Selecting Benchmark Indicators and Tools

Lauren Supplee, PhD
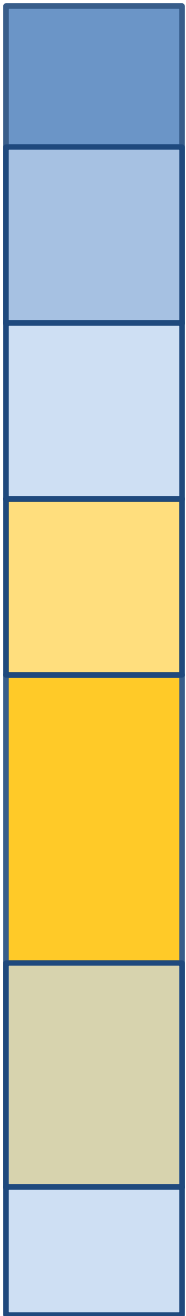*Administration for Children & Families*
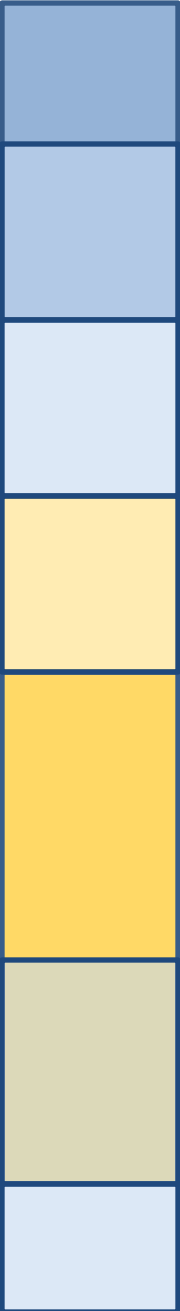
Jill Filene, MPH
James Bell Associates

Nancy Whitesell, PhD
Center for American Indian and Alaskan Native Health,
University of Colorado, Denver

DOHVE: <u>D</u>esign <u>O</u>ptions for Maternal, Infant, and Early Childhood <u>H</u>ome <u>V</u>isiting <u>E</u>valuation

- Working with US-DHHS to support the federal Maternal, Infant, and Early Childhood Home Visiting Program

- DOHVE:

  - ✓ Design options for a federal evaluation of evidence-based home visiting programs

  - ✓ Evaluation-related Technical Assistance (TA) for "promising approaches"

  - ✓ TA for grantees' continuous quality improvement, Management Information Systems (MIS), and benchmarks

DOHVE Evaluation TA Team

- James Bell Associates (JBA)
- MDRC
- Cincinnati Children's Hospital Medical Center & Every Child Succeeds

Goals of Session

- Provide overview of legislative requirements for measuring & reporting on benchmarks

- Issues to consider when identifying data sources & instruments to assess participant outcomes
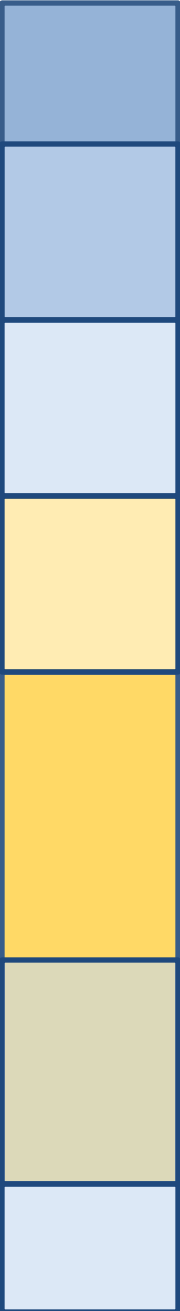
Benchmarks in Legislation

- Grantees will establish, subject to approval of HRSA/ACF, quantifiable, measureable benchmarks for demonstrating program results in improvements for eligible families participating in the program.

Benchmarks in Legislation (cont.)

- Six benchmark domains:
    - Improved maternal and newborn health

    - Prevention of child injuries, child abuse, neglect or maltreatment, and reductions of emergency department visits

    - Improvement in school readiness and achievement

    - Reduction in crime or domestic violence

    - Improvements in family economic self-sufficiency

    - Improvements in the coordination and referrals for other community resources and supports

Data Collection Sources

- To determine the best data source, consider:
  - *What source is likely to provide the most accurate information?*
  - *What source is the least costly or time consuming?*
  - *Will collecting information from a particular source pose an excessive burden on that person?*

- Direct data source
  - E.g., program participants, program partners

- Archival data sources
  - E.g., administrative data, case records

(Program Manager's Guide to Evaluation, 2010)

Data Collection Methods

- Use multiple data collection methods
  - Self report by client
    - Paper and pencil survey
    - Interview
    - Focus groups
  - Home visitor collection
    - Observation
  - Case record review
  - Abstraction of administrative data
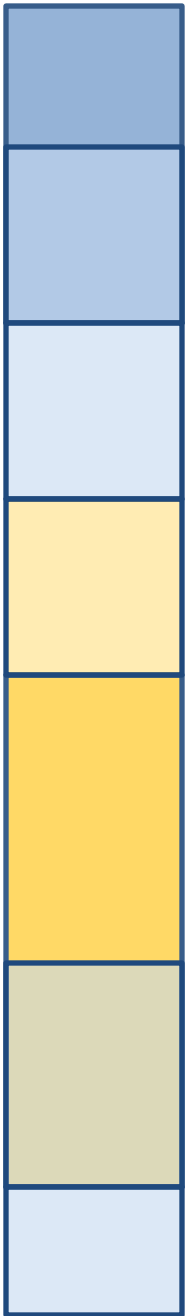
Considerations in Choosing Specific Measurement Tools

- Appropriateness
- Cost
- Time to administer
- Sensitivity
- Specificity
- Standardization of measure
- Reliability
- Validity
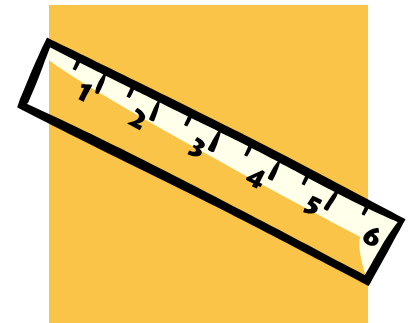
Appropriateness of Measurement Tools

- Is the tool appropriate for the children/family participating in your program?
  - Culturally appropriate
  - Developmentally/age appropriate
 - Language
  - Literacy levels

Costs & Time Associated with Measurement Tools

- How much will the tool cost?
  - Purchase of the instrument
  - Training people to use it
  - Data collectors
    - Internal staff activities or external staff/evaluator
- How much time will it take to use the tool?

Sensitivity & Specificity of Measurement Tools

- Sensitivity of the instrument
  - The degree to which an instrument correctly identifies those individuals who have a specific condition

- Specificity of the instrument
  - The degree to which an instrument correctly "screens out" those individuals who do not have a specific condition
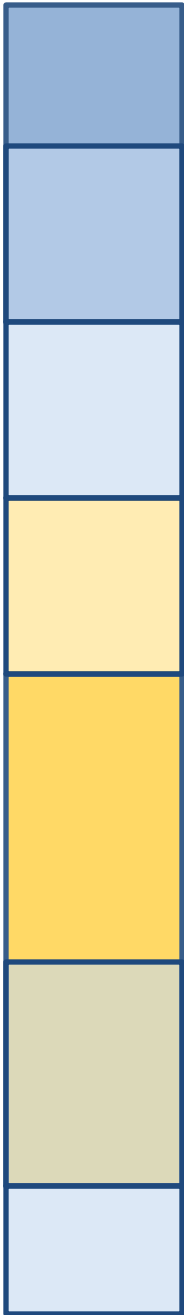
Standardization of Measurement Tools

- Is it a standardized instrument?
  - It's been used with a very large population, so that we know how children/parents typically score

Reliability of Measurement Tools

- Is the instrument reliable?
  - How consistently (i.e., similar results) does the instrument measure a construct- over time, by different people, or with different items?
    - Test-retest reliability*
    - Inter-rater reliability*
    - Split-half reliability
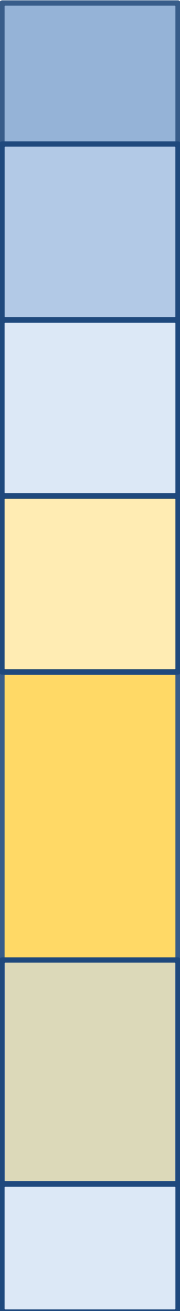   - Internal consistency

Validity of Measurement Tools

- Is the instrument valid?
  - Does it measure the concept it's supposed to measure?
    - Criterion-related validity
    - Content validity
    - Construct validity
    - Convergent validity
    - Discriminant validity
    - Face validity

Identifying Instruments Appropriate for AIAN Populations

- Review the literature
- Contact other researchers working in Indian Country (e.g., other Tribal home visiting grantee kickoff meeting participants)
- Get community input
- Pilot test

Identifying Instruments Appropriate for AIAN Populations (cont.)

- Seek technical assistance as needed

- Examples:
  - DOHVE Evaluation TA Team
  - Centers for American Indian and Alaska Native Health http://www.ucdenver.edu/academics/colleges/PublicHealth/research/centers/CAIANH/Pages/caianh.aspx
  - One Sky Center: The American Indian/Alaska Native National Resource Center for Substance Abuse and Mental Health Services http://www.oneskycenter.org/

Next Steps

Stay tuned for additional webinars, individualized TA, and other information from the Evaluation TA team on:
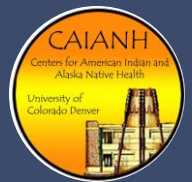
✓ Selecting an evaluation design

✓ Data systems

✓ CQI systems

✓ And more…

Selecting Benchmark Indicators and Tools

Working with Evaluation Partners to Find the Right Measures for Your Programs

Nancy Rumbaugh Whitesell, Ph.D.
Michelle Sarche, Ph.D.
Centers for American Indian and Alaska Native Health
Colorado School of Public Health, University of Colorado Denver

Acknowledgements

- Tribal partners and colleagues in the American Indian and Alaska Native Head Start Research Center

- Tribal partners and colleagues in the Centers for American Indian and Alaska Native Health

Goals of this Session

- Describe a process for **identifying potential measures**

- Describe a process for **reviewing and adapting selected measures**

- Describe a process for **evaluating the selected/adapted measures with your families**

Benchmarks

- The six benchmarks –
  - Health, abuse/neglect, school readiness, crime/domestic violence, economic self-sufficiency, coordination of referrals

- Should we think of these benchmarks the same in tribal communities as in other communities? Do they have the same meaning?
  - Example: School readiness
- Are there benchmarks important in your community that are missing?
  - Example: Cultural engagement

Benchmark Indicators

- Indicators – things we can observe that show progress on benchmarks

  - Example: Language development is a marker of school readiness

Benchmark Indicators

Are indicators of benchmarks in other communities good indicators of those same benchmarks in tribal communities?

Examples from the Peabody Picture Vocabulary Test (PPVT)

Benchmark Indicators

Some indicators may be straightforward and may work the same in tribal communities as in other communities

Benchmark Indicators

Some may need to be thought of in context

- Example: Language development in bilingual/Native-speaking families or in remote reservation communities.

Benchmark Indicators

Some indicators need to be considered through a cultural lens

- Example: Cultural norms may influence the way children view the pictures in the PPVT.

Benchmark Indicators

In addition to considering the appropriateness of the CONTENT of indicators, it will be important to consider the appropriateness of the PROCESS administering measures.

For example – PPVT administered by outside experts, trained on the measure.

- Adult stranger
- Often Non-Native
- Native from different community/tribe
- Protocol is to work hard to engage child through animated conversation – potential cultural mismatch?

Think carefully about how to adapt testing protocol to fit with cultural ways

Identifying Potential Measures

What resources are available to help you find the right measures?

1. Your evaluation partner
   - Expertise in measurement
   - Access to scientific literature

2. Your team
   - Expertise in your community and culture
   - Connections to other tribal communities

Identifying Potential Measures

Standardized measures – "standardized" for who?

Few measures of benchmarks have been shown to work (or even tested at all) with Native families

- Does "below average" or "above average" performance by Native children on a standardized measure – compared to "standard" scores – really mean anything?
    - Example: "Below average" could simply reflect consistent mismatches between Native contexts and test items (like *helicopter* and *fountain*)

Identifying Potential Measures

Need to think outside the box, work with evaluation partner to determine appropriate use of standardized measures

- Standardized measures may need to be used as is, but interpretation may need to reflect knowledge of the shortcomings

- Standardized measures may need to be adapted or supplemented

- Some combination of both approaches may be needed in your community – perhaps using standardized tests with both standard and alternative scoring

Identifying Potential Measures

Short-term assessment of long-term benchmarks

Depending on your program implementation design, you may need to evaluate early – before they can fully unfold.

- Example: School readiness in 2-year-olds

Work with evaluation partner to identify potential "mediators", based on existing research, that can be assessed in the short term

- Example: Early language development may be an important marker of the kind of developmental path that will result in school

Identifying Potential Measures

Feasibility and Cost

In choosing measures, you will also have to consider staff and financial realities.

- Availability of trained personnel (some measures require extensive training to administer)
- Cost of measures ("proprietary" measures can be expensive, need to budget for these
- Time to administer measures (direct assessments of children, observations of families can be lengthy)
- "Participant burden" (children have short attention spans, parents are busy)

Identifying Potential Measures

You will need to make some hard choices about what is most essential and most practical in your community – to get the most important information within the limits of the available time and money.

- Evaluation partners can help with this, using their experience with measuring outcomes, using assessment tools, and doing evaluation with many children and families.

- You can help with this, using your specific experience with children and families in your community,

Reviewing and Adapting Measures

Steps in the process:

- Community review

- Suggested adaptations

- Maintaining reliability and validity

- Pre-testing adaptations

Reviewing and Adapting Measures
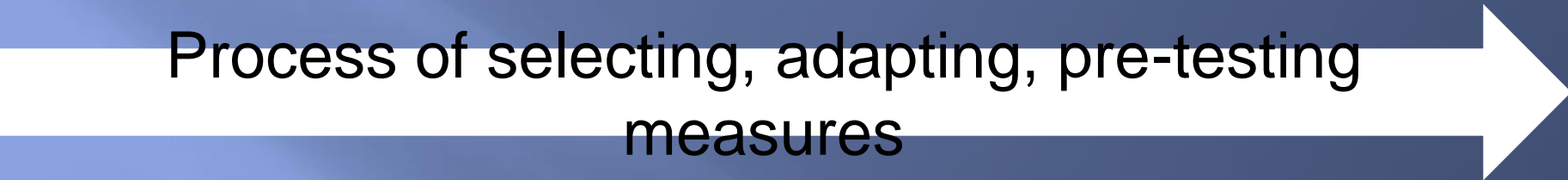
Community Review

- Item-by-item within measure-by-measure review
- Identifying items that seem incongruent with community context – can result in unreliability and/or invalidity (helicopter)
- Identifying items that are culturally inappropriate
    - Cultural taboos (accident)
    - Mismatch between cultural ways and assessment – e.g., evidence of parental warmth

Reviewing and Adapting Measures

Examples of community review

- **Team review** – project leaders sit down with evaluation partners to go through potential measures
  - Take the time to do this step carefully and thoroughly

- **Community Review** – find broader input where needed
  - Community advisory board, local teachers or health service providers who can help review measures for local applicability and appropriateness, focus groups

Process of selecting, adapting, pre-testing measures →

Evaluation Partners

Project Team

Community Advisors

Reviewing and Adapting Measures

Deciding when to adapt

- Standardized assessment tools – tread carefully (slide 18) – rely on evaluation partners

- Weigh pros and cons of each adaptation

- Consider supplementing standardized measures and benchmark outcomes to provide a fuller/more accurate picture of outcomes in your community

Reviewing and Adapting Measures

Considering ways to adapt

- Rewording of survey items, to include local language (without changing content)

- Changing content of survey items to reflect local practices

Maintaining measure integrity

Think about reliability and validity in the adaptation process:

- Reliability – measuring *something* well
  - Reliability without validity – eye color as a measure of intelligence
- Validity – measuring the *right* thing, the thing you wanted to measure
  - Validity without reliability – e.g., language assessment that includes words that don't fit culture or context

Reviewing and Adapting Measures

Pre-testing measures you intend to use

- Pilot testing of child assessments

- "Cognitive testing" of final survey items

- Final revisions

Evaluating Outcomes

FIRST, confirm effectiveness of measures:

- Evaluate both standard and adapted measures

- Tests of reliability and validity

- Evaluation partners can work with you to confirm that both of these important criteria are met

Evaluating Outcomes

THEN, interpret findings

- Ongoing, circular process of analysis, review, discussion

- Involves both evaluation of data and thoughtful consideration

- Consider how measures used may affect outcomes observed

- Community involvement in interpretation of data is important

Contact information:

nancy.whitesell@ucdenver.edu

michelle.sarche@ucdenver.edu

For more information…

Jill Filene, James Bell Associates (JBA)
filene@jbassoc.com

Lauren Supplee, Administration for Children and Families
lauren.supplee@acf.hhs.gov

Carol Gage, Administration for Children and Families
carol.gage@acf.hhs.gov

Carlos Cano, Health Resources and Services
Administration
ccano@hrsa.hhs.gov

Questions and Comments