Development and testing of a brief EBP implementation intentions scale using Rasch analysis

Joanna Moullin, Ph.D.
Postdoctoral scholar, UC San Diego
Department of Psychiatry
jcmoullin@gmail.com
Funding agencies

USA National Institutes of Health (NIH) (R01) supported by National Institute of Mental Health (NIMH) and National Institute on Drug Abuse (NIDA) [R01DA038466] for funding the Leadership and Organizational Change for Implementation (LOCI) study

Team members

Mark G Ehrhart, Amy E Green, Elisa M Torres, Gregory A Aarons

Plus

Community-based organizations, counsellors, case managers, and supervisors that collaborate with us to make this work possible.
Science is constantly evolving

Methods change over time
Policy-makers and providers expect ‘useful’ results
We need to keep up...

#ImpSci
To be useful measures need to be...
An alternative way to develop and test measurement scales

**Item Response Theory**
In the next 15 minutes...

1. Basic principles of Rasch measurement model
2. Intentions scale and its psychometric properties
3. Practical considerations and future possibilities
Item Response Theory (IRT)

• Family of latent trait models

• Apply mathematical models to estimate parameters that represent the ‘locations’ of persons and items along an underlying, latent, continuum
  • Produces an interval scale

• Analysis assesses items’ strengths and weaknesses
  • Each response option should occupy a distinct portion of the continuum

• Focus is on the probability of individuals agreeing with an item given their responses to the other items
Differences between CTT and IRT

Classical Test Theory

• Focuses on composite scale
• Assumes items are equal indicators of the latent variable
• Similar items selected to consistently measure the latent variable
• Items have same response format
• Large number of items increases reliability
• Scale/test only applies to the specific sample. Needs always to be renormed
• Full scale must be administered

Item Response Theory

• Focuses on items
• Assesses each item for its contribution to measuring the latent variable
• Items selected that tap different levels of the latent variable
• Items may have different response format
• Scales may be shorter
• Scale/test independent of sample
• Need large heterogeneous sample to show item consistency and independence
• Subset of items can be administered
Rasch model

- One parameter IRT model for unidimensional scales
- Originally developed for use in educational settings for achievement/ability testing
- Now used in a variety of discipline areas
  - Psychology, medicine, health, business
- Rasch modelling is interested in modelling people’s ability or level of latent variable using items of varying difficulty
- As a person’s ability or level of latent variable increases, the likelihood that they will pass/agree with each item increases
- Data is assessed against the Rasch model and discrepancies among the items or the persons are identified
Rasch analysis

Aim is to identify a set of items that represent a range of levels of the attribute being measured

Used to assess and adjust all aspects of a scale

- Response format
- Items
- Persons
Measuring Intentions

Intentions are hypothesized to mediate the relationship between a provider’s attitudes and use of an EBP
  • Tested in theories of behavior

Need a scale to capture the level of plans, desire, and scope aspects of intentions to implement a specific EBP
Methods

1. Define construct & item generation
2. Administer to sample
3. Descriptive statistics
## Define construct & item generation

### Plans
- I plan to use Motivational Interviewing with my clients
- I am going to apply my training in Motivational Interviewing to address my clients’ needs
- I intend to use Motivational Interviewing when appropriate for my clients

### Desire
- Using Motivational Interviewing is a high priority for me
- I strive to apply Motivational Interviewing principles in working with my clients
- I will consider using Motivational Interviewing with new clients

### Scope
- I will use all aspects of Motivational Interviewing with my clients
- I will use parts of Motivational Interviewing with my clients.
- I will use certain Motivational Interviewing strategies with my clients
Methods

1. Define construct & item generation
2. Administer to sample
3. Descriptive statistics
4. Transfer from SPSS to RUMM2030
5. Analysis of all items, all cases (persons)
6. Scale refinement
   • Removal of cases
   • Change response format of items
   • Removal of items

• Overall model fit
• Overall item fit
• Overall person-fit
• Person separation index
• Individual person fit
• Individual item fit
• Thresholds & category probability curves
• Differential item functioning (DIF)
• Dimensionality
## Original scale – summary statistics

### Summary Statistics for Analysis Name INTENT1

<table>
<thead>
<tr>
<th>ITEMS</th>
<th>Location</th>
<th>Mean</th>
<th>Fit Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std Dev</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kurtosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PERSONS</th>
<th>Location</th>
<th>Mean</th>
<th>Fit Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std Dev</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kurtosis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Item-Trait Interaction

- Total Item Chi Square: 53.6696
- Degrees of Freedom: 8
- Chi Square Probability: 0.00000

### Person Reliability Indices

- PerSepIdx
  - intent1
  - * with extms: 0.84563
  - * NO extms: 0.81823

- CoefAlpha
  - * with extms: 0.88353
  - * NO extms: 0.84031

### Power of Analysis of Fit

This display is intended as a guide ONLY and should be used in conjunction with other analysis indicators.

<table>
<thead>
<tr>
<th>Index</th>
<th>Person/Item</th>
<th>Excellent</th>
<th>Good</th>
<th>Reasonable</th>
<th>Low</th>
<th>Too Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9781</td>
<td>0.91891</td>
<td>Excellent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9895</td>
<td>0.97087</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1980</td>
<td>0.367918</td>
<td>Reasonable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Key Statistics
- Overall model fit
- Overall item fit
- Overall person-fit
- Person separation index
- Individual person fit
- Individual item fit
- Thresholds & category probability curves
- Differential item functioning (DIF)
- Dimensionality
### Original scale – Persons fit

#### Overall model fit
- Overall item fit
- Overall person-fit
- Person separation index
- Individual person fit
- Individual item fit
- Thresholds & category probability curves
- Differential item functioning (DIF)
- Dimensionality

<table>
<thead>
<tr>
<th>ID</th>
<th>Total Exps</th>
<th>Max Exps</th>
<th>Items</th>
<th>ExtM</th>
<th>Location</th>
<th>SE</th>
<th>Fit Resid</th>
<th>Deg Free</th>
<th>Data Pts</th>
<th>ID</th>
<th>PF agency</th>
<th>PF sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>24</td>
<td>32</td>
<td>8</td>
<td>2.389</td>
<td>0.554</td>
<td>3.674</td>
<td>6.7</td>
<td>8</td>
<td>1137</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>72</td>
<td>28</td>
<td>32</td>
<td>8</td>
<td>3.777</td>
<td>0.649</td>
<td>3.434</td>
<td>6.7</td>
<td>8</td>
<td>1075</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>69</td>
<td>28</td>
<td>32</td>
<td>8</td>
<td>3.777</td>
<td>0.649</td>
<td>3.434</td>
<td>6.7</td>
<td>8</td>
<td>1109</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>70</td>
<td>26</td>
<td>32</td>
<td>8</td>
<td>3.037</td>
<td>0.590</td>
<td>2.505</td>
<td>6.7</td>
<td>8</td>
<td>1075</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>1.541</td>
<td>0.514</td>
<td>2.003</td>
<td>6.7</td>
<td>8</td>
<td>1036</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>3.777</td>
<td>0.649</td>
<td>1.608</td>
<td>6.7</td>
<td>8</td>
<td>1020</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>45</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>3.392</td>
<td>0.641</td>
<td>1.405</td>
<td>6.7</td>
<td>8</td>
<td>1047</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>2.705</td>
<td>0.571</td>
<td>1.557</td>
<td>6.7</td>
<td>8</td>
<td>1026</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>48</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>3.777</td>
<td>0.649</td>
<td>0.979</td>
<td>6.7</td>
<td>8</td>
<td>1014</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>57</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>2.389</td>
<td>0.554</td>
<td>0.979</td>
<td>6.7</td>
<td>8</td>
<td>1020</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>58</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>0.444</td>
<td>0.492</td>
<td>1.381</td>
<td>6.7</td>
<td>8</td>
<td>1014</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>59</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>-0.675</td>
<td>0.492</td>
<td>1.381</td>
<td>6.7</td>
<td>8</td>
<td>1014</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>54</td>
<td>21</td>
<td>32</td>
<td>8</td>
<td>0.263</td>
<td>0.492</td>
<td>1.381</td>
<td>6.7</td>
<td>8</td>
<td>1014</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**Mean**
- ExtM: 2.708
- Location: -0.429

**Std Dev**
- ExtM: 1.785
- Location: 1.529

**Separation Index**
- 0.84563

**Coefficient Alpha**
- 0.88353

**Mean Error Variance**
- 0.492

**Est. True Variance**
- 2.696

**Display Control**
- Exclude Extreme Persons: Yes
- Include Extreme Persons: No
- Sort Persons by: Fit Residual Order (Descent)

**File Text Format**
- Fixed

**Save Options**
- Frequency
- Person-by-Item
Original scale – Persons fit
Original scale – Threshold map

- Overall model fit
- Overall item fit
- Overall person-fit
- Person separation index
- Individual person fit
- Individual item fit
- Thresholds & category probability curves
- Differential item functioning (DIF)
- Dimensionality
Original scale – Category probability curves
Removal of 4 cases = ordered thresholds

- Overall model fit
- Overall item fit
- Overall person-fit
- Person separation index
- Individual item fit
- Individual person fit
- Thresholds & category probability curves
- Differential item functioning (DIF)
- Dimensionality
Methods: Analytic Process

1. All 9 items all 106 cases
2. All 9 items minus 1 case (104/1138)
3. All 9 items minus 4 cases (104/1138, 72/1086, 89/1114, 70/1079)
4. 8 items (minus item 9), minus 4 cases (104/1138, 72/1086, 89/1114, 70/1079)
5. 8 items (minus item 8), all cases 7 items (minus items 8 & 9)
6. 7 items (minus items 8 & 9), minus 1 case (104/1138)
7. 6 items (minus items 2, 8, 9), minus 1 case (104/1138)
8. 5 items [1, 4, 5, 6, 7] (minus items 2, 3, 8, 9)
9. 4 items [1, 4, 5, 7] (minus items 2, 3, 6, 8, 9)
10. 3 items [3, 5, 7] (minus items 2, 3, 4, 6, 8, 9)
Final scale – summary statistics

<table>
<thead>
<tr>
<th>ITEM-PERSON INTERACTION</th>
<th>PERSONS</th>
<th>ITEM-TRAIT INTERACTION</th>
<th>PERSON RELIABILITY INDICES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ITEMS</strong></td>
<td><strong>PERSONS</strong></td>
<td><strong>ITEM-TRAIT INTERACTION</strong></td>
<td><strong>PERSON RELIABILITY INDICES</strong></td>
</tr>
<tr>
<td>Mean</td>
<td>Location</td>
<td>Fit Residual</td>
<td>Mean</td>
</tr>
<tr>
<td>0.0000</td>
<td>1.1295</td>
<td>0.1034</td>
<td>3.6894</td>
</tr>
<tr>
<td>Std Dev</td>
<td>Std Dev</td>
<td>Skewness</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>1.1295</td>
<td>0.5847</td>
<td>-0.0902</td>
<td>-2.3333</td>
</tr>
<tr>
<td>Skewness</td>
<td>Kurtosis</td>
<td>Correlation</td>
<td>Location/StdResidual</td>
</tr>
<tr>
<td>-0.0902</td>
<td>-2.3333</td>
<td>-0.5225</td>
<td>Location/StdResidual</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>Correlation</td>
<td>Location/StdResidual</td>
<td></td>
</tr>
<tr>
<td>0.0000</td>
<td>-0.5225</td>
<td>Location/StdResidual</td>
<td></td>
</tr>
</tbody>
</table>

**SEPARATION INDICES**

- **Index:** 0.993266
- **Variance:** 22.49258
- **Error:** 0.15162

**POWER OF ANALYSIS OF FIT**

- **Excellent**
- **Good**
- **Reasonable**
- **Low**
- **Too Low**
Please answer the following questions about the extent to which you intend to use Motivational Interviewing.

**Plans**
- I am going to apply my training in Motivational Interviewing to address my clients' needs.

**Desire**
- Using Motivational Interviewing is a high priority for me.

**Scope**
- I will use all aspects of Motivational Interviewing with my clients.

Anchors: 0=Not at all ... 4=To a very great extent.
Next steps...

• Use of Rasch model to scientifically determine the number of cut-offs and the value of cut-offs for scale scores
  • E.g. low, moderate, high intentions

• Test predictive validity (use of EBP)
Practical implications

Organization
- Executive/Management Attitudes
- Adoption decision

Team
- Team-level Implementation climate
- Providers' Attitudes
- Intentions
- Use

Provider

Implementation strategies
- Tailored
Thoughts, questions, ideas?

Joanna Moullin, Ph.D.
Postdoctoral scholar, UC San Diego
Department of Psychiatry
jcmoullin@gmail.com
Thank you

Joanna Moullin, Ph.D.
Postdoctoral scholar, UC San Diego
Department of Psychiatry
jcmoullin@gmail.com
A systems approach towards the identification of implementation success

Arno Parolini
Department of Social Work

2 October, 2017
Significant progress in implementation research over the last decade:

- Barriers & Facilitators to EBP (e.g. Solomons & Spross, 2011)

- Scale development for measurement (Chaudoir, 2013)

- Frameworks to put these factors into the context of implementation (Fixsen et.al, 2005; Damschroder, 2009; Aarons, 2011).
Focus shift towards evidence-based systems of care

1. Scaling
2. Change at multiple levels of system
3. Multiple EBPs

(Chambers, 2012; Proctor, 2014)

Implementation success ➔ Intervention and implementation effectiveness critical
(Proctor et al., 2011)

Often occur at different levels in the system and influence each other!

➔ Causal pathways through which implementation works are of interest for implementation research ➔ Leverage points
Framing implementation process as a decision model within the context of multi-level system.

Draw on advances in causal inference, econometrics, statistics, economics, psychology, computer science, systems biology (e.g., Pearl, 2009; Heckman & Vytlacil, 2007; White & Lu, 2011)

Integrate implementation frameworks (e.g. Aarons et al., 2011) with structural choice models where each phase is linked to the next through a decision.

Using existing implementation frameworks as a starting point for collecting the right information combined with existing evidence (quantitative and qualitative) and key stakeholder knowledge (e.g. practitioners, clients)
Multiple-Level Decision Juncture Model

This model is formulated as a system of structural equations, where each equations represents a causal relationship between the dependent variable and covariates.

These models are influenced by existing knowledge, theory and qualitative research.

Source: Parolini, Tan & Shlonsky (2017)
Three-stage process of analysing causal models

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Defining the set of hypotheticals or counterfactuals</td>
<td>A scientific theory</td>
</tr>
<tr>
<td>2</td>
<td>Identifying parameters (causal or otherwise) from hypothetical population data</td>
<td>Mathematical analysis of point or set identification</td>
</tr>
<tr>
<td>3</td>
<td>Identifying parameters from real data</td>
<td>Estimation and testing theory</td>
</tr>
</tbody>
</table>

Source: Heckman & Vytlacil (2007, p.4785)

Precise research question is crucial!!
A hypothetical case study
• Hypothetical parenting intervention in child welfare system
• Randomization not feasible
• Assumptions: SUTVA, no feedback, no measurement errors
• Outcome variable is parenting score
Simplified scenario with sequential binary choices can be extended to the more general case of ordered and nominal response models with state-dependent outcomes.

Two Questions to demonstrate the interrelation between intervention and implementation:

1. What is the effect of the implementation strategy on treatment assignment?

2. What is the effect of treatment on client outcomes for organisations that chose to implement?
Organisational variables:
Z_1.....Intra-organisational networks
Z_2.....Management style
Z_3.....Organisational structure (hierarchical, flat)

Intervention characteristic:
W_1.....Cost of implementation strategy 1
W_2.....Cost of implementation strategy 2
W.....Weighted average of strategy costs
B_1.....Empirical evidence of effectiveness

Caseworker variables:
X_1.....Age
X_2.....Perceived Leadership in Organisation
X_3.....Tenure in child welfare

Client variables:
V_1.....Age
V_2.....Gender
V_3.....Socio Economic Status
Y_1.....Outcome at baseline
Y_2.....Outcome at follow up

Decision junctures:
D_1.....Organisation decides to implement
D_2.....Organisation chooses implementation strategy
T_1.....Caseworker chooses to assign client to intervention
Simulation exercise:

- Data is modelled based on results reported in studies (e.g. Ehrhart et al., 2016)
- Assumed linearity and separability in the latent index model
- \( N = 10,000 \) (\( N_O = 100 \), \( N_{CW,O} = 20 \), \( N_{CL,CW,O} = 5 \))
- 10,000 replications
- Software: Stata SE 14.2

2 October, 2017
## Implications for policy and practice

- Clearly defining the knowledge about the system in terms of assumptions → Efficient variable selection
- Linking implementation, intervention and system outcomes through causal mechanisms.
- Estimating causal and structural parameters from observational data
- Identifying leverage points for interventions
- Accommodating implementation sustainment and CQI
- Facilitating collaboration between researchers and practitioners in model design and evaluation
- More intuitive presentation in form of choice processes rather than experimental ideal based on unconfoundedness or “as good as randomized”

## Generalizability

- Learning about causal and structural parameters → What works how
- Predicting existing policy effects in new settings
- Predicting effects of new policies in existing settings
References

• Fixsen DL, Naoon SF, Blase KA, Friedman RM, Wallace F. Implementation research: A synthesis of the literature. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, the National Implementation Research Network; 2005.
• Proctor E. Dissemination and implementation research. In: Encyclopedia of Social Work, the National Association of Social Workers (NASW Press) and Oxford University Press (OUP); 2014.
Advancing the Pragmatic Measures Construct

Cameo Stanick, ¹ Byron J. Powell, ² Heather Halko, ³ Caitlin Dorsey, ⁴ Bryan Weiner, ⁵ & Cara Lewis ⁴

¹HATHAWAY-SYCAMORES CHILD AND FAMILY SERVICES
²UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
³UNIVERSITY OF MONTANA
⁴GROUP HEALTH RESEARCH INSTITUTE - KAISER PERMANENTE
⁵UNIVERSITY OF WASHINGTON
Need for Pragmatic Measures

- Valid and reliable measures are needed to:
  - Assess barriers and facilitators to implementing EBPs
  - Inform the development, selection, and tailoring of implementation strategies
  - Evaluating implementation outcomes
- However, stakeholders are not likely to use these measures if they are not pragmatic.
Recent Call for Pragmatic Measures

Pragmatic Measures
What They Are and Why We Need Them
Russell E. Glasgow, PhD, William T. Riley, PhD

Required Criteria:
• Important to stakeholders
• Low burden (for respondents and staff)
• Actionable
• Sensitive to change

Recommended Criteria:
• Broadly applicable
• Use for benchmark
• Unlikely to cause harm
• Psychometrically strong
• Related to theory or model

Gaps:
• Little consensus on how to define pragmatic measures
• Lack of stakeholder input on pragmatic constructs
Present Study

Aim 1: Establish a stakeholder-driven operationalization of pragmatic measures and develop reliable, valid rating criteria for assessing pragmatic strength.

DOI 10.1186/s13012-015-0287-0

STUDY PROTOCOL

Advancing implementation science through measure development and evaluation: a study protocol

Cara C. Lewis1,2*, Bryan J. Weiner3, Cameo Stanick4 and Sarah M. Fischer1

NIMH: 1R01MH106510-01
Proposal:
◦ Develop a stakeholder-driven operationalization and rating system to assess ‘pragmatic’ measure qualities

Approach:
◦ Recruit frontline stakeholders from diverse mental health settings to participate in qualitative and quantitative data collection to define and refine the pragmatic construct
◦ Conduct a systematic literature search to inductively develop pragmatic construct dimensions
◦ Crosswalk the literature search results with results from the stakeholder interviews to develop a comprehensive list of terms/phrases relevant to pragmatic measurement
◦ Utilize cross-walk results to inform both quantitative and qualitative next steps
Aim 1 Approach Phases

- Literature search
- Stakeholder interviews
- Concept mapping
- Delphi
- Reliability and known-groups validity testing

NIMH: 1R01MH106510-01
Literature Review

- PsycINFO and PubMed to capture mental health and health
- Titles and abstracts reviewed
- Articles that made it past title and abstract review were accessed and read more thoroughly
- All synonyms (e.g., ‘simple) and/or dimension terms/phrases (e.g., ‘ease to score’) were noted, including the field from which the terms were extracted (e.g., mental health treatment outcomes, geriatric rehabilitation, etc.)
- Longer phrases edited to approximately 4 words or less
Results: Literature Review

Table 1. Search Strings

1. pragmatic
AND
2. assessment* OR measures*
NOT
4. language

- PsycINFO returned 1,572 articles
- PubMed returned 20,000+
- N = 7 (NO articles relevant within first 200 results of PubMed)
Results: Literature Review

32 unique terms/phrases identified after reduced for redundancies

<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Scientific Discipline</th>
<th>Terms Relevant to the Pragmatic Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfieffer (1996)</td>
<td>Treatment Outcomes</td>
<td>confirm efficacy of intervention, efficient, lead to treatment planning, low complexity, low cost, provide a clinical cut-off score</td>
</tr>
<tr>
<td>Slade (1999)</td>
<td>Treatment Outcomes</td>
<td>able to be photocopied, acceptable, allows feedback data to be returned in a positive way, available, benefits outweigh costs, brief, compliments clinical judgment, contains readable indices, easily accessible, easy to use</td>
</tr>
</tbody>
</table>

NIMH: 1R01MH106510-01
Qualitative Interviews

- 7 stakeholders recruited and interviewed
  - **Purposeful sampling**
  - Specific areas represented:
    - Outpatient community mental health
    - School-based mental health
    - State mental health department
    - Residential treatment
    - Psychiatry
  - Interviews took approximately 1 hour to complete
  - Coded by 2 authors
    - All terms/phrases relevant to pragmatic measure construct extracted

NIMH: 1R01MH106510-01

1. Palinkas, Horwitz, Green, Wisdom, Duan, & Hoagwood, 2013
Results: Qualitative Interviews

- Reduced/edits terms/phrases for redundancy
- Removed terms that exclusively referred to clinical outcomes (as opposed to implementation measures/outcomes)
- **39 terms were identified in initial coding procedure**
  - Reduced phrases to be approximately 4 words \( n = 14 \)
  - E.g.:
    - Original: Fits within the sphere and scope of activities that are done
    - Revised: Fits **organizational activities**
  - 25 **domains** of pragmatic measures identified (e.g., cost)
  - 11 **dimensions** of those domains identified (e.g., less than $1.00 per measure)
  - Identified 16 **antonyms** of what a pragmatic measure is (e.g., costly)

NIMH: 1R01MH106510-01
Combining Qualitative Interview and Literature Search Results

- Literature review and interview terms/phrases were combined, further edited for redundancy \((N = 47 \text{ terms})\)
- Final list of terms/phrases submitted into a concept mapping activity
  - 20 stakeholders recruited to complete concept mapping
  - Stakeholders asked to group/categorize terms/phrases, as well as rate each term/phrase for *clarity* and *importance*
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsive</td>
<td>Results provide suggestions for clinicians about adherence or fidelity</td>
</tr>
<tr>
<td>Staff believes in it</td>
<td>Can be done over the phone</td>
</tr>
<tr>
<td>Accessible language</td>
<td>Can be extracted from EMRs</td>
</tr>
<tr>
<td>Language is clear</td>
<td>Measurable</td>
</tr>
<tr>
<td>Tied to reimbursement</td>
<td>Successfully measures things for you that aren’t working</td>
</tr>
<tr>
<td>Sensitive to change</td>
<td>Results provide suggestions about next clinical steps</td>
</tr>
<tr>
<td>Transparent</td>
<td>Tells what clinical interventions work for certain types of behaviors</td>
</tr>
<tr>
<td>Easy to use</td>
<td>Doesn’t require being an expert to administer</td>
</tr>
<tr>
<td>Useable results</td>
<td>Can be scored and/or interpreted by someone else</td>
</tr>
<tr>
<td>Accessible</td>
<td>Uses an easy scale</td>
</tr>
<tr>
<td>Important to clinical care</td>
<td>Reliable and valid results</td>
</tr>
<tr>
<td>Feasible</td>
<td>Successfully measures things for staff/organization that are not working</td>
</tr>
<tr>
<td>Meaningful</td>
<td>Results provide suggestions for clinicians about adherence or fidelity</td>
</tr>
<tr>
<td>Short</td>
<td>Natural output of clinical work and quality improvement activities</td>
</tr>
<tr>
<td>Fits within the sphere and scope of activities that are done by an organization</td>
<td></td>
</tr>
</tbody>
</table>
Concept Mapping

Participants:

◦ 24 community stakeholders with expertise implementing EBPs in mental health settings

Process:

◦ Structured sorting task involving all 47 criteria from interviews and lit search
◦ Rating each criterion in terms of its “clarity” and “importance” on a ten point scale

Analysis:

◦ Multidimensional scaling, hierarchical cluster analysis, & descriptive statistics

NIMH: 1R01MH106510-01

Note: Data collected and analyzed in Concept System® Global Max©
Cluster Map

NIMH: 1R01MH106510-01
## Categories and Criteria for Pragmatic Measures

### Useful (N = 15 criteria)

<table>
<thead>
<tr>
<th></th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>informs decision making</td>
</tr>
<tr>
<td>2</td>
<td>fits organizational activities</td>
</tr>
<tr>
<td>5</td>
<td>provides a cut-off score leading to an intervention or treatment plan</td>
</tr>
<tr>
<td>6</td>
<td>connects to clinical outcomes</td>
</tr>
<tr>
<td>11</td>
<td>produces reliable and valid results</td>
</tr>
<tr>
<td>13</td>
<td>reveals problems/issues in process or outcomes</td>
</tr>
<tr>
<td>14</td>
<td>informs adherence of fidelity</td>
</tr>
<tr>
<td>33</td>
<td>confirms efficacy of interventions</td>
</tr>
<tr>
<td>35</td>
<td>has a meaningful score distribution</td>
</tr>
<tr>
<td>46</td>
<td>informs clinical intervention selection</td>
</tr>
<tr>
<td>7</td>
<td>important to clinical care</td>
</tr>
<tr>
<td>19</td>
<td>sensitive to change</td>
</tr>
<tr>
<td>18</td>
<td>assesses organizational progress over time</td>
</tr>
<tr>
<td>38</td>
<td>optimizes patient care</td>
</tr>
</tbody>
</table>

### Compatible (N = 6 criteria)

<table>
<thead>
<tr>
<th></th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>applicable</td>
</tr>
<tr>
<td>16</td>
<td>the output of routine activities</td>
</tr>
<tr>
<td>12</td>
<td>focused</td>
</tr>
<tr>
<td>45</td>
<td>offers a compatible format to setting/user</td>
</tr>
<tr>
<td>10</td>
<td>easy to interpret</td>
</tr>
<tr>
<td>40</td>
<td>non-duplicative</td>
</tr>
<tr>
<td>8</td>
<td>efficient</td>
</tr>
<tr>
<td>37</td>
<td>generates data that provides a positive feedback loop (not used for staff punishment)</td>
</tr>
<tr>
<td>14</td>
<td>informs adherence of fidelity</td>
</tr>
<tr>
<td>33</td>
<td>confirms efficacy of interventions</td>
</tr>
<tr>
<td>35</td>
<td>has a meaningful score distribution</td>
</tr>
<tr>
<td>46</td>
<td>informs clinical intervention selection</td>
</tr>
</tbody>
</table>

### Easy (N = 19 criteria)

<table>
<thead>
<tr>
<th></th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>uses accessible language</td>
</tr>
<tr>
<td>31</td>
<td>accessible by phone</td>
</tr>
<tr>
<td>30</td>
<td>offers flexible administration time</td>
</tr>
<tr>
<td>10</td>
<td>easy to interpret</td>
</tr>
<tr>
<td>15</td>
<td>creates low assessor burden (ease of training, scoring, administration time)</td>
</tr>
<tr>
<td>20</td>
<td>not wordy</td>
</tr>
<tr>
<td>21</td>
<td>completed with ease</td>
</tr>
<tr>
<td>23</td>
<td>requires no expertise</td>
</tr>
<tr>
<td>26</td>
<td>of low complexity</td>
</tr>
<tr>
<td>32</td>
<td>brief</td>
</tr>
<tr>
<td>39</td>
<td>simple</td>
</tr>
<tr>
<td>47</td>
<td>low burden</td>
</tr>
<tr>
<td>17</td>
<td>easy to administer</td>
</tr>
<tr>
<td>34</td>
<td>intuitive</td>
</tr>
<tr>
<td>41</td>
<td>easy to use</td>
</tr>
<tr>
<td>42</td>
<td>easy to score</td>
</tr>
<tr>
<td>44</td>
<td>offers automated scoring or can be scored elsewhere</td>
</tr>
<tr>
<td>43</td>
<td>low cost</td>
</tr>
</tbody>
</table>

### Acceptable (N = 7 criteria)

<table>
<thead>
<tr>
<th></th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>creates a low social desirability bias</td>
</tr>
<tr>
<td>25</td>
<td>tied to reimbursement</td>
</tr>
<tr>
<td>28</td>
<td>relevant</td>
</tr>
<tr>
<td>30</td>
<td>offers relative advantage over existing methods</td>
</tr>
<tr>
<td>22</td>
<td>transparent</td>
</tr>
<tr>
<td>24</td>
<td>acceptable (to staff and clients)</td>
</tr>
<tr>
<td>43</td>
<td>low cost</td>
</tr>
<tr>
<td>26</td>
<td>of low complexity</td>
</tr>
</tbody>
</table>

**NIMH:** 1R01MH106510-01
Pattern Match
Go Zone

**Important & Clear:**
1. able to inform decision making
2. fits organizational activities
6. connects to clinical outcomes
7. important to clinical care
8. efficient
10. easy to interpret
11. produces reliable and valid results
15. creates low assessor burden (ease of training, scoring, administration time)
17. easy to administer
21. completed with ease
24. acceptable (by staff and clients)
27. uses accessible language
37. generates data that provides a positive feedback loop (not used for staff punishment)
41. easy to use
42. easy to score
43. low cost
46. able to inform clinical intervention selection

**Important, Less Clear:**
3. applicable
13. able to reveal problems/issues in process or outcomes
18. able to assess organizational progress over time
28. relevant
29. meaningful
30. offers relative advantage over existing methods
33. can confirm efficacy of interventions
36. feasible
38. optimizes patient care
47. low burden

**Clear, Less Important:**
5. provides a cut-off score leading to an intervention or treatment plan
20. not wordy
25. tied to reimbursement
31. accessible by phone
32. brief
44. offers automated scoring or can be scored elsewhere

**Less Clear, Less Important:**
4. creates a low social desirability bias
9. offers flexible administration time
12. focused
14. able to inform adherence of fidelity
16. the output of routine activities
19. sensitive to change
22. transparent
23. requires no expertise
26. of low complexity
34. intuitive
35. has a meaningful score distribution
39. simple
40. non-duplicative
45. offers a compatible format to setting/user
Delphi

Investigator team further reduced term/phrase count following concept mapping

17 terms/phrases

Acceptable = 5 criteria

Compatible = 5 criteria

Easy = 8 criteria

Useful = 2 criteria
Delphi methodology ($N=20$) to achieve consensus on the dimension priorities (relative weights)

Assign each phrase a score from 0-100. Your assigned points must be equal to 100 in total. Each score should represent how important/relevant the phrase is to a measure being "Easy".

<table>
<thead>
<tr>
<th>Easy</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses accessible language</td>
<td></td>
</tr>
<tr>
<td>Efficient</td>
<td></td>
</tr>
<tr>
<td>Feasible</td>
<td></td>
</tr>
<tr>
<td>Easy to interpret</td>
<td></td>
</tr>
<tr>
<td>Creates low assessor burden (ease of training, scoring, administration time)</td>
<td></td>
</tr>
<tr>
<td>Items not wordy</td>
<td></td>
</tr>
<tr>
<td>Completed with ease</td>
<td></td>
</tr>
<tr>
<td>Brief</td>
<td></td>
</tr>
</tbody>
</table>
Delphi Results

Consensus achieved within and across 16 of the 17 criteria

Utilizing Delphi results, further refined to 11 terms/phrases for final inclusion
  - e.g., Acceptable; Uses accessible language

Divided into two groups:

Objective Pragmatic Rating Criteria
Apply to published measures/literature by objective rater

Stakeholder-Facing Pragmatic Criteria
Applied by stakeholders when faced with the measure
Sample Criteria and Dimensions

### OBJECTIVE PRAGMATIC RATING CRITERIA

#### Easy Domain

**Brief**

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>Poor (P): The measure has greater than 200 items.</td>
</tr>
<tr>
<td>0</td>
<td>None (N): The measure is not available for use in the public domain.</td>
</tr>
<tr>
<td>1</td>
<td>Minimal/Emerging (M): The measure has greater than 100 items but fewer than 200 items.</td>
</tr>
<tr>
<td>2</td>
<td>Adequate (A): The measure has greater than 50 items but fewer than 100.</td>
</tr>
<tr>
<td>3</td>
<td>Good (G): The measure has greater than 10 items but fewer than 50.</td>
</tr>
<tr>
<td>4</td>
<td>Excellent (E): The measure has fewer than 10 items.</td>
</tr>
</tbody>
</table>

### STAKEHOLDER-FACING RATING CRITERIA

#### Useful Domain

**Fits Organizational Activities**

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>Poor (P)</td>
</tr>
<tr>
<td>0</td>
<td>None (N): The organizational fit of the measure was not assessed</td>
</tr>
<tr>
<td>1</td>
<td>Minimal/Emerging (M)</td>
</tr>
<tr>
<td>2</td>
<td>Adequate (A)</td>
</tr>
<tr>
<td>3</td>
<td>Good (G)</td>
</tr>
<tr>
<td>4</td>
<td>Excellent (E)</td>
</tr>
</tbody>
</table>
Future Directions: Shorter-Term

- Test-retest and inter-rater reliability assessments of the emergent rating system

- Conduct reliability and known-groups validity testing of the top 3 prioritized pragmatic criteria

- Test and assess performance of stakeholder-facing pragmatic rating criteria
Future Directions: Long-Term

- Combine pragmatic construct rating criteria with evidence-based assessment rating system
- Apply refined rating system to existing 450+ implementation measures
- Influence reporting standards for implementation measurement to include pragmatic dimensions, rather than solely focusing on psychometric strength
We now have a set of criteria that stakeholders deem relevant and important for implementation measurement!

- Informs development and evaluation of existing measures
- Nudges measure developers toward use of measures that are valid, reliable, and pragmatic
- Final results – via SIRC IRP website – will guide stakeholders toward measures that meet their criteria

Aligns researcher goals with stakeholder needs for measuring practice and progress of implementation projects

NIMH: 1R01MH106510-01
Psychometric Assessment of Three Newly Developed Implementation Outcome Measures

PRESENTER: DR. BRYAN WEINER
Domain Delineation – Conceptual Definition

- Process of defining what a concept is and what it is not.

- **Acceptability**: the quality or state of meeting one’s needs, preferences, or expectations.

- **Appropriateness**: the quality or state of being fitting, suitable, or proper for a particular purpose, person, condition, occasion, or place.

- **Feasibility**: the state or degree of being easily or conveniently done.
## Domain Delineation – Concept Differentiation

<table>
<thead>
<tr>
<th>Construct</th>
<th>Fit</th>
<th>Antecedent Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptability</td>
<td>Innovation-individual fit</td>
<td>Attitudes, preferences</td>
</tr>
<tr>
<td>Appropriateness</td>
<td>Innovation-social fit</td>
<td>Norms, values</td>
</tr>
<tr>
<td></td>
<td>Innovation-task fit</td>
<td>Efficacy (means-ends)</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Innovation-system fit</td>
<td>Practicality</td>
</tr>
</tbody>
</table>
Domain Delineation - Nomological Network*

A representation of the constructs of interest in a study, their observable manifestations, and the interrelationships among and between these.

- Openness
- Trialability
- Norms/values
- Perceived Efficacy
- Resource Availability
- Complexity

- Acceptability
- Appropriateness
- Feasibility

- Adoption Implementation

* Focus on identifying distinctive/differential determinants

NIMH: 1R01MH106510-01
Item Generation – Acceptability*

- This EBP seems fine.
- This EBP seems good enough.
- This EBP will do.
- This EBP meets my approval.
- This EBP meets my needs.
- This EBP is fine by me.
- This EBP is pretty good.
- This EBP is satisfactory.
- I have no objection to this EBP
- I have no concerns about this EBP
- This EBP is appealing.
- I like this EBP
- I welcome use of this EBP

* Items cover range from neutral to positive feeling
Item Generation – Appropriateness*

- This EBP seems reasonable
- This EBP seems right
- This EBP seems fitting
- This EBP seems suitable
- This EBP seems applicable

- This EBP seems right on the button
- This EBP seems proper
- This EBP seems apt
- This EBP seems like a good match.
- This EBP seems well aligned

* Items could be tailored to purpose, person, condition, occasion, or place
Item Generation – Feasibility

- This EBP seems practical.
- This EBP seems realistic.
- This EBP seems workable.
- This EBP seems possible.
- This EBP seems viable.
- This EBP seems doable.
- This EBP seems easy to use.
- This EPB seems implementable.
- This EPB seems challenging
Substantive & Discriminant Content Validity

- **Sample**: convenience sample of 36 implementation scientists and 27 implementation experienced mental health professionals
- **Data Collection**: web-based survey with respondents assigning items to construct(s) and rating their confidence in their assignments
- **Measures**: weighted item assignments
- **Data Analysis**:
  - ICCs from 2-way mixed ANOVA to assess agreement in item assignments
  - Wilcoxon one-sample signed rank test to determine whether the item represents the intended construct more so than the other constructs
  - EFA and CFA to formally test validity of conceptual model and trim the number of items

NIMH: 1R01MH106510-01
Results Summarized

- Inter-rater reliability high (.82-.94) for all participants, all items, and all constructs. No subgroup differences noted.

- All but 6 items exhibited substantive and discriminant content validity.

- Acceptability: this EBP is good enough, this EBP will do, this EBP is pretty good

- Appropriateness: this EBP seems right, this EBP seems right on the button, this EBP seems reasonable

- EFA and CFA resulted in trimmed scales (5 items), with good model fit and good scale reliability ($\alpha = .87$ to $.89$). Scales highly correlated.
Structural & Known-Groups Validity

- **Sample:** convenience sample of 326 counselors belonging to the American Mental Health Counselors Association (AMHCA)
- **Data Collection:** $2^3$ factorial between-subjects design using web-based survey. Vignettes of therapist considering adopting measurement based care (MBC). Manipulated information about hypothesized determinants. Rated from therapist’s perspective.
- **Measures:** 15 items from the trimmed CFA
- **Data Analysis:**
  - Scale refinement (construct specific CFAs and inter-item consistency)
  - Structural validity (3-factor, 2-factor, and 1-factor CFAs)
  - Known-groups validity (ANOVA)
Results Summarized

- Scale refinement: construct-specific CFAs produced trimmed (4-item) scales with good reliability ($\alpha = .85 \text{ to } .91$)

- Structural validity: 3-factor CFA model fit was good: $\text{CFI} = 0.96$ and $\text{RMSEA}= 0.08 \ (CI, 0.06-0.09)$. 2-factor CFA model fit and 1-factor CFA model fit were poor.

- Discriminant validity: acceptability and appropriateness scales highly correlated ($r = .77$), but possibly inflated due to survey design error.

- Known-groups validity: medium-size main effects based on known differences in vignettes. Incomplete separation of acceptability and appropriateness, probably due to survey design error.
Test-Retest Reliability & Sensitivity to Change

- **Sample:** convenience sample of 192 AMHCA counselors
- **Data Collection:** Half randomly assigned to receive same vignette; other half randomly assigned to receive opposite.
- **Measures:** 12 items from the trimmed structural validity CFA
- **Data Analysis:**
  - Scale assessment: inter-item consistency
  - Test-retest reliability: Pearson correlation corrected for measurement error
  - Sensitivity to change: regression analysis of difference scores
Results Summarized

- Scale assessment: Chronbach α’s were 0.85 for acceptability, 0.91 for appropriateness, and 0.89 for feasibility.

- Test-retest reliability: Pearson correlation coefficients corrected for measurement error were 0.80 for acceptability, 0.73 for appropriateness, and 0.88 for feasibility.

- Sensitivity to change: Regression coefficients for acceptability (0.76, -0.90), appropriateness (0.68, -1.18), and feasibility (0.92, -1.26) were significant and signed as expected for Low→High and High→Low.
Results Summarized

<table>
<thead>
<tr>
<th>Psychometric Property</th>
<th>Directly measure construct</th>
<th>Show some discriminant validity</th>
<th>Could be used for EBPs or implementation processes</th>
<th>Item referents can be tailored (e.g., appropriateness)</th>
<th>Short in length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content validity</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discriminant content validity</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural validity</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural invariance</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Known-groups validity</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictive validity</td>
<td>Not yet tested</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsiveness to change</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Scales

<table>
<thead>
<tr>
<th>Acceptability</th>
<th>Appropriateness</th>
<th>Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBC meets her approval.</td>
<td>MBC seems fitting.</td>
<td>MBC seems implementable.</td>
</tr>
<tr>
<td>MBC is appealing to her.</td>
<td>MBC seems suitable.</td>
<td>MBC seems possible.</td>
</tr>
<tr>
<td>She likes this MBC.</td>
<td>MBC seems applicable.</td>
<td>MBC seems doable.</td>
</tr>
<tr>
<td>She welcomes MBC.</td>
<td>MBC seems like a good match.</td>
<td>MBC seems easy to use.</td>
</tr>
</tbody>
</table>
Implications, Limitations, & Directions

- New measures demonstrate substantive validity, discriminant content validity, reliability, structural validity, known-groups validity, test-retest reliability, and sensitivity to change.
- Measures are brief (pragmatic)
- Discriminant validity remains unclear – further assessment required, might not be necessary to field all three in the same study
- Survey design error limited assessment of discriminant validity.
- Replication needed with different samples and materials.
- Next up: predictive validity assessment