Beyond the Rubber Ruler
Creating standardized instruments for assessing energy interventions

Beth Karlin
SEE Change Institute
Cindy Frantz
Oberlin College

Agenda

• Welcome and Introductions
• **Toolkit Background:** Why a unified approach to measurement?
• **Toolkit Development:** What is worth measuring?
• **Toolkit Testing:** How do you create a good measure?
• **Toolkit Implementation:** How will this work in the field?
• Q&A / Discussion
Toolkit Background

Why a unified approach to measurement?
The IMPORTANCE of Asking Questions

The **IMPORTANCE** of Measurement Consistency

<table>
<thead>
<tr>
<th>Stanford Binet Intelligence Scale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genius</strong></td>
<td><strong>Over 140</strong></td>
</tr>
<tr>
<td><strong>Very Superior</strong></td>
<td><strong>120 - 139</strong></td>
</tr>
<tr>
<td><strong>Superior</strong></td>
<td><strong>110 - 119</strong></td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>90 - 109</strong></td>
</tr>
<tr>
<td><strong>Dull</strong></td>
<td><strong>80 - 89</strong></td>
</tr>
<tr>
<td><strong>Borderline Deficiency</strong></td>
<td><strong>70 - 79</strong></td>
</tr>
<tr>
<td><strong>Moron</strong></td>
<td><strong>50 - 69</strong></td>
</tr>
<tr>
<td><strong>Imbecile</strong></td>
<td><strong>20 - 49</strong></td>
</tr>
<tr>
<td><strong>Idiot</strong></td>
<td><strong>Below 20</strong></td>
</tr>
</tbody>
</table>
**Reviewing Ten Years of Data Collection**

- Literature search conducted to identify all studies between 2003 and 2013
  - **315 behaviour-based energy intervention studies identified**

- Review of four key criteria resulted in 230 papers excluded from analysis
  - **85 behaviour-based energy intervention studies retained**

- Coding sheet developed and each study coded according to the same criteria

<table>
<thead>
<tr>
<th>Intervention Type</th>
<th>Measures</th>
<th>General Info</th>
<th>Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commitment</td>
<td>Context</td>
<td>Year published</td>
<td>Surveys</td>
</tr>
<tr>
<td>Audits</td>
<td>Behaviour</td>
<td>No. of participants</td>
<td>Interviews</td>
</tr>
<tr>
<td>Workshops</td>
<td>Attitudes</td>
<td>Quantitative data</td>
<td>Focus Groups</td>
</tr>
<tr>
<td>Media Campaigns</td>
<td>Knowledge</td>
<td>Qualitative data</td>
<td>Participants</td>
</tr>
<tr>
<td>Feedback</td>
<td>User Experience</td>
<td></td>
<td>Type of Data</td>
</tr>
<tr>
<td>Incentives</td>
<td>Specific Scales</td>
<td></td>
<td>When Collected</td>
</tr>
</tbody>
</table>

Most Reports Are Not Sharing Instruments

How awesome is our blog content?

- Out of this world awesome
- Pretty awesome
- I enjoy it a lot
- I like it
- I don’t like it
Rate how much you agree with the following

- Helping people and the environment is important
- My car and my house are energy efficient
- I turn off lights and unplug appliances when not in use
The IMPORTANCE of Measurement Validation

Close Ended Question
How much time do you spend studying?
A) 1-8 hours B) 9-18 hours C) > 18 hours

Open Ended Question
Tell me about your study habits...
The IMPORTANCE of Measurement Validation

<table>
<thead>
<tr>
<th>Issue</th>
<th>Open-ended</th>
<th>Closed-ended</th>
</tr>
</thead>
<tbody>
<tr>
<td>The economy</td>
<td>35%</td>
<td>58%</td>
</tr>
<tr>
<td>The war in Iraq</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Health care</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Terrorism</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Energy policy</td>
<td>*</td>
<td>6</td>
</tr>
<tr>
<td>Other</td>
<td>43</td>
<td>8</td>
</tr>
<tr>
<td>Candidate mentions</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>Moral values/social issues</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>Taxes/dist. of income</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>Other issues</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>Other political mentions</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Change</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Other</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>Don’t know</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

1 Unprompted first response to open-ended question
2 First choice from 5 options read to respondents
The IMPORTANCE of Measurement Validation

OR

OR
Psychometrics

- Theory and technique of measurement:
  Knowledge, abilities, attitudes, traits

- Construction and validation of instruments:
  Questionnaires, tests, assessments

<table>
<thead>
<tr>
<th>Right Thing</th>
<th>Wrong Thing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right Way</td>
<td>Wrong Way</td>
</tr>
</tbody>
</table>
The PROCESS of Psychometrics

Theoretically relevant constructs identified

Item generation

Item pre-screening and pre-testing

Item reduction & construct validation

Factor structure confirmation

Field testing

Have Think Do
Our Project

IEA DSM Task 24

Consultation

Feedback

Initial concept paper

Methods review

Theory review

Toolkit

Pilot testing

Member country testing

SCE Psychometric testing

Report for IEA Karlin et al., 2015

Field testing
Toolkit Development

What is worth measuring?
Energy Program

Why is it working?

Changes in Energy Behavior

Who does it work for?

HOW Does it work?
Energy Cultures Framework

Material Culture

Norms & Aspirations

Energy Practices
Energy Cultures Framework

HAVE

THINK

DO
Material Culture (Have)

- What time of dwelling do you live in?
- Which of the following appliances do you own? (E.g., space heater, dishwasher, central AC, etc.)
Toolkit Development

Beliefs (Think)
Connection and Concern
Norms (Personal and Social)
Efficacy (Performance and Response)

Motivation
Behavioral Intention

THINK

Habits and traditions
Social expectations and aspirations
Environmental connection and concern
Motives to engage
Beliefs (Think)

- I can invest the time and effort to make changes towards reducing my energy use.
- If enough people use less energy, we can benefit the natural environment.

Efficacy (Performance and Response)

Motives to engage

Habits and traditions

Social expectations and aspirations

Environmental connection and concern

THINK
One-time and Recurring Behaviors (Do)

- Please indicate whether you have done each of the following since moving into your home: (e.g., installed insulation, replaced incandescent bulbs with CFLs or LEDs, etc.)
- How frequently do you: (e.g., limit time in shower, turn off lights when not needed, etc.)

Turning on heater

Drawing curtains

Putting on sweater

Maintaining heating technologies
User Experience (UPscale)

Ease of Use
- I feel very confident interpreting the information provided to me.
- A person would need to learn a lot in order to understand this __________.

Engagement
- I do not find this __________ to be useful.
- I think that I would like to use this ___________ frequently.
Toolkit Testing

How do you create a good measure?
Toolkit Testing - Methods

Potential survey questions identified from literature and prior work

Step 1: Preliminary Toolkit split into 4 instruments

- Have/Do
- Think 1
- Think 2
- Ux/Intention

Step 2: Each Instrument tested for clarity (Testing round 1)

Step 3: Questions revised based on round 1 results

Step 4: Full Toolkit psychometric testing (Testing round 2)

Step 5: Analysis and suggested revisions

Potential survey questions identified from literature and prior work

Step 1: Preliminary Toolkit split into 4 instruments

- Have/Do
- Think 1
- Think 2
- Ux/Intention

Step 2: Each Instrument tested for clarity (Testing round 1)

Step 3: Questions revised based on round 1 results

Step 4: Full Toolkit psychometric testing (Testing round 2)

Step 5: Analysis and suggested revisions
• Online survey (Amazon Mechanical Turk)

• Testing Round 2: 520 participants ($2 compensation)
  o Most lived in houses (67.1%) followed by an apartment (4.2%)
  o About half owned their home (47.9%) or rented their home (47.7%)
  o Average # of people in home was 2.8 (SD = 1.4)
  o More than half ranged from 25-45 years old (59.2%)
  o 50.2% were female and most identified as White (79.4%)

---

Step 4: Full Toolkit psychometric testing (Testing round 2)

Step 5: Analysis and suggested revisions
Toolkit Testing - Results

Psychometric Properties

1. Factor Analysis/Item Reduction
   - Are measures of each construct distinct?
   - Which items best represent the construct?

2. Reliability
   - Do items hang together/interrelate strongly?

3. Criterion Validity
   - Does it measure what it's supposed to?
   - Does it relate to/predict relevant constructs?
1. Factor Analysis/Item Reduction

- Motivation: Self-comfort, Prosocial, Cost
- Energy Literacy: Awareness, Knowledge
- Connection & Concern: Connection, Concern
- Norms: Social, Personal
- Efficacy: Performance, Response
- Utility Qs: Response to peak demand, Anti-programming, Utility perceptions
- User Experience: Ease of use, Engagement, Trust
## Toolkit Testing - Results

### 1. Factor Analysis

- **Motivation - cost**: 1 item, Alpha = N/A
- **Motivation - prosocial**: 3 items, Alpha = 0.801
- **Motivation - self-comfort**: 3 items, Alpha = 0.796
- **Energy Literacy - awareness**: 2 items, Alpha = 0.861
- **Connection**: 2 items, Alpha = 0.769
- **Concern**: 2 items, Alpha = 0.843
- **Personal Norms**: 3 items, Alpha = 0.808
- **Social Norms**: 2 items, Alpha = 0.738
- **Performance Efficacy**: 2 items, Alpha = 0.820
- **Response Efficacy**: 3 items, Alpha = 0.828
- **Utility Questions: Response to peak demand**: 4 items, Alpha = 0.655
- **Utility Questions: Anti-programming**: 3 items, Alpha = 0.443
- **Utility Questions: Utility perceptions**: 2 items, Alpha = 0.505
- **UPscale - Ease of Use**: 7 items, Alpha = 0.870
- **UPscale - Engagement**: 9 items, Alpha = 0.900
- **UPscale - Trust**: 2 items, Alpha = 0.924
- **Satisfaction**: 2 items, Alpha = 0.870

### 2. Reliability

Alpha > .70 is good
1. Factor Analysis

2. Reliability

3. Criterion Validity

Convergent: Scales should relate to other relevant concepts

Divergent: Scales should NOT relate to certain concepts

- Big Five Personality Inventory
- Connectedness to Nature Scale (CNS)
- New Environmental Paradigm (NEP)
- Environmental Attitudes Scale (EAS)
- Frugality Scale

Toolkit Testing - Results
3. Criterion Validity

Convergent: Scales should relate to other relevant concepts

** Significant at the p < .01 level

(Exemplary Evidence of)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Intrinsic Environmental Attitudes</th>
<th>Extrinsic Environmental Attitudes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation - prosocial</td>
<td>.451**</td>
<td>.388**</td>
</tr>
<tr>
<td>Motivation - self comfort</td>
<td>-0.033</td>
<td>-0.06</td>
</tr>
<tr>
<td>Motivation - cost</td>
<td>.258**</td>
<td>0.023</td>
</tr>
<tr>
<td>Energy Literacy - awareness</td>
<td>.421**</td>
<td>.273**</td>
</tr>
<tr>
<td>Connection</td>
<td>.540**</td>
<td>.356**</td>
</tr>
<tr>
<td>Concern</td>
<td>.570**</td>
<td>.280**</td>
</tr>
<tr>
<td>Personal norms</td>
<td>.701**</td>
<td>.297**</td>
</tr>
<tr>
<td>Social norms</td>
<td>0.026</td>
<td>.294**</td>
</tr>
<tr>
<td>Performance efficacy</td>
<td>.390**</td>
<td>.150**</td>
</tr>
<tr>
<td>Response efficacy</td>
<td>.560**</td>
<td>.237**</td>
</tr>
<tr>
<td>UPscale - Ease of use</td>
<td>-.430**</td>
<td>-.234**</td>
</tr>
<tr>
<td>UPscale - Engagement</td>
<td>.428**</td>
<td>0.053</td>
</tr>
<tr>
<td>UPscale - Trust</td>
<td>.254**</td>
<td>0.053</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>.255</td>
<td>.010</td>
</tr>
</tbody>
</table>
Toolkit Implementation

How will this work in the field?
Toolkit Implementation

Next Steps

Initial concept paper
Methods review
Theory review
Toolkit
Pilot testing
Member country testing
SCE Psychometric testing
Field testing
Report for IEA
Karlin et al., 2015
Consultation
Feedback
Thank you!

Beth Karlin
SEE Change Institute

Cindy Frantz
Oberlin College