

# Process Evaluation: Is it Working?

Validity and Accountability: Planning Ahead  
*July 2012*

Kenneth Hinson, PhD  
Nancy Carey, PhD  
Paul F. Parker, PhD



**Prince William County**

**PUBLIC SCHOOLS**

*Providing A World-Class Education*

# Agenda

- **Process Evaluation**
- **Virginia Tiered System of Supports: Benchmarks**
  - **Ways to Evaluate the Processes Involved in Attaining Benchmarks**
  - **Final Comments on Process Evaluation**
- **Validity**
  - **Aspects of Validity**
  - **Final Comments on Validity**
- **Activity on Developing a Plan for Evaluating the Implementation of a Division's Tiered System of Supports**

# What Is Evaluation and What Is Process Evaluation?

- **Evaluation** – involves systematic inquiry to help improve programs, products, and personnel, as well as the human actions associated with them.
- **Process Evaluation** – focuses on early stages of implementation, which involve activities such as:
  - Defining goals and objectives
  - Assigning roles and responsibilities
  - Conducting training
  - Monitoring efforts
  - Collecting feedback on how things are going for stakeholders that manage the program, receive benefits from the program, or have a vested interest in the program.

# Virginia's Tiered System of Supports (TSS) Benchmarks

## **To what extent is professional development that supports TSS provided across administrative, operational, instructional, and support levels?**

*(Leadership - Sustaining Stage)*

- **Are professional development plans available for administrative, operational, instructional, and support levels?**
- **What is the availability of professional development?**
- **Is professional development differentiated to meet unique needs of stakeholders?**
- **Who attended what training and when?**
- **Is professional development content relevant to stakeholders' needs?**

**To what extent is professional development that supports TSS provided across administrative, operational, instructional, and support levels?**  
*(Leadership Continued – Sustaining Stage)*

- **After receiving professional development, is it used to inform or adjust classroom instruction or adjust practices in administrative and operational environments?**
- **Should changes or improvements be made to professional development that supports TSS?**

# To what extent is professional development that supports TSS provided across administrative, operational, instructional, and support levels? (*Leadership Continued – Sustaining Stage*)

- **How do you collect data to answer key questions?**
  - Review of professional development plans, attendance logs, etc.
  - Observations of practices, meetings, etc.
  - Stakeholders' perceptions based on responses to items and questions used in interviews, focus groups, and/or surveys
    - Professional Development is relevant to my professional needs: strongly agree, agree, disagree, or strongly disagree.
    - Knowledge and skills gained from professional development are used all of the time, most of the time, some of the time, or none of the time to adjust instructional practices.
    - How can professional development be improved?

**To what extent does the Division conduct regular needs assessments and student and staff surveys to assess climate toward improved academic performance of students?**  
*(Organizational Culture and Structure – Emerging Stage)*

- **What stakeholders are surveyed?**
  - Administrators
  - Teachers
  - Students
  - Parents
- **What areas are covered on surveys?**
  - Quality of instructional programs, services, and opportunities
  - Teaching, learning, and working environment
  - Family involvement in decision-making
  - Teacher quality

**To what extent does the Division conduct regular needs assessments and student and staff surveys to assess climate toward improved academic performance of students?**  
*(Organizational Culture and Structure Continued – Emerging Stage)*

- **How do you collect data to answer key questions?**

**Students**

- Adults at my school care about me: strongly agree, agree, disagree, or strongly disagree.
- I feel safe from physical violence from other students at school: strongly agree, agree, disagree, or strongly disagree.
- My teachers expect all students to do their best: strongly agree, agree, disagree, or strongly disagree.
- I feel that I am getting a good education at my school: strongly agree, agree, disagree, or strongly disagree.
- What could be done to improve your school?

**To what extent does the Division conduct regular needs assessments and student and staff surveys to assess climate toward improved academic performance of students?**  
*(Organizational Culture and Structure Continued – Emerging Stage)*

- **How do you collect data to answer key questions?**

**Teachers**

- My school offers a wide variety of opportunities for parent/community involvement in decision-making: strongly agree, agree, disagree, or strongly disagree.
- I am satisfied with the instructional programs at my school: strongly agree, agree, disagree, or strongly disagree.
- I have a wide variety of delivery options for staff development available to me: strongly agree, agree, disagree, or strongly disagree.
- I am satisfied with the teaching and learning environment at my school: strongly agree, agree, disagree, or strongly disagree.
- What improvements do you feel should be made at your school?

**To what extent does the Division conduct regular needs assessments and student and staff surveys to assess climate toward improved academic performance of students?**  
*(Organizational Culture and Structure Continued – Emerging Stage)*

- **How do you collect data to answer key questions?**

**Parents**

- I have good relationships with staff at my child's school: strongly agree, agree, disagree, or strongly disagree.
- Overall, I am satisfied with my child's school environment : strongly agree, agree, disagree, or strongly disagree.
- My child has good teachers: strongly agree, agree, disagree, or strongly disagree
- Overall, my child is getting a good education: strongly agree, agree, disagree, or strongly disagree.
- There are many different ways that I can be involved in decision-making at my child's school: strongly agree, agree, disagree, or strongly disagree.
- I feel welcome at my child's school: strongly agree, agree, disagree, or strongly disagree.
- In what way could your child's school be improved?

**To what extent does the Division use a formal system of monitoring student progress that includes universal screening, formative assessment, and other diagnostic processes?**  
*(Monitoring Student Progress – Sustaining Stage)*

- **What is the formal system of monitoring student progress?**
  - Curriculum Based Measurement
    - AIMSweb
    - Dynamic Indicators of Basic Early Literacy Skills (DIBELS)
    - mClass Math
    - Basic Skill Builders
  - STAR Math, STAR Early Literacy, etc.
  - Other formative assessments
  - Combination of various tools/assessments

**To what extent does the Division use a formal system of monitoring student progress that includes universal screening, formative assessment, and other diagnostic processes?**  
*(Monitoring Student Progress)*

- **How do you know a formal system is being used?**
  - Reviewing reports
  - Observing practices
  - Getting stakeholders' perceptions based on responses to items and questions used in interviews, focus groups, and/or surveys
    - Are tools for monitoring progress adequate?
    - Are there too many tools?
    - Are tools standardized across the division?
    - What seems to be working and what is not working?
    - How can monitoring student progress be improved?

**To what extent are all decisions made using a divisionwide problem solving process which is consistently implemented with fidelity?**

*(Problem Solving – Sustaining Stage)*

- **Is there a divisionwide problem solving process available?**
  - PDSA Model
    - Plan – develop objectives, questions, predictions
    - Do – carry out the plan, document problems, analyze data
    - Study – continue analyzing data, compare objectives to predictions
    - Act – determine what changes should be made

**To what extent are all decisions made using a divisionwide problem solving process which is consistently implemented with fidelity?**

*(Problem Solving – Sustaining Stage)*

- **Is there a divisionwide problem solving process available?**
  - Problem Solving Process related to Response to Intervention
    - Define Problem – determine what makes up the problem.
    - Analyze Problem – analyze data to determine cause of problem.
    - Develop Plan – develop a plan to address the problem.
    - Implement Plan – put the plan in motion and monitor progress.
    - Evaluate Plan – determine effectiveness of plan.
  - Baldrige model of continuous improvement

## **Are all decisions made using a divisionwide problem solving process which is consistently implemented with fidelity?**

*(Problem Solving – Sustaining Stage)*

- **How do you know the process is working or being used?**
  - Observation of practices
  - Stakeholders' perceptions based on responses to items and questions used in interviews, focus groups, and/or surveys
    - A divisionwide problem solving process is used to make decisions all of the time, most of the time, some of the time, or none of the time.
    - A divisionwide problem solving process is implemented with fidelity: strongly agree, agree, disagree, or strongly disagree.
    - Quality decisions are made using the problem solving process: strongly agree, agree, disagree, or strongly disagree.
    - How can the problem solving process be improved?

# To what extent are curriculum/intervention resource decisions made using progress monitoring data and are they used with fidelity?

*(Curriculum – Sustaining Stage)*

- **How should curriculum/intervention resource decisions be made?**
- **Who makes the decisions:**
  - Central Office?
  - Teacher leaders?
  - Individual teachers?
  - Professional Learning Communities?
  - Grade-level teams?
  - Content area teams?
  - Combination of stakeholders?

# To what extent are curriculum/intervention resource decisions made using progress monitoring data and are they used with fidelity?

*(Curriculum Continuing – Sustaining Stage)*

- **How do you know decisions are being made in accordance with established guidelines/procedures?**
  - Observation of practices, meetings, etc.
  - Review of documents/data
  - Stakeholders' perceptions based on responses to items and questions used in interviews, focus groups, and/or surveys
    - Curriculum/resource decisions are made using progress monitoring data all of the time, most of the time, some of the time, or none of the time.
    - Curriculum/resource decisions are used with fidelity: strongly agree, agree, disagree, or strongly disagree.
    - Please discuss any changes that should be made on how curriculum/intervention resource decisions are made.

# To what extent are evidence-based instructional practices that are an appropriate match for the needs of the student used with fidelity?

*(Instruction – Sustaining Stage)*

- **Identify and/or confirm evidence-based instructional practices that are appropriate to meet students' needs.**
  - Proven as a result of robust objective methods
  - Source for confirmation: Institute of Education Sciences (IES), US Department of Education - **What Works Clearinghouse**

# To what extent are evidence-based instructional practices that are an appropriate match for the needs of the student used with fidelity?

*(Instruction – Sustaining Stage)*

- **How do you know practices are evidence-based and appropriate?**
  - Review of documents
  - Observations of practices, meetings, etc.
  - Stakeholders' perceptions based on responses to items and questions used in interviews, focus groups, and/or surveys
    - Evidence-based instructional practices are used in accordance with established protocol or guidelines: strongly agree, agree, disagree, or strongly disagree.
    - Evidence-based instructional practices are used all of the time, most of the time, some of the time, or none of the time.

# To what extent do parents view themselves as partners and advocates in the School Division's Tiered System of Supports?

*(Family and Community Partnerships – Sustaining Stage)*

- How have parents been informed of the Division's TSS?
- Have parents been invited to play active roles in the TSS?
- How does the division/school view parents as partners and advocates?
- How do parents view themselves as partners and advocates?

# To what extent do parents view themselves as partners and advocates in the school division's Tiered System of Supports?

*(Family and Community Partnerships Continuing – Sustaining Stage)*

- **How do you know?**

- Observations of practices, meetings, etc.
- Stakeholders' perceptions based on responses to items and questions used in interviews, focus groups, and/or surveys
  - Mark all the ways you have been made aware of the Division's TSS: School Board meeting, newsletter, email, phone call, Division website, school website, back-to-school nights, parent-teacher meetings, etc.
  - I was invited to participate in the Division's TSS by: newsletter, email, phone call, back-to-school night, parent/teacher meeting, etc.
  - The role that I fulfill for the Division's TSS are (mark all that apply): classroom volunteer, tutor, support team member, child's advocate.

# To what extent do Division data demonstrate that the TSS model has led to improved outcomes for students according to multiple measures?

*(Evaluation Plan – Sustaining)*

- To say that the TSS has led to improved outcomes implies that TSS caused the improvement, which is difficult to show without being substantiated by robust objective methods.

# To what extent do Division data demonstrate that the TSS model has led to improved outcomes for students according to multiple measures?

*(Evaluation Plan Continued – Sustaining)*

- **Analyze data for indications or trends of improved outcomes.**
  - Frequently reviewed results from interim formative assessments
  - Annual results from summative assessments – SOL tests, SAT, etc.
  - In the past, we compared the current year’s cohort performance with cohorts in the previous one to two years.
  - With the state’s new growth model, comparisons of students’ current-year performance with their performance from the previous year can be made to determine relative academic growth.

# Final Comments on Process Evaluation

- Process Evaluation places emphasis on early implementation plans, processes, and procedures.
- Choose more important aspects of implementation to evaluate/monitor if there is a scarcity of resources.
- Although emphasis is placed on the early stages of implementation, there is a need to continue process evaluation to a lesser extent throughout the life of the project, program, or intervention.

# Validity

- ... the degree to which evidence and theory support the interpretation and use of test scores entailed by proposed uses of the test (AERA et al., 1999)
  - Validity is the most fundamental consideration for [assessment] development and score use (AERA et al., 1999).
  - Every [assessment] development activity should contribute validity evidence.
- ***How do you evaluate the validity of the interpretation and use of student performance on assessments?***
  - Messick (1995) described six aspects that serve as general validity criteria for all educational measurement.

# Aspects of Validity

1. **Content** – relevance and representativeness of content and technical quality
2. **Structural** – appraises the fidelity of an assessment's scoring structure to the structure of the construct domain
3. **Substantive** – degree to which theories about content and processing models explain observations of students within the context of the assessment

# Aspects of Validity

4. **Generalizability** – degree to which assessment performance maintains its meaning across various contexts
5. **External** – degree to which assessment performance is related to external measures of the same trait, similar traits and other traits
6. **Consequential** – focuses on the value implications of the interpretation of performance as a source for action

# 1. Content Aspect

- ***Question: Is assessment content relevant, or irrelevant, and representative of the subject area?***
- **Explore the processes through which assessment components are developed and evaluate their technical quality.**
- **Activities for exploring validity evidence**
  - Purpose
  - Assessment specifications
  - Development process (items, scoring rubric)
  - Expert reviews
    - Content reviews
    - Sensitivity reviews
    - Map/reverse map items on test form to an assessment blueprint
  - Pilot and field tests
  - Technical quality

## 2. Structural Aspect

- ***Question: Is the scoring model of the assessment consistent with known relationships inherent in the content and behavioral domains?***
- **Explore whether intended relationships associated with the trait of interest manifest in student observations.**
- **Activities for exploring validity evidence**
  - Correlation analyses
  - Dimensionality analyses

# 3. Substantive Aspect

- ***Question: Are the processes engaged by students when completing the assessment representative of, and relevant to, the trait of interest?***
- **Explore whether student responses are logically consistent with the cognitive processes required to perform in the theoretical domain.**
- **Activities for exploring validity evidence**
  - Behavioral Observations (e.g., Student Think Alouds, Review Scratch Paper)
  - Distracter Functioning
  - Confirmation of Item Difficulty Hierarchies
  - Person Fit

## 4. Generalizability Aspect

- ***Question: Does the trait of interest maintain its meaning and interpretability across a variety of samples, contexts, etc?***
- **Explore whether measures maintain their interpretability and usefulness across various contexts.**
- **Activities for exploring validity evidence**
  - Reliability studies
  - Confirm item invariance (e.g., explore DIF)
  - Differential prediction
  - Validity generalization

# 5. External Aspect

- ***Question: Are scores generated by the assessment consistent with theories about the trait of interest?***
- **Explore whether measures exhibit relationships that are consistent with the theoretical framework.**
- **Activities for exploring validity evidence**
  - Multitrait-multimethod matrices
  - Group comparisons
  - Changes of individuals
  - Cross-group item calibrations

# 6. Consequential Aspect

- ***Question: Are the social consequences of performance interpretation and use consistent with the goals and purpose of the program?***
- **Explore the consistency of uses of the measures and outcomes as they relate to your purpose.**
- **Activities for exploring validity evidence**
  - Standard setting method
  - Rater/Scorer qualifications
  - Quality of ratings
  - Classification studies
  - Criterion selection

# Final Comments on Validity

- Evaluating the validity of decisions about students extends beyond the context of any single assessment to include additional and equally important contexts such as alternative modes of assessment, frequency of assessment, appropriateness of assessment, etc.
- Validation is an iterative and continuous process.
- Various sources of validity evidence should be triangulated to support the decisions we make about our students.

# References

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed.). New York: American Council on Education.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.  
Expanded edition 1980, with Forward and Afteward by B. D. Wright.  
Chicago: University of Chicago Press.

Wolfe, E.W., & Smith, E.V., Jr. (2007). Instrument development tools and activities for measure validation using Rasch Models: Parts I & II. In Smith, E. V. & Smith, R. M. (Eds.), *Rasch measurement: Advanced specialized applications*. Maple Grove, MN: Journal of Applied Measurement Press.

# Virginia's Tiered System of Supports (TSS) Benchmarks

## *(Evaluation Plan Group Activity – 30 Minutes)*

- Using the handout showing the “Emerging Stage” of all benchmarks, please concentrate on the following benchmark from the Evaluation Plan.
  - ***The evaluation plan has multiple components and includes assessing both short and long-term impact of TSS implementation and impact.***
- Think about your School Division’s evaluation plan. Consider information and processes shared in both presentations to determine questions that could/should be raised to allow you to know that the benchmark is being met. If possible, in groups of about 5 to 7 persons each, discuss your questions.
- For about 10 minutes groups/individuals will be asked to share their results.

# Contact Information

## Prince William County Public Schools

Kenneth Hinson, PhD  
Supervisor of Program Evaluation  
Phone: 703.791.7277  
Fax: 703.791.7366  
Email: [hinsonke@pwcs.edu](mailto:hinsonke@pwcs.edu)

Paul Parker, PhD  
Supervisor of Testing  
Phone: 703.791.8308  
Fax: 703.791.7366  
Email: [parkerpf@pwcs.edu](mailto:parkerpf@pwcs.edu)

Nancy Carey, PhD  
Coordinator of Program Evaluation  
Phone: 703.791.8346  
Fax: 703.791.7366  
Email: [CareyNL@pwcs.edu](mailto:CareyNL@pwcs.edu)