scores.

Question: *How can we compare performance of our boys v. girls, socioeconomic groups, etc.*

Score(s) to use: Normal Curve Equivalent Scores.

Question: *How can we report individual student performance to parents and students.*

Score(s) to use: Percentiles

Scale scores on norm-referenced tests:

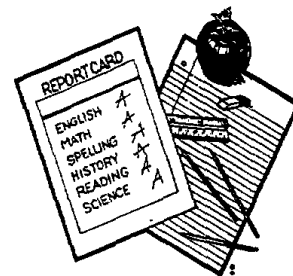
Scale scores are used to allow the comparison across different age or grade levels within the same test battery. Different test publishers use different names for their scale scores. For example, they are known as Growth Scale Values, Standard Scores, Converted Scores, Scaled Scores, and Developmental Scale Scores, to name a few. Norm-referenced test scale scores are arbitrary and vary from subtest to subtest even within the same test battery. This means that you cannot aggregate or average norm-referenced test scale scores.

Do's and Don't's When Using Norm-Referenced Results

Do

Don't

- | | |
|--|---|
| <ul style="list-style-type: none"> ◆ Check for alignment of the assessment to the expectations in the local or Vermont standards. (See Vermont Department of Education Norm-Reference Test Review.) ◆ Use percentiles to report to parents. ◆ Use stanines and quartiles to identify focus areas related to high and low performance. ◆ Use normal curve equivalents for comparing performance on tests. ◆ Use normal curve equivalents to compare groups, such as boys v. girls. ◆ Use scale scores for longitudinal studies within a content area. | <ul style="list-style-type: none"> ◆ Do not use percentiles for data analysis. ◆ Do not perform mathematical operations with percentiles, grade equivalents, stanines, or scale scores. ◆ Do not use grade level equivalents (except for very limited purposes). ◆ Do not compare one grade to another in a given year with anything but a scale score. |
|--|---|



Using Non-Test Data in Action Planning

What are examples of non-test data that might be used in action planning?

There are countless data sources that can be used in conjunction with test scores in action planning. Here are a few:

Curriculum alignment and implementation: Is the curriculum aligned with Vermont standards? Is it articulated across classrooms, grades and schools? Is the official curriculum the one being implemented? Are there different curriculums for different groups of students?

Teaching practices: Are standards-based teaching practices being implemented uniformly throughout the school?

Resource allocation: Is there sufficient time allotted to planning and professional development? Are teachers provided appropriate and adequate materials needed to teach a standards-based curriculum? Is proper technology available and in use? Are staff utilized effectively?

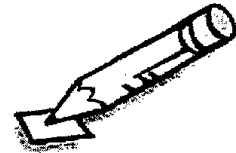
Attendance rate: What is the average number of days students are in school? How many students missed 10 days of school or more? Are there differences in attendance patterns at various grade levels? Are attendance patterns different for boys and girls?

Dropout rates: What percentage of students entering high school in their freshman year graduate four years later? How many students drop out as freshmen, sophomores, etc.? Is there a relationship between dropping out and test scores? Grades? Attendance? Socioeconomic status? Migration?

Grades: What is the relationship between test scores and grades? What is the distribution of grades for freshmen, sophomores, etc.? What is the distribution in math, science, classes, etc.?

Migration: What is the in-migration/out-migration rate? Is this rate different at different age or grade levels?

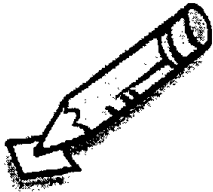
Key idea



Analysis of other factors relating to performance may lead to changes in policies, curriculum, teaching practices, or an analysis of school and community expectations.

Equity and Excellence

THE Key Idea



The goal is to reduce differences in performance between groups while increasing excellence for all.

How can data be used to assure equity and excellence?

“Excellence” can be measured by decreasing the gap between present achievement and the performance target.

However, excellence can only be achieved when we analyze results for equity and close the gaps between different groups of students.

It is possible, although not desirable, for overall performance to appear to increase while gaps widen between different groups of students.

Analyzing data for groups of students who differ on characteristics such as, but not limited to, gender, race, socioeconomic status, and disability will allow a school to assure equity.

“...Disaggregation is a practical, hands-on process that allows a school’s faculty to answer two critical questions: Effective at what? Effective for whom? It is not a problem solving process but a problem finding process.”

Lezotte and Jacoby, 1992

The goal is to reduce differences in performance between groups while increasing excellence for all.

“To keep Vermont’s democracy competitive and thriving, Vermont students must be afforded substantially equal access to a quality basic education. However, one of the strengths of Vermont’s education system lies in its rich diversity and the ability for each local school district to adapt its educational program to local needs and desires.”

- Vermont’s Equal Educational Opportunity Act

What are different ways to disaggregate data to assure equity?

In Vermont, the two most critical characteristics that need to be examined to assure equity in regard to student performance are gender and socioeconomic status. However, each school should look at its demographic patterns and disaggregate data accordingly. Here is a list of variables that may be considered depending upon your school's demographics:

- ➡ gender
- ➡ socio-economic status
- ➡ mobility
- ➡ race/ethnicity
- ➡ special education/disability
- ➡ English as a second language
- ➡ enrollment patterns in courses like physics or algebra (see templates C.8 to C.11)
- ➡ successful completion of specific courses (see templates C.12 to C.15)
- ➡ participation in programs such as gifted and talented classes and special education classes
- ➡ special education count
- ➡ number of students who score 3 or higher on Advanced Placement exams (see template C.16)
- ➡ risk behaviors
- ➡ other

A Vermont Case Study : A gender gap to consider

There was a 36-point difference in performance between male and female Vermont students on the College Board Mathematics SAT in 1996, with males outperforming females. However, according to the data released by the College Board, Vermont females take more higher-level mathematics courses than Vermont males. This gap should be closely studied by every Vermont high school. Some questions to ask...

- ☞ Does the same gap in performance exist at your high school?
- ☞ Does the same relationship between gap in performance and course taking patterns exist?

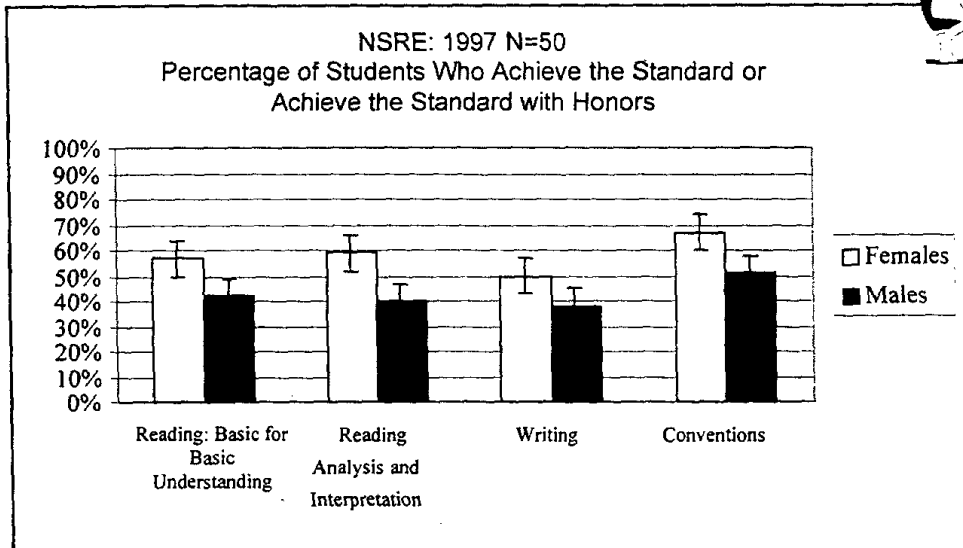
If the answer is yes, then a closer analysis of teaching practices should be conducted at the high school and the feeder schools.

Example A:

The data in this graph alerted a school to focus on gaps in the performance between males and females.



See template C.2



Example B

Using a model like the one below, one school tracked the increase of overall performance – excellence – while also tracking the gap between male and female performance – equity.

Standards-based example
Reading: Basic Understanding

Percent of students meeting or exceeding the standards

	Males Meeting or Exceeding Standards	Females Meeting or Exceeding Standards	Gender Gap	Overall Performance
1997-98 N _f =20 N _m = 20	35%	55%	20%	45%
1998-99 N _f =25 N _m = 28	43%	56%	13%	49%
1999-2000 N _f =22 N _m = 25	52%	59%	7%	55%

The school has decreased the gap between male and female performance while increasing the number for students Achieving the Standard or Achieving the Standard with Honors. In this case the school is striving for excellence through equity.

Tracking Forward and Backward

What does it mean to “track forward” or “track backward?”

Tracking backward can provide information about the program students have received before coming to your school. Tracking forward allows you to see how successfully your program prepared students for the next set of expectations.

What data should be tracked forward and how can this type of data be used in action planning?

There are 11 different configurations of schools in Vermont. Students’ success as they move from one school to the next is very important. A K-2 school should be very interested in how its students perform in third and fourth grade. A middle school should be very interested in how successful students are in high school. A high school should be interested in how students who attend technical centers perform. A high school should also be interested in students’ success in the workforce or in higher education.

What types of questions might be asked as students move from one school level to the next?

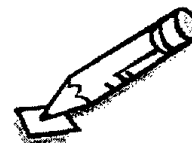
An example:

Middle School to High School:

- What percentage of the students from your school succeed in high school algebra, foreign language, science, other?
- What courses do the students in your school elect to take in high school?
- How do the students from your school perform on standardized assessments related to standards?
- What percentage of the students from your school make the honor roll during the first and second quarter of high school?

In action planning, each school should determine which measures of success it wants to track as students move to the next level and then track those variables over time.

Key ideas



Tracking forward to determine the success of graduates is very important at all grade levels.

Which indicators to “track forward” is a local decision.

Stay focused

This chapter has presented many ways that data can be looked at to help build action plans. As you go through the process of analyzing your data, it is important to remember to look at each set of data on its own in all possible ways. Remember, also, to look across different sets of data on the same standards or content area. Only by looking at all the data you have, in all the ways you can, will you be able to build strong plans for improving student performance.

Most educators agree that to see change requires a focus on data by those closest to the students. Classroom assessments must be built right into the instructional program to provide ongoing information to teachers and students as they progress toward established goals.

It will be the use of the combination of annual indicators for action plans and the meaningful use of an array of assessment information throughout the year that will drive improvement. Then you will be able to see the desired results of the implementation of your action plans.

