

Balancing Utility & Research Criteria in 'Conditions for Learning' Measurement

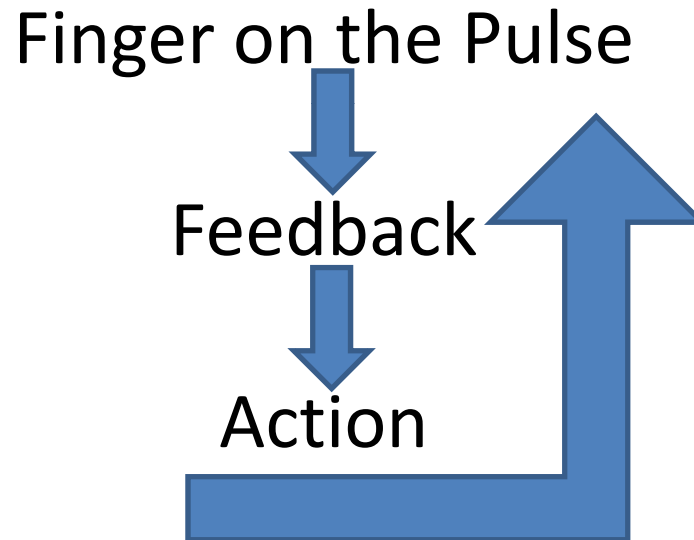
Presented by:
Ron Avi Astor Ph.D.
USC School of Social Work and Education
&
Rami Benbenishty Ph.D.
Bar Ilan University, Israel

Outline of presentation

- Introduction to monitoring and its strengths
- Balancing utility and research criteria
- Major considerations and dilemmas in planning monitoring
 - Who should be monitored
 - What should be Monitored
 - A ‘safety score’
- Reliability and validity in the context of monitoring
- Monitoring and evaluation
- Providing feedback

Monitoring Conditions for Learning

A social feedback system:



- Systematic
- Ongoing
- Overtime

Balancing Utility & Research Criteria in Monitoring 'Conditions for Learning'

- A false contrast:
 - Scientific research vs. ongoing monitoring
- Good monitoring is:
 - Based on best available scientific knowledge and
 - Makes a contribution to scientific knowledge

Balancing Utility & Research Criteria

- What we are looking for is strong and relevant evidence to support our understanding of reality and inform our decisions
- Strength of evidence and its utility depends on:
 - Asking relevant questions
 - Sample
 - Design
 - Measurement
 - Analysis
 - Appropriate interpretation
- **Research is always a balance between multiple demands and realistic constraints**

Inherit Strengths of State-Wide Monitoring of School Climate

- Relevance and utility
- Adaptive and responsive to changing needs
- The power of large samples
 - Detects even small variations, if they are consistent
 - Respects and amplifies within-group variations
 - Generalizability
 - Convincing
 - ‘Outliers’ – an important finding

Inherit Strengths of State-Wide Monitoring of School Climate (cont.)

- Replications as a major scientific tool
 - Across different groups and contexts
 - Over time
- Both similarities and differences in replications are important
 - Similar distributions, with reasonable variations
 - Different distributions between different contexts and cultural groups
 - Similar structures – different levels

A (very) few examples of the scientific contributions of state-wide monitoring

- The relationships between school climate and achievement, among different socio-cultural groups
- Gangs in California
- Clusters of drug use as they are associated with clusters of violence perpetration and victimization
- Foster children in the educational system
- Differences **within** ethnic groups
- Differential changes over time in districts that emphasized different health behaviors

Who should we monitor?

Current practice and funding:

- ‘Eligible Schools’; ‘Schools in needs for improvement’

We suggest that **all schools** be included

- Change of focus to more positive and promotional – all students could benefit, examples of good practice highlighted
- Put things in a wider perspectives
- ‘Eligible schools’ – A shifting ground

Students and more

S3 Invitational Priority:

- ***Family and Staff Inclusion*** *in Needs Assessments Measuring School Engagement*
- Teachers
- *Parents*
- *Principals (?)*

Multiple levels of monitoring

- National
- State
- (County)
- District
- School
- Grade-level, class (?)

‘Valid and Reliable Instruments’

An uneasy (and important) look at a mantra

- Who does not want reliable and valid instruments? Mom & Apple pie
- The false comfort of ‘Valid & Reliable’ measures
- Research-Based vs. ‘State-Grown’ instruments

Valid & Reliable Measures

- Research-based instruments come with a price;
- They need to be considered and their advantages and limitations weighted carefully.
 - Relevance to local needs and circumstances
 - Length/complexity
 - Developed for a different context
 - Level of detail not appropriate

Reliability considerations

- The continuum between Narrow- and Wide-Band instruments
- Where is the right balance between:
 - extensive coverage of a very small number of issues
 - An omnibus of ‘single-item’ topics
- **Internal consistency**- how important and when is it less important, or even meaningless?
- Sometimes a ‘behavior is a behavior’, important in and of itself, and not as an indicator of a latent construct (did you consider committing suicide in the last thirty days?)
- **Examine** reliability as you develop your monitoring system:
 - Test-retest on a grand scale
 - Compare reliabilities in different contexts

Validity Considerations

- Do we measure what we intended to? The importance of relevance
- Testing validity in the right context:
 - Cross validating with
 - Other instruments (within the monitoring system)
 - Other information (e.g., incident reports)
 - ‘Known Groups’ (e.g., gender, age)
- Consider ‘formal’ validation of ‘State-Grown’ instrument in a large enough pilot.

What to include in a survey?

The California Healthy Kids example

- The **CORE**, **Supplemental** & **Custom** Modules
- Core
 - **Demographics** (central for identifying subgroups)
 - Alcohol and Other Drug Use, Tobacco
 - School climate & connectedness
 - School, home and neighborhood assets
 - Violence
 - Safety (in school and risk behaviors)
 - Physical health

Supplemental Modules

- [Resilience Supplemental Module](#)
- [AOD \(Alcohol and Other Drugs\), Violence & Suicide Module](#)
- [Tobacco Module](#)
- [Physical Health & Nutrition Module](#)
- [Sexual Behavior Module](#)
- [District Afterschool Module \(DASM\)](#)
- [Military Connected School Module](#)
- [Gang Risk Awareness Module](#)
- [Service Learning Module](#)
- [Closing the Achievement Gap \(CTAG\) Module](#)

School Safety Score

Is it a good idea?

- Advantage – a simple way of tracking
- Disadvantage – a simple way of tracking
 - Complex realities cannot be encapsulated
 - Does not direct interventions
 - Encourages mostly meaningless ‘League Tables’
 - Hard to interpret changes
- When presented, show not only formula but also ‘raw scores’ that generated the score; they will be useful for policy and practice

Monitoring- Not only surveys

- Use, improve and create administrative data
 - Attendance, Truancy
 - Major Incidence Report
 - Police Data
 - Emergency room reports
- Mapping of schools and routes to and from schools

Design: Can you evaluate programs with monitoring?

- Clinical trials- A gold standard in evaluation of programs- Do they always provide 'Strong Evidence'?
- **Strong evidence (S3)** means evidence from studies with designs that can support causal conclusions (i.e., studies with **high internal validity**), and studies that, in total, include enough of the range of participants and settings to support scaling up to the State, regional, or national level (i.e., studies with **high external validity**).
- In reality, very very few clinical trials in this area have both high internal and high external validity.
- Major concerns about implementation
- The development of **translational science**.

Hence,

1. When assessing the limitations of any evaluation design, consider the alternatives-
We are not comparing with a perfect alternative, but with realistic alternative designs
2. Whatever you chose to do, enhance and **continuously improve.**

Using monitoring to inform evaluation

- As an alternative to ‘clean’ designs (hardly possible anyway)
- Look for/create ‘experiments in nature’
- Use statistical controls, leverage large sample size
- Emphasize ‘rich’ documentation
- Share findings and get corrective feedback
- Keep conclusions tentative
- Keep monitoring; it is not a one-shot deal
- Aim for local generalization; If it repeats itself, a good candidate for a wider generalization

Measurement then what?

Corrective Feedback is Essential

- **Dissemination to all constituents** in appropriate and multiple formats
- Building Capacity to utilize the feedback
 - ‘Specialists’
 - Educating and engaging all constituents in making sense of the data and promoting new ideas
- Connecting between findings and:
 - Scientific knowledge
 - Local best practices
 - Evidence-Based programs

S3, LEA'S and Academia

(a shameless plug for universities)

- Collaboration with universities could be mutually beneficial. Academia could be useful to:
 - Identify and design instruments
 - Analysis of findings
 - Building capacity to interpret findings
 - Connect with knowledge and existing EBP
 - Generate new knowledge
 - Educate researchers about real life issues
 - Create a fertile environment of creative disagreements and critical look at both sides