A Clinically Meaningful Theory of Outcome Measurement in Rehabilitation

Robert W. Massof, PhD

Professor of Ophthalmology and Neuroscience

Lions Vision Research & Rehabilitation Center

Wilmer Eye Institute

Johns Hopkins University School of Medicine

Financial Disclosure

I have no financial interests related to the topic of this presentation

Other interests:

- I serve as a reviewer for the Food and Drug Administration (SGE)
- Research supported by grant EY022322 from the National Eye Institute, National Institutes of Health

Measuring clinical outcomes of behavioral and symptom-targeted interventions

- Mental health workers, rehabilitation therapists, nurses, and many other types of health care service providers often treat their patients with the aim of changing aspects of the patient's behavior, feelings, symptoms, or daily functioning.
- In these cases, the treatment is goal-directed and targets specific problems experienced by the patient.
- Goals of treatment are <u>individualized</u> to meet each patient's personal priorities and to be feasible given each patient's capabilities.
- How does one measure clinical outcomes when the treatment is customized for each patient?

Goal Attainment Scaling (GAS)

- Need an outcome measure that recognizes and accommodates treatment plans targeted to the multiple personal goals of intervention and different capabilities of individual patients/clients
- GAS first developed in the late 1960’s by Thomas Kiresuk, a clinical psychologist, and Robert Sherman, a statistician, to serve this need

Kiresuk TJ, Sherman RE. Goal Attainment Scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health Journal* 1968;**4**(6):443-453.

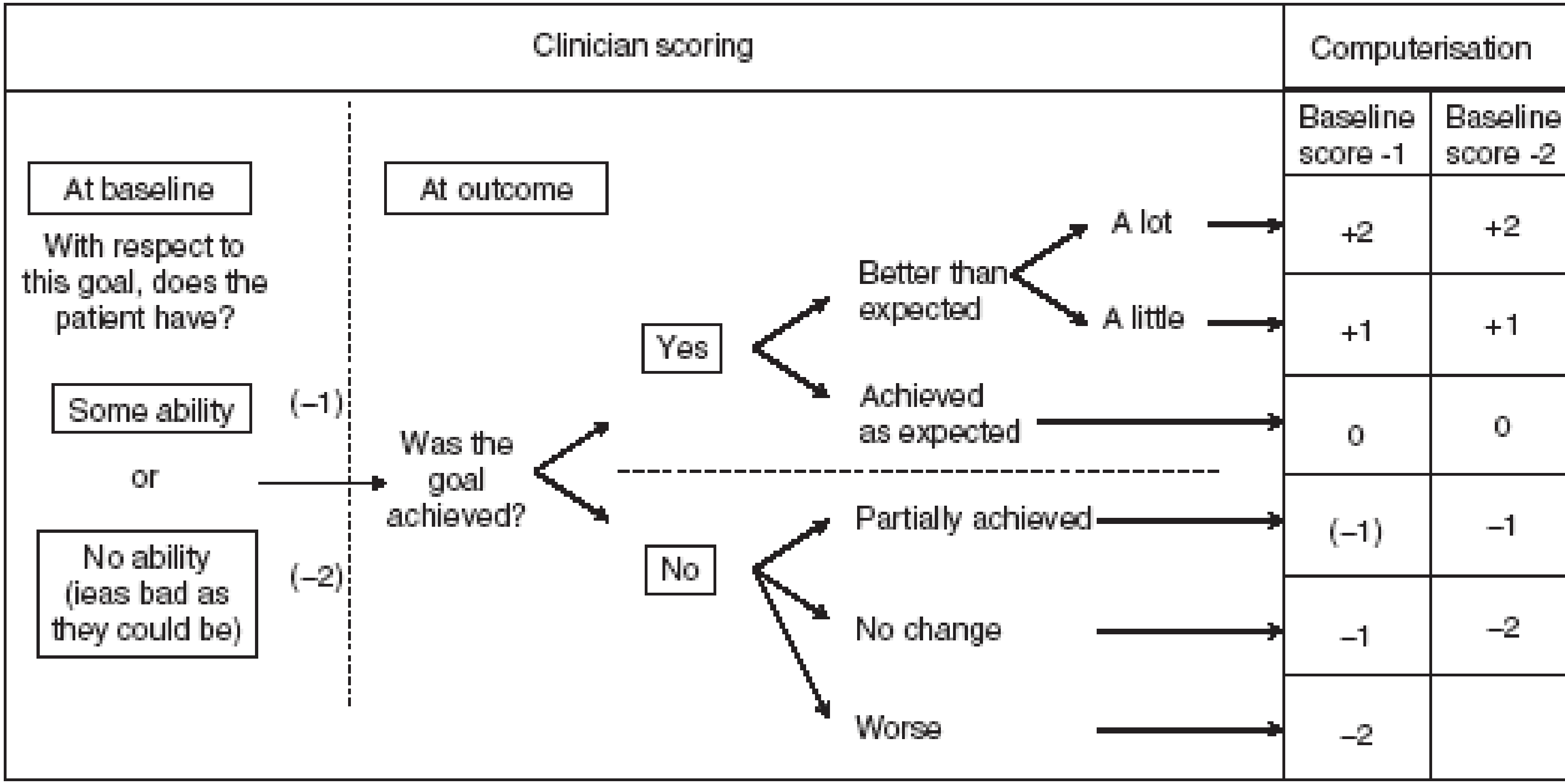


Figure 1 Algorithm for converting verbal scoring by clinicians to the 5-point goal attainment scores. This algorithm allows clinicians to record goal attainment without reference to the numeric scores, and so avoids the perceived negative connotations of zero and minus scores. Providing the level at baseline is known, 'partial achievement', 'no change' and worse can be translated by computerization outside the clinical arena. This incidentally offers the opportunity to compare the effect of using different scoring systems such as (–1, –2 and –3) or (–0.5, –1 and –2) and this work is currently underway.

Turner-Stokes L. Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clinical Rehabilitation* 2009;**23:**362-370.

Problems with GAS

- GAS relies on therapist ratings, which necessarily incorporate therapist-specific biases
- Outcomes are scaled relative to the choice of goals in the rehabilitation plan
  - A rating of 0 means the patient is at the goal
  - Not all goals are the same, so the meaning of the scale changes across goals
- Need a theory that explicitly identifies all relevant variables and can be reduced to a valid measurement model

Outline of Presentation

- Overview of how to model outcome measurements from rating scale responses (start with more familiar patient self-report)
- Intervention-specific differential item functioning (DIF)
  - Modeling effects of interventions that target item difficulty
  - Examples of intervention-specific DIF and combined effects from a low vision rehabilitation RCT
- Critical Analysis of Goal Attainment Scaling (GAS)
- New approach to GAS

Rating scale questionnaires produce conjoint observations

- Patient-reported functional ability questionnaires consist of a set of items, each of which describes an activity.
- The person responds with an ordered category.
- The items serve as the standard references against which we will compare each person.

VF-14

- Read small print such as labels on medicine bottles, a telephone book, or food labels
- Read ordinary newsprint
- Read large-print book, or large-print newspaper, or numbers on a telephone
- Recognize people when they are close to you
- See steps, stairs, or curbs
- Read traffic signs, street signs, or store signs
- Do fine handwork like sewing, knitting, crocheting, or carpentry
- Write checks or fill out forms
- Play games such as bingo, dominos, card games, or mah-jongg
- Take part in sports like bowling, handball, tennis, or golf
- Cook
- Watch TV
- Drive During the daytime
- Drive at night

No difficulty	Extreme difficulty
Some difficulty	Unable to do
Moderate difficulty	Not applicable

Measuring functional ability

- Functional ability is a latent variable (trait of the person)
- Each person has some level of functional ability called the "person measure": P_n for person n
- Each activity requires some level of functional ability to be performed with ease called the "item measure": I_j for item j
- Functional reserve = difference between person's functional ability and ability required by the activity: $R_{nj} = P_n - I_j$
- Perceived difficulty of performing the activity is expected to depend on functional reserve
- To respond with difficulty rating "x", functional reserve must fall in the interval for x: $C_x < R_{nj} < C_{x+1}$ where C_x is the criterion functional reserve for responding with rating category x

But P_n, I_j, C_x are fixed variables

- Deterministic measurements
  - Functional ability is a fixed property of the person P_n
  - Required functional ability is a fixed property of the item I_j
  - The response threshold, C_x, is a fixed property of the interval x
- In the real world these variables are inferred from the observations and there is uncertainty about the inferred values

Item 5	Score
Obs 1	1
Obs 2	0
Obs 3	0
Obs 4	0

Item 5 score = 1

Items I_j

Categories C_x

Assumptions of measurement theory

- P_n is a fixed trait of person n
- I_{nj} is person n's estimate of required functional ability of item j
- I_j is the expected required functional ability of item j (average value of I_{nj} across people in the target population):

$$I_j = \sum_{n=1}^{N} \frac{I_{nj}}{N}$$

- e_{nj} is a random between person and item variable:

$$e_{nj} = I_{nj} - I_j$$

Assumptions of measurement theory

- C_{nx} is person n's response criterion for using rating category x
- C_x is the expected response criterion for response category x (average value of C_{nx} across people)
- e_{nx} is a random between person and category variable $e_{nx} = C_{nx} - C_x$

Assumptions of measurement theory

- Person *n* uses difficulty ratings to estimate the magnitude of his own functional reserve for item *j*

$$R_{nj} = P_n - I_j - e_{nj}$$

- To respond with rating category *x*, functional reserve must be greater than the threshold for *x* and less than the threshold for *x*+1

$$C_x + e_{nx} < P_n - I_j - e_{nj} < C_{x+1} + e_{nx+1}$$

- Item Response Theory (IRT) models assume that the response thresholds are fixed, i.e.,

$$e_{nx} = 0$$

Assumptions of measurement theory

- To respond with rating category *x*, functional reserve must be greater than the threshold for *x* and less than the threshold for *x*+1

$$C_x + e_{nx} < P_n - I_j - e_{nj} < C_{x+1} + e_{nx+1}$$

- Define a new random term

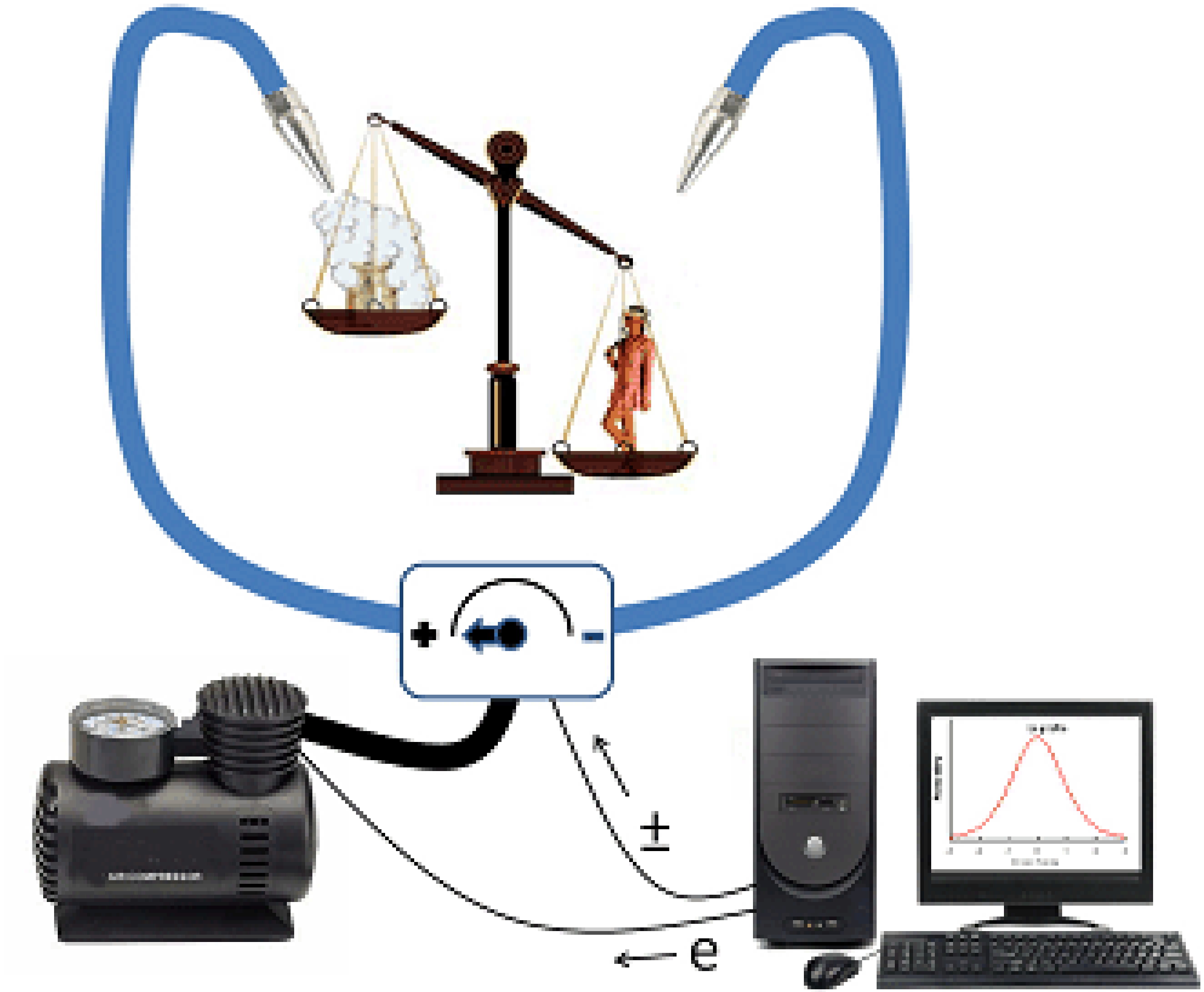
$$e_{njx} = e_{nj} + e_{nx}$$

- Therefore, the simplified measurement theory is

$$C_x + e_{njx} < P_n - I_j < C_{x+1} + e_{njx+1}$$

- Rasch theory assumes statistical independence of e_{njx}

Addition of randomly generated error



Rating scale questionnaires produce conjoint observations

- Patient-reported functional ability questionnaires consist of a set of items, each of which describes an activity.
- The person responds with an ordered category.
- The items serve as the standard references against which we will compare each person.

VF-14

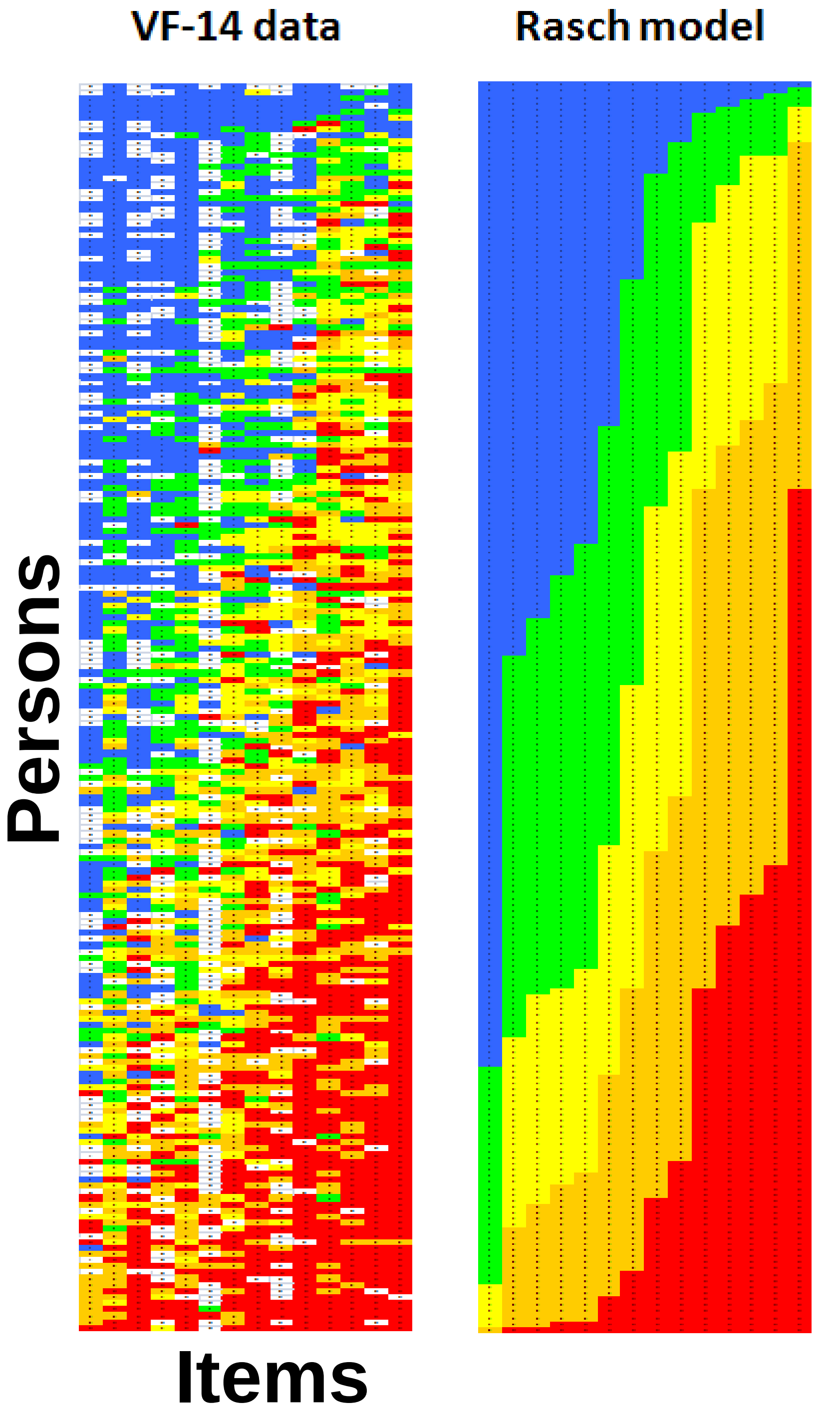
- Read small print such as labels on medicine bottles, a telephone book, or food labels
- Read ordinary newsprint
- Read large-print book, or large-print newspaper, or numbers on a telephone
- Recognize people when they are close to you
- See steps, stairs, or curbs
- Read traffic signs, street signs, or store signs
- Do fine handwork like sewing, knitting, crocheting, or carpentry
- Write checks or fill out forms
- Play games such as bingo, dominos, card games, or mah-jongg
- Take part in sports like bowling, handball, tennis, or golf
- Cook
- Watch TV
- Drive During the daytime
- Drive at night

No difficulty	Extreme difficulty
Some difficulty	Unable to do
Moderate difficulty	Not applicable

Maximum Likelihood Estimation of Fixed Variables

- P for each person
- I for each item
- C for each threshold

No difficulty
Some difficulty
Moderate difficulty
Extreme difficulty
Unable to do
Not applicable

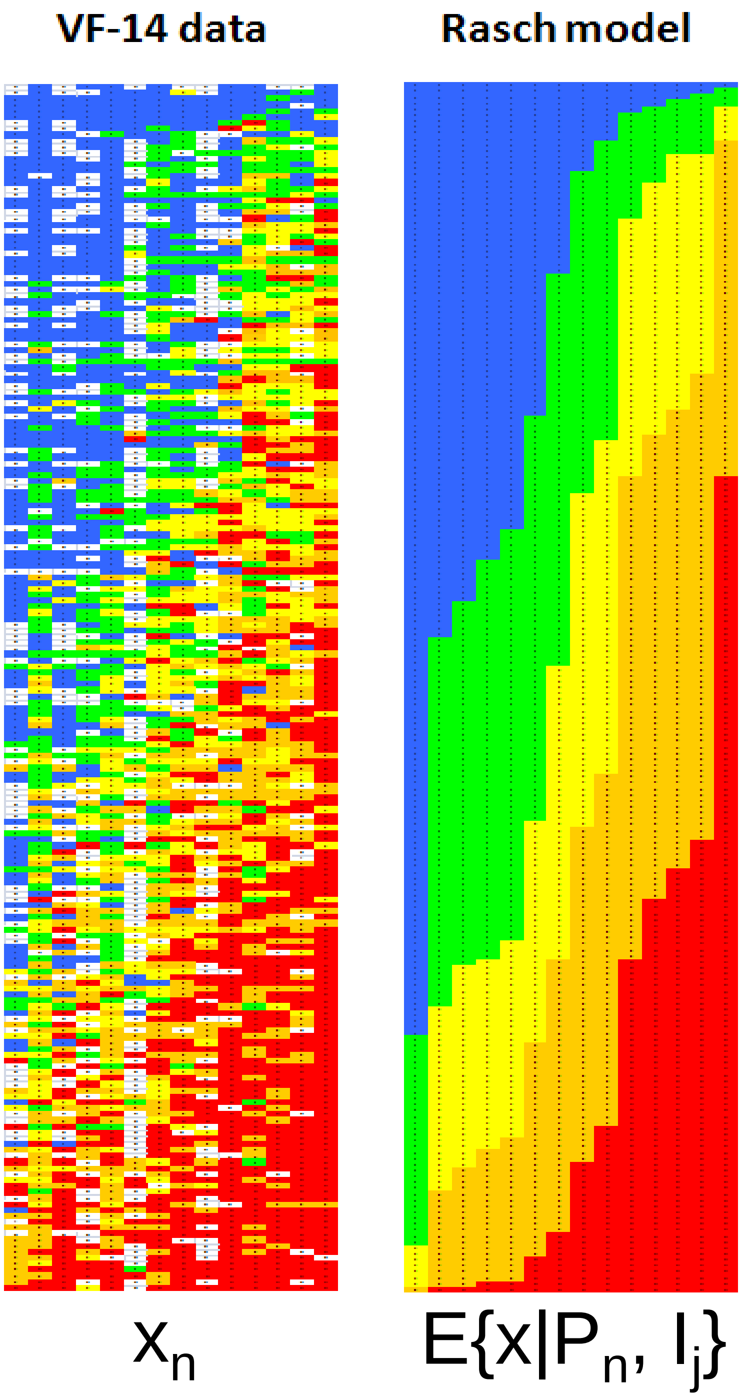


Validity
(Accuracy of assumptions)

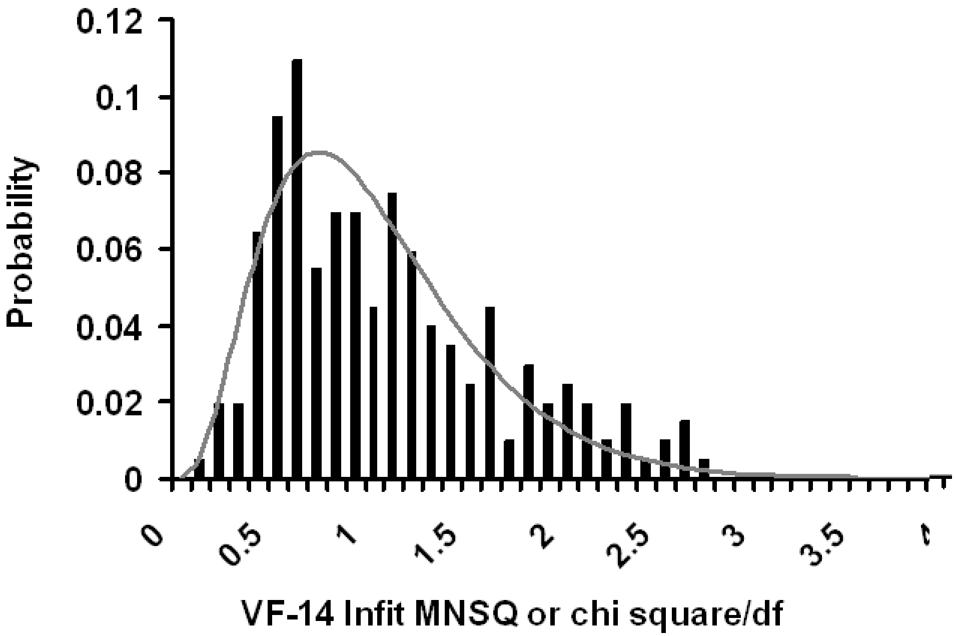
Mean square fit statistic for each person (tests assumption that all stochastic variance can be attributed to a single source, viz., e_{njx})

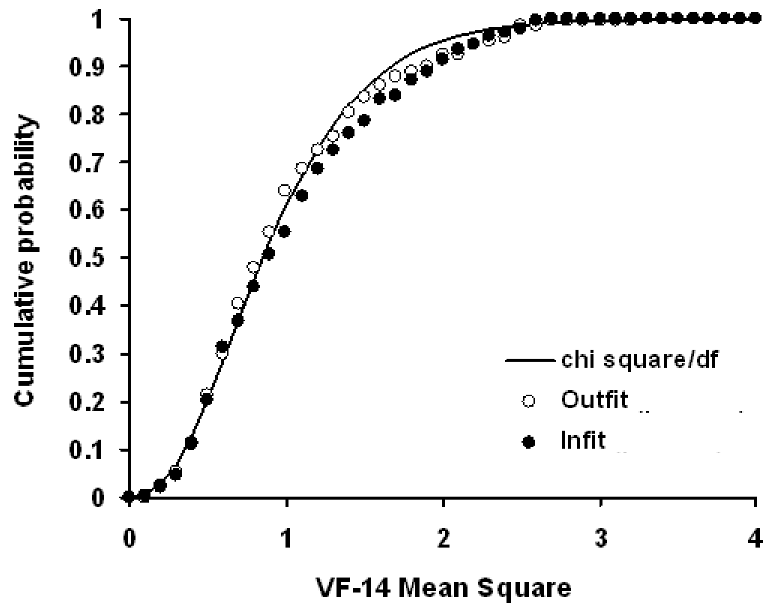
$$\frac{\text{Mean square residual for each person}}{\text{Model's predicted variance for each person}}$$

$$\frac{\sum_{j=1}^{J}\left(x_{nj}-E\left\{x \mid P_n, I_j\right\}\right)^2}{\sum_{j=1}^{J}E\left\{x^2 \mid P_n, I_j\right\}-E\left\{x \mid P_n, I_j\right\}^2}=\frac{\chi^2}{df}$$



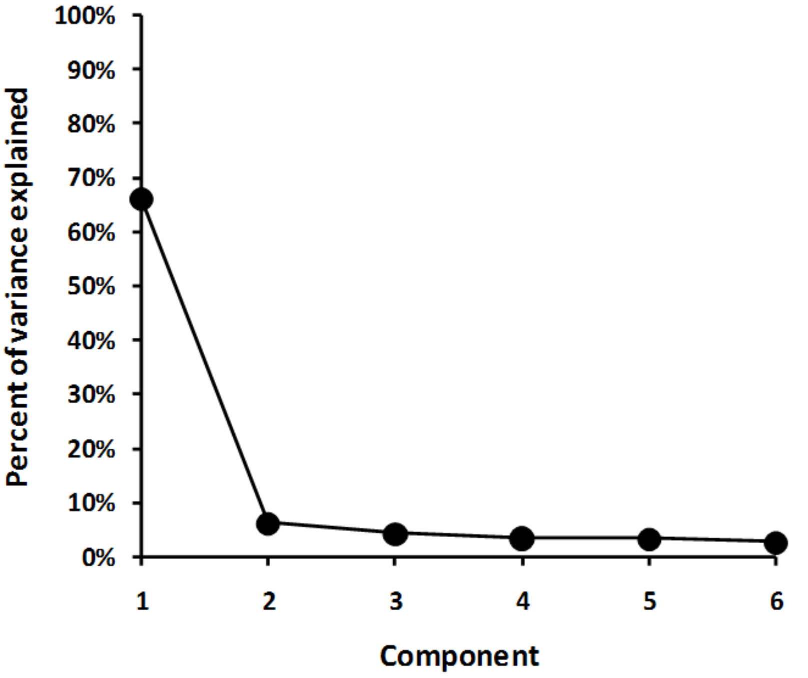
Testing validity of measure of visual ability in low vision patients with VF-14





Principal components analysis of residuals

- Person measure is first principal component (explains 67% of variance)
- Remaining variance is random noise (e_{njx}), which is expected by the model



Possible effects of intervention

- Change the person measure:

$$P_n(t_0) \rightarrow P_n(t_0) + \Delta P_n$$

$$\Delta P_n = P_n(t) - P_n(t_0)$$

- Change the item measure:

$$I_j(t_0) \rightarrow I_j(t_0) + \Delta I_{nj}$$

$$\Delta I_{nj} = I_{nj}(t) - I_j(t_0)$$

- Change in the person's response bias:

$$C_x(t_0) \rightarrow C_x(t_0) + B_n$$

$$B_n = \sum_{x=1}^{m} \frac{\Delta C_x}{m}$$

Change in functional reserve

$$C_x(t_0)+e_{njx} < P_n(t_0)-I_j(t_0) < C_{x+1}(t_0)+e_{njx+1}$$

$$C_x(t_0)+e_{njx} < R_{nj}(t_0) < C_{x+1}(t_0)+e_{njx+1}$$

$$C_x(t_0)+e_{njx}+B_n < R_{nj}(t_0)+\Delta P_n-\Delta I_{nj} < C_{x+1}(t_0)+e_{njx+1}+B_n$$

$$C_x(t_0)+e_{njx} < R_{nj}(t_0)+\Delta P_n-\Delta I_{nj}-B_n < C_{x+1}(t_0)+e_{njx+1}$$

$$C_x(t_0)+e_{njx} < R_{nj}(t_0)+\Delta R_{nj} < C_{x+1}(t_0)+e_{njx+1}$$

$$\Delta R_{nj}=\Delta P_n-\Delta I_{nj}-B_n$$

Single outcome measure
Average change in functional reserve

$$\Delta R_{nj} = \Delta P_n - \Delta I_{nj} - B_n$$

$$\overline{\Delta R_n} = \sum_{j=1}^{J} \frac{\Delta R_{nj}}{J} = \sum_{j=1}^{J} \frac{\Delta P_n - \Delta I_{nj} - B_n}{J}$$

$$\overline{\Delta R_n} = \Delta P_n - B_n - \overline{\Delta I_n}$$

$$\overline{\Delta I_n} = \sum_{j=1}^{J} \frac{\Delta I_{nj}}{J}$$

Anchor item measures and response category thresholds to baseline values

$$\overline{\Delta R_n} = \Delta P_n - B_n - \overline{\Delta I_n}$$

$$\Delta P_n = \overline{\Delta R_n} + B_n + \overline{\Delta I_n}$$

$$I_{nj}(t) \equiv I_j(t_0) \therefore \overline{\Delta I_n} \equiv 0$$

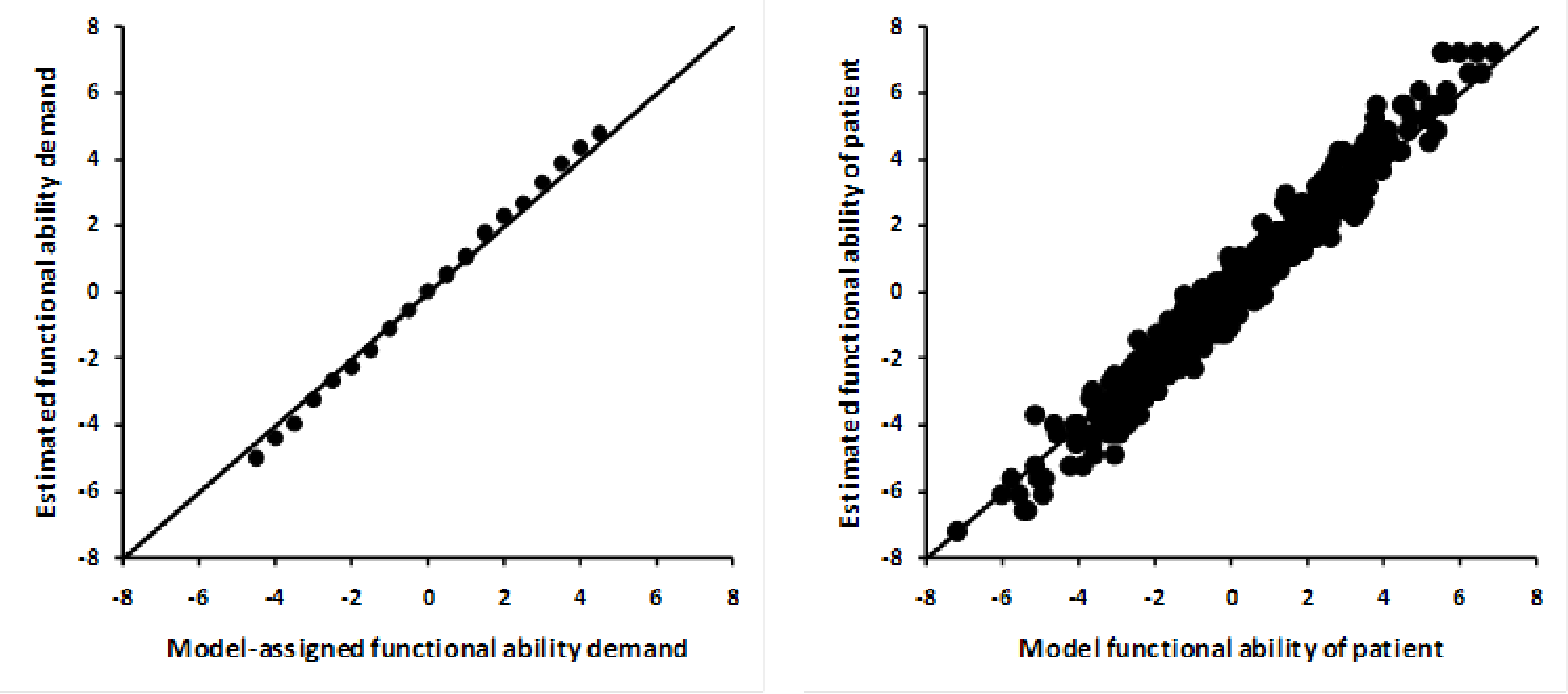
$$C_x(t) \equiv C_x(t_0) \therefore B_n \equiv 0$$

$$\Delta P_n = \overline{\Delta R_n}$$

Simulation

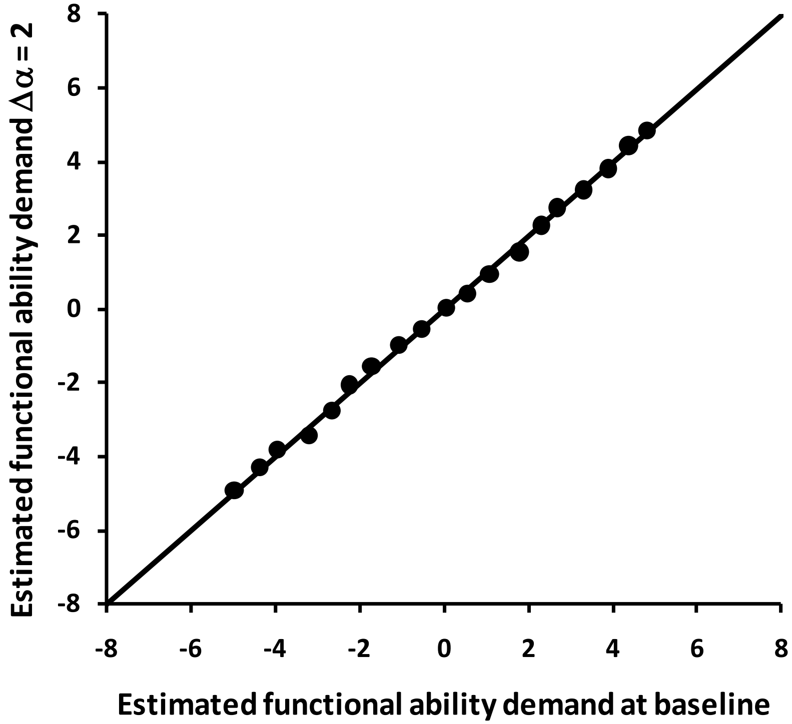
- 500 persons, 19 items, 4 response categories
- $P_n(t_0)$ is normally distributed with mean = 0 logit and sd = 2.5 logit
- $I_j(t_0)$ ranges from -4.5 to 4.5 logits in 0.5 logit steps
- $C_1(t_0) = -2; C_2(t_0) = 0; C_3 = 2$
- e_{njx} is normally distributed, $\sim N\left(0, \pi/\sqrt{3}\right)$, with a constant diagonal covariance matrix
- $C_x(t_0) + e_{njx} < P_n(t_0) - I_j(t_0) < C_{x+1}(t_0) + e_{njx+1}$

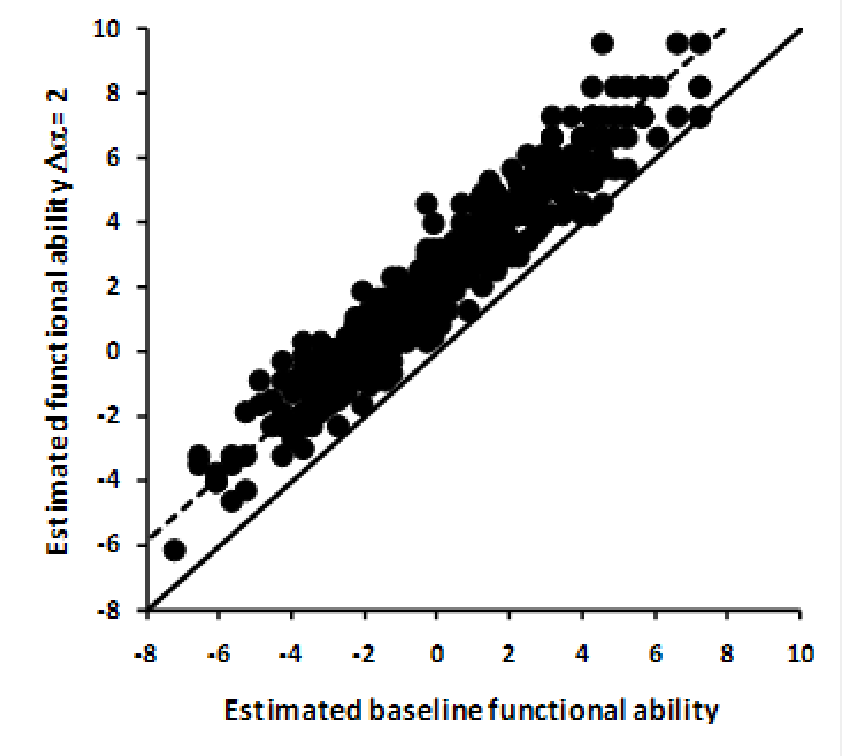
Simulation of baseline responses



$$C_x(t_0) + e_{njx} < P_n(t_0) - I_j(t_0) < C_{x+1}(t_0) + e_{njx+1}$$

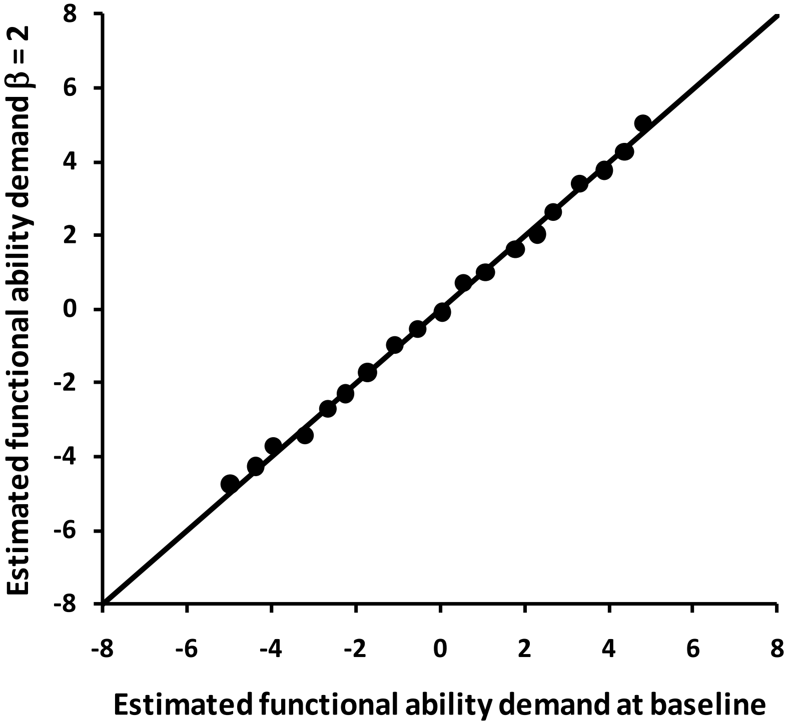
Simulation $\Delta P_n = 2$

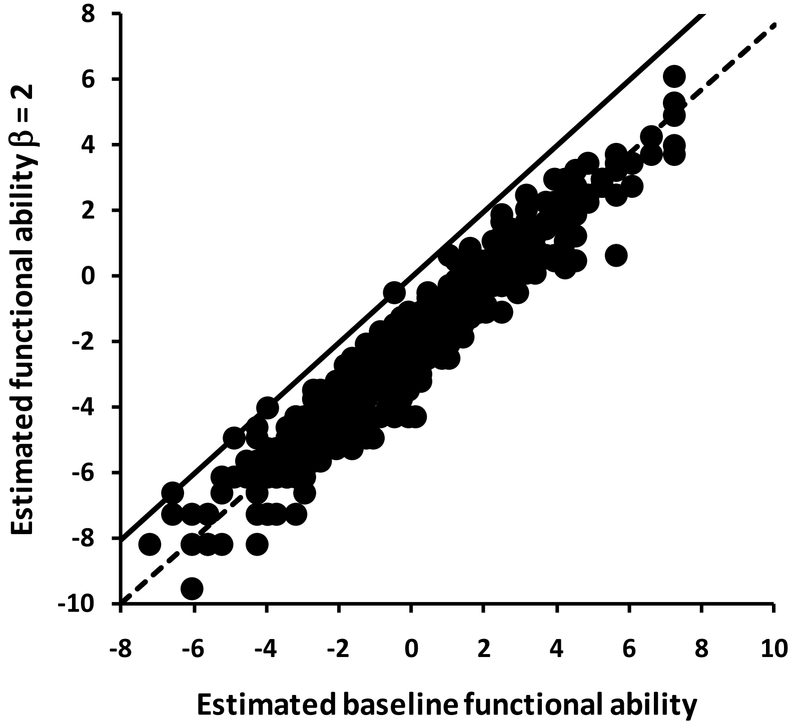




$$C_x(t_0) + e_{njx} < P_n(t_0) + \Delta P_n - I_j(t_0) < C_{x+1}(t_0) + e_{njx+1}$$

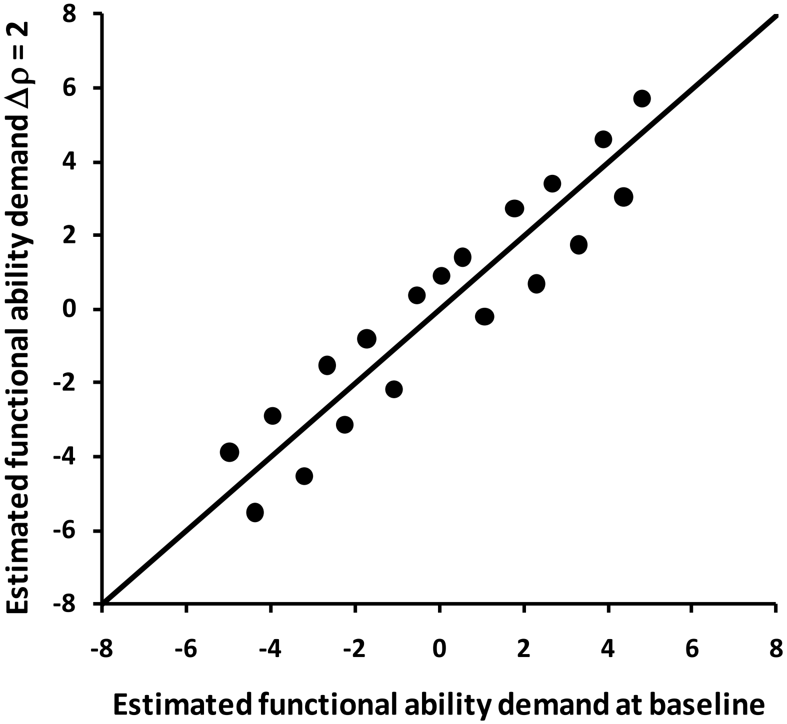
Simulation $B_n = 2$

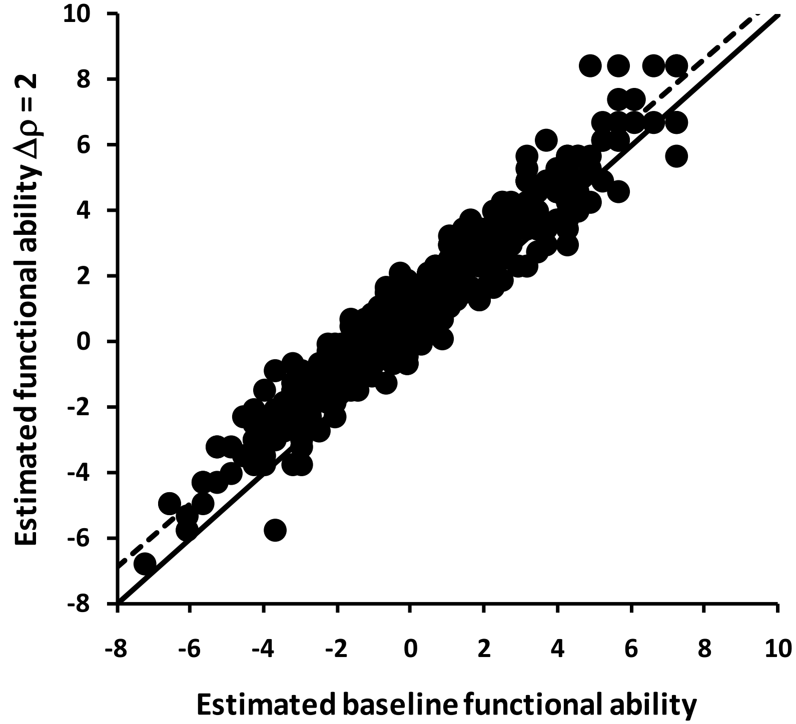




$$C_x(t_0) + e_{njx} + B_n < P_n(t_0) - I_j(t_0) < C_{x+1}(t_0) + e_{njx+1} + B_n$$

Simulation $\Delta I_j = -2$ for 8 items and $\Delta I_j = 0$ for 11 items

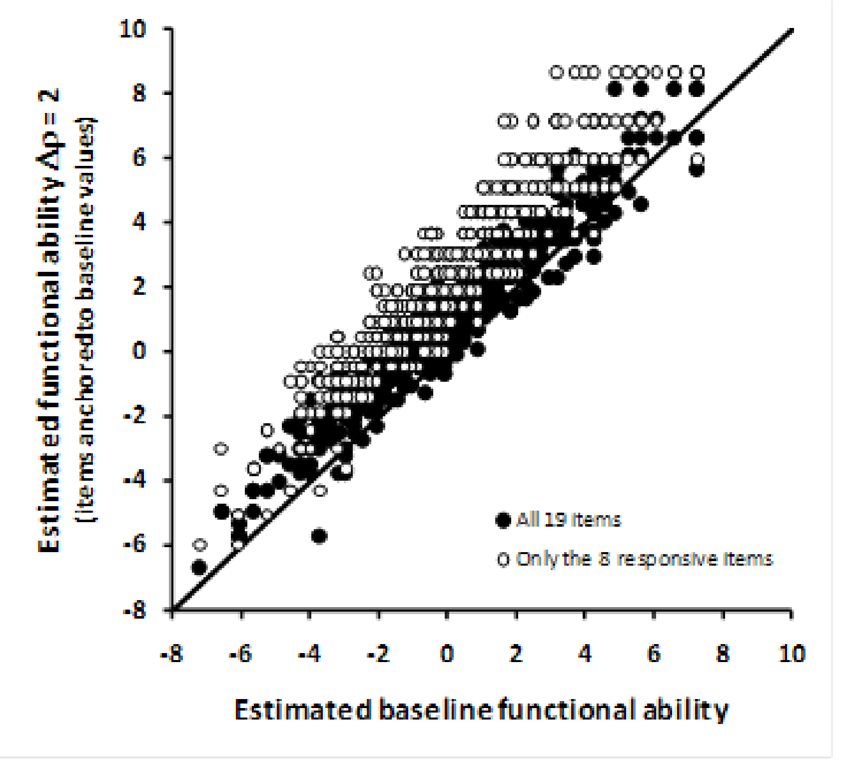




$$C_x(t_0) + e_{njx} < P_n(t_0) - I_j(t_0) - \Delta I_{nj} < C_{x+1}(t_0) + e_{njx+1}$$

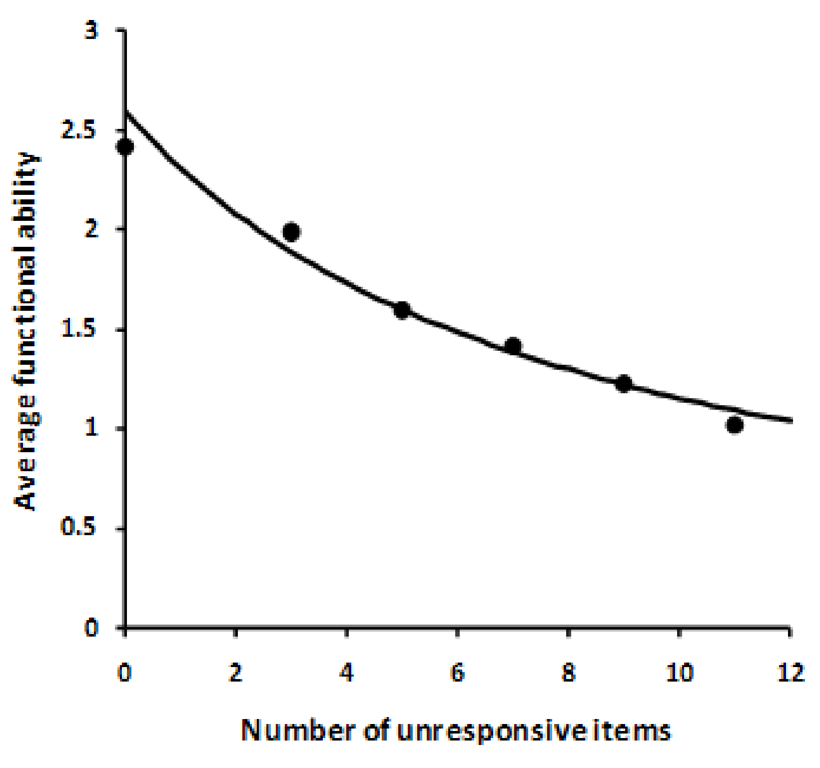
Simulation $\Delta I_j = -2$ for 8 items and $\Delta I_j = 0$ for 11 items

- Rasch analysis performed with item measures and category thresholds anchored to baseline values
- Filled circles: simulated responses to all 19 items included in analysis
- Open circles: only the 8 responsive items



$$C_x(t_0) + e_{njx} < P_n(t_0) - I_j(t_0) - \Delta I_{nj} < C_{x+1}(t_0) + e_{njx+1}$$

Unresponsive items dilutes effect of intervention



$$\overline{\Delta I_n} = \sum_{j=1}^{J} \frac{\Delta I_{nj}}{J} = \sum_{k=1}^{K} \frac{\Delta I_{nk}}{U+K} \quad \text{where} \quad U + K = J$$

$$\Delta I_j = \sum_{n=1}^{N} \frac{\Delta I_{nj}}{N}$$

$$\Delta R_{nj} = \Delta P_n - \Delta I_j$$

- Removing a cataract → ΔP_n
- Providing a magnifier → ΔI_j
- $\Delta I_j \neq 0$ indicates intervention-specific differential item functioning (DIF)
- Usually DIF is considered bad, in this case DIF is an indicator of a positive outcome

Low Vision Intervention Trial (LOVIT)

- RCT of the effectiveness of outpatient low vision rehabilitation in the VA for elderly legally blind veterans from visual acuity loss (20/200 to 20/500)
- Treatment group received vision assistive equipment (e.g., magnifiers), visual skills instruction, and adaptive skills training
- Control group received supportive telephone calls while they were on the wait list

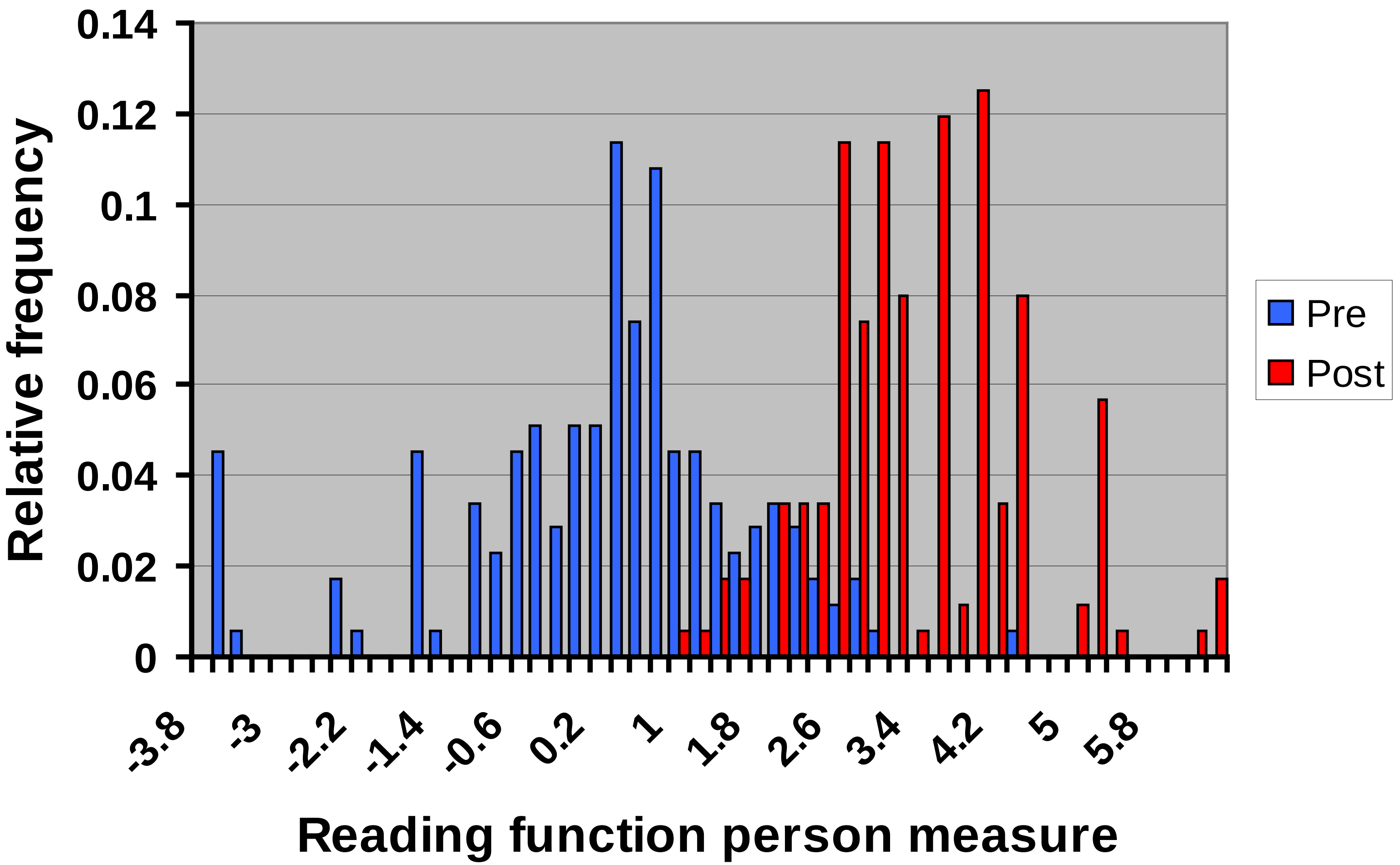
Stelmack JA, Tang XC, Reda DJ, Rinne S, Mancil RM, Massof RW. Outcomes of the Veterans Affairs Low Vision Intervention Trial (LOVIT). Arch Ophthalmol 2008;126:608-617.

Low Vision Intervention Trial (LOVIT)

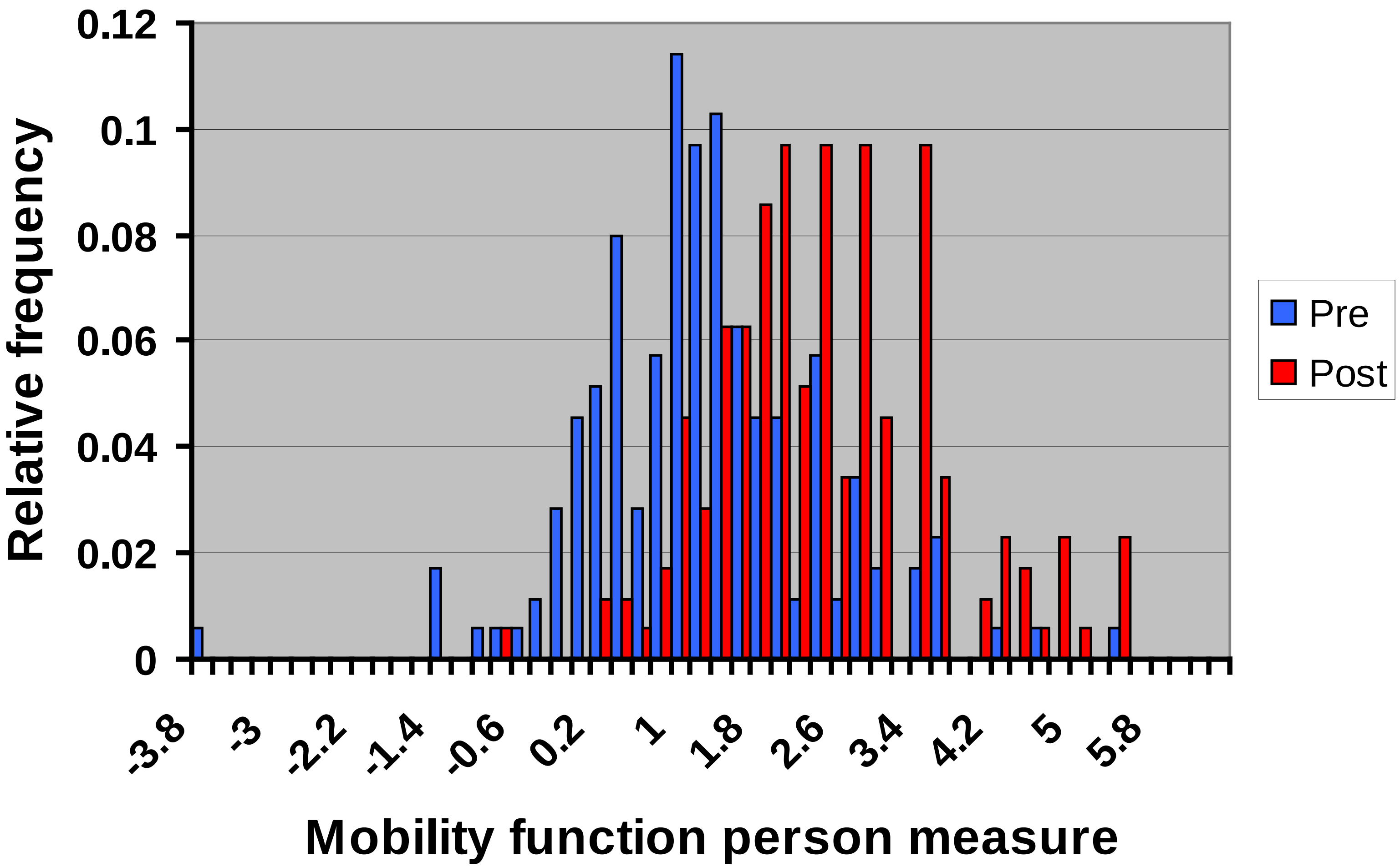
- A 48-item VFQ was administered at pre-intervention baseline and again 4 months later (approximately 2 months after the completion of intervention)
- Person measures from Rasch analysis of item difficulty ratings by participants for reading, mobility, visual perception, and visual motor function (from different subsets of items)
- Item measures and category threshold measures were <u>anchored</u> to pre-calibrated baseline values

Stelmack JA, Szlyk JP, Stelmack TR, Demers-Turco P, Williams RT, Moran D, Massof RW. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. Invest Ophthalmol Vis Sci. 2004;45:3919-3928.

Histograms of person measure for reading function



Histograms of person measures for mobility function

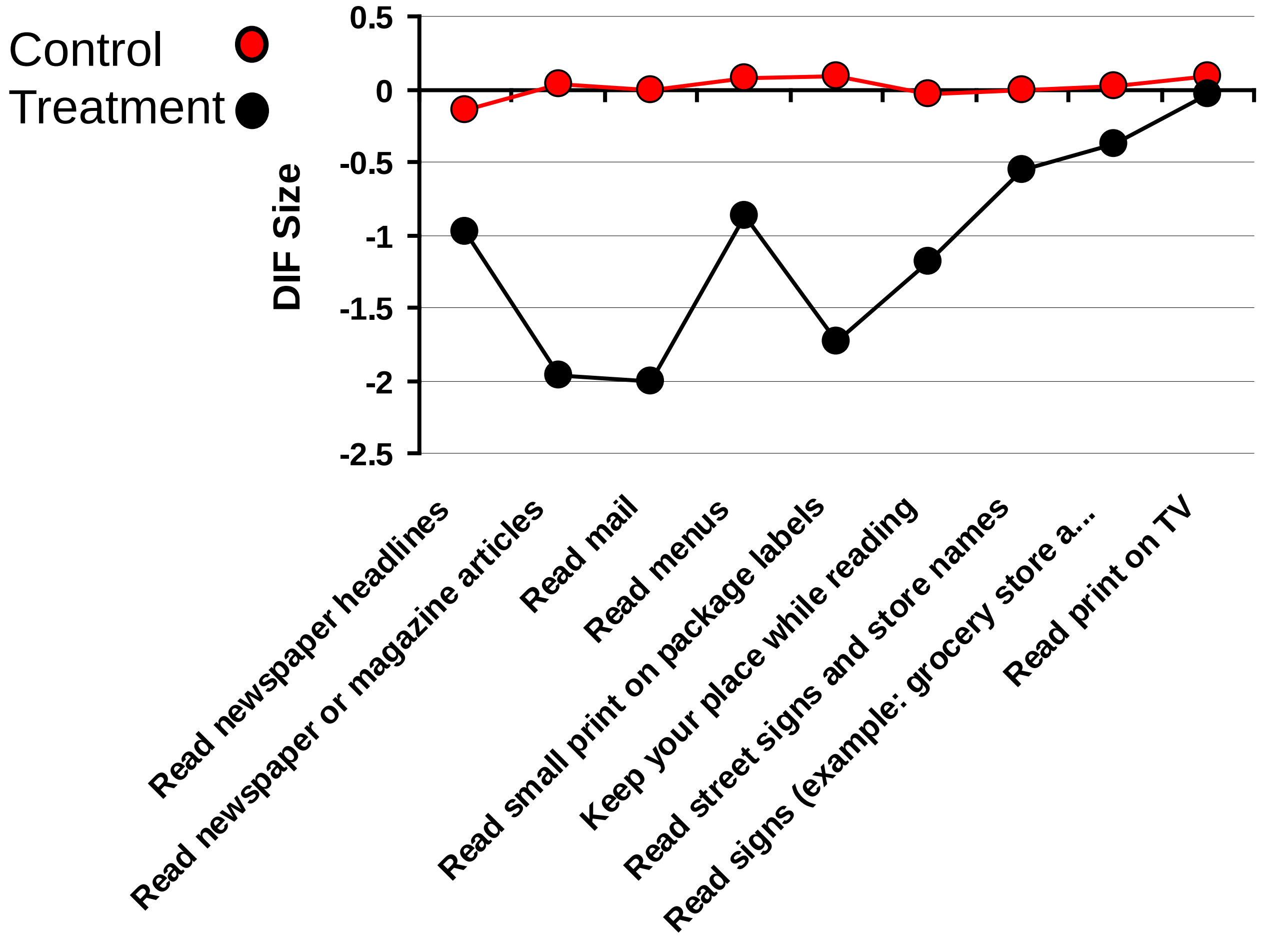


Are all effects of intervention in LOVIT changes in the person?

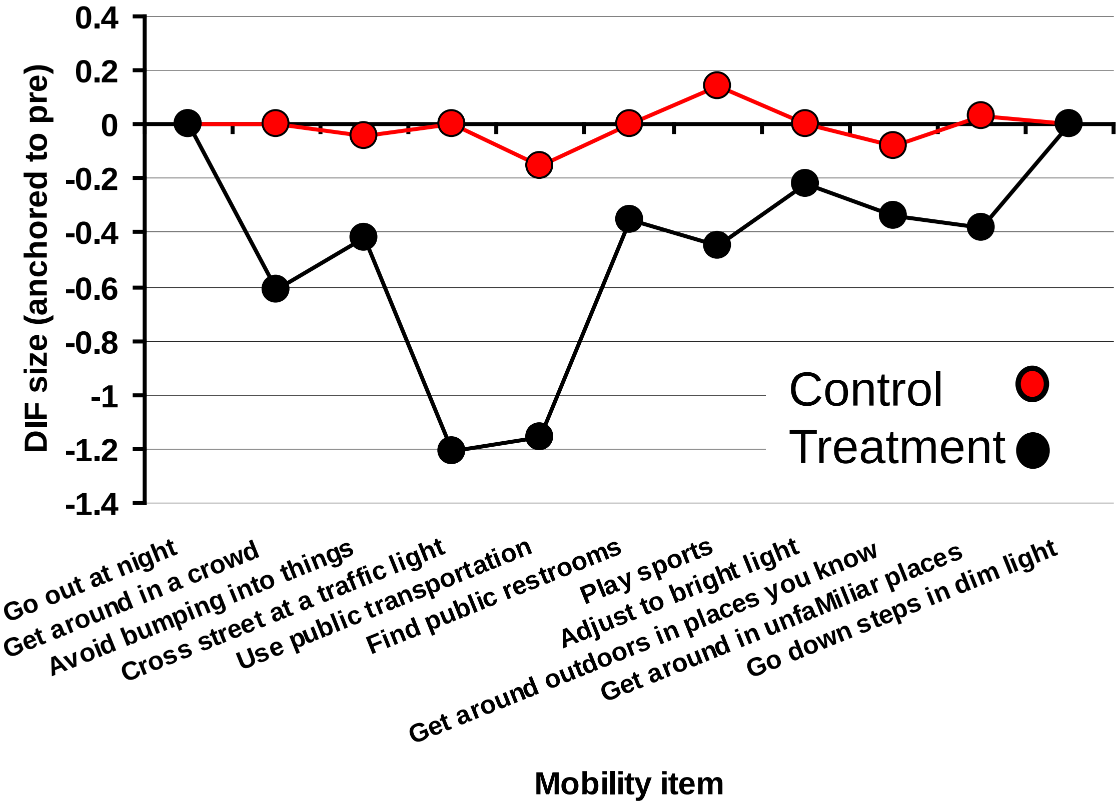
- By anchoring item measures to pre-calibrated baseline values, we have forced all effects of rehabilitation to manifest as ΔP_n
- Is there evidence of intervention-specific DIF?

Stelmack JA, Szlyk JP, Stelmack TR, Demers-Turco P,Williams RT, Moran D, Massof RW. Measuring outcomes of vision rehabilitation with the Veterans Affairs Low Vision Visual Functioning Questionnaire. Invest Ophthalmol Vis Sci. 2006;47:3253-3261.

Intervention-specific DIF for reading



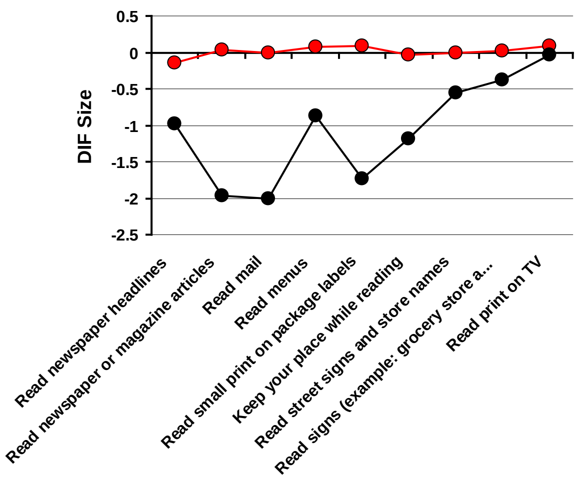
Intervention-specific DIF for mobility



Effect of intervention

$$\Delta R_{nj} = \Delta P_n - \Delta I_j$$

$$\Delta R_n = \sum_{j=1}^{J} \frac{\Delta P_n - \Delta I_j}{J} = \Delta P_n - \sum_{j=1}^{J} \frac{\Delta I_j}{J}$$



$$\Delta R_n = 3.09$$

$$\sum_{j=1}^{J} \frac{\Delta I_j}{J} = -1.08$$

$$\Delta P_n = 2.01$$

Patient-generated outcome measures

- If there is intervention-specific DIF, measured outcomes of intervention will depend on the choice of items
- Items must be important to the person and not be at the response ceiling at baseline
- If items not targeted by intervention or items that have no room for improvement are included in the outcome measure, the measure will not change and they will dilute the measured effect

of intervention by dragging down $\sum_{j=1}^{J} \frac{\Delta I_{nj}}{J}$

Goal Attainment Scaling (GAS)

- Most rating scale questionnaires have a fixed set of items, which can lead to underestimates of treatment effects because of intervention-specific DIF
- Need an outcome measure that recognizes and accommodates treatment plans targeted to the multiple personal goals of intervention and different capabilities of individual patients/clients

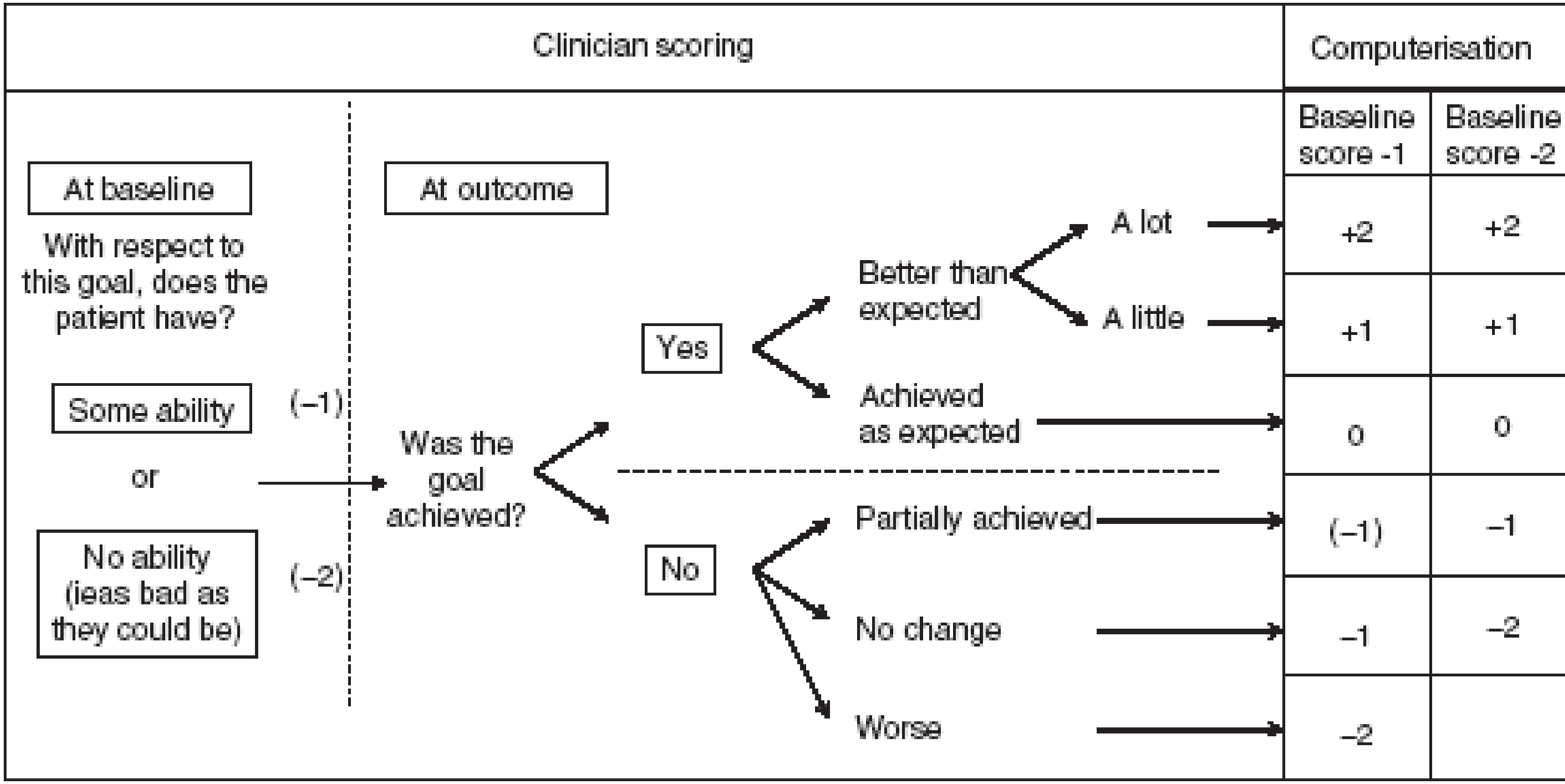


Figure 1 Algorithm for converting verbal scoring by clinicians to the 5-point goal attainment scores. This algorithm allows clinicians to record goal attainment without reference to the numeric scores, and so avoids the perceived negative connotations of zero and minus scores. Providing the level at baseline is known, 'partial achievement', 'no change' and worse can be translated by computerization outside the clinical arena. This incidentally offers the opportunity to compare the effect of using different scoring systems such as (–1, –2 and –3) or (–0.5, –1 and –2) and this work is currently underway.

Turner-Stokes L. Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clinical Rehabilitation* 2009;**23:**362-370.

Calculate a T-score from the ratings

- Service provider k's rating, X_{njk}, for person n and goal j are weighted, w_{njk}, summed across goals, normalized, and added to 50 to generate a T score for the person. (The correlation, r, usually is set to 0.3)

$$T = 50 + \frac{10\sum_{j=1}^{J} w_{njk} X_{njk}}{\sqrt{(1-r)\sum_{j=1}^{J} w_{njk}^2 + r\left(\sum_{j=1}^{J} w_{njk}\right)^2}}$$

Theoretical Interpretation of Standardized GAS Score

1. Service provider k estimates the state of person n with respect to a particular target of intervention j at time t with bias $B_k(t)$ in terms of functional reserve

$$R_{njk}(t) = P_n(t) - I_j(t) + B_k(t)$$

2. Service provider k defines a goal outcome value of R_{nj} that accommodates the person's potential for improvement in state as a result of intervention, ΔR_{njk}

$$R'_{njk} = R_{njk}(t_0) + \Delta R_{njk}$$

3. Estimate the person's proximity to the goal at time t

$$X_{njk}(t) = f\{R_{njk}(t) - R'_{njk}\} = f\{\Delta P_n(t) + \Delta B_k(t) - \Delta I_j(t) - \Delta R_{njk}\}$$

Problems with GAS

1. Raw scores for X_{njk} are not on an interval scale and we do not know the service provider's response category criteria, C_{kx}
2. Origin floats for different intervention targets and different persons because we have no estimate of I_j for the target of intervention and the service provider controls the estimate of ΔR_{njk}
3. Service provider might choose goals that represent different latent variables – i.e., violates unidimensionality
4. Service provider can have bias that changes over time – i.e., $\Delta B_k(t)$

Outcome measures

- The effect of intervention is the change in R_{nj}

$$E_{nj} = R_{nj}(t) - R_{nj}(t_0) = P_n(t) - I_j(t) - \left(P_n(t_0) - I_j(t_0)\right)$$

$$E_{nj} = P_n(t) - P_n(t_0) - \left(I_j(t) - I_j(t_0)\right)$$

$$E_{nj} = \Delta P_n - \Delta I_j$$

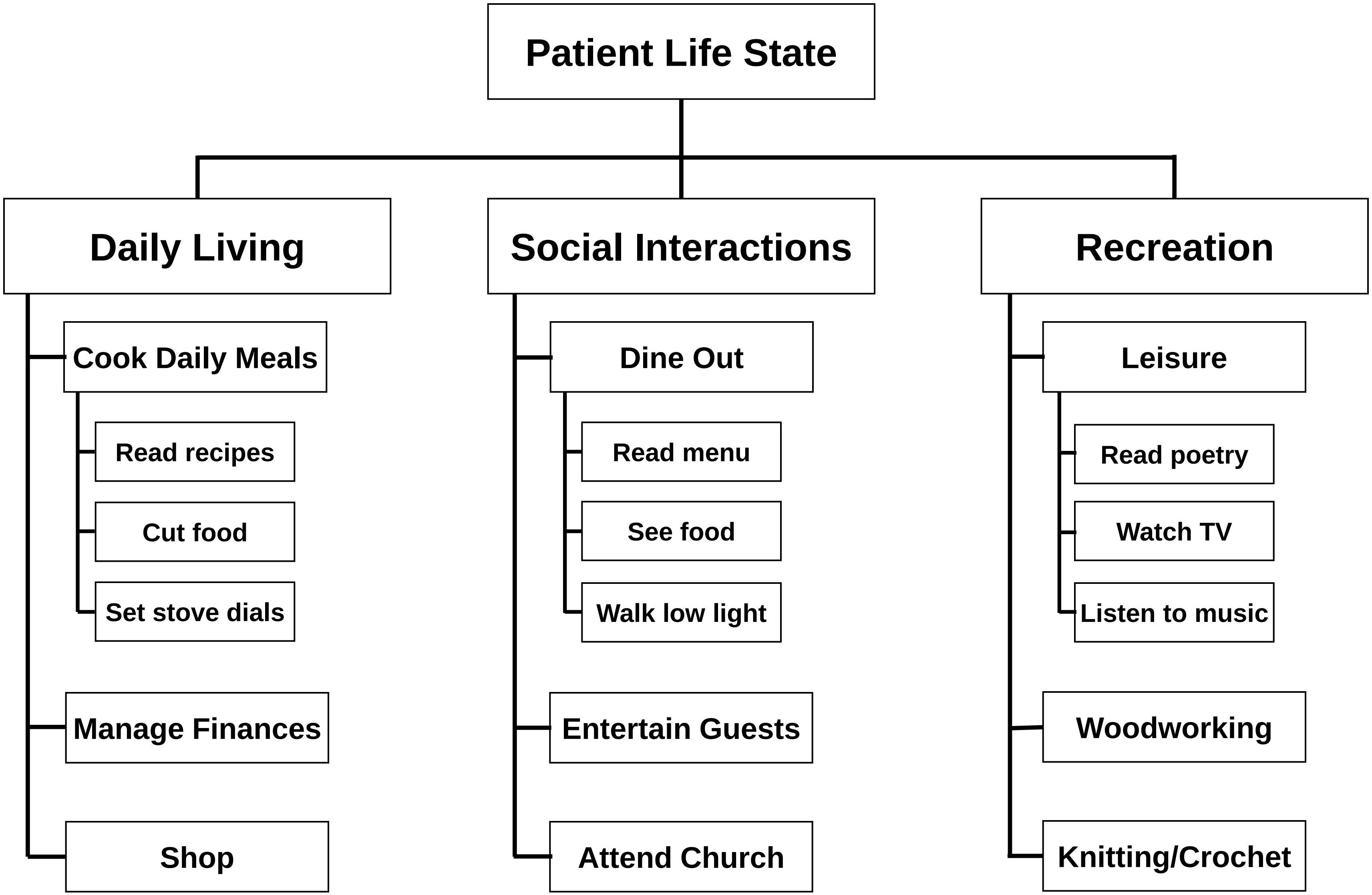
- Within GAS framework, the effects of interventions are confounded by biases in service provider judgments

$$E_{njk} = \Delta P_n - \Delta I_j + \Delta B_k - \Delta R_{njk}$$

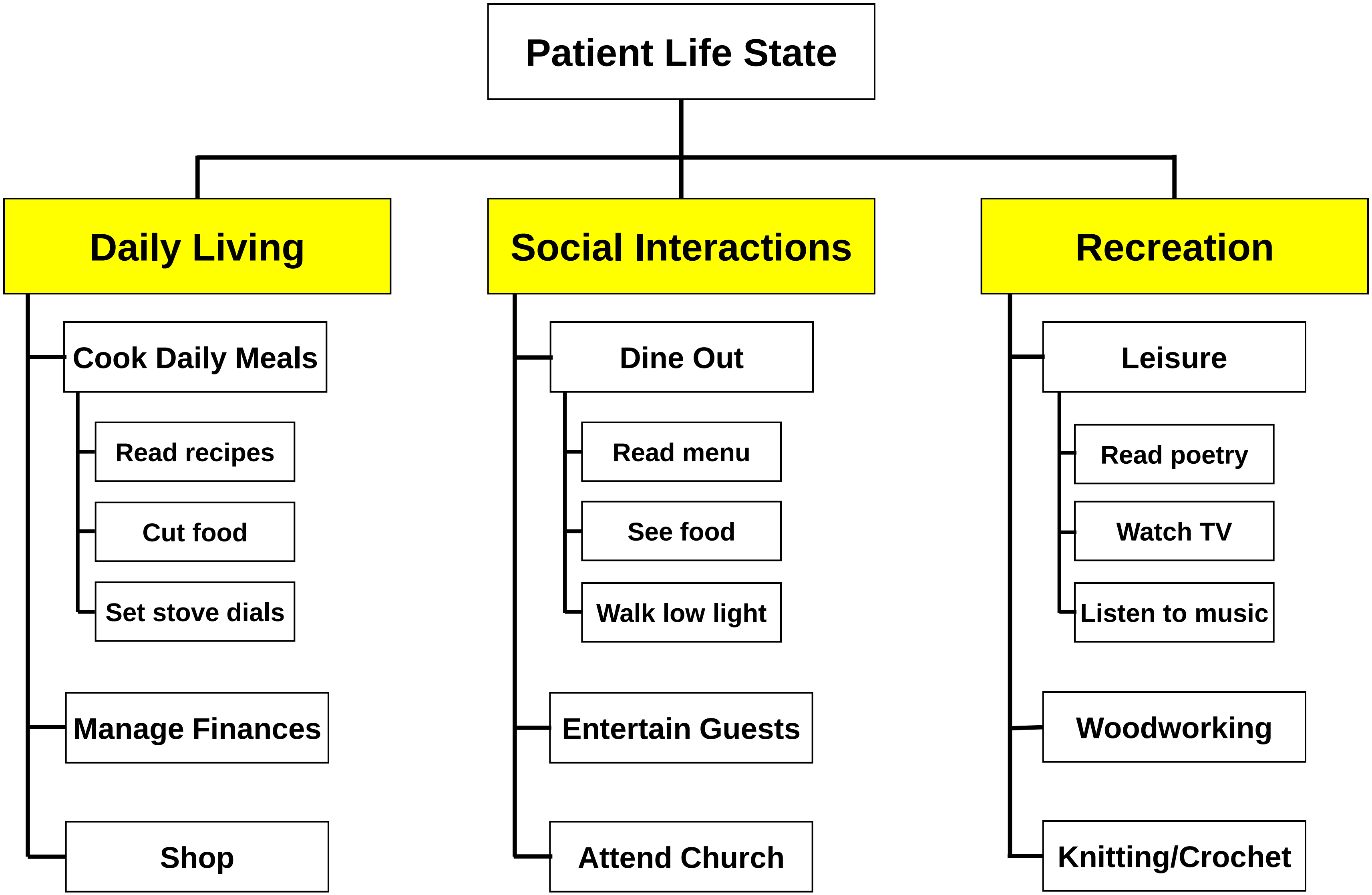
An alternative approach to GAS

- Develop a calibrated item bank of common goals of intervention for the target population (items calibrated to baseline values, so we know I_j for each goal)
- Filter the items using the person's importance ratings of the items (patient defines goals, not the service provider).
- Filter out items for which the baseline response is at the ceiling (remove goals that do not need to be included in the rehabilitation plan)
- Obtain objective measure of person state at baseline using self-report or other accepted method to prevent service provider bias

Activity Breakdown Structure (ABS)



Activity Breakdown Structure (ABS)



Activity Breakdown Structure (ABS)

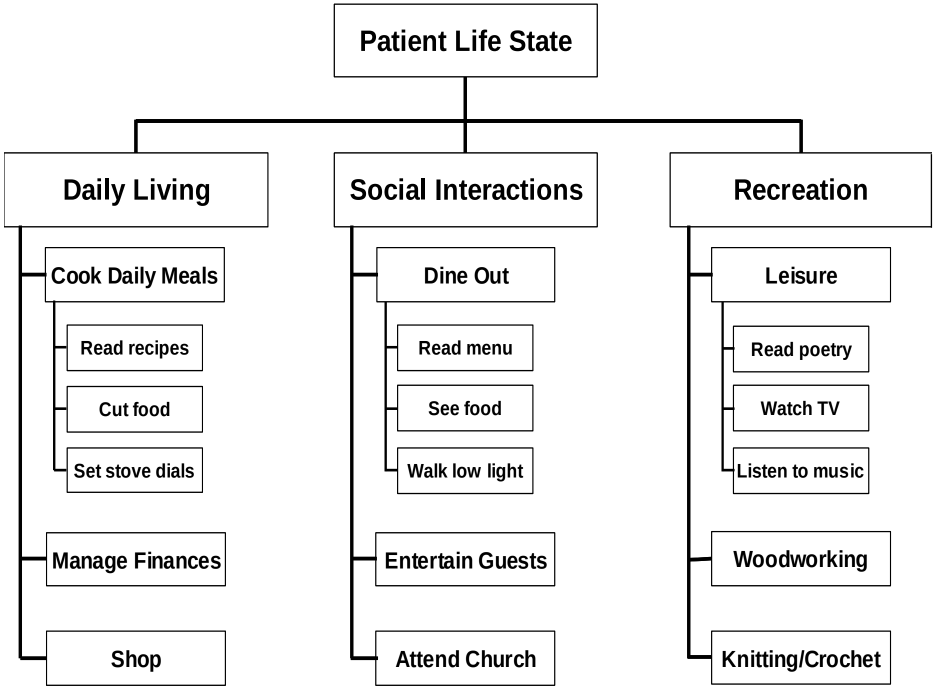
- **Patient Life State**
  - **Daily Living**
    - **Cook Daily Meals**
      - Read recipes
      - Cut food
      - Set stove dials
    - **Manage Finances**
    - **Shop**
  - **Social Interactions**
    - **Dine Out**
      - Read menu
      - See food
      - Walk low light
    - **Entertain Guests**
    - **Attend Church**
  - **Recreation**
    - **Leisure**
      - Read poetry
      - Watch TV
      - Listen to music
    - **Woodworking**
    - **Knitting/Crochet**

Activity Breakdown Structure (ABS)

- **Patient Life State**
  - **Daily Living**
    - Cook Daily Meals
      - Read recipes
      - Cut food
      - Set stove dials
    - Manage Finances
    - Shop
  - **Social Interactions**
    - Dine Out
      - Read menu
      - See food
      - Walk low light
    - Entertain Guests
    - Attend Church
  - **Recreation**
    - Leisure
      - Read poetry
      - Watch TV
      - Listen to music
    - Woodworking
    - Knitting/Crochet

Activity Inventory (AI)

- AI is an adaptively administered rating scale questionnaire
- Design and administration guided by the Activity Breakdown Structure
- 50 standard activity goals which commonly are reported within the low vision population
- 460 tasks nested under the 50 goals
- Goal and task item measures anchored to values calibrated from the baseline responses of over 3500 low vision patients made before intervention



Schematic of the Activity Breakdown Structure (ABS). The patient's life state is broken down into daily living, social interactions and recreation objectives. Each objective is broken down into the goals of activities (e.g., cook daily meals, manage finances, and shop under daily living). Each goal is broken down into subsidiary tasks that must be performed to achieve the goal (or may be deemed not applicable). Examples of tasks are read menu, see food, and walk in low light under the dine-out goal.

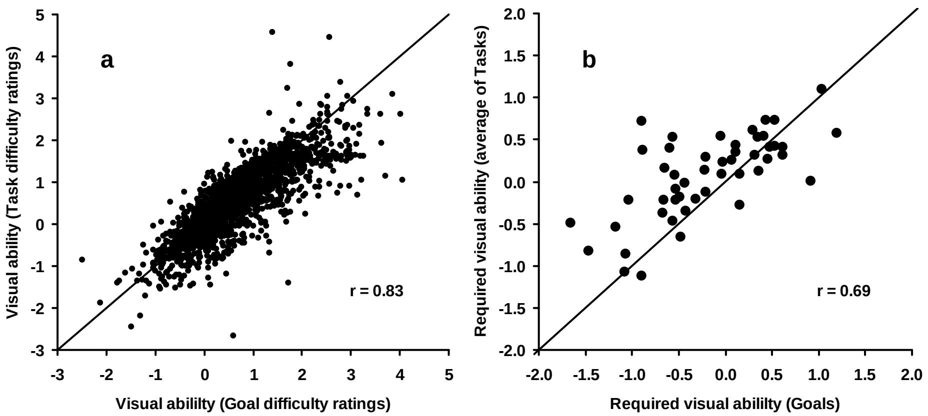
Massof RW, Ahmadian L, Grover LL, Deremeik JT, Goldstein JE, Rainey C, Epstein C, Barnett GD. The Activity Inventory: An adaptive visual function questionnaire. Optom Vis Sci 2007;84:763-774.

Adaptive administration of the AI

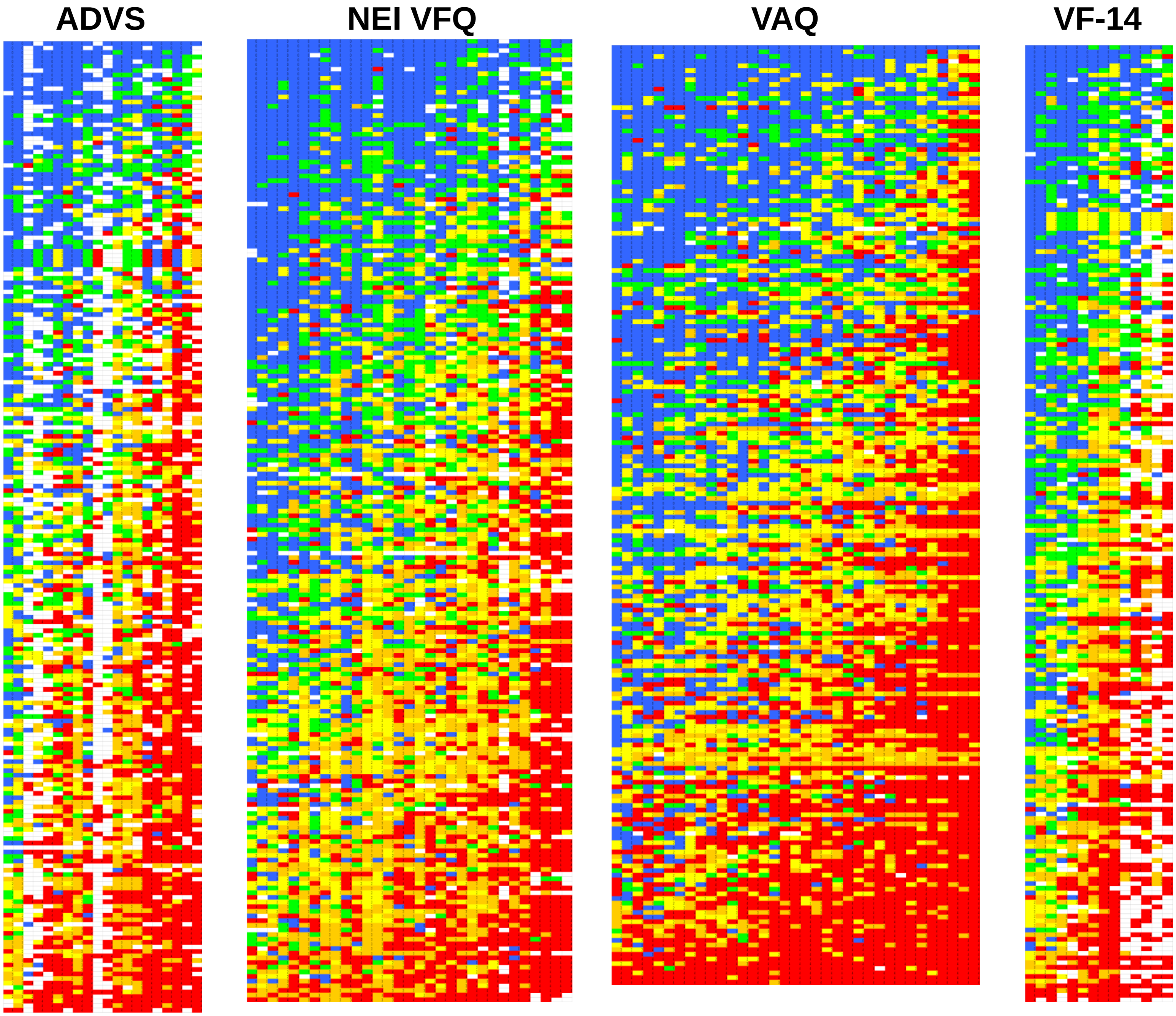
- Patient rates the importance of each goal
- Patient rates the difficulty of goals that exceed a criterion level of importance
- Patient rates the difficulty of tasks under goals that exceed a criterion level of difficulty, or responds that the task is not applicable (tagged as missing data)

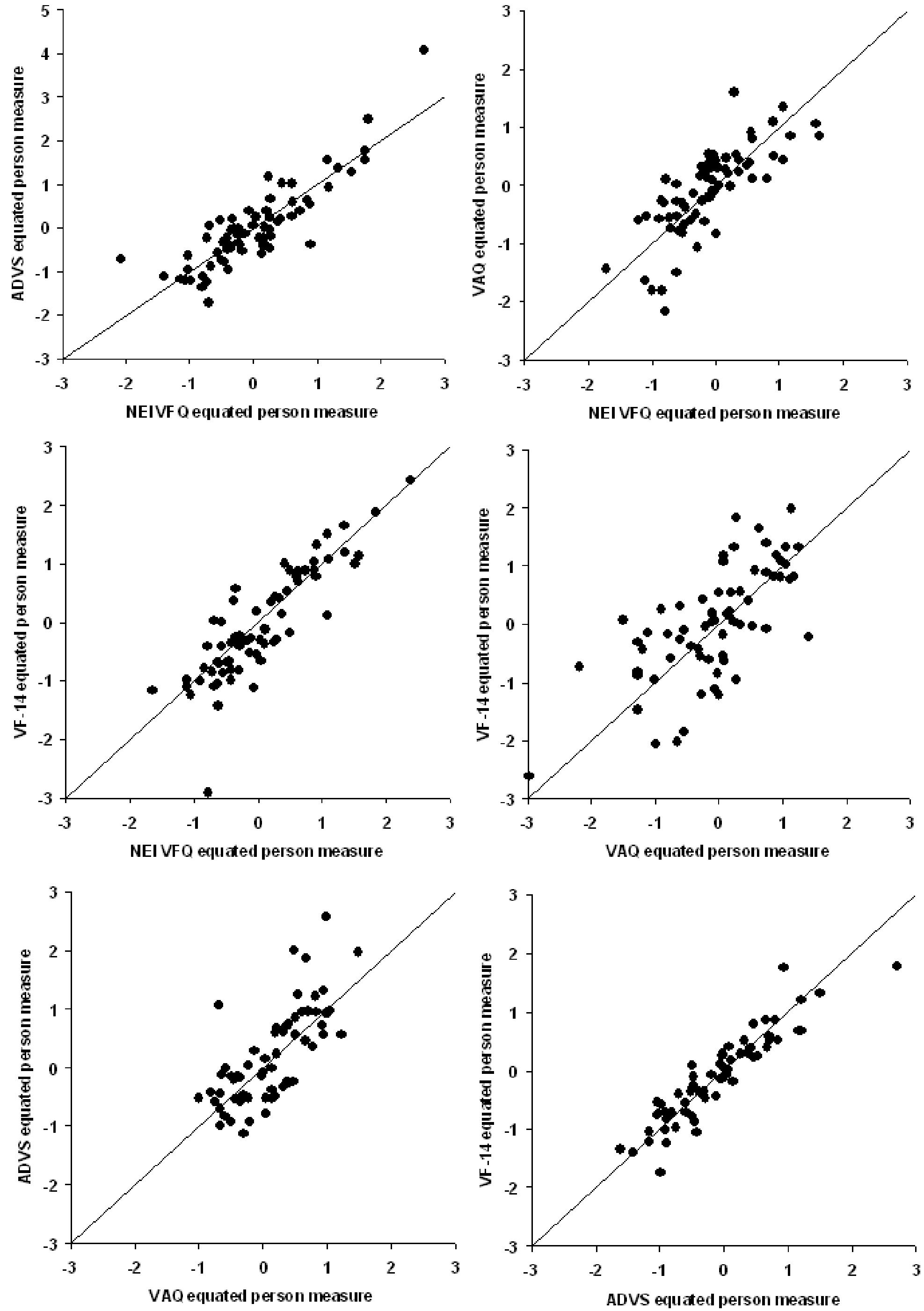
Activity Breakdown Structure (ABS)

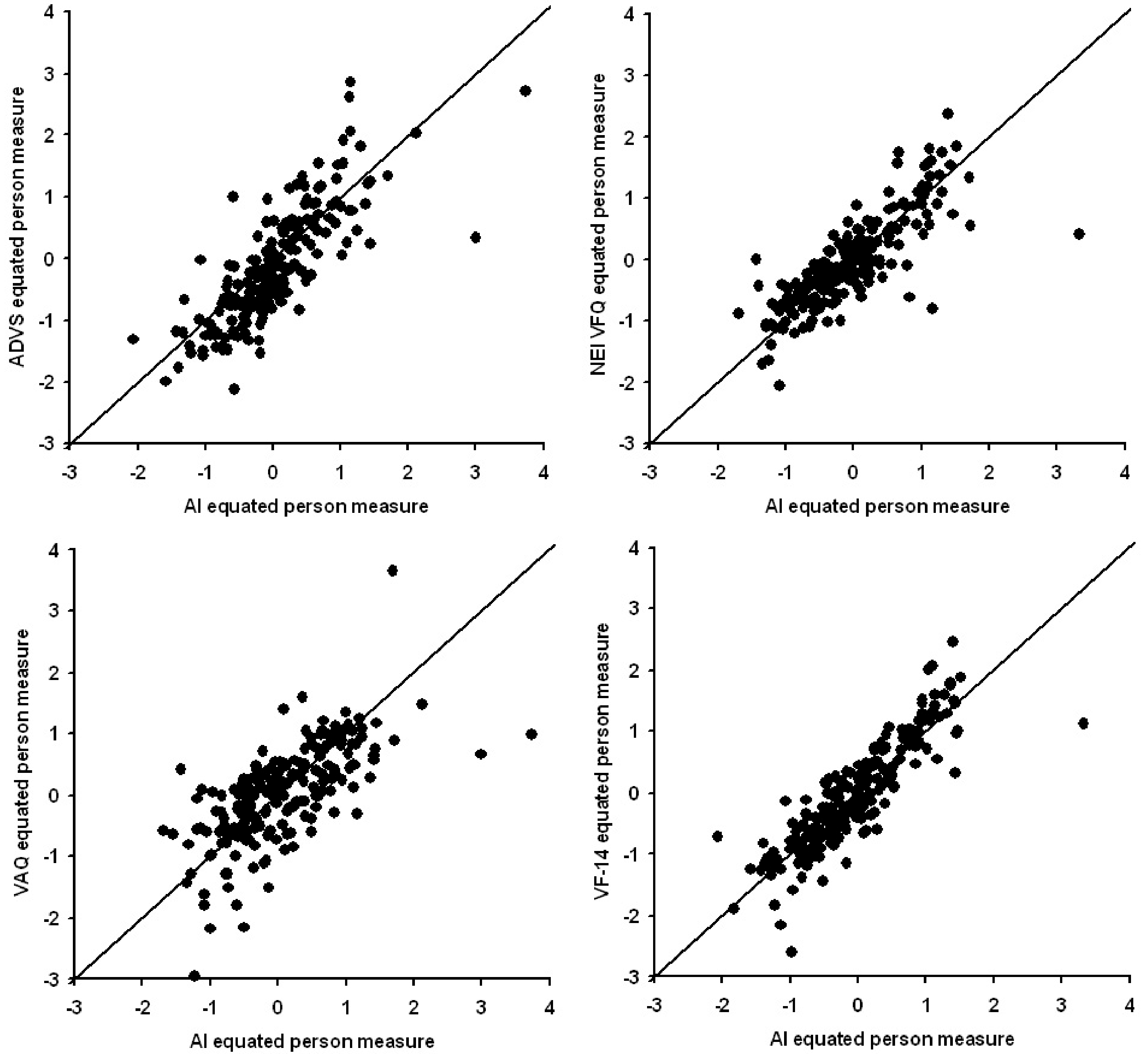
- Baseline person measures estimated from difficulty ratings of tasks agree with baseline person measures estimated from difficulty ratings of goals
- The goal item measure is well approximated by the average item measure of subsidiary tasks



a) Visual ability person measures estimated by Rasch analysis from task difficulty ratings in the AI vs visual ability person measures estimated from AI goal difficulty ratings. Solid line – identity line. Pearson correlation is 0.83. **b)** Average of required visual ability across tasks that serve the same goal in the AI vs required visual ability of the goal. Each point represents a different goal. Solid line – identity line. Pearson correlation is 0.69.

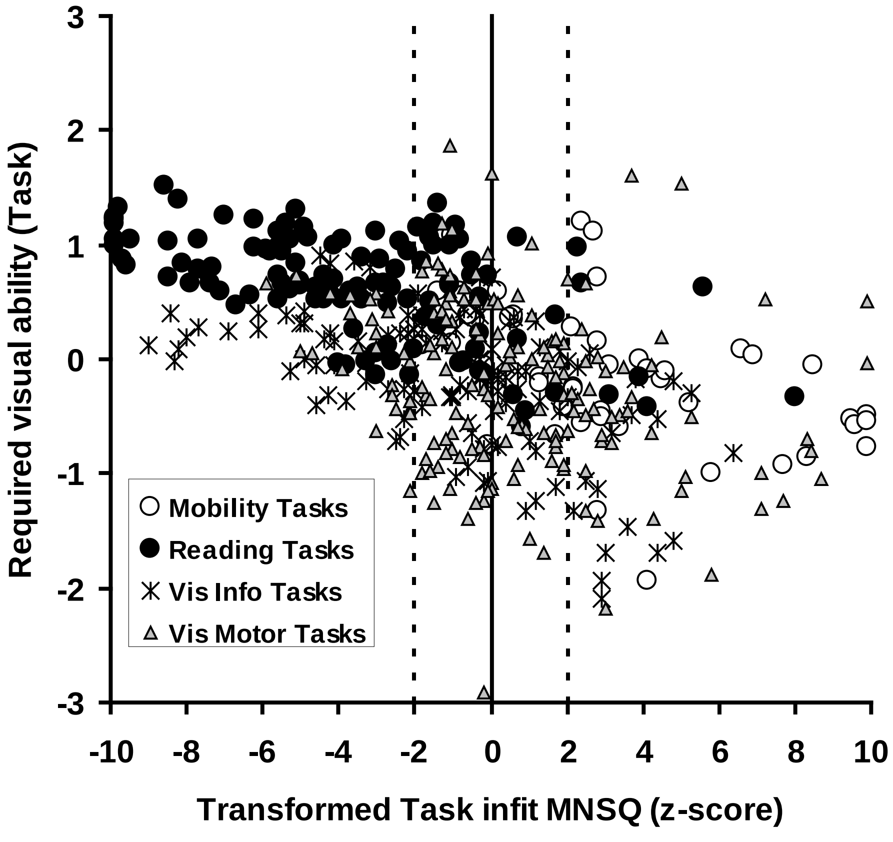

ADVS
NEI VFQ
VAQ
VF-14


ADVS equated person measure
NEI VFQ equated person measure
VAQ equated person measure
NEI VFQ equated person measure
VF-14 equated person measure
NEI VFQ equated person measure
VF-14 equated person measure
VAQ equated person measure
ADVS equated person measure
VAQ equated person measure
VF-14 equated person measure
ADVS equated person measure

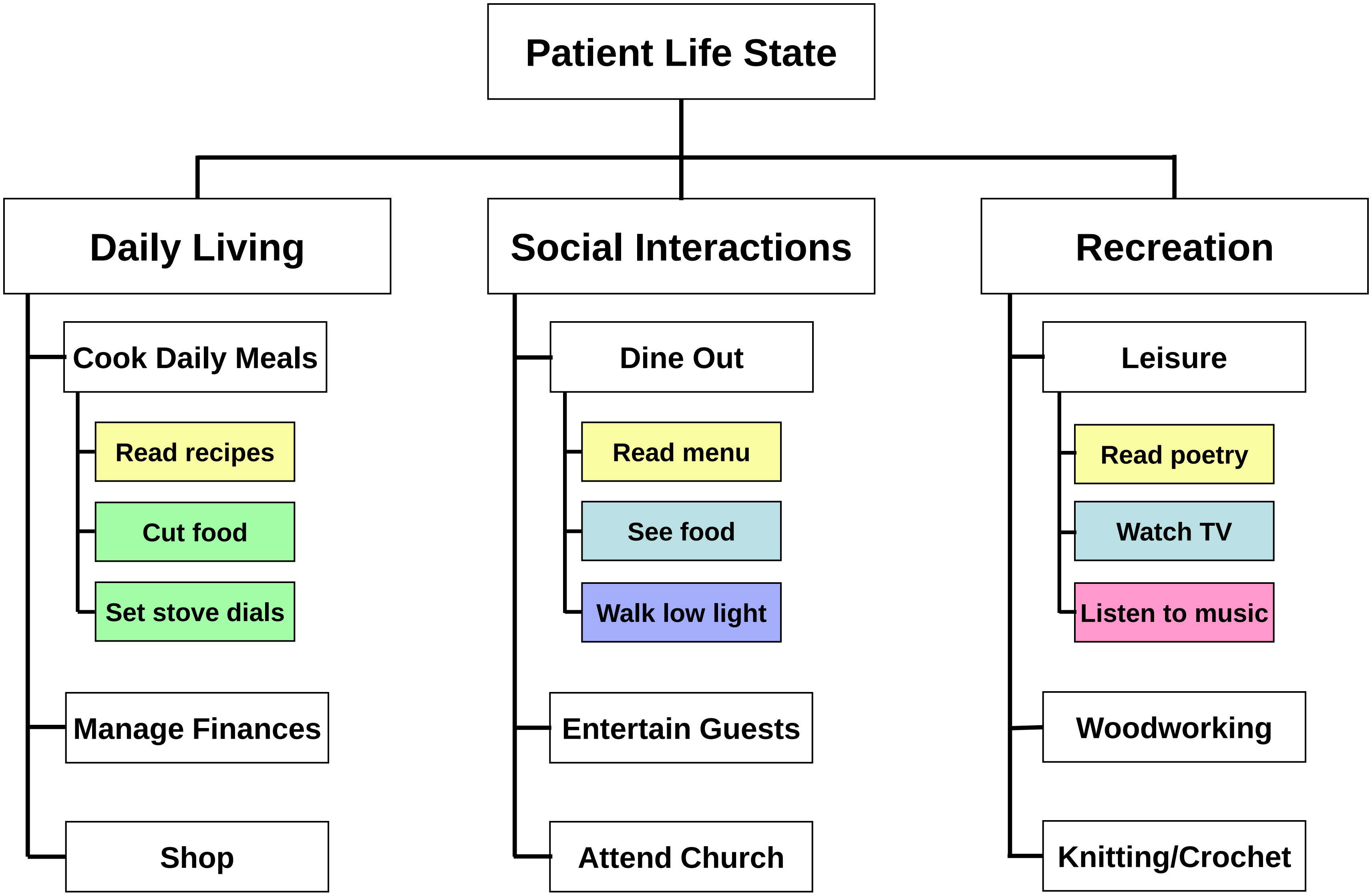


ADVS equated person measure
AI equated person measure
NEI VFQ equated person measure
AI equated person measure
VAQ equated person measure
AI equated person measure
VF-14 equated person measure
AI equated person measure
-3
-2
-1
0
1
2
3
4

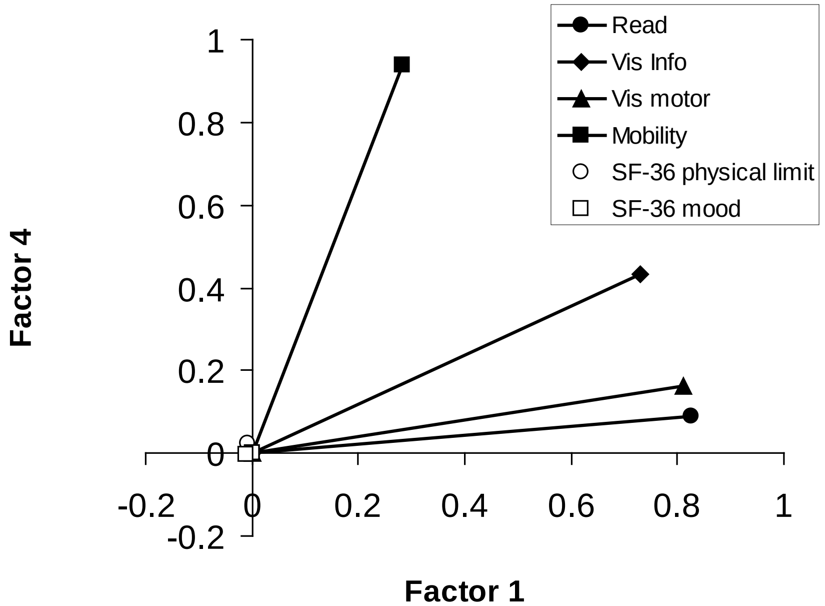
Mean square fit statistic transformed to z-score (standard normal deviate)

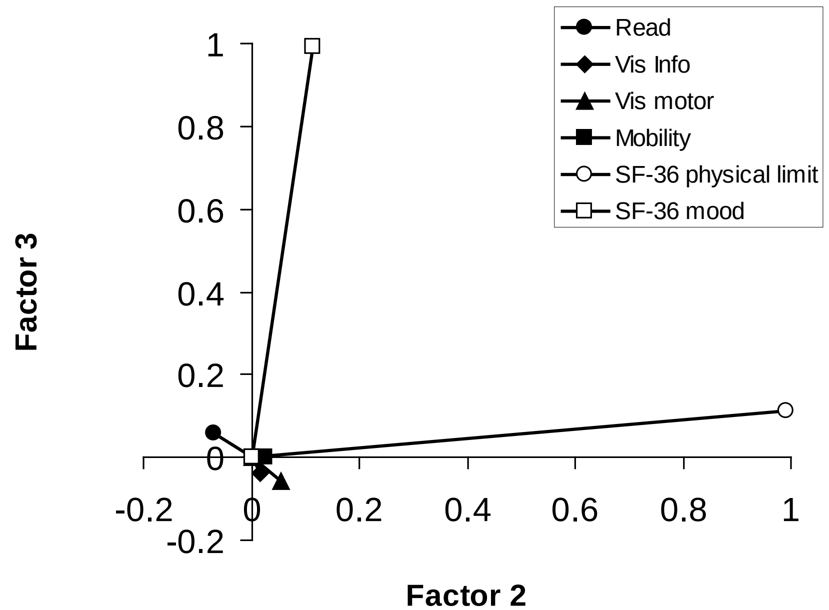


Activity Breakdown Structure (ABS)



Factor analysis of person measure estimates





Effects of intervention

- Rehabilitation helps patients achieve goals by
  - Improving the patient's vision (e.g., refractive error correction) $\Delta P \uparrow$
  - Improving patient's confidence and psychological state $\Delta P \uparrow$
  - Enhancing vision to make tasks easier to perform (e.g., visual skills, VAE) $\Delta I \uparrow$
  - Modify environment (e.g., lighting, contrast) $\Delta I \uparrow$
  - Adapt tasks so they are easier to perform without depending on vision $\Delta I \uparrow$
  - Develop new strategies using easier tasks so that goals can be achieved without performing the usual and customary tasks (tasks become N/A at follow-up and are filtered from AI) $J \downarrow$ so $\sum_{j=1}^{J} \frac{\Delta I_{nj}}{J} \uparrow$
- If rehabilitation potential is low
  - Counsel patient to devalue goal and obtain assistance to achieve the goal's larger objective (tasks are filtered from AI by goal's low importance ratings at follow-up) $J \downarrow$ so $\sum_{j=1}^{J} \frac{\Delta I_{nj}}{J} \uparrow$

Conclusions on an alternative approach to GAS

- Outcome of intervention should be judged by the patient, not by the service provider
- Estimate objective outcome measurements using item measures anchored to baseline calibrations for the served population
- If outcome measures include intervention-specific DIF, employ item filtering so that outcome measures are based on items that are important and relevant to the patient and are targeted by the intervention
- Intervention affects outcome measures by changing the person, changing the item difficulty (which causes DIF), or changing item filtering