



# Field testing

Henrik Hein Lauridsen

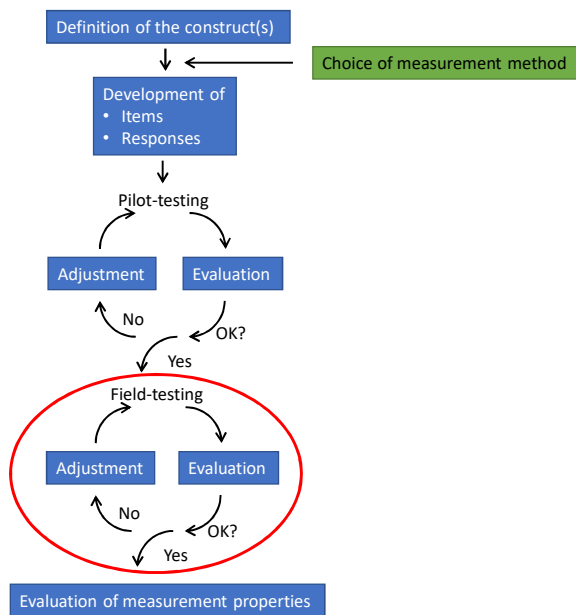
Associate professor

Research Unit for Clinical Biomechanics

University of Southern Denmark

1

## Questionnaire development



Department of Sports Science and Clinical Biomechanics



2

# What is field testing?

*"A further refinement of our instrument after pilot testing"*

- **Aim**
  - Item reduction
  - Insights into data structure

# Requirements

- **Formative models**
  - **Item importance**
    - Structured interviews
  - **Item characteristics**
    - Item scores, distribution, response curves
- **Reflective models**
  - **Item characteristics**
    - Item scores, distribution, response curves
  - **Dimensionality**
    - Factor analyses (EFA, CFA)
  - **Reproducibility**
    - Internal consistency
    - Reliability
    - Floor/ceiling effect

# Requirements

## • Formative models

- **Item importance** → **Target population ("rules of thumb")**
  - Structured interviews $N \geq 100$
- **Item characteristics** →  $N \geq 100$  (item scores & distribution)  
 $N = 300-500$  (response curves)
  - Item scores, distribution, response curves

## • Reflective models

- **Item characteristics** → See above
  - Item scores, distribution, response curves
- **Dimensionality** → 4-10 subjects/variable, min. 100<sup>1</sup>
  - Factor analyses (EFA, CFA)
- **Reproducibility** →  $N \geq 50$ 
  - Internal consistency
  - Reliability
  - Floor/ceiling effect

de Vet et al., 2005; Raczek et al., 1998

Department of Sports Science and Clinical Biomechanics



5

# Item importance

## • Structured interviews

- **Asthma QoL Questionnaire (impairment with asthma)**<sup>1</sup>
  - Symptoms, emotional problems, troublesome environmental stimuli, avoidance etc.
- **Two extra questions:**
  - Has 'item' been troublesome in the past year? (yes/no)
  - How important is the 'item'? (1 = not very important... 5 = extremely important)
- **Impact score**

**Impact = % endorsing item x importance (range 1 – 5)**

E.g. Shortness of breath: Troublesome = 92%, importance avg. = 3.60

Impact = 0.92 x 3.60 = 3.31

<sup>1</sup>Juniper et al., 1992

Department of Sports Science and Clinical Biomechanics



6

## Item characteristics

- Item scores, distribution and response curves

### *Oswestry Disability Index*

- Questionnaire measuring 'pain-related function' in low back pain patients
- 10 items scored on an ordinal scale from 0-5
- Score range [0-50] where 0 = no disability and 50 = high disability

7

## ODI item scores

Score range [0 - 5]

Item	Content	Mean (N=147)	Median (N=147)
1	Pain intensity	2.60	3
2	Personal care (washing, dressing etc)	1.02	1
3	Lifting	2.33	2
4	Walking	1.27	1
5	Sitting	1.99	2
6	Standing	1.92	2
7	Sleeping	1.33	1
8	Sex life (if applicable)	1.52	1
9	Social life	1.77	2
10	Travelling	1.64	1

Mean/median is low for all variables except 'pain intensity'

8

## ODI item scores

Score range [0 - 5]

Item	Content	Mean (N=147)	Median (N=147)	Missing items (% of 147)
1	Pain intensity	2.60	3	0.0
2	Personal care (washing, dressing etc)	1.02	1	1.4
3	Lifting	2.33	2	1.4
4	Walking	1.27	1	1.4
5	Sitting	1.99	2	1.4
6	Standing	1.92	2	0.7
7	Sleeping	1.33	1	2.7
8	Sex life (if applicable)	1.52	1	15.0
9	Social life	1.77	2	2.7
10	Travelling	1.64	1	0.7

What is acceptable?

Rule of thumb<sup>1</sup>:

< 3% = acceptable

> 15% = unacceptable

<sup>1</sup>Measurement in Medicine, de Vet et al. 2011

Department of Sports Science and Clinical Biomechanics



9

## ODI item distribution

Item	Content	Distribution (%) of response options					
		0	1	2	3	4	5
1	Pain intensity	0	6.8	40.8	38.8	12.9	0.7
2	Personal care (washing, dressing etc)	27.9	47.6	19.1	2.7	0	1.4
3	Lifting	4.8	26.5	18.4	29.9	18.4	0.7
4	Walking	30.6	33.3	16.3	15.0	2.7	0.7
5	Sitting	9.5	18.4	44.2	18.4	6.8	1.4
6	Standing	10.2	30.6	25.2	23.1	10.2	0.7
7	Sleeping	12.9	52.4	21.1	8.8	2.0	0
8	Sex life (if applicable)	24.5	25.9	14.3	9.5	8.2	2.7
9	Social life	16.3	21.8	31.3	23.8	4.1	2.7
10	Travelling	9.5	40.8	30.6	13.6	3.4	1.4

→ Looks like a normal distribution

→ Looks like very skewed distributions

Department of Sports Science and Clinical Biomechanics

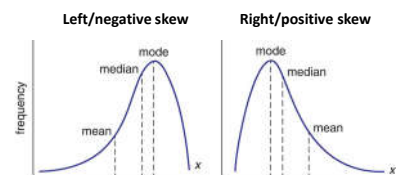


10

# ODI item distribution

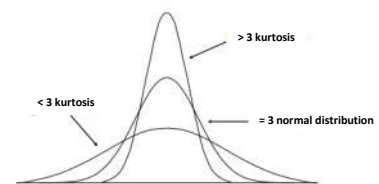
## • Skewness

- A measure of the lack of symmetry of a distribution
  - 0 = symmetric distribution
  - - coefficient  $\rightarrow$  median  $>$  mean  $\rightarrow$  left or negative skew
  - + coefficient  $\rightarrow$  median  $<$  mean  $\rightarrow$  right or positive skew



## • Kurtosis

- A measure of peakedness of a distribution
  - $<$  3  $\rightarrow$  flatter distribution
  - 3 = normal distribution
  - $>$  3  $\rightarrow$  more peaked distribution



Department of Sports Science and Clinical Biomechanics



11

# ODI item distribution

- = left skewed  
+ = right skewed

3 = normal distribution  
 $<$ 3 = flat curve  
 $>$ 3 = peaked curve

Distribution (%) over response options										
Item	Content	0	1	2	3	4	5	Skewness	Kurtosis	
1	Pain intensity	0	6.8	40.8	38.8	12.9	0.7	0.2	2.7	$\rightarrow$ $\approx$ normal distrib.
2	Personal care (washing, dressing etc)	27.9	47.6	19.1	2.7	0	1.4	1.3	6.7	$\rightarrow$ $\approx$ right skewed and peaked
3	Lifting	4.8	26.5	18.4	29.9	18.4	0.7	-0.1	1.9	$\rightarrow$ $\approx$ normal distrib. and flat
4	Walking	30.6	33.3	16.3	15.0	2.7	0.7	0.7	2.7	
5	Sitting	9.5	18.4	44.2	18.4	6.8	1.4	0.2	3.1	
6	Standing	10.2	30.6	25.2	23.1	10.2	0.7	0.1	2.1	
7	Sleeping	12.9	52.4	21.1	8.8	2.0	0	0.8	3.6	
8	Sex life (if applicable)	24.5	25.9	14.3	9.5	8.2	2.7	0.8	2.6	
9	Social life	16.3	21.8	31.3	23.8	4.1	2.7	-0.1	2.1	
10	Travelling	9.5	40.8	30.6	13.6	3.4	1.4	0.1	3.6	

Department of Sports Science and Clinical Biomechanics



12

## ODI item response curves (IRT)

### Item threshold

Is the point where the probability of a response in either one of two adjacent categories is 50%

### We expect

Patients with a high amount of the measured trait should have a higher probability of endorsing say category 3 compared to category 2

### Oswestry Disability Index

#### Afsnit 1: Smerter

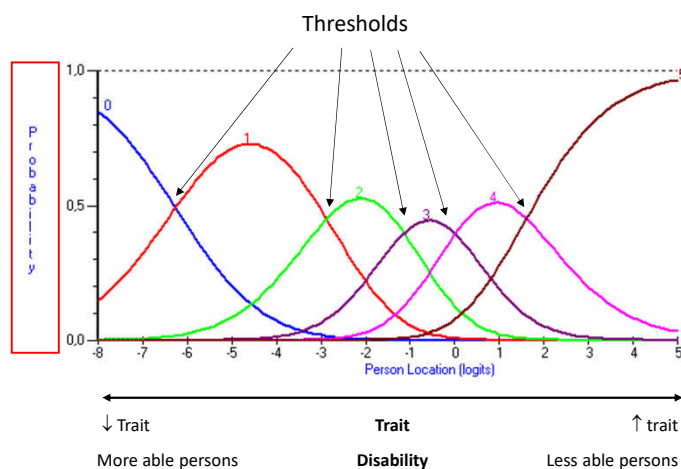
- 0  Jeg har ingen smerter for øjeblikket
- 1  Smerterne er meget svage for øjeblikket
- 2  Smerterne er moderate for øjeblikket
- 3  Smerterne er forholdsvis kraftige for øjeblikket
- 4  Smerterne er meget kraftige for øjeblikket
- 5  Smerterne er de værst tænkelige for øjeblikket

Department of Sports Science and Clinical Biomechanics



13

## ODI item response curves



### Oswestry Disability Index

#### Afsnit 1: Smerter

- Jeg har ingen smerter for øjeblikket
- Smerterne er meget svage for øjeblikket
- Smerterne er moderate for øjeblikket
- Smerterne er forholdsvis kraftige for øjeblikket
- Smerterne er meget kraftige for øjeblikket
- Smerterne er de værst tænkelige for øjeblikket

Department of Sports Science and Clinical Biomechanics

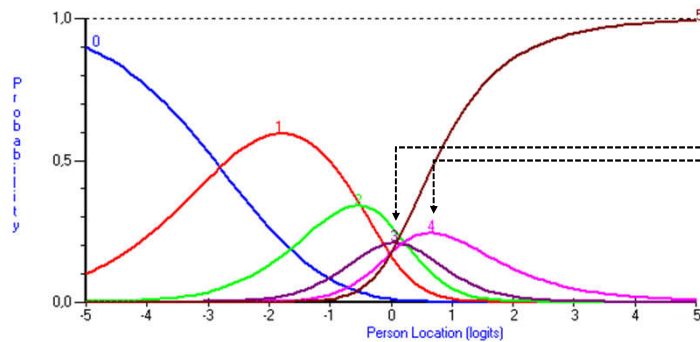


14

## ODI item response curves

### Afsnit 10: Rejse

- 0  Jeg kan rejse hvorhen jeg vil uden smerter
- 1  Jeg kan rejse hvorhen jeg vil, men det giver mig flere smerter
- 2  Smerterne er slemme, men jeg kan godt klare over 2 timers rejse
- 3  Smerterne begrænser mine rejser til mindre end 1 time
- 4  Smerterne begrænser mine rejser til korte, nødvendige rejser under 30 minutter
- 5  Smerterne hindrer mig i at rejse, undtagen for at få behandling



Disordered thresholds

Category 3 and 4 are unlikely to be endorsed

Department of Sports Science and Clinical Biomechanics

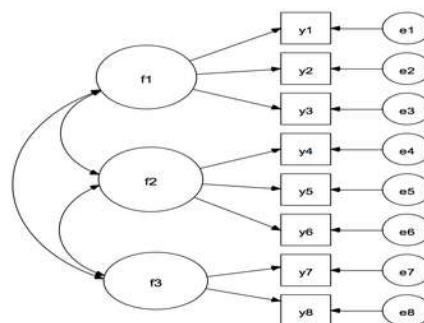


15

## Dimensionality

- Factor analyses (EFA, CFA)

- Eva Ørnbøl



Department of Sports Science and Clinical Biomechanics



16



## Reproducibility

- Internal consistency
- Reliability
- Floor/ceiling effect
- Done AFTER factor analysis



## Internal consistency

*“Degree of inter-relatedness among items”*

*(COSMIN 2011)*

- Used on multi-item questionnaires
- Assumes that all items measure the same concept
  - Item scores are correlated
  - No items are irrelevant
  - Unidimensionality
- Cronbach's  $\alpha$

# Alpha output

```
. alpha osw*, std item label
```

```
Test scale = mean(standardized items)
```

Item	Obs	Sign	item-test corr.	item-rest corr.	interitem corr.	alpha	Label
osw1	147	+	0.6645	0.5659	0.4044	0.8594	Pain intensity
osw2	145	+	0.7125	0.6250	0.3949	0.8545	Personal care
osw3	145	+	0.6149	0.5071	0.4143	0.8643	Lifting
osw4	145	+	0.6742	0.5774	0.4026	0.8584	Walking
osw5	145	+	0.6434	0.5400	0.4088	0.8616	Sitting
osw6	146	+	0.6768	0.5805	0.4022	0.8582	Standing
osw7	143	+	0.6078	0.4986	0.4158	0.8649	Sleeping
osw8	125	+	0.6842	0.5960	0.3991	0.8567	Sexlife (if applicable)
osw9	143	+	0.7753	0.7026	0.3833	0.8484	Social life
osw10	146	+	0.7446	0.6650	0.3884	0.8511	Travelling
Test scale					0.4014	0.8702	mean(standardized items)

Average interitem correlation      Cronbach's  $\alpha$

<http://methods.sagepub.com/reference/encyc-of-research-design/n201.xml>

Pearson corr. calculated for pairs of scores where one item of each pair is an item score and the other item is the total test score.

Item correlation with sumscore of the rest of the items

Avg. interitem correlation with all other items excluding itself.

Should be roughly the same for all items

Rule of thumb: Should not be > 0.8 → sign of multicollinearity

If substantially higher than avg. interitem correlation → maybe misfitting

Alpha excluding the one item

If substantially higher than Chronbach's  $\alpha$  → maybe misfitting

19

# Alpha output – item removal

```
alpha osw*, std item label
```

```
Test scale = mean(standardized items)
```

Item	Obs	Sign	item-test corr.	item-rest corr.	interitem corr.	alpha	Label
osw1	147	+	0.6645	0.5659	0.4044	0.8594	Pain intensity
osw2	145	+	0.7125	0.6250	0.3949	0.8545	Personal care
osw3	145	+	0.6149	0.5071	0.4143	0.8643	Lifting
osw4	145	+	0.6742	0.5774	0.4026	0.8584	Walking
osw5	145	+	0.6434	0.5400	0.4088	0.8616	Sitting
osw6	146	+	0.6768	0.5805	0.4022	0.8582	Standing
osw7	143	+	0.6078	0.4986	0.4158	0.8649	Sleeping
osw8	125	+	0.6842	0.5960	0.3991	0.8567	Sexlife (if applicable)
osw9	143	+	0.7753	0.7026	0.3833	0.8484	Social life
osw10	146	+	0.7446	0.6650	0.3884	0.8511	Travelling
Test scale					0.4014	0.8702	mean(standardized items)

Lets remove item 9

20

# Alpha output

alpha osw1 osw1-osw8 osw10 if baop==1, std item label

	<b>Interitem</b>	
	<b>corr</b>	<b>alpha</b>
<b>Item 9:</b>	<b>0.3833</b>	<b>0.8484</b>

Test scale = mean(standardized items)

Item	Obs	Sign	item-test corr.	item-rest corr.	interitem corr.	alpha	Label
osw1	147	+	0.6750	0.5653	0.3830	0.8324	Pain intensity
osw2	145	+	0.7218	0.6238	0.3723	0.8260	Personal care
osw3	145	+	0.6228	0.5012	0.3949	0.8392	Lifting
osw4	145	+	0.6815	0.5727	0.3816	0.8316	Walking
osw5	145	+	0.6476	0.5313	0.3893	0.8360	Sitting
osw6	146	+	0.6758	0.5652	0.3828	0.8323	Standing
osw7	143	+	0.6165	0.4938	0.3963	0.8401	Sleeping
osw8	125	+	0.6724	0.5686	0.3819	0.8317	Sexlife (if applicable)
osw10	146	+	0.7420	0.6504	0.3678	0.8231	Travelling
Test scale					<b>0.3833</b>	<b>0.8484</b>	mean(standardized items)

21

# Reliability

- Item reliability measured by ICC version 2,1 - agreement
- If < 0.50 consider removing item

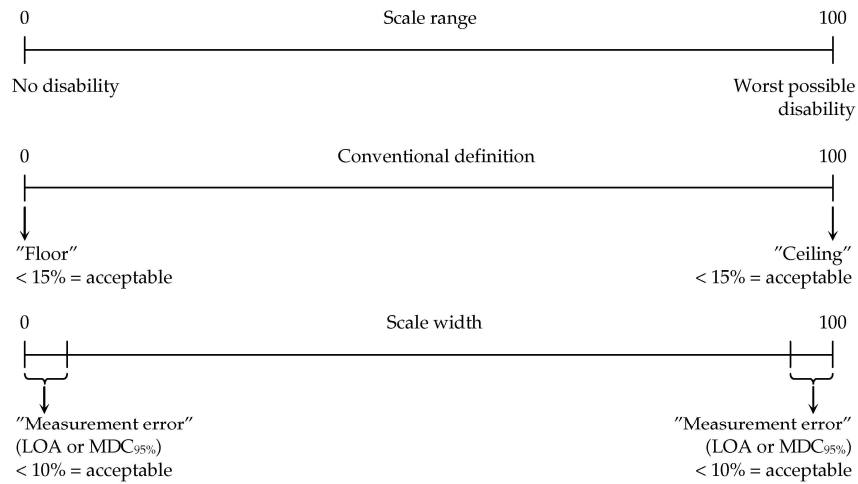
Item	Content	ICC 2,1 (agreement)
1	Pain intensity	0.54
2	Personal care (washing, dressing etc)	0.61
3	Lifting	0.67
4	Walking	0.79
5	Sitting	0.66
6	Standing	0.74
7	Sleeping	0.74
8	Sex life (if applicable)	0.85
9	Social life	0.68
10	Travelling	0.66

Low but acceptable

Fine

22

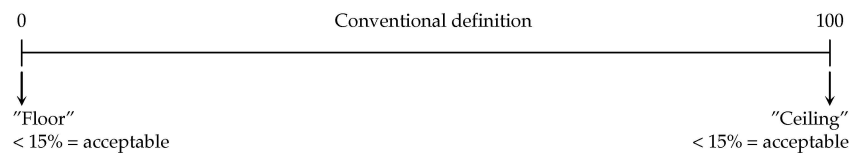
## Floor and ceiling effect



Davidson et al. 2002

23

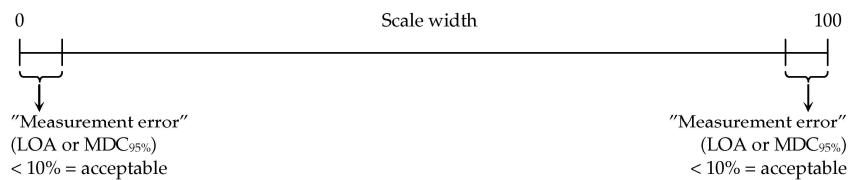
## Floor and ceiling effect



	N	Oswestry Disability Index (%)
Floor effect	147	0 (0.0)
Ceiling effect	147	0 (0.0)

24

## Floor and ceiling effect



	N	Oswestry Disability Index (%)
Floor effect	147	9 (6.1)
Ceiling effect	147	1 (0.7)

25

## Requirements

- **Formative models**

- Item importance
  - Structured interviews
- Item characteristics
  - Item scores, distribution, response curves

- **Reflective models**

- Item characteristics
  - Item scores, distribution, response curves
- Dimensionality
  - Factor analyses (EFA, CFA)
- Reproducibility
  - Internal consistency
  - Reliability
  - Floor/ceiling effect

de Vet et al., 2005; Raczek et al., 1998

Department of Sports Science and Clinical Biomechanics



26

# Questions

