

Understanding and Choosing Self-Report (Survey) Measures: Part Two

Anita Stewart
Institute for Health & Aging
Center for Aging in Diverse Communities

Content of Measurement Lectures

- Importance of concepts
 - Process of selecting measures
 - Reviewing measures
 - Get to know the measure
 - Appropriateness
 - Conceptual and psychometric adequacy
 - Practicality
- Part 1
- Part 2
-

PROCESS of Selecting Measures for Your Studies

Describe context: population & study constraints

Define concept (variable)

Identify potential measures

Review measures for:

- conceptual and psychometric adequacy
- appropriateness for your sample
- practical considerations

Pretest best measure

If problematic:
modify and pretest again

Final measure

Tool for Reviewing Measures

- Template for reviewing measure for appropriateness for your study
- Systematic (checklist) approach
- Goal: find best measure for your study
 - May not be the most popular measure

Categories on Template for Reviewing Measures

- At the top
 - Concept defined by you, name of measure
- Categories for reviewing measure
 - Get to know the measure
 - Appropriateness for your sample
 - Conceptual adequacy
 - Psychometric adequacy
 - Practical issues

Define Concept (from Part One)

- Research questions stated in terms of concepts (latent variables)
- Ability to answer research questions depends on how well the measures reflect the concepts
- Defining concepts before selecting measures increases chance of observing true associations

Categories on Template for Reviewing Measures

- At the top
 - Concept defined by you, name of measure
- Categories for reviewing measure
 - **Get to know the measure**
 - Appropriateness for your sample
 - Conceptual adequacy
 - Psychometric adequacy
 - Practical issues

What's in a Label?

- Do not depend on name
 - Label may not reflect what it actually contains
- May find measure of your concept with the right name
 - But it measures something different than you need

Get to Know the Measure

- Original publication
 - Subsequent studies of measure (especially in your population)
- Obtain copy of questionnaire
- Measurement model – structure
- Review actual instrument
- Interpretability of scores

Things to Consider in Looking Over Actual Questionnaire

- Instructions
- Format for responding
- Time frame
- Item stems
- Response choices

Handouts: Four Questionnaires

- Patient Health Questionnaire (PHQ-9)
- Center for Epidemiological Studies Depression Scale (CES-D)
- Perceived Stress Scale (PSS)
- Jefferson Scale of Empathy – Physician/Health Professional (HP-version)

Poor Format/Presentation = High Respondent Burden

- Instructions wordy, unclear
- Complicated way of responding
- Difficult to navigate
 - Crowded, hard to track across page
- Hard to read
 - Poor contrast, small font

Composition of a Survey Item

During the past month,
how much of the time
have you felt tired?

- 1 Never
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

Composition of a Survey Item

During the past month,
how much of the time
have you felt tired?



**Item
stem**

- 1 Never
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

Composition of a Survey Item

During the past month,
how much of the time
have you felt tired?

Time frame

**Item
stem**

Item content

- 1 Never
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

Composition of a Survey Item

During the past month,
how much of the time
have you felt tired?

- 1 Never
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time



**Question
being asked
about content**

Composition of a Survey Item

During the past month,
how much of the time
have you felt tired?

- 1 Never
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

**Question
being asked
about content**

**Response
choices**

Composition of a Survey Item

During the past month,
how often have you felt
tired?

**Question
being asked
(how often)**

- 1 - Never
- 2 - Once or twice
- 3 - A few times
- 4 - Fairly often
- 5 - Very often

Composition of a Survey Item

During the past month,
how often have you felt
tired?

**Question
being asked
(how often)**

- 1 - Never
- 2 - Once or twice
- 3 - A few times
- 4 - Fairly often
- 5 - Very often

**Frequency
response
choices**

Intensity Response Scale Choices

In the past month, how tired have you been?

1 – Not at all

2 - A little

3 - Fairly

4 - Quite

5 - Extremely

1 – Not at all

2 - A little

3 - Somewhat

4 - Very much

Types of Response Scales

- | | |
|-------------|--|
| State/level | Level or state of feeling, symptom <ul style="list-style-type: none">- Frequency, intensity, severity of pain |
| Evaluative | Value attached to level or state <ul style="list-style-type: none">- Satisfaction with health- Bothered by pain |
| Comparative | Comparison to prior time <ul style="list-style-type: none">- Pain compared to before treatment |
| Agreement | Level of agreement with statement <ul style="list-style-type: none">- "My health is excellent" |

Ways to Ask about Symptoms

Depressive symptoms

- How much of the time (proportion of time)
- How often (frequency)
- How severe (intensity)
 - On average? At its worst?
- How much bothered by (evaluative)

Ways to Ask about Symptoms

Depressive symptoms

- How much of the time (proportion of time)
- How often (frequency) (CES-D)
- How tired (intensity)
 - On average? At its worst?
- How often bothered by (evaluative) (PHQ-9)

What Makes a Good Item

- Short and concise item stem
- Only one “concept” per item
- No conditional item stems
- Response choices match item stem
- Clear format/instructions

Not Short and Concise, Two Concepts in One Item

PHQ-9 #8

Over the last 2 weeks, how often have you been bothered by

- ... Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual.

Conditional Items

PSS #9

- In the last month, how often have you been angered because of things that were outside of your control?

Conditional Items

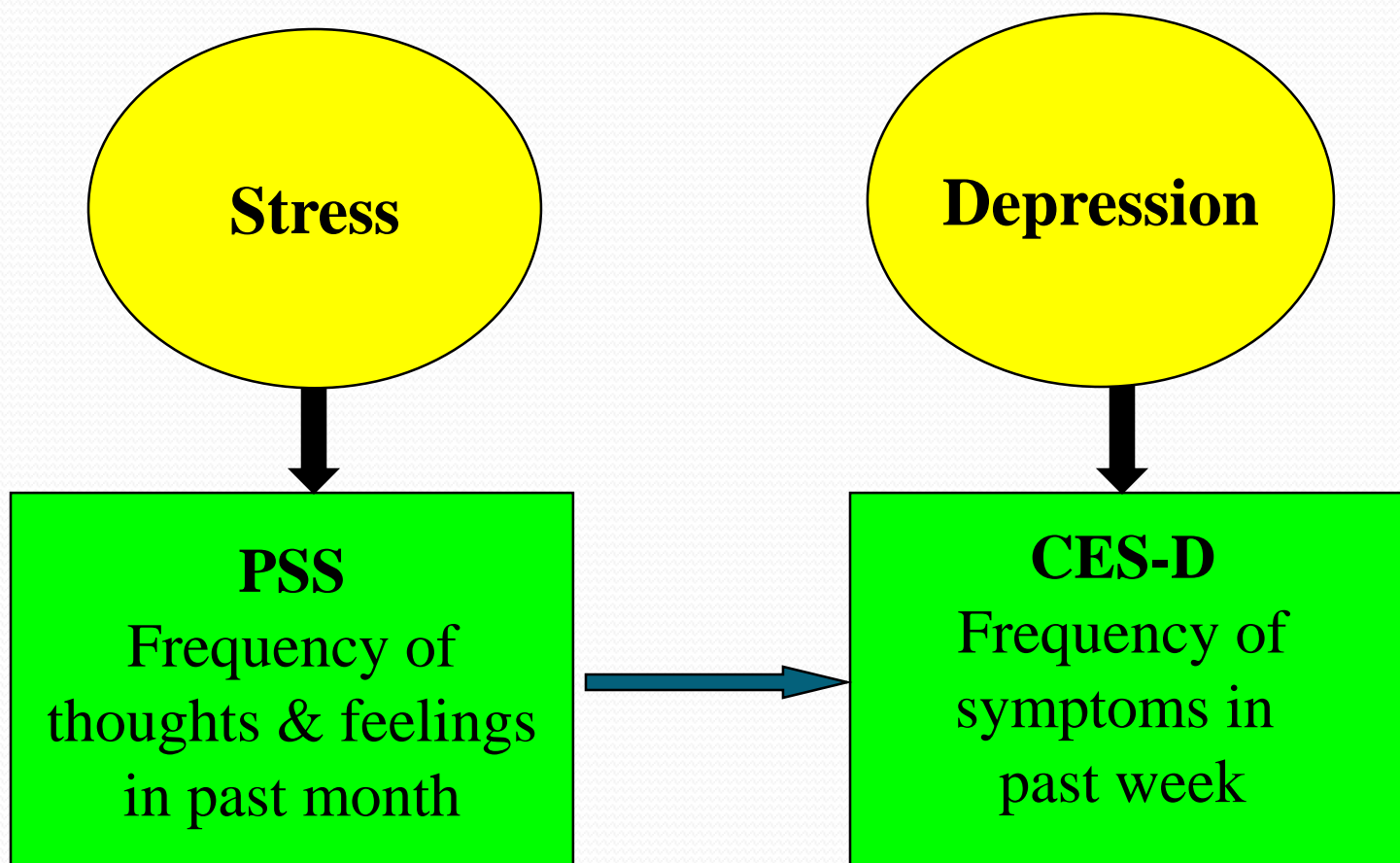
PSS #9

- In the last month, how often have you been angered because of things that were outside of your control?
 - Report frequency of being angered **only if (conditional on)** the cause is something out of their control

Take Away on Items

- How items are written matters!
- Item wording, time frame, and response choices all affect respondents' answers
- Their answers are YOUR RAW DATA
- Optimal measures: Items have been carefully developed and tested

Concept Being Measured Depends on All Features of Items



Interpretability: Another Part of Getting to Know Your Measures

- What is the possible range?
- What does a high score mean?
(direction of scoring)
- What is minimum and maximum possible?
 - Enables interpretation of mean score

Possible Range

- Multi-item scales: no inherent meaning to scores
- When item scores are added, possible range depends on # of items and # of response choices
 - 5 items, 4 response choices = 5 to 20
 - 3 items, 5 response choices = 3 to 15

Categories on Template for Reviewing Measures

- At the top
 - Concept defined by you, name of measure
- Categories for reviewing measure
 - Get to know the measure
 - **Appropriateness for your sample**
 - Conceptual adequacy
 - Psychometric adequacy
 - Practical issues

Applied or Tested in Group Similar to Yours?

- Some measures have been evaluated in minority groups (CES-D, SF-36, PSS, PHQ-9)
 - Psychometric evaluation the focus
 - Applied in research but psychometrics reported

What Makes a Good Item – Additional Criteria in Disparities Populations

- Short and concise item stem
- Only one “concept” per item
- No conditional items
- Response choices match item stem
- Clear format/instructions
- Reading level – 5th or 6th grade
- Translatable
- No jargon or colloquialisms
 - Universally understood

Reading Level

- Is reading level appropriate for your target population?
 - Special concern - lower SES, limited English proficiency
- If reading level not known
 - Make your own judgment
 - Pretest with target population

Respondent Burden

- Real burden
 - Length, convenience, time to complete
- Perceived burden
 - A function of item difficulty, distress due to content, perceived value of survey, expected length
- Some population subgroups have more difficulty, take longer to complete

Categories on Template for Reviewing Measures

- At the top: concept defined by you, name of measure
- Categories for reviewing a measure
 - Get to know the measure
 - Appropriateness for your sample
 - **Conceptual adequacy**
 - Psychometric adequacy
 - Practical issues

Conceptual Adequacy: Three Questions

- What concept is being measured?
 - As defined by developers
 - As determined by you in item review
- Has concept been explored in your target population?
 - Is concept different than mainstream?
- Does concept match the one you defined?

Approaches to Explore Conceptual Adequacy in a Diverse Group

- Literature reviews of concept in diverse groups
- In-depth interviews and focus groups
 - Discuss concept, obtain their views
- Expert review (from diverse group)
 - Review concept definitions
 - Rate relevance of items

Example of Inadequate Concept

- Patient satisfaction conceptualized in mainstream populations in terms of
 - Access, technical care, communication, continuity, interpersonal style
- In minority and low income groups, additional relevant domains include
 - Discrimination by health professionals
 - Sensitivity to language barriers

MN Fongwa et al., *Ethnicity Dis*, 2006;16(3):948-955.

Example: In-Depth Interviews

Exploring Depression in Diverse Groups

- Sample: White, South Asian, and Black Caribbean individuals in U.K.
- Explored nature and causes of depression, barriers to treatment
- Questions:
 - What does the word depression mean to you?
 - Do you consider depression to be an illness?
 - What do you think might cause depression?

V Lawrence et al., *Gerontologist*, 2006;46:23-32

Results of Lawrence et al., 2006

- Causes and symptoms varied by ethnic group
 - Mentally weak
 - Loss of independence
 - Sleep problems, loss of appetite
 - Social withdrawal
 - Loss of motivation
- Implications for choice of measures of depression

Resources for Exploring Concepts in Diverse Populations – CADC website

- Qualitative methods
 - Focus groups
 - Annotated bibliography of examples and how-to guidelines
- Includes our own work as methodological guidelines

<https://cadc.ucsf.edu/testing-methods>

Categories on Template for Reviewing Measures

- At the top
 - Concept defined by you, name of measure
- Categories for reviewing measure
 - Get to know the measure
 - Appropriateness for your sample
 - Conceptual adequacy
 - **Psychometric adequacy**
 - Practical issues

Review for Psychometric Adequacy

- Minimal standards met:
 - Sufficient variability
 - Adequate reliability/reproducibility
 - Evidence of construct validity
 - Evidence of sensitivity to change
- In original population and in samples similar to your target group

Indicators of Variability

- Range of scores
- Mean, median, mode
- Standard deviation (or standard error)
- Skewness statistic
- % at floor (lowest possible score)
- % at ceiling (highest possible score)

Understanding Variability of Possible Measure

In samples similar to yours...

- Where does the mean score fall in the range of possible scores?
 - Toward top, bottom?
 - In the middle?
- Are there published norms?

SF-36 Scores in Patients with Chronic Conditions

	Physical function	Role- physical	Mental health	Vitality (energy)
<u>MOS patients</u>				
Mean (SD)	80 (27)	75 (41)	71 (21)	54 (22)

JE Ware et al, *SF-36 Health Survey Manual and Interpretation Guide*, The Health Institute, 1993.

SF-36 Scores in Patients with Chronic Conditions

	Physical function	Role- physical	Mental health	Vitality (energy)
<u>MOS patients</u>				
Mean (SD)	80 (27)	75 (41)	71 (21)	54 (22)

Scores range from 0-100, higher scores indicate better health

SF-36 Scores in Patients with Chronic Conditions

	Physical function	Role- physical	Mental health	Vitality (energy)
<hr/>				
<u>MOS patients</u>				
Mean (SD)	80 (27)	75 (41)	71 (21)	54 (22)
<u>NORMS</u>				
Gen pop	84 (23)	81 (34)	75 (18)	61 (21)
Age 75+	53 (30)	45 (42)	74 (20)	50 (24)

Is There Room for Improvement?

- Know where your sample falls along the distribution (pretest)
- If mean score too high, little room for improvement
- PHQ-9 scores in choir study:
 - Mean 4.3 (possible range 0-24, higher is more depression)
 - Cut point for possible depression >10)

Importance of Reliability

- Necessary for validity
 - Low reliability (high measurement error) attenuates correlations with other variables
 - May conclude that two variables are not related when they are
- Greater reliability = greater power
 - The more reliable, the smaller sample size you need to detect an association

Internal Consistency Reliability: Cronbach's Alpha

- Requires multiple items hypothesized to measure same construct to calculate
- Extent to which all items measure the same construct (same latent variable)

Minimum Standards for Internal Consistency Reliability

- For group comparisons (e.g., regression, correlational analyses)
 - .70 or above is minimum
 - .80 is optimal
- Any information on reliability in samples similar to yours

JC Nunnally, *Psychometric Theory* 3rd ed, McGraw-Hill, 1994

Construct Validity Basics

- Does measure relate to other measures in hypothesized ways?
 - Do measures “behave as expected”?
- 3-step process
 - State hypothesis: direction and magnitude
 - Calculate correlations
 - Do results confirm hypothesis?

Review Potential Measures for Evidence of Validity

- Original publication of measure
 - Preliminary evidence of validity
- Subsequent applications
 - Measure performs “as expected”
 - Added evidence of validity
- Focus on validity in samples similar to yours

Measuring Sensitivity to Change

- Score is stable in those who are not changing
- Score changes in those who are actually changing (true change)
- One method
 - Identify groups “known” to change
 - Compare changes in measure across these groups

Sensitivity to Change Evidence for PHQ-9

- Classified patients with major depression (DSM-IV criteria) over time as:
 - Persistent depression
 - Partial remission
 - Full remission
- Examined change scores in these “known groups”

Löwe B et al. *Med Care*, 2004;42:1194-1201

Changes in PHQ-9 by Change in Depression at 6 Months

	<u>Mean change</u>	<u>Effect size</u>
Persistent depression	-4.4	-0.9
Partial remission	-8.8	-1.8
Full remission	-13.0	-2.6

Categories on Template for Reviewing Measures

- At the top
 - Concept defined by you, name of measure
- Categories for reviewing measure
 - Get to know the measure
 - Appropriateness for your sample
 - Conceptual adequacy
 - Psychometric adequacy
 - **Practical issues**

Obtain Permission to Use Measures!

- Public domain measures
 - Usually don't need permission
- Private or proprietary measures
 - Write to author or distributor
 - Allow 4-6 weeks to obtain permission
- Permission statements often found at source of measure

Review for Practical Considerations

- Method of administration
 - Matches context (funds, staffing)
- Cost to use or to score
 - Matches study resources
- Scoring instructions clear?
 - In publication? Scoring codebook?
 - Computer software?

Two CADDC Publications on Measurement Issues in Health Disparities Research

- Advancing health disparities research: can we afford to ignore measurement issues?
 - Stewart AL & Nápoles-Springer A, *Medical Care*, 2003;41(11):1207-1220.
- Health-related quality of life assessments in diverse population groups in the U.S.
 - Stewart AL & Nápoles-Springer A, *Medical Care*, 2000;38(9 Suppl):II102-124.



-

Thank you!