

SAMPLE APPROACHES FOR USING ASSESSMENT DATA AS PART OF A RESULTS-DRIVEN ACCOUNTABILITY SYSTEM

August 24, 2012

This document was prepared by the National Center on Educational Outcomes (NCEO) Core Team for consideration by OSEP. NCEO is supported through a Cooperative Agreement (#H326G110002) with the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education.

Overview of Sample Approaches

The National Center on Educational Outcomes (NCEO) was asked by the Office of Special Education Programs (OSEP) to work with a team to provide input on the use of assessment data that to review states on the performance results of their students with disabilities who receive special education services. NCEO established two groups to work on this task. A Core Team was established to identify measures of the academic performance results of students with disabilities who receive special education services. A Resource Group was established to support and serve as a “critical friend” to the Core Team to critique and support the refinement of the proposed measures. The Core Team first identified several framing considerations to guide its work. These included:

- The importance of public transparency and understandability
- The need for multiple measures
- The importance of creating incentives for values such as inclusion in the general assessment
- The use of measures to flag areas to look into more deeply
- The need for a plan to evaluate and improve measures that are used
- The need to have measures that are reliable
- The goal of no increased burden on states in terms of collecting any new data

With these in mind, the Core Team generated suggestions for OSEP’s consideration and a possible reporting format. The Core Team recognized that some measures that might be desired (for example, disaggregated assessment results by category of disability, placement, or demographic factors) are not possible at this time as these data are not currently collected. The Team also recognized the importance of definitions that are used (such as the pool of students reflected in the denominator for a measure), as well as decisions that are made about thresholds for categorizing states.

The Core Team’s work identified critical variables for OSEP to consider. The Core Team also noted that creating a formula to make decisions that prompt different levels of ratings and support, which is the goal of the OSEP accountability system, is a complex and challenging endeavor. To develop this type of rating-and-differentiated reviewing system, the Core Team strongly recommended that stakeholders, experts, and OSEP needed to be involved in the discussion about policy judgments and technical decisions on how these multiple measures are used for making decisions for reviewing purposes.

This document represents a beginning toward the next step – making policy judgments about how to look at the variables and displays suggested by the Core Team. NCEO generated sample approaches for OSEP to consider in its efforts to include assessment data as part of a results-driven accountability system. Two approaches are described here, each with a sample set of data for states and a beginning list of pros and cons:

1. Decision Matrix Approach
 - 1a. Includes state proficiency target
 - 1b. Does not include state proficiency target
 - 1c. Does not include state proficiency target; includes alternate assessment data
2. Decision-Making Steps Approach

The sample approaches included in this document give anonymous state data. State names have been removed and replaced with the identifier used in the report, *Using Assessment Data as Part of a Results-Driven Accountability System*. States are labeled consistently across the sample approaches included in this document.

Sample Approach 1a: Decision Matrix (Includes State Proficiency Target)

(Using Core Team Tables 1-3)

NCEO's Core Team suggested that the data in Tables 1-3 in its report, *Using Assessment Data as Part of a Results-Driven Accountability System*, were the primary data that should be considered in reviewing assessment results. When these data raise concerns, additional data should be considered before making decisions based on the review. Because it is difficult to process multiple sets of data simultaneously, a decision matrix approach can be helpful. Sample Approach 1a presents a possible decision making matrix that is formed by combining multiple sets of data.

The data for Sample Approach 1a include **five elements** that are derived from the data in Tables 1-3 of the Core Team report:

1. **Participation rate** of students with disabilities in the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 3)
2. **Improvement in percent proficient** or above of students with disabilities (see Core Team report, Tables 1 and 2, column 5)
3. **Gap in proficiency** between students with disabilities and students without disabilities on the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 4)
4. **Percent proficient or above** of students with disabilities on the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 2)
5. **Gap between actual and target proficiency rate** of students with disabilities on all assessments (see Core Team report, Table 3, columns 4 and 7)

Each element would result in a score for each state. For example, if the state is exceeding expectations on the given element it receives 1 point; if it is meeting expectations, it receives 0 points; if it is not meeting expectations, it receives a -1. (**There should, of course be a discussion around what is meant by “exceeding expectations,” “meeting expectations,” and “not meeting expectations.”**) The highest score a state could receive is a 5 and the lowest is a -5.¹

If a green, yellow, red model was used, a state receiving a 2 or higher might be considered “green flag”; a state receiving a score between +1 and -1 might receive a “yellow flag”; and a state receiving a -2 or lower might receive a “red flag.” This scoring would occur for both mathematics and reading (separately).

Possible ways to designate scores on the selected elements are provided below (see pages 7-10 for examples of calculations using this approach).

¹ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds for “exceeding expectations,” “meeting expectations,” and “not meeting expectations.”

Element 1: Participation rate on general reading and mathematics assessments

This indicates the percent of students with disabilities participating in the general assessment. To use this element, a decision needs to be made on what is an acceptable participation rate of students with disabilities in the general assessment (90% is a suggested target expectation, based on about 1% participating in an alternate assessment based on alternate achievement standards).

Possible Scoring: Above 90% receives a +1; 85-90% receives a 0; below the 85% threshold receives a -1.²

Element 2: Improvement in percent proficient

This element demonstrates the improvement that students with disabilities are making across years. This element can be used only if the standards assessed are the same across the years that are compared or a valid comparability measure is applied.

Possible Scoring: If a state has demonstrated an increase in the percent proficient on the general assessment of at least 0.5% or above, it receives a +1; if the state has neither increased nor decreased (i.e., 0.0% to 0.5%), it receives a 0; if the state has had a decrease in the percent of students with disabilities proficient and above, it receives a -1.³

Element 3: Gap in proficiency between students with disabilities and students without disabilities

This element will help shed light on whether the students with disabilities' instruction in a given state is being prioritized at a level that impacts their performance. A very large gap between the two might indicate a lack of priority on the use of evidence-based instructional interventions with students with disabilities.

Possible Scoring (two options): *Option 1* – If a state has demonstrated that the gap is 25% or less, it receives a +1, if gap is above 25% but less than 40%, it receives a 0, and if the gap is above 40%, the state receives a -1. *Option 2* - Rank the states based on their gap and divide them into thirds. The top third receives a +1; middle third receives a 0; bottom third receives a -1.⁴
(Note—Option 2 was used in the example.)

² Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds.

³ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds.

⁴ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds. They will also need to consider whether Element 3 should take into account situations where all students are performing poorly but the gap is small.

Element 4: Percent of students with disabilities proficient on general state assessment

The success of a state in terms of the results of students with disabilities should consider how many students with disabilities are proficient on the general assessment. It is also important that the rigor of the general assessment is considered when using this element.

Possible Scoring: To do this effectively, scoring needs to consider both the percent proficient for students with disabilities and the rigor of the general assessment (using the ranking via the NAEP equivalent score). One way to do this is to simply multiply the rank of the state on the NAEP scale by the percent proficient (Rank times % proficient). This product then could be used to rank states from 1 through 50 (50 is the number of states that have a NAEP equivalent score on reading). The top third would receive a +1; the middle third would receive a 0; the bottom third would receive a -1 (or a different threshold could be established and used to indicate the score a state would receive). To explain this further, please consider the following example:

For reading, the state with the most rigorous proficiency standard according to the NAEP equivalent score would receive a 50. If that state had 17.3% proficient on the general assessment, its score would be $50 \times 17.3 = 865$. The state with the second most rigorous standard would receive a rank of 49. If it had 28.3% proficient on its assessment, it would receive a score of $49 \times 28.3 = 1386.7$. After calculating a score for all of the states, they could be separated into thirds and given a score +1, 0, -1, as explained above.⁵

Element 5: Difference between state proficiency target and actual proficiency rate

The success of a state could be measured in terms of how well the state reaches its targets for student assessment results.

Possible Scoring. The actual proficiency rate in a particular year is subtracted from the target percent proficient for that year (so that a positive number indicates that the actual percent scoring proficient or above, is above the target for the percent scoring proficient or above). States that exceed their target (positive number) would receive a 1; states that meet their target would receive a 0; and states below target (negative number) would receive a -1.⁶

Final Step

The total score a state could receive using these elements in each content area (reading, math) ranges from +5 to -5. These two scores for each state would be put into a matrix (see below).⁷

⁵ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds.

⁶ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds.

⁷ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds for “green,” “yellow,” and “red.”

The combination of elements that the state receives would provide an indication of whether there additional data should be explored before possible Technical Assistance.

Decision Matrix (Option 1a: Includes State Proficiency Target)

MATH	READING		
	Green (2 to 5)	Yellow (1 to -1)	Red (-2 to -5)
Green (2 to 5)	Green Flag	Green Flag	Red Flag
Yellow (1 to -1)	Green Flag	Yellow Flag	Red Flag
Red (-2 to -5)	Red Flag	Red Flag	Red Flag

Some benefits (pros) and challenges (cons) to using this approach are:

Pros	Cons
<ul style="list-style-type: none"> Several variables that are difficult to look at separately are combined to provide a “portrait” of the state. It is easy to see where a state falls (assuming appropriate judgment criteria) across reading and math. 	<ul style="list-style-type: none"> This approach is more complex and may lack desired transparency. Each element has issues that need to be carefully considered: <ul style="list-style-type: none"> Element 1—a threshold for acceptable participation needs to be established and justified Element 2—change cannot be used any time the “proficient” cut-off score is changed; if a state has a high percent proficient, it will be difficult to show change (in these cases, as desired percent proficient might be used instead) Element 3—does not consider the level of performance Element 4—not all states have a NAEP equivalent score to standardize the gap information Element 5—not all states have a target score; in 2008-09 the target score was not aligned to ESEA (but it will be starting 2009-10)

Sample Approach 1a

(See page 10 for example of calculations for a single state—State S17)

Code	TA Need	Reading Score Total	Math Score Total	Reading Score Components					Math Score Components				
				Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	State Target	Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	State Target
S1	NA	NA	NA	-1	NA	0	NA	-1	-1	-1	NA	-1	
S2 ^a	NA	NA	NA	1	NA	-1	NA	NA	1	-1	-1	NA	NA
S3 ^a	Yellow Flag	-1	-1	-1	-1	1	0	NA	-1	0	0	NA	NA
S4	Red Flag	-2	-2	-1	0	1	-1	-1	-1	0	1	-1	-1
S5	Red Flag	-2	-2	0	0	-1	0	-1	0	0	-1	0	-1
S6	NA	NA	NA	1	NA	1	NA	-1	1	-1	-1	NA	1
S7	Red Flag	-2	0	0	-1	1	-1	-1	-1	-1	1	0	1
S8	Red Flag	-1	-2	1	-1	1	-1	-1	1	-1	0	-1	-1
S9	NA	NA	NA	0	-1	1	-1	NA	1	0	1	-1	NA
S10	Red Flag	-4	-4	-1	-1	-1	0	-1	-1	-1	-1	0	-1
S11	Green Flag	1	2	1	0	1	0	-1	1	0	1	1	-1
S12	NA	NA	NA	-1	NA	1	NA	-1	-1	-1	1	NA	-1
S13	Red Flag	-3	-3	0	-1	-1	0	-1	0	-1	-1	0	-1
S14	NA	NA	NA	1	-1	1	1	NA	1	-1	1	0	NA
S15	NA	NA	NA	1	-1	1	0	NA	1	-1	1	0	NA
S16	NA	NA	NA	1	1	-1	NA	-1	1	1	-1	NA	-1
S17 ^a	Red Flag	-3	-1	-1	-1	1	-1	-1	-1	0	1	0	-1
S18	NA	NA	NA	0	-1	1	1	NA	0	-1	1	0	NA
S19	NA	NA	NA	1	-1	1	1	NA	1	0	-1	1	NA
S20 ^a	NA	NA	NA	-1	-1	-1	0	NA	-1	-1	1	-1	NA
S21	NA	NA	NA	0	-1	1	1	NA	1	0	1	1	NA
S22	Yellow Flag	-1	-1	1	-1	1	-1	-1	1	-1	1	-1	-1
S23	Yellow Flag	-1	0	0	-1	0	1	-1	0	-1	1	1	-1
S24	NA	NA	NA	0	-1	1	0	NA	0	-1	0	-1	NA
S25	Yellow Flag	1	1	1	-1	1	1	-1	1	-1	1	1	-1

Code	TA Need	Reading Score Total	Math Score Total	Reading Score Components					Math Score Components				
				Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	State Target	Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	State Target
S26	NA	NA	NA	1	-1	-1	0	NA	1	-1	-1	0	NA
S27	NA	NA	NA	0	-1	-1	1	NA	0	-1	-1	1	NA
S28	Yellow Flag	1	1	1	-1	1	1	-1	1	-1	1	1	-1
S29	Yellow Flag	1	1	0	0	1	1	-1	0	0	1	1	-1
S30	NA	NA	NA	-1	-1	-1	-1	NA	-1	-1	-1	-1	NA
S31	NA	NA	NA	1	0	1	1	NA	1	0	1	1	NA
S32	NA	NA	NA	1	-1	-1	1	NA	1	0	-1	1	NA
S33	Red Flag	-3	-3	-1	-1	-1	-1	1	-1	-1	-1	-1	1
S34	Red Flag	-3	-3	-1	-1	-1	1	-1	-1	-1	-1	1	-1
S35	NA	NA	NA	1	-1	1	-1	NA	1	-1	1	-1	NA
S36	Yellow Flag	1	1	0	-1	1	0	1	0	0	1	-1	1
S37	Green Flag	3	4	1	-1	1	1	1	1	0	1	1	1
S38	NA	NA	NA	1	-1	1	1	NA	1	0	1	-1	NA
S39	Red Flag	-3	-2	0	-1	-1	0	-1	0	-1	0	0	-1
S40	NA	NA	NA	1	-1	-1	1	NA	1	-1	1	1	NA
S41	NA	NA	NA	1	-1	0	0	NA	1	-1	1	0	NA
S42	Yellow Flag	-1	0	0	-1	1	0	-1	0	0	1	0	-1
S43	Red Flag	-2	-1	1	-1	-1	-1	0	1	-1	-1	0	0
S44	Green Flag	2	4	1	0	1	1	-1	1	0	1	1	1
S45	NA	NA	NA	1	-1	1	0	NA	0	0	-1	1	NA
S46	Red Flag	-2	1	1	-1	-1	-1	NA	1	-1	1	0	NA
S47	Green Flag	2	0	0	-1	1	1	1	0	-1	1	1	-1
S48	Red Flag	-2	-4	0	-1	1	-1	-1	0	-1	-1	-1	-1
S49	Green Flag	2	3	1	-1	1	0	1	1	-1	1	1	1
S50 ^a	Red Flag	-3	-4	-1	-1	-1	-1	1	-1	-1	-1	0	-1
S51 ^a	NA	NA	NA	-1	0	1	0	NA	-1	-1	1	0	NA
S52	Red Flag	-2	-2	0	-1	1	-1	-1	0	-1	1	-1	-1
S53	Green Flag	2	-1	1	0	1	-1	1	1	0	0	-1	-1

Code	TA Need	Reading Score Total	Math Score Total	Reading Score Components					Math Score Components				
				Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	State Target	Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	State Target
S54	Green Flag	2	1	1	1	1	0	-1	1	1	1	-1	-1
S55	Yellow Flag	1	0	1	1	1	-1	-1	1	0	1	-1	-1
S56	NA	NA	NA	1	1	1	NA	1	0	1	1	NA	NA
S57	NA	NA	NA	1	-1	1	NA	NA	1	-1	1	NA	NA

^a State reported AA-MAS participation and performance data.

Example 1a. Sample State Profile for Decision Matrix—State S17**Red Flag**

Criterion	Reading		Math	
	Rate/Target/Percentage	Decision Matrix Score	Rate/Target/Percentage	Decision Matrix Score
Participation Rate	77.4	-1	77.4	-1
Improvement in Proficiency (2006//07-2008-09)	43.6	-1	39.3	0
Proficiency Gap	4.3	1	4.4	1
% Proficient times NAEP Difficulty	Bottom third	-1	Middle third	0
% Proficient	33.5		36.5	
NAEP Rank	19		20	
State Target	Bottom third	-1	Bottom third	-1
Target %Proficient	53.5		59.7	
Actual % Proficient	33.5		36.5	
Gap	-20.0		-21.4	
Score by Content Area		-3		-1

Sample Approach 1b: Decision Matrix (Without State Proficiency Target)

(Using Core Team Tables 1-2)

NCEO's Core Team suggested that the data in Tables 1-2 in its report, *Using Assessment Data as Part of a Results-Driven Accountability System*, were the primary data that should be considered in reviewing assessment results. When these data raise concerns, additional data should be considered before making decisions based on the review. Because it is difficult to process multiple sets of data simultaneously, a decision matrix approach can be helpful. Sample Approach 1b presents a possible decision making matrix that is formed by combining multiple sets of data.

The data for Sample Approach 1b include **four elements** that are based on the data in Tables 1-3 of the Core Team report:

1. **Participation rate** of students with disabilities in the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 3)
2. **Improvement in percent proficient** or above of students with disabilities (see Core Team report, Tables 1 and 2, column 5)
3. **Gap in proficiency** between students with disabilities and students without disabilities on the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 4)
4. **Percent proficient or above** of students with disabilities on the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 2)

Each element would result in a score for each state. For example, if the state is exceeding expectations on the given element it receives 1 point; if it is meeting expectations, it receives 0 points; if it is not meeting expectations, it receives a -1. (**There should, of course be a discussion around what is meant by “exceeding expectations,” “meeting expectations,” and “not meeting expectations.”**) The highest score a state could receive is a 5 and the lowest is a -5.⁸

If a green, yellow, red model was used, a state receiving a 2 or higher might be considered “green flag”; a state receiving a score between +1 and -1 might receive a “yellow flag”; and a state receiving a -2 or lower might receive a “red flag.” This scoring would occur for both mathematics and reading (separately).

Possible ways to designate scores on the selected elements are provided below (see pages 15-18 for examples of calculations using this approach).

⁸ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds for “exceeding expectations,” “meeting expectations,” and “not meeting expectations.”

Element 1: Participation rate on general reading and mathematics assessments

This indicates the percent of students with disabilities participating in the general assessment. To use this element, a decision needs to be made on what is an acceptable participation rate of students with disabilities in the general assessment (90% is a suggested target expectation, based on about 1% participating in an alternate assessment based on alternate achievement standards).

Possible Scoring: Above 90% receives a +1; 85-90% receives a 0; below the 85% threshold receives a -1.

Element 2: Improvement in percent proficient

This element demonstrates the improvement that students with disabilities are making across years. This element can be used only if the assessed standards are the same across the years that are compared.

Possible Scoring: If a state has demonstrated an increase in the percent proficient on the general assessment of at least 0.5% or above, it receives a +1; if the state has neither increased nor decreased (i.e., 0.0% to 0.5%), it receives a 0; if the state has had a decrease in the percent of students with disabilities proficient and above, it receives a -1.⁹

Element 3: Gap in proficiency between students with disabilities and students without disabilities

This element will help shed light on whether the students with disabilities in a given state are receiving the same amount of attention as are students without disabilities. A very large gap between the two might indicate a lack of priority in establishing proficient students with disabilities.

Possible Scoring (two options): *Option 1* – If a state has demonstrated that the gap is 25% or less, it receives a +1, if gap is above 25% but less than 40%, it receives a 0, and if the gap is above 40%, the state receives a -1. *Option 2* - Rank the states based on their gap and divide them into thirds. The top third receives a +1; middle third receives a 0; bottom third receives a -1.¹⁰
(Note – Option 2 was used in the example.)

⁹ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds.

¹⁰ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds. They will also need to consider whether Element 3 should take into account situations where all students are performing poorly but the gap is small.

Element 4: Percent students with disabilities proficient on general state assessment

The success of a state in terms of the results of students with disabilities should consider how many students with disabilities are proficient on the general assessment. It is also important that the rigor of the general assessment is considered when using this element.

Possible Scoring: To do this effectively, scoring needs to consider both the percent proficient for students with disabilities and the rigor of the general assessment (using the ranking via the NAEP equivalent score). One way to do this is to simply multiply the rank of the state on the NAEP scale by the percent proficient (Rank times % proficient). This product then could be used to rank states from 1 through 50 (50 is the number of states that have a NAEP equivalent score on reading). The top third would receive a +1; the middle third would receive a 0; the bottom third would receive a -1 (or a different threshold could be established and used to indicate the score a state would receive). To explain this further, please consider the following example:

For reading, the state with the most rigorous proficiency standard according to the NAEP equivalent score would receive a 50. If that state had 17.3% proficient on the general assessment, its score would be $50 \times 17.3 = 865$. The state with the second most rigorous standard would receive a rank of 49. If it had 28.3% proficient on its assessment, it would receive a score of $49 \times 28.3 = 1386.7$. After calculating a score for all of the states, they could be separated into thirds and given a score +1, 0, -1, as explained above.¹¹

Final Step

The total score a state could receive using these elements in each content area (reading, math) ranges from +4 to -4.¹² These two scores for each state would be put into a matrix (see below). The combination of elements that the state receives would provide an indication of whether there additional data should be explored before possible Technical Assistance.

Decision Matrix (Option 1b: Without State Proficiency Target)

MATH	READING		
	Green (2 to 4)	Yellow (1 to -1)	Red (-2 to -4)
Green (2 to 4)	Green Flag	Green Flag	Red Flag
Yellow (1 to -1)	Green Flag	Yellow Flag	Red Flag
Red (-2 to -4)	Red Flag	Red Flag	Red Flag

¹¹ Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds.

¹² Different thresholds could be established. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds for “green,” “yellow,” and “red.”

Some benefits (pros) and challenges (cons) to using this approach are:

Pros	Cons
<ul style="list-style-type: none"> • Several variables that are difficult to look at separately are combined. • It is easy to see where a state falls (assuming appropriate judgment criteria) across reading and math. 	<ul style="list-style-type: none"> • This approach is more complex and may lack desired transparency. • Without the state proficiency target data, students who participate in alternate assessments are not explicitly included. • Each element has issues that need to be carefully considered: <ul style="list-style-type: none"> Element 1—a threshold for acceptable participation needs to be established and justified Element 2—change cannot be used any time the “proficient” cut-off score is changed; if a state has a high percent proficient, it will be difficult to show change (in these cases, as desired percent proficient might be used instead) Element 3—does not consider the level of performance Element 4—not all states have a NAEP equivalent score to standardize the gap information

Sample Approach 1b

(See page 18 for example of calculations for a single state – State S17)

Code	TA Need	Reading Score Total	Math Score Total	Reading Score Components				Math Score Components			
				Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty
S1	NA	NA	NA	-1	NA	0	NA	-1	-1	-1	NA
S2 ^a	NA	NA	NA	1	NA	-1	NA	1	-1	-1	NA
S3 ^a	NA	-1	NA	-1	-1	1	0	-1	0	0	NA
S4	Yellow Flag	-1	-1	-1	0	1	-1	-1	0	1	-1
S5	Yellow Flag	-1	-1	0	0	-1	0	0	0	-1	0
S6	NA	NA	NA	1	NA	1	NA	1	-1	-1	NA
S7	Yellow Flag	-1	-1	0	-1	1	-1	-1	-1	1	0
S8	Yellow Flag	0	-1	1	-1	1	-1	1	-1	0	-1
S9	Yellow Flag	-1	1	0	-1	1	-1	1	0	1	-1
S10 ^a	Red Flag	-3	-3	-1	-1	-1	0	-1	-1	-1	0
S11	Green Flag	2	3	1	0	1	0	1	0	1	1
S12	NA	NA	NA	-1	NA	1	NA	-1	-1	1	NA
S13	Red Flag	-2	-2	0	-1	-1	0	0	-1	-1	0
S14	Green Flag	2	1	1	-1	1	1	1	-1	1	0
S15	Yellow Flag	1	1	1	-1	1	0	1	-1	1	0
S16	NA	NA	NA	1	1	-1	NA	1	1	-1	NA
S17 ^a	Red Flag	-2	0	-1	-1	1	-1	-1	0	1	0
S18	Yellow Flag	1	0	0	-1	1	1	0	-1	1	0
S19	Green Flag	2	1	1	-1	1	1	1	0	-1	1
S20	Red Flag	-3	-2	-1	-1	-1	0	-1	-1	1	-1
S21	Green Flag	1	3	0	-1	1	1	1	0	1	1
S22	Yellow Flag	0	0	1	-1	1	-1	1	-1	1	-1
S23	Yellow Flag	0	1	0	-1	0	1	0	-1	1	1
S24	Red Flag	0	-2	0	-1	1	0	0	-1	0	-1

Code	TA Need	Reading Score Total	Math Score Total	Reading Score Components				Math Score Components			
				Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty
S25	Green Flag	2	2	1	-1	1	1	1	-1	1	1
S26	Yellow Flag	-1	-1	1	-1	-1	0	1	-1	-1	0
S27	Yellow Flag	-1	-1	0	-1	-1	1	0	-1	-1	1
S28	Green Flag	2	2	1	-1	1	1	1	-1	1	1
S29	Yellow Flag	2	2	0	0	1	1	0	0	1	1
S30	Red Flag	-4	-4	-1	-1	-1	-1	-1	-1	-1	-1
S31	Green Flag	3	3	1	0	1	1	1	0	1	1
S32	Yellow Flag	0	1	1	-1	-1	1	1	0	-1	1
S33 ^a	Red Flag	-4	-4	-1	-1	-1	-1	-1	-1	-1	-1
S34 ^a	Red Flag	-2	-2	-1	-1	-1	1	-1	-1	-1	1
S35	Yellow Flag	0	0	1	-1	1	-1	1	-1	1	-1
S36	Yellow Flag	0	0	0	-1	1	0	0	0	1	-1
S37	Green Flag	2	3	1	-1	1	1	1	0	1	1
S38	Green Flag	2	1	1	-1	1	1	1	0	1	-1
S39	Red Flag	-2	-1	0	-1	-1	0	0	-1	0	0
S40	Green Flag	0	2	1	-1	-1	1	1	-1	1	1
S41	Yellow Flag	0	1	1	-1	0	0	1	-1	1	0
S42	Yellow Flag	0	1	0	-1	1	0	0	0	1	0
S43	Red Flag	-2	-1	1	-1	-1	-1	1	-1	-1	0
S44	Green Flag	3	3	1	0	1	1	1	0	1	1
S45	Yellow Flag	1	0	1	-1	1	0	0	0	-1	1
S46	Red Flag	-2	1	1	-1	-1	-1	1	-1	1	0
S47	Yellow Flag	1	1	0	-1	1	1	0	-1	1	1
S48	Red Flag	-1	-3	0	-1	1	-1	0	-1	-1	-1
S49	Green Flag	1	2	1	-1	1	0	1	-1	1	1
S50 ^a	Red Flag	-4	-3	-1	-1	-1	-1	-1	-1	-1	0
S51 ^a	Yellow Flag	0	-1	-1	0	1	0	-1	-1	1	0

Code	TA Need	Reading Score Total	Math Score Total	Reading Score Components				Math Score Components			
				Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty	Participation Rate	Improvement in Proficiency	Proficiency Gap	% Proficient *NAEP Difficulty
S52	Yellow Flag	-1	-1	0	-1	1	-1	0	-1	1	-1
S53	Green Flag	1	0	1	0	1	-1	1	0	0	-1
S54	Green Flag	3	2	1	1	1	0	1	1	1	-1
S55	Green Flag	2	1	1	1	1	-1	1	0	1	-1
S56	NA	NA	NA	1	1	1	NA	0	1	1	NA
S57	NA	NA	NA	1	-1	1	NA	1	-1	1	NA

^a State reported AA-MAS participation and performance data.

Example 1b. Sample State Profile for Decision Matrix—State S17**Red Flag**

Criterion	Reading		Math	
	Rate/Target/Percentage	Decision Matrix Score	Rate/Target/Percentage	Decision Matrix Score
Participation Rate	77.4	-1	77.4	-1
Improvement in Proficiency (2006//07-2008-09)	43.6	-1	39.3	0
Proficiency Gap	4.3	1	4.4	1
% Proficient times NAEP Difficulty	Bottom third	-1	Middle third	0
% Proficient	33.5		36.5	
NAEP Rank	19		20	
Score by Content Area		-2		0

Sample Approach 1c: Decision Matrix (Without State Proficiency Target)

(Using Core Team Tables 1-2 and additional alternate assessment data)

The data for Sample Approach 1c include **five elements** that are based on the data in Tables 1-2 of the Core Team report as well as additional alternate assessment performance data:

1. **Participation rate** of students with disabilities in the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 3)
2. **Improvement in percent proficient** or above of students with disabilities (see Core Team report, Tables 1 and 2, column 5)
3. **Gap in proficiency** between students with disabilities and students without disabilities on the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 4)
4. **Percent proficient or above** of students with disabilities on the general state reading and mathematics assessments (see Core Team report, Tables 1 and 2, column 2)
5. **Gap between percent of students proficient on regular assessment and percent of students proficient on the alternate assessment based on alternate achievement standards** of students with disabilities on all assessments (additional data will be needed to calculate this element.)

Detailed information about this approach is not included in this document due to the need for additional data to calculate Element 5, but a decision matrix approach similar to that presented in Examples 1a and 1b would be used.

Some benefits (pros) and challenges (cons) to using this approach are:

Pros	Cons
<ul style="list-style-type: none"> • Several variables that are difficult to look at separately are combined. • It is easy to see where a state falls (assuming appropriate judgment criteria) across reading and math. • Explicitly includes students who participate in the alternate assessment based on alternate achievement standards. 	<ul style="list-style-type: none"> • This approach is more complex and may lack desired transparency. • Each element has issues that need to be carefully considered: <ul style="list-style-type: none"> Element 1—a threshold for acceptable participation needs to be established and justified Element 2—change cannot be used any time the “proficient” cut-off score is changed; if a state has a high percent proficient, it will be difficult to show change (in these cases, as desired percent proficient might be used instead) Element 3—does not consider the level of performance Element 4—not all states have a NAEP equivalent score to standardize the gap information Element 5—additional data will be needed to calculate the gap between percent of students proficient on regular assessment and percent of students proficient on the alternate assessment based on alternate achievement standards

Sample Approach 2: Decision-Making Steps

(Using Data in Tables 1-6)

A decision-making approach, with specific steps to follow, could be used as a way to consider many of the variables in the six tables suggested by the NCEO Core Team. This sample approach includes five steps, each of which is intended to enable the review a specific area of concern relative to student performance. Thus, each step represents an area that may require additional data exploration or possibly technical assistance by OSEP. The decision-making approach is not intended to create a cumulative score through which technical assistance decisions are made, but rather to flag specific areas in which technical assistance may be needed (see examples for two states on pages 24-25).

Step 1: Participation in the General Assessment. Examine the participation rate in the state's reading and mathematics general state assessments. Determine whether the state has less than 90% of its students in the general state assessments. If the state has 90% or more of its students in the general assessment, move on to Step 3. If the state has less than 90% of its students in the general state assessments for either subject area, flag the state as possibly needing further exploration and possibly technical assistance.¹³ If participation is low on the general assessment this may indicate higher participation rates on the state alternate assessment(s). In addition to noting the need for further data and possibly technical assistance, it may be informative to record the participation rate on alternate assessments for this step.

Step 2: Gap in Performance between Students with Disabilities and Students without Disabilities. Examine the gap between students with and without disabilities. Determine whether the gap is greater than 30 percentage points.¹⁴ Flag these states, particularly if the difficulty of the assessment is very high or very low.

Step 3: Proficiency Rates on the General Assessment in Relation to the Difficulty of the State's Assessment. An examination of the state's proficiency rates on the general assessment must be considered in relation to the difficulty of the state's assessment. The difficulty of the state assessment can be calculated using the NAEP equivalent score. This step, then, involves a two-part process.

3A: Difficulty of State Assessment. Examine the difficulty of the state's reading and mathematics general state assessments as a contextual variable. Determine whether the state is in

¹³Different thresholds could be established. In states with an alternate assessment based on modified achievement standards (AA-MAS), a lower percentage of the students would participate in the general assessment. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds.

¹⁴Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds. They will also need to consider whether Step 2 should take into account situations where all students are performing poorly but the gap is small.

the bottom third, the middle third, or the top third of states in terms of difficulty, as defined by the NAEP equivalent score. Note the level, and move on to 3B. This information may provide insights when interpreting the rates of participation and proficiency observed within a state. NAEP scores represent the relative level of rigor for each state assessment against a common scale. The observed highest, lowest, and average NAEP scores for each assessed subject are reported below:

NAEP Score	Reading	Math
Low	190.5	212.0
High	248.0	277.5
Average	212.8	218.1

3B: Proficiency Rates. Examine the rates of proficient or above scores on the state's reading and mathematics general state assessments, and consider these in relation to the difficulty of the test as determined in 3A. The difficulty of the assessment should also be considered in relation to the participation rate in the general assessment (Step 1) and the gap between the performance of students with disabilities and students without disabilities (Step 2). If the proficient rates are high, but the other variables are low (difficulty, participation rate) or high (gap), flag the state as possibly needing further exploration and possibly technical assistance; also indicate the same need if the proficiency rate is low.

Review the recorded rate of participation in the state alternate assessments. If it was greater than 10% consider reviewing the percent proficient on the alternate here as well.¹⁵ While higher rates of proficiency on alternate assessments may suggest a higher overall rate of proficiency among assessed students in that state, policy goals emphasize participation in and proficiency on the general assessment. Thus, noting a higher frequency of participation on alternate assessments, including possibly higher rates of proficiency on that assessment, could suggest an additional need for further exploration of data and technical assistance.

Step 4: Check other Variables for Okay States. For those states not flagged via Steps 1-4, do a quick check on performance on the alternate assessment based on alternate achievement standards (AA-AAS), as well as a check on the difference between actual performance on all assessments compared to the target performance on all assessments. Note the relation between the percent proficient on the alternate assessment based on alternate achievement standards (AA-AAS) compared to the percent proficient on the general assessment as an element of the difficulty of the AA-AAS. Note large differences between actual and target performance, as well as the context of the rigor of the target for performance.

¹⁵ Different thresholds could be established. In states with an alternate assessment based on modified achievement standards (AA-MAS), a lower percentage of the students participate in the general assessment. Stakeholders, experts, and OSEP will need to be involved in determining appropriate thresholds.

Possible Decision-Making Recording Sheet

Use this table to flag areas that may need further exploration of data and possibly technical assistance based on review of state data when following the above steps. If a state meets the above mentioned criteria for technical assistance in either reading or mathematics mark a “TA” in the appropriate cell below.

State Name: _____	Reading	Mathematics
Step 1: General Assessment Participation		
Step 2: General Assessment Gap		
Step 3: General Assessment Proficiency		
Step 4: Alternate Participation and Proficiency		

Some benefits (pros) and challenges (cons) to using this approach are:

Pros	Cons
<ul style="list-style-type: none"> • Considers all of the reporting tables recommended by the Core Team. • Provides greater transparency than some other approaches. • Number of steps or their focus can be adjusted depending on shifts in policy imperatives. 	<ul style="list-style-type: none"> • Decision making approach needs to be completed one state at a time. • Approach is more subjective than some other approaches. • If different people look at different states, the comparability across states might be questionable.

Sample Approach 2—Sample Profiles for Two States

STATE NAME: S3

Regular Assessment			
Reading			
State % Proficient	All States' Avg. % Proficient	State % Participating	State Proficiency Gap between SWD & SWOD
13.9%	33.1%	60.7%	40.3%

Relative NAEP Difficulty in Reading	
State Relative NAEP Score	All States' Avg. NAEP Score
230.5	212.8

Alt. Assessment		
Reading		
State % Proficient	All States' Avg. % Proficient	States % Participating
21.9%	7.6%	35.6%

Regular Assessment			
Math			
State % Proficient	All States' Avg. % Proficient	State % Participating	State Proficiency Gap between SWD & SWOD
17.2%	31.0%	65.8%	38.5%

Relative NAEP Difficulty in Math	
State Relative NAEP Score	All States' Avg. NAEP Score
NA	218.1

Alt. Assessment		
Math		
State % Proficient	All States' Avg. % Proficient	States % Participating
12.7%	7.3%	29.8%

Descriptive Notes about Results:

 = Bottom Third of All States in terms of Percent of students proficient on the Regular Assessment

 = Middle Third of All States in terms of Percent of students proficient on the Regular Assessment

 = Top Third of All States in terms of Percent of students proficient on the Regular Assessment

NAEP Scores represent the relative level of rigor for each state assessment against a common scale. The observed highest, lowest, and average NAEP scores for each assessed subject are reported below:

	Reading	Math
Low	190.5	212
High	248	277.5
Average	212.8	218.1

STATE NAME: S24

Regular Assessment			
Reading			
State % Proficient	All States' Avg. % Proficient	State % Participating	State Proficiency Gap between SWD & SWOD
28.9%	33.1%	88.3%	0.7%

Relative NAEP Difficulty in Reading	
State Relative NAEP Score	All States' Avg. NAEP Score
225.5	212.8

Alt. Assessment		
Reading		
State % Proficient	All States' Avg. % Proficient	States % Participating
2.3%	7.6%	9.2%

Regular Assessment			
Math			
State % Proficient	All States' Avg. % Proficient	State % Participating	State Proficiency Gap between SWD & SWOD
38.3%	31.0%	88.7%	0.3%

Relative NAEP Difficulty in Math	
State Relative NAEP Score	All States' Avg. NAEP Score
232.5	218.1

Alt. Assessment		
Math		
State % Proficient	All States' Avg. % Proficient	States % Participating
3.0%	7.3%	9.2%

Descriptive Notes about the above results:

 = Bottom Third of All States in terms of Percent of students proficient on the Regular Assessment

 = Middle Third of All States in terms of Percent of students proficient on the Regular Assessment

 = Top Third of All States in terms of Percent of students proficient on the Regular Assessment

NAEP Scores represent the relative level of rigor for each state assessment against a common scale. The observed highest, lowest, and average NAEP scores for each assessed subject are reported below:

	Reading	Math
Low	190.5	212
High	248	277.5
Average	212.8	218.1