A Brief Guide to Selecting and Using Pre-Post Assessments

Sara Sanders, Ed.D.

The National Technical Assistance Center for the Education of Neglected or Delinquent Children and Youth

This document was developed by the National Technical Assistance Center for the Education of Neglected or Delinquent Children and Youth (NDTAC), which is funded by a contract awarded by the U.S. Department of Education to the American Institutes for Research (AIR) in Washington, D.C. The mission of NDTAC is to improve educational programming for youth who are neglected, delinquent, or at risk of academic failure. NDTAC's mandates are to provide information, resources, and direct technical assistance to States and those who support or provide education to youth who are neglected or delinquent, to develop a model and tools to assist States and providers with reporting data and evaluating their services, and to serve as a facilitator to increase information-sharing and peer-to-peer learning at the State and local levels. For additional information on NDTAC, visit https://neglected-delinquent.ed.gov/.

Suggested Citation

This guide is designed to assist States, agencies, and/or facilities who work with youth who are neglected, delinquent, or at-risk (N or D). The information in the guide will benefit those who are (a) interested in implementing pre-posttests, (b) in the process of identifying an appropriate pre-posttest, or (c) ready to evaluate current testing procedures.

Within this guide, you will find basic information on what pre-posttests are, why facilities should implement pre-posttest procedures, characteristics of different pre-posttests, and how agencies can use the information from their assessment practice. With an increasing number of different pre-posttests available, it is important that facilities understand how to identify the pre-posttest best suited for the needs of their unique youth population

What are pre-posttests?

Pre-posttests are academic achievement tests (e.g., math, reading, writing) designed to assess youth progress over a predetermined period of time. For youth who are N or D and are served in an alternative education setting, pre-posttesting may occur upon entry and exit of the facility. In these situations, the pretest can give facilities a baseline, or information about a youth's current academic level. The posttest, given when the youth exits the facility, then provides information about the youth's academic progress. This allows facilities to identify the effectiveness of their current academic programming as well as to share academic data with the youth's next placement.

Assessing pre-posttests upon entry and exit is not the only time assessments can be administered. For example, facilities who serve youth long term may choose to assess youth on a predetermined schedule (e.g., every 60 days, every 90 days). Regular posttesting allows facilities to continuously monitor youth progress and the effectiveness of the current educational programming. If a facility chooses to implement regular posttesting of academic progress, it is important to choose a pre-posttest with multiple forms of the tests so that youth are not given the same test repeatedly. More details about choosing an

appropriate pre-posttest are discussed later in this guide.

Why should we pre-posttest?

Youth who are N or D frequently display academic deficits. In fact, it is estimated that 30 percent of youth have an identified learning disability, and 48 percent enter facilities with academic skills below grade level.[1] There is an established correlation between a lack of education and involvement in the juvenile justice system and an understanding that education is one factor that helps prevent recidivism.[2] Collecting regular academic data using pre-posttests is one way to monitor the progress of youth with or at-risk for academic deficits, ensuring that youth receive appropriate supports and programming to benefit academically. Ultimately, the data from pre-posttests benefit *youth, teachers,* and *facilities and agencies.*

Youth benefit from pre-posttesting both upon entry to the facility and after they move to their next placement. Pretesting provides staff with current information on academic levels and functioning, thereby matching youth with appropriate educational programming. This approach allows youth to begin benefitting from educational services immediately and can prevent youth from becoming too frustrated (if they are functioning below grade level) or bored (if they score above grade level). Posttest scores allow youth to leave with updated academic records, which can ultimately support postsecondary goals and outcomes.

Pre-posttesting can also be valuable to *teachers* because it provides teachers with baseline information when beginning instruction. Although the pretest should not be the sole source of information when determining academic level and instructional methods, it can be a valuable source of information. In facilities in which assessments are given multiple times during placement, teachers can use theses scores as a way to monitor progress (progress monitoring) youth improvement and make any necessary changes. Finally, posttest scores enable teachers to measure overall progress when youth leave the facility.

[1] The National Technical Assistance Center for the Education of Neglected or Delinquent Children and Youth. (2015). *Fast facts.* https://neglected-delinquent.ed.gov/fast-facts/united-states

[2] The National Technical Assistance Center for the Education of Neglected or Delinquent Children and Youth. (2015). *Fast facts.* https://neglected-delinquent.ed.gov/fast-facts/united-states

Facilities and *administrators* can also benefit from information collected as result of pre-posttesting. First, scores can assist with the evaluation of current educational programming, allowing facilities to determine whether changes need to be made. Another way scores can be used is through a comparison of two different programs. It is important to note that if facilities want to compare two types of educational programming, the same pre-posttests should be used to evaluate the results. Scores can also improve the transition process of youth between facilities and programs by providing current educational data with incoming and outgoing youth. Finally, data collection focusing on educational progress is often mandated by both State and Federal agencies.

Why is mandated, yearly testing (e.g., National Assessment of Educational Progress [NAEP] assessments) not enough?

Although mandated, yearly testing (e.g., NAEP assessments) does provide some information on youth progress, its data is insufficient when assessing educational services provided to youth. First, the nature of restrictive settings means that youth stays vary greatly and are often less than one year, making yearly testing an inefficient assessment method. Second, because mandated testing occurs once a year, facilities cannot compare youth progress over different time periods. Finally, waiting more than one year to assess and modify programming is not the most responsive method of addressing youth deficits. As States have historically struggled to provide appropriate supports and services—with data indicating that youth frequently receive fewer overall hours of educational programming, fewer hours of math and science instruction, and are held to less rigorous standards—more frequent assessment procedures can help facilities evaluate current educational practices.

Every Student Succeeds Act and Mandated Testing

The Every Student Succeeds Act (ESSA), passed in 2016, improves upon previous No Child Left Behind legislation by allowing States and local education agencies (LEAs) to develop individual accountability systems rather than relying on a "one size fits all"

approach. ESSA encourages the collaboration of State policymakers, LEAs, and juvenile justice facilities to develop and implement an accountability system for the education programming within facilities. These individualized plans allow facilities to account for unique characteristics and context, thereby enabling facilities to effectively monitor progress and promote improvement efforts. Three priorities should be addressed within the accountability plan:

1. *Data collection and information should be shared between State and local education systems and juvenile justice facilities:* Educational services, especially in long-term juvenile justice facilities, may be the responsibility of multiple agencies. Therefore, developing a cohesive data collection plan agreed upon by all stakeholders is crucial to facilitating and supporting youth educational success. Sharing education data across systems as youth move between facilities and schools increases the likelihood that youth will receive the appropriate services and supports. Pre-posttest scores are part of this valuable data. Pre-posttesting supports a smooth transfer of education records and allows for the tracking of youth progress.

2. *There should be an accountability system that includes education services within juvenile justice facilities:* The accountability systems used by juvenile justice facilities can be the same as those used by States and LEAs, modified to accommodate the differences found in alternative settings or distinct systems that are aligned with goals of involved agencies. Pre-posttest data can be an important piece of this accountability system because continuous monitoring of youth progress can ensure that the education services provided are rigorous and appropriate.

3. *Outcome measures that hold education programs/schools accountable should be identified:* Recidivism is a primary outcome variable that is typically measured by agencies and facilities. However, when assessing educational programming, it can be helpful to assess other outcome measures such as educational gains, credential attainment, readiness for the workforce, and/or other postsecondary opportunities. Pre-posttest scores can be one way that facilities can measure educational gains.

Title I, Part D Legislation

Title I, Part D legislation requires programs who serve N and D youth and receive Federal funding to monitor the academic progress of youth to ensure that youth are receiving adequate educational services and supports. Pre-posttest data can be useful when evaluating programming.

Section 1431 (Program Evaluations) states:

(a) Each State agency or local education agency that conducts a program under subpart 1 or 2 shall evaluate the program, disaggregating the data on participation by gender, race, ethnicity, and age, not less than once every 3 years, to determine the program's impact on the ability of students:

> (1) to maintain and improve educational achievement;
>
> (2) to accrue school credits that meet State requirements for grade promotion and secondary school graduation;
>
> (3) to make the transition to a regular program or other education program operated by a local educational agency;
>
> (4) to complete secondary school (or secondary school equivalency requirements) and obtain employment after leaving the correctional facility or institution for neglected and delinquent children and youth; and
>
> (5) as appropriate, to participate in postsecondary education and job training programs.

(b) Exception: The disaggregation required under subsection (a) shall not be required in a case in which the number of students in a category is insufficient to yield statistically reliable information or the results would reveal personally identifiable information about an individual student.

(c) In conducting each evaluation under subsection (a), a SA or LEA shall use multiple and appropriate measures of student progress.

Choosing a Pre-Posttest

A number of different pre-posttests are available for use. It is important to consider many different factors when choosing a pre-posttest because they are not all created equal and are not always interchangeable.

Different pre-posttests are created for different purposes (e.g., measure reading comprehension versus reading oral fluency), different populations (e.g., elementary versus secondary), and different settings. It is important to know how to examine the characteristics of pre-posttests to determine which pre-posttest is right for your facility.

Reliability and Validity

The terms reliability and validity are frequently used when describing an assessment or test. When choosing a pre-posttest for your facility, it is important to have an understanding of both terms. The following section defines and describes both terms as well as provides an example of each.

Reliability: Reliability is the overall consistency of a measurement or tool. In terms of assessment and testing, it indicates the degree to which a test measures a construct accurately and consistently. Reliable tests will repeatedly produce consistent results.

One good example of a reliable instrument is a scale. If you weigh a suitcase at one point in time and then, without removing or adding any items, weigh the suitcase again, you should get a very similar (if not exactly the same) results. In this case, we can say the scale is a reliable measurement tool.

Similarly, if we were to give a student a math assessment in the morning and then another version of the assessment later in that afternoon, without providing any time of instruction in between, we would expect to see very similar scores. This result would indicate that the assessment is reliable.

Validity: Validity measures how well a test or tool fits the function for which it is used. In terms of assessment and testing, it measures the extent to which a test is measuring what it is supposed to measure. For example, if a math test included word problems that were above the reading level of the target audience, students may have difficulty understanding what the math problem was asking. Even if students had the mathematical abilities to solve the problem, they may not be able to show their understanding of the concept because they cannot read the math problem correctly. In this case, this assessment would *not* be valid because it is primarily assessing reading comprehension, not mathematical processes.

When choosing a pre-posttest, you want one that is both reliable and valid. One way to imagine what that looks like is through this visual of an archery target:

The first target helps us visualize what a reliable test looks like. The arrows are clustered together, showing consistent results. However, the arrows are not in the at the center of the target and, therefore, are not valid.

The second target would be considered valid but not reliable. Although the arrows seldom hit the center, if you "averaged" each of the arrows, you would hit the center (which is your goal). However, the arrows are spread out all over the target, meaning that the test is not reliable (you are hitting a different part of the target each time).

The third target shows us what it means to be neither reliable nor valid. It is not reliable because the arrows are spread out, hitting a different part of the target each time. It is also not valid. Even if you averaged the arrows, you would not be hitting the center of the target.

The fourth target is both reliable and valid. The arrows are clustered together, hitting the same part of the target over and over. It is also valid because the arrows are clustered around the center of the target.

Test publishers often report reliability and validity in different ways. The following table defines six ways publishers frequently report reliability and validity, explains how to interpret the scores, and what this means for your facility as you evaluate different pre-posttest options.
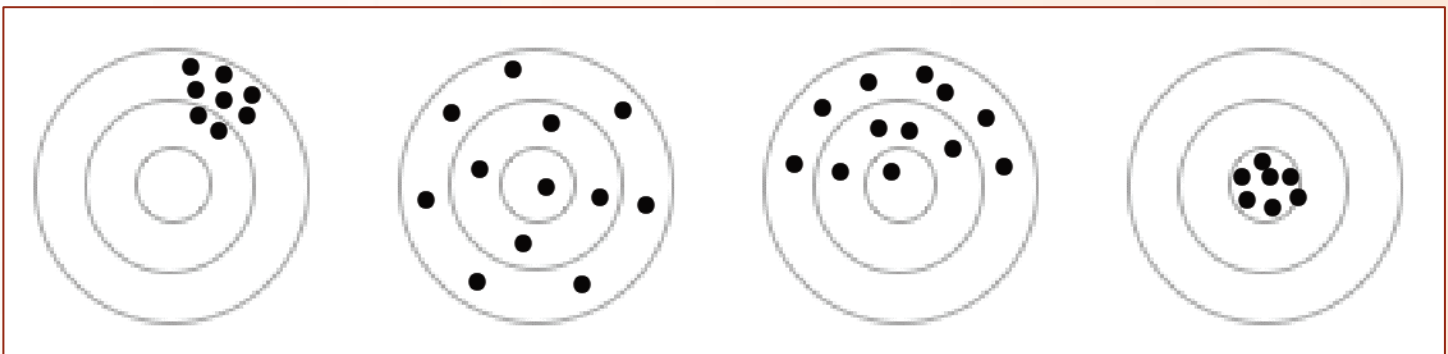
| 1. Reliable | 2. Valid | 3. Not Reliable or Valid | 4. Reliable and Valid |

Table 1. How Publishers Report Reliability and Validity

Types of Reliability and Validity	Definition	What does this mean for my facility?
Test-Retest Reliability	Test-retest reliability indicates whether a test is reliable over time. To assess test-retest reliability, students are given the same tests at different times, and then the scores are correlated.	Look for a test with a high correlation between tests. High correlation is an indicator that the test is reliable.
Alternate Form Reliability	Alternate form reliability shows whether two forms of a test are equivalent or reliable. To assess, students are given both forms of the test close together, and the scores are correlated.	Look for a test with a high correlation between tests. High correlation indicates that the tests are reliable, and any change/improvement seen in scores is due to improved skills and ability, not to a difference in test difficulty.
Content Validity	Content validity shows the extent to which a test includes a representative sample of items about a certain topic. Tests should reflect both the content/topic and cognitive processes and abilities and should be examined for inclusion of all desired topics.	Content validity is particularly relevant on academic achievement testing. It is often shown through a topic and process grid. Different topics and processing skills are listed in a table, and questions from the test are marked.
External Validity	External validity is the extent to which the results of one test from one group of students are relevant to other groups of students. Publishers will develop testing norms and establish validity and reliability based on a sample of students who took the test.	It is important to know information about the norm group (e.g., setting, disabilities). If the norm group is significantly different from your population, you may not be able to compare scores. Look for a diverse norm group or a norm group that shares characteristics that are similar to your population.
Construct Validity	Construct validity shows how well a test measures the content and skills (the "construct"). Examples of constructs include aggression, intelligence, and reading comprehension. This can be difficult to establish because constructs are not easily defined.	Look for a test with a high correlation with an already established test.

Types of Reliability and Validity	Definition	What does this mean for my facility?
	• Convergent validity is one type of construct validity. It shows the extent to which scores on a test are related to scores of a similarly existing test that measures the same or similar constructs. • For example, a reading pre-posttest may report convergent validity when compared with an established test (e.g., Woodcock Johnson IV reading subtest).	
Criterion Validity	Criterion validity shows how well a test reflects a set of current abilities or future abilities. It is how well the test relates to an outcome. There are two types of criterion validity: predictive and concurrent. • Predictive validity: Does the test accurately predict what it is designed to predict (future status)? • Concurrent validity: Do the tests scores on one measure relate to a criterion measured at the same time?	Criterion validity is particularly relevant on academic achievement testing. Look for a test with a reported high correlation. Some common examples of predictive validity include: • Performance tests for a job predict how well an applicant will perform on a job • GRE predicts how well a student will do in graduate school

Additional Pre-Posttest Characteristics to Consider

Once you have established that a test is reliable and valid, there are additional pre-posttest characteristics to consider:

Does the pre-posttest allow for multiple administrations?

Pre-posttests should be designed to be administered more than once a year. If a test is designed to be administered annually, it is not appropriate to be used for a pre-posttest. These types of tests are generally not sensitive enough to pick up smaller changes in youth skills and ability.

Does the pre-posttest include multiple forms?

Because pre-posttests should be given more than once per year, they should include multiple forms of the tests to give to the same youth. It is important that the same version of a test is not given for both pre-

posttests because any improvement may be the result of advance knowledge and/or recognition of the test questions rather than a change in ability. For example, if a youth is given the same reading passage and reading comprehension questions for both the pretest and posttest, the youth may remember parts of the passage or questions.

Many pre-posttests only have two forms (e.g., Form A and Form B), but some tests have more. For example, computer-adaptive testing (CAT) often allows the same youth to take a test more than two times. This may be an attractive option for facilities looking to administer pre-posttests on a set schedule (e.g., every 60 days) and in which youth will be assessed more than twice during their stay.

What is the test design?
Pre-posttests are designed for specific demographics (e.g., age, grade). This is important to note because it is important to choose a pre-posttest that accurately assesses the skill level of youth. For example, if your population has severe academic deficits, you want to make sure that the pre-posttest you select will still identify their current level of academic functioning.

It is also important to note whether the pre-posttest has been normed with N and D youth, alternative populations, and/or students with disabilities. Although it may be difficult to find a test that meets these criteria, it should be a consideration.

What is the test content?
The content assessed by the pre-posttest should align with the content your facility is looking to monitor (e.g., reading comprehension, writing skills, computations skills). Math and reading are some of the most frequently assessed academic skills and are often a part of Federal reporting requirements, so there are many different pre-posttests that assess these skills. However, different subskills can be measured in reading and math (e.g., reading comprehension versus reading fluency). Some pre-posttests may assess multiple content areas and skills, and some may have individual tests for different content areas. Make sure that the pre-posttest you have selected aligns with what you want to measure.

Other recommendations when considering test content include the following:

- Do not use IQ tests for pre-posttests.

- Do not use written or language assessments to measure reading comprehension.

- Be careful about the amount of reading that is required in math testing. If youth in your facility struggle with reading, choosing a math test with a lot of word problems may result in lower scores that do not accurately reflect ability.

What are the different test types?
The format in which a test is delivered is another important consideration. Format includes how youth are presented with the question *and* how they are required to respond. The individual characteristics of both your youth and staff population should be considered when identifying the best test format. If an incompatible format is chosen, it can severely impact youth scores, providing inaccurate information. For example, youth who cannot read fluently or write will struggle to complete a written test.

There are four common test formats to consider:

1. *Oral Administration:* This refers to both how the test is given and how the youth responds. Typically, a staff member would administer the test by reading the questions aloud. Then, depending on individual needs, youth would either answer the question orally or write it down. This is an ideal format for youth who struggle to read and write. However, it does require one-on-one administration, and staff should be trained on testing procedures prior to administering the test.

2. *Group/Individual:* This refers to the setting during the administration of the test. Testing can be completed in a group, either small (e.g., two to five youth) or large (e.g., the entire class). Group testing does not require as many staff members for administration, and multiple youth can be given the assessment at the same time. Individual testing is conducted in a one-on-one setting. Individual testing may be beneficial with youth who experience test anxiety or have identified disabilities. It does require more time from staff because only one youth can be tested at a time.

It is important to note that some pre-posttesting is specifically designed to be administered in a group or individual setting. If a pre-posttest is designed to

be given in an individual setting, that is how it should be administered. Some pre-posttests are designed for group administration, although, for these types of pre-posttests, you could administer the test individually if needed.

How your facility chooses to set up pre-posttesting may also influence whether you choose to administer tests in a group or individual setting. For example, if your facility decides to give pre-posttests at entry and exit, individual testing may be more appropriate. However, if your facility wants to conduct testing at set intervals, it may be more prudent to identify a pre-posttest that can be given in a group setting.

3. *Paper-Pencil Versus Computer*: This refers to how youth respond to the test. Some tests are designed to be administered using paper and pencil, while others are presented on a computer and the youth respond using the computer.

Computer testing can be advantageous for the following reasons: (a) multiple youth can take the assessment simultaneously, (b) it provides immediate scoring of the assessment, and (c) more and more youth (and staff) are comfortable with using technology. There are also considerations to keep in mind if your facility is interested in adapting computer testing: (a) make sure your facility has the appropriate software and operating systems that your identified pre-posttest requires to run, (b) ensure that your facility will be able to maintain the technology so that you can continue to administer the pre-posttests, and (c) be aware that some youth or staff may be unfamiliar or uncomfortable using the technology and may require additional training prior to administration of the pre-posttests.

4. *Computer-Adaptive Testing, or CAT*: CAT is a type of computer testing where the computer program adapts to the youth's present level of performance. As a youth gets answers correct, the questions become increasingly difficult. If a question is answered incorrectly, an easier question is presented next. CAT can help youth functioning below grade level to not become frustrated with overly difficult questions, while also ensuring that youth functioning at or above grade level are not bored with easy questions. CAT also typically results in fewer overall questions, while still accurately assessing current academic functioning. If you choose to use a CAT for your pre-posttest, be sure that the test has a large bank of questions so that youth are not getting the same questions over and over.

How are pre-posttests scored?

Pre-posttests are typically either norm or criterion referenced. This refers to how pre-posttest scores are evaluated. Many pre-posttests allow for evaluation using both methods, but it should be a factor considered by facilities when choosing a pre-posttest.

Norm-referenced pre-posttests compare the performance of an individual with the performance of a larger group. We call this group a "norm group." For example, if a 15-year-old girl is taking a reading comprehension assessment, her score would be compared to the scores of other 15-year-old girls who took the same assessment. It is important to note that the validity of this comparison depends on the norm group. If the normed group does not include students with disabilities or N and D youth, it may threaten the validity of the test when used in alternative settings.

In norm-referenced tests, the test score itself has no meaning until it is compared to the norm group. Two types of scores can then be generated.

1. *Grade Equivalent Scores*: To create grade equivalent scores, the raw scores from a large sample of students are collected and scaled, showing the average scores of students at set grade levels. These scores allow facilities to compare the scores of youth with what is expected of students at each grade.

2. *Percentile Scores*: A percentile score indicates how well a student did compared with a larger group. For example, if a student has a percentile score of 80, it means they did better than 80 percent of students who took the test and only 20 percent of students did better. Sometimes percentiles are created based on a norm group and other times are created using other students who took the test in a given time period (e.g., all the students who took the annual State reading assessment). Therefore, a youth's percentile scores can change based on how the percentiles are created.

Criterion-referenced tests compare youth achievement to a preset criterion. Often, criteria are expressed in proficiency or achievement levels (e.g., basic, proficient, advanced) and are set by test makers to evaluate knowledge of a specific content area or skill acquisition. These levels indicate what a youth is expected to know at certain grade levels or points in time. Criterion-referenced tests may provide more information than norm-referenced tests when tracking progress of N and D youth.

Considerations When Implementing Pre-Posttests

Integrating Pre-Posttests Into the Facilitywide Positive Behavioral Interventions and Supports Framework

As an increasing number of facilities adopt the facilitywide positive behavioral interventions and supports (FW-PBIS) framework, facilities may wonder how pre-posttesting fits into this framework. FW-PBIS is an evidence-based, data-driven tiered framework in which universal supports are delivered at Tier 1, additional supports and interventions are provided to some at Tier 2, and Tier 3 provides intensive supports to a few. Pre-posttesting can be integrated into this framework, with some facilities using pretest scores as one indicator when determining what tier of services youth need.

Academic testing information can be added to the youth profile, which may be useful when addressing other areas of youth development. Facilities frequently assess social-emotional and behavioral functioning at intake and throughout a youth's stay. Using pre-posttesting to add information about academic functioning to a youth's profile can be helpful to staff. For example, the behavior of a youth in math class may escalate more quickly if a youth has math deficits and is asked to perform academic tasks above current functioning levels. This full profile, which includes academic information, can then allow for a better, universal understanding of the youth in the facility.

Implementation Procedure

Pre-posttests should be implemented using similar procedures for consistency. To accomplish this, facilities should establish guidelines and procedures for when and how tests are administered to youth. Staff members responsible for conducting the testing should be trained on these implementation procedures, and fidelity of implementation should be collected (e.g., through a checklist).

Within the implementation procedures, it can be helpful to provide youth a reason for testing. Explaining that these assessments are designed to help teachers in the facility understand what a youth already knows can help establish buy-in. When youth take the posttest, staff should remind youth that their scores will be compared with the pretest to assess progress. When appropriate, sharing the results of the pre-posttests can be reinforcing for youth.

Facilities may also want to put off pre-testing incoming youth until youth have been at the facility for a few days or longer. Youth may need time to adjust to new environments and some may be angry about their situation and not in the right mind-set to take an assessment. In addition, it is important not to use assessments as a punishment for bad behaviors.

Accommodations

Accommodations are changes in the test format or the method in which it is administered. It does not change the content of the test or the cognitive processing expected by youth but, rather, alters either the environment of the test or the test format (e.g., orally answering questions). Common testing accommodations include:

- Testing individually or in a small group

- Reading written direction aloud

- Reading test questions aloud (including choices if applicable)

- Frequent breaks (may need breaks within one test or between different tests)

- Extended time

- Manipulatives (e.g., scratch paper, hundreds chart, calculator)

- Scribe (for written answers)

Facilities should have a system for checking for and implementing accommodations. Youth entering with an identified disability and an individualized

education program (IEP) may have testing accommodations listed in the document. By law, youth should be provided with these accommodations when testing. Remember, the purpose of the test is to allow youth to demonstrate their knowledge of the tested construct(s). For some youth, accommodations are necessary to demonstrate current academic functioning. Not providing these accommodations can invalidate pre-posttest scores.

Considerations When Interpreting Pre-Posttests

Finally, once a pre-posttest is selected, a number of other factors may impact test results and should be taken into consideration when evaluating youth scores:

1. *Anxiety*: Youth who are diagnosed with an anxiety disorder or who present with symptoms of anxiety may struggle when completing pre-posttests. For some youth, their anxiety may interfere with their ability to demonstrate what they know and lower their test scores. Providing an alternate quiet setting, extended time, and frequent breaks are some accommodations that facilities may want to consider offering.

2. *Fatigue*: Youth who are tired or hungry are unlikely to perform at their best, and it could have a negative impact on scores. This factor can be addressed by altering the circumstances in which the test is given. Breaking the tests into smaller sections or providing breaks can help alleviate fatigue. In addition, if possible, facilities should avoid testing right before lunch or dinner.

3. *Motivation*: If youth are not motivated to do well on a pre-posttest, scores are likely to be lower than a youth's actual current academic level. Encouraging youth to do their best and providing them with a reason for the testing are both ways to increase youth motivation. It is important to understand that if youth are not motivated to do their best, scores will be impacted. For example, if a youth enters a facility and is not motivated to do their best on the pretest, the score is likely to be lower than what the youth is capable of. Then if that same youth takes the posttest prior to exiting and is motivated to do their best, their score is likely to be a more accurate representation. However, the difference between the pre-posttest may look like the youth made more progress than

in reality because of the change in motivation. That is why considering youth motivation and effort is important when interpreting scores.

4. *Cultural Considerations*: Some test questions on pre-posttests may be culturally biased. For example, if a reading comprehension test is referencing a subway, youth from a rural Midwest town may lack the background knowledge to understand nuances about the subway system. However, a youth from an urban city may have a better understanding of this concept. Other questions may be biased based on culture, ethnicity, or income.

Pre-Posttest Examples

It is important to realize that you might not be able to find a test that fits all of your desired criteria. Few tests are specifically designed for N and D youth. However, by identifying what you want measured using your pre-posttest and taking into account the unique characteristics of the students in your facility, you can find the best suited test for pre- and posttesting youth in your facility. It is also important to reevaluate your current pre-posttests every three or four years in order to determine whether a new or different pre-posttest is more applicable to your facility.

The following are examples of pre-post assessments used by some Title I, Part D programs across the country. Remember, prior to choosing an assessment, your facility should carefully examine the pros and cons of each assessment and choose one that best fits the youth served and their needs. The inclusion of a pre-post assessment here should not be considered an endorsement by NDTAC or the U.S. Department of Education.

Test of Adult Basic Education (TABE)

The Test of Adult Basic Education (TABE) is designed to assess skills that adults need to succeed in the workplace. TABE can be used to track student progress using a pre-post format, identify student strengths and weaknesses, and guide the decision-making process regarding employment.

TABE is aligned with the College and Career Readiness Standards. Two forms are available,

making it feasible to use as a pre-post assessment. Subtests assess reading, math, and language skills. TABE is administered via paper and pencil or in an online format.

Renaissance STAR Assessments

Renaissance STAR Assessments are designed to assess K–12 student progress in four areas: early literacy, reading, reading in Spanish, and math. One or more of these assessments can be utilized to assess a student's current level and to monitor progress.

STAR Assessments are computer-adaptive tests (CATs), meaning that the test continually adjusts the level of questions based on a student's previous response. STAR Assessments can be given in a short amount of time, taking anywhere from 10 to 20 minutes to complete depending on the subject. Student scores can be assessed using criterion- and/or norm-referenced scores. Also, because STAR Assessments are CAT, they can be given more than two times to the same student, making it ideal for facilities that want to assess progress over multiple points in time.

Wide Range Achievement Test (WRAT)

The Wide Range Achievement Test (WRAT) assesses reading, spelling, and math skills of people ages 5–85. Reading, spelling, and math are assessed within three different subtests, allowing facilities to choose the subtests that are needed. Two different forms of each subtest can then be used to assess current academic functioning and assess progress over a period of time.

WRAT typically takes 30–45 minutes to complete and can be administered via paper and pencil or digitally. Norm-referenced scoring is used to evaluate student tests.

Comprehensive Adult Student Assessment System (CASAS)

The Comprehensive Adult Student Assessment System (CASAS) uses one system to assess a variety of learners, including students, English language learners, high school diploma candidates, and vocational students. Subtests include reading, listening, math, writing, and speaking.

Multiple testing formats are available, including paper and pencil, computer-delivered, and online testing. Following the administration of the pretest, CASAS

provides instructional resources for teachers based on student scores. It is recommended that posttests be given after a minimum of 40 instructional hours to best assess student progress.

Northwest Evaluation Association (NWEA): Measure of Academic Progress (MAP)

The Northwest Evaluation Association (NWEA) is a nonprofit research-based organization dedicated to providing effective assessments that allow teachers to measure growth and proficiency. They have developed the MAP (Measure of Academic Progress) Suite assessment system, which allows teachers to screen students, measure growth, project proficiency, and assess mastery. In addition, it supports teachers as they plan and differentiate instruction in the classroom.

MAP assessments are computer-adaptive tests (CATs) that adjust the level of difficulty based on student responses. These assessments (which include math, reading, language usage, and science) are able to assess students who are below, on, or above grade level. In addition, because MAP assessments are CATs, they can be given up to four times per year. Each subtest takes approximately 45 minutes.

MAP scoring uses the RIT scale. This score is not meant to be compared with other students' scores. Instead, scores are used to track progress over a period of time by measuring a student's score against their scores on previous tests. Score ranges based on grades are provided to identify whether a student is functioning below, at, or above grade level.

Woodcock-Johnson IV Test of Achievement

The Woodcock-Johnson IV (WJ-IV) Test of Achievements allow for the screening and progress monitoring of reading, writing, and mathematics skills across all age levels (K–12). The WJ-IV allows for the administration of specific subtests, the standard battery (11 subtests), or the extended battery (20 subtests) depending on the scope of the assessment. This flexibility allows for individualization based on facility testing goals and/or youth needs.

There are three forms of the WJ-IV achievement tests, which allows for progress monitoring over time. The WJ-IV also offers web-based scoring and reporting for easy analysis of student scores. Norm-referenced

scoring is used to evaluate student tests, with norms based on age and grade included.

HMH Math and Reading Inventory

The HMH Math Inventory (formally, the Scholastic Math Inventory) assesses math skills in students from kindergarten through Algebra II and is designed to prepare students to be career and college ready. It is an adaptive, group-administered test that takes approximately 40 minutes and can be given three to five times per year. The test also includes read-aloud audio in English and Spanish, positive student messaging integrated within the test, and the ability to skip questions. Performance-level reporting allows teachers to identify readiness levels of individual students as well as track student progress.

The HMH Reading Inventory (formally, the Scholastic Reading Inventory) assesses reading skills in students from kindergarten through college readiness. Like the math inventory, it is an adaptive, group-administered test that can be given three to five times per year. The Reading Inventory takes approximately 30 minutes and includes a mix of both literary and informational texts. The Reading Inventory uses the Lexile Framework to measure reading ability.

Pre-Posttest Checklist

Choosing a Pre-Posttest:

☐ Identify what topic you want to pre-posttest: _____

☐ Does the pre-posttest report reliability? What types of reliability have been reported?

_____

☐ Does the pre-posttest allow for multiple administrations? _____

☐ Does the pre-posttest include multiple forms? _____

☐ Does the pre-posttest provide information on the test design? _____

☐ Does the test content match what you want to assess? _____

☐ What test format(s) are you planning to use?

_____

_____

☐ How is your pre-posttest scored? _____

Implementing Pre-Posttests:

☐ What is your implementation procedure (e.g., when are they given, who gives them)

_____

_____

_____

☐ How will you identify necessary accommodations?

_____

_____

_____

☐ Are there additional considerations that need to be addressed (e.g., fatigue, anxiety, motivation)? How will they be addressed?

_____

_____

_____

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: Author.

American Youth Policy Forum. (2018). *Leveraging the Every Student Succeeds Act to improve educational services in juvenile justice facilities.* Washington, DC: Author.

Every Student Succeeds Act of 2015, Pub. L. 114-95 114 § Stat. 1177 (2015-2016).

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Boston, MA: Allyn & Bacon.

Jolivette, K., Kimball, K. A., Boden, L. J., & Sprague, J. R. (2016). The utility of a multi-tiered behavioral system in juvenile corrections: The positive behavior interventions and supports (PBIS) framework. *Corrections Today, 78*, 42–47.

McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators.* Thousand Oaks, CA: Corwin Press.

National Evaluation and Technical Assistance Center for the Education of Children and Youth Who Are Neglected, Delinquent, and At-Risk. (2006). *A brief guide to selecting and using pre-post assessments.* Washington, DC: Author. Retrieved from https://neglected-delinquent.ed.gov/sites/default/files/docs/guide_prepost.pdf

Salkind, N. J. (2017). *Tests & measurements for people who (think they) hate tests & measurement* (3rd ed.). Thousand Oaks, CA: SAGE.

The National Technical Assistance Center for the Education of Neglected or Delinquent Children and Youth. (2015). *Fast facts.* Retrieved from https://neglected-delinquent.ed.gov/fast-facts/united-states

Urbina, S. (2014). *Essentials of psychological testing* (2nd ed.). Hoboken, NJ: Wiley.

neglected-delinquent.ed.gov

The National Technical Assistance Center
for the Education of Neglected or Delinquent Children and Youth

American Institutes for Research
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835

For more information, please contact NDTAC at ndtac@air.org or visit our
Web site at https://neglected-delinquent.ed.gov/.

February 2019

Improving educational
programming for youth
who are neglected or delinquent

AIR
AMERICAN INSTITUTES FOR RESEARCH®