

# Chapter 5

## An Adaptive Strategy for Measuring Patient-Reported Outcomes: Incorporating Patient Preferences Relevant to Cost-Benefit Assessments of Vision Rehabilitation



Robert W. Massof  and Chris Bradley

**Abstract** Vision rehabilitation aims to improve daily functioning in patients with chronic disabling vision impairments – low vision – by providing vision assistive equipment and teaching patients effective use of the equipment; teaching patients adaptations to obviate dependence on vision; modifications of the patient’s environment to improve visibility and increase safety; and psychosocial counseling and education to help patients cope with the stress of impaired vision. The interventions that constitute vision rehabilitation are tailored to each patient’s functional goals and personal preferences. This chapter describes the theory, structure, and mechanics of an adaptive patient-reported outcome measure – the Activity Inventory (AI). The AI was developed to identify each patient’s unique set of functional *goals* within a hierarchical framework, to specify the cognitive and motor activities (*i.e.*, *tasks* in different functional domains) the patient customarily performs to achieve his or her personal goals, and to elicit ordinal patient ratings of the importance and difficulty of each goal and the difficulty of relevant subsidiary tasks under goals that meet or exceed defined importance and difficulty criteria. The theoretical assumptions of a Rasch psychometric model that is employed to estimate measures of the patient’s overall visual ability from goal difficulty ratings and ability in each functional domain from task difficulty ratings before and after vision rehabilitation are presented and the results for a large group of low vision patients are reviewed. This chapter then explores the issue of how individual patient preferences must be factored into the outcome measures to determine the utility of functional improvements from vision rehabilitation. A theoretical framework for incorporating the utility of individual vision rehabilitation outcomes is offered and its application to functional outcome data is demonstrated.

---

R. W. Massof (✉) · C. Bradley  
Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, USA  
e-mail: [bmassof@jhmi.edu](mailto:bmassof@jhmi.edu)

© The Author(s) 2023  
W. P. Fisher, Jr., S. J. Cano (eds.), *Person-Centered Outcome Metrology*,  
Springer Series in Measurement Science and Technology,  
[https://doi.org/10.1007/978-3-031-07465-3\\_5](https://doi.org/10.1007/978-3-031-07465-3_5)

**Keywords** Low vision rehabilitation · Personalized care · Adaptive instrument administration

## 5.1 Introduction

Rasch rating scale models have three free parameters to be estimated from rating scale questionnaire responses: (1) person measures; (2) item measures; and (3) rating category thresholds. When developing and validating patient-reported outcome measures (PROM), practitioners of Rasch analysis typically regard differential item functioning (DIF), differential person functioning (DPF), mistargeting of item measures to person measures, and disordering of rating category thresholds to be distressing. These indicators, taken at face value, are interpreted as signs of metrological trouble with the data that warrant modifying the list of items and rating categories to improve measurement purity.

In this chapter, we describe our approach to developing a PROM, the *Activity Inventory* (AI), for adaptively measuring the outcomes of rehabilitation of daily functioning by people with chronic disabling vision impairments – low vision. Vision rehabilitation aims to help people with chronic vision impairments overcome vision disabilities through behavioral and environmental modifications, the use of vision assistive equipment, education, and psychosocial counseling. The multifaceted and idiosyncratic nature of vision rehabilitation poses significant challenges to measuring its outcomes. Consequently, with respect to the tenets of Rasch analysis (specific objectivity, sufficiency of raw scores, and invariant comparison), our approach to overcoming the challenges to measuring outcomes of vision rehabilitation may seem iconoclastic. The most heretical features of our approach are:

- Vision rehabilitation requires different interventions designed to target different activities. Because patients respond to different subsets of items and the different daily activities define the content of different items in the AI, a positive outcome of vision rehabilitation is expected to manifest as *intervention-specific DIF*, which when taken literally appears to challenge specific objectivity.
- Official definitions of vision disability traditionally have been based on a criterion visual acuity in the better seeing eye [1] (e.g.,  $\leq 20/200$  [USA] or  $< 3/60$  [WHO]) that cannot be improved with eyeglasses or contact lenses) or by peripheral vision loss specified by the horizontal extent of the visual field (e.g.,  $\leq 20^\circ$  [USA] or  $\leq 10^\circ$  [WHO]). Given that these two types of vision impairment can occur separately, or in varying degrees together, and they differentially affect the person's ability to perform different activities that define the AI item content, we expect to see *DPF*, which also appears to challenge specific objectivity.
- The endpoint of vision rehabilitation is the attainment of activity-performance goals that are defined and prioritized by the individual's personal preferences. Consequently, the AI items must be administered adaptively, drawn from a calibrated item bank having anchored item measures. However, the choice of

items is driven by the visually impaired person's preferences, not by item administration efficiency to achieve a criterion level of precision in the estimation of the person measure. Thus, the items drawn from the item bank parallel the content of the *individualized rehabilitation plan* and, unlike the strategy of computer adaptive testing (CAT), are *not necessarily well-targeted to the neighborhood of the individual's person measure*. Contrary to this strategy, students of Rasch analysis are taught that a well-designed instrument has a rectangular distribution of item measures that spans the distribution of person measures, or at least an item measure distribution that matches the person measure distribution so that measurement precision is highest where the density of persons to be discriminated is highest [2].

- The respondent to the AI is asked to rate the difficulty of performing activities described by the item content. Particularly in light of intervention-specific DIF, to be measurable the outcome of vision rehabilitation for each activity must manifest as a change in the difficulty rating of the corresponding item. The responsiveness of the AI to a change in difficulty depends on the number of rating categories and on the sizes of the concatenated rating category intervals, which are separated by category thresholds. However, the polytomous Rasch models most often used to estimate measures from responses to rating scale questionnaires routinely estimate disordered thresholds, especially as the number of rating categories is increased in an attempt to improve resolution on the measurement scale. Because *disordered rating category thresholds are illogical and an unacceptable output of a valid rating scale instrument*, it has been assumed that the problem is with the data. Consequently, instrument developers are advised to reduce the number of rating categories after the fact through mergers of neighboring categories until the offending threshold disordering is eliminated [2]. A more rational approach is to assume that respondents understand what "ordered" means and the analyst employs a *measurement model in which proper ordering of category thresholds is axiomatic*.

## 5.2 A Person-Centered Measure for Vision Rehabilitation

The primary aim of vision rehabilitation is to improve the visually impaired person's functional ability on an activity-by-activity basis by ameliorating functional limitations caused or exacerbated by the person's vision impairment. Functional ability is a multidimensional construct (*e.g.*, cognitive, motor, psychological, sensory, etc.). Vision rehabilitation targets one dimension of functional ability – visual ability (*i.e.*, the ability to perform activities that depend on vision), which in turn may be multidimensional given the different types of vision impairments inherent in official definitions of vision disability that can occur independently (*e.g.*, reduced visual acuity, visual field loss, impaired color discrimination).

### 5.2.1 A Measurement Model for Visual Ability

For visually impaired people, each activity described by items in a visual function rating scale questionnaire [3] requires some amount of visual ability to be performed – a fixed latent variable called the *item measure*. In this chapter, we will denote latent variables with Greek letters and manifest variables (as well as constants, indices, functions, and operators) with Latin letters. Accordingly, we use the lower-case Greek letter rho and the subscript  $j$  to represent the average amount of visual ability required by the visually impaired population to perform the  $j$ th item ( $\rho_j$ ). However, the actual amount of visual ability required by an item can vary randomly between persons depending on available technology, customary practices, environmental factors, etc. Thus, for the  $n$ th person, the actual visual ability required to perform the  $j$ th item ( $\rho_{n,j}$ ) is expected to deviate from the average item measure for the population of interest by an amount for each person that is randomly distributed in the specified population ( $\varepsilon_{n,j}$ ), such that  $\rho_{n,j} = \rho_j + \varepsilon_{n,j}$ . By definition,  $\varepsilon_{n,j}$  can be positive or negative and the mean of epsilon values across all persons is zero for each item. Ignoring within-person variance in the deviate for the time being, the standard deviation of epsilon between persons for the  $j$ th item is denoted as  $\sigma_j$  (although technically the standard deviation is not a latent variable, it is referring to the distribution of a latent variable, and sigma is the conventional notation for standard deviation).

Each person has some amount of visual ability – a trait of the person represented by a latent variable called the *person measure*. Thus,  $\alpha_n$  denotes the amount of visual ability possessed by the  $n$ th person. If person  $n$  has far more ability (the value of alpha) than is required to perform the activity described by item  $j$ , *i.e.*,  $\alpha_n > \rho_{n,j}$ , then the person is likely to report that the activity is “not difficult” to do. If person  $n$  has far less ability than is required to perform the activity described by item  $j$ , *i.e.*,  $\alpha_n < \rho_{n,j}$ , then the person is likely to report that the activity is “impossible” to do. Between these two extremes, person  $n$  could use an ordinal rating scale to estimate the level of difficulty she or he has performing the activity described by item  $j$  (*e.g.*, 1 – “very difficult”; 2 – “moderately difficult”; 3 – “somewhat difficult”; and to complete the scale we would add the extremes: 4 – “not difficult” and 0 – “impossible to do”). Theoretically, we interpret the chosen rating as the person’s magnitude estimate of his or her *functional reserve* for the activity described by the item content [4]. Functional reserve, denoted as  $\varphi_{n,j}$  for person  $n$  relative to item  $j$ , is simply the difference between the person measure and the measure for the item that person is responding to,  $\varphi_{n,j} = \alpha_n - \rho_{n,j}$ . From the definition of  $\rho_{n,j}$ , functional reserve also can be written as  $\varphi_{n,j} = \alpha_n - \rho_j - \varepsilon_{n,j}$ , which incorporates the average item measure for the specified population,  $\rho_j$ , and the deviation from the average for person  $n$ ,  $\varepsilon_{n,j}$ .

We can now think of creating a functional reserve ruler in units of the latent variable phi. Although all persons are given the same ordinal difficulty rating categories, each person divides the  $\varphi$  ruler into his or her own set of intervals – the only thing persons agree on is that the interval for rating category 0 comes before

and is concatenated with the interval for rating category 1, which comes before and is concatenated with the interval for rating category 2, etc. Although the sizes and locations of these intervals on the  $\varphi$  ruler are likely to be unique to the person, they must be in the order of the  $\varphi$  magnitude estimates they represent and separated from their neighboring intervals by boundaries located at positions unique to the person. The boundaries between ordered intervals on any given trial are called *response thresholds*, and their locations in  $\varphi$  units for the example we have been using are  $\tau_{n,1}$ ,  $\tau_{n,2}$ ,  $\tau_{n,3}$ ,  $\tau_{n,4}$  for person  $n$ . The  $\varphi$  scale is open-ended so that the lower bound for interval 0 is negative infinity and the upper bound for interval 4 is positive infinity. People most likely do not agree with one another, and each person may be inconsistent from trial to trial in the positions of the different response thresholds on the phi ruler, but by definition the thresholds must be ordered on every trial.

As we did with the item measures, we can estimate a population-based average threshold for each interval, which are fixed values, and define the threshold for the  $x$ th interval as the average threshold ( $\tau_x$ ) plus a person-specific deviate ( $\eta_{n,x}$ ) that is randomly distributed between persons, *i.e.*,  $\tau_{n,x} = \tau_x + \eta_{n,x}$ , for  $x = 1$  to 4. The average value of eta across persons for each threshold is zero and the standard deviation of eta between persons is  $\sigma_x$ . With these definitions and explanations, the use of a rating scale to make magnitude estimates implies that for person  $n$  to assign a rating of  $x$  to item  $j$ , functional reserve ( $\varphi_{n,j}$ ) must be greater than or equal to person  $n$ 's threshold for category  $x$  and less than person  $n$ 's threshold for category  $x + 1$ , and all intervals must be ordered:

$$\cdots < \tau_{n,x-1} < \tau_{n,x} \leq \varphi_{n,j} < \tau_{n,x+1} < \tau_{n,x+2} < \cdots$$

Stated more precisely, we assume a real line is partitioned by ordered thresholds into ordered intervals called rating categories, where the rating categories are defined using half-open intervals so that every point on the real line corresponds to precisely one rating category. Substituting terms in this expression with sums of fixed variables and randomly distributed deviates, we obtain

$$\cdots < \tau_{x-1} + \eta_{n,x-1} < \tau_x + \eta_{n,x} \leq \alpha_n - \rho_j - \varepsilon_{nj} < \tau_{x+1} + \eta_{n,x+1} < \tau_{x+2} + \eta_{n,x+2} < \cdots$$

The deviate  $\varepsilon_{n,j}$  can be added to each term in the above expression, and defining a new deviate,  $\zeta_{n,j,x} = \eta_{n,x} + \varepsilon_{n,j}$ , that is randomly distributed between persons, we obtain

$$\cdots < \tau_{x-1} + \zeta_{n,j,x-1} < \tau_x + \zeta_{n,j,x} \leq \alpha_n - \rho_j < \tau_{x+1} + \zeta_{n,j,x+1} < \tau_{x+2} + \zeta_{n,j,x+2} < \cdots \quad (5.1)$$

But there are two ways to interpret the error term,  $\zeta_{n,j,x}$ , on the average rating category thresholds used in expression (5.1): (1) the threshold error term can be identified as a deviate or (2) it can be identified as a random variable. As a deviate we can describe the trial-to-trial distribution of threshold values while enforcing

threshold ordering [5], whereas random variables can take on any value, including those that result in threshold disordering. To represent this second interpretation of rating category thresholds, expression (5.1) is modified to be less specific about the ordering of thresholds:

$$\{\dots, \tau_{x-1} + \zeta_{n,j,x-1}, \tau_x + \zeta_{n,j,x}\} \leq \alpha_n - \rho_j < \{\tau_{x+1} + \zeta_{n,j,x+1}, \tau_{x+2} + \zeta_{n,j,x+2}, \dots\} \quad (5.2)$$

Expression (5.2) says that for person  $n$  to respond with rating category  $x$  on a given trial, functional reserve must be greater than or equal to all the average thresholds (plus a random variable) on the left and be less than all of the average thresholds (plus a random variable) on the right (*i.e.*, on each trial the person's thresholds are segregated into two subsets, one in which all elements do not exceed the functional reserve and the other in which all elements are greater than the functional reserve). Expression (5.1) is simply one form that the more general expression (5.2) can take.

Mathematically, the assumptions built into expression (5.1) require average thresholds to be ordered whereas the assumptions built into expression (5.2) permit average thresholds to be disordered. In other words, expression (5.1) emphasizes the intervals – they must be ordered and concatenated on every trial for every person, but their sizes and how they are centered as a group on the  $\varphi$  scale are free to vary between persons and between trials by the deviate zeta. For expression (5.1), the locations of the response category thresholds are identified as the boundaries of the concatenated intervals. Thus,  $\tau_{x-1} + \zeta_{n,j,x-1}$  is always less than  $\tau_x + \zeta_{n,j,x}$ , which in turn is always less than  $\tau_{x+1} + \zeta_{n,j,x+1}$ , etc.

For expression (5.2), the definition of “threshold” is subtly changed, even though when casually described it continues to be used as if it refers to the boundaries of an interval in which  $\varphi_{nj}$  falls. But for expression (5.2) the ordinal value assigned to the response refers to a count of the number of thresholds pre-labeled with ordered numbers that are less than  $\varphi_{n,j}$ . That is, on each trial for each person the response category thresholds are identified and permanently labeled (like a number on a soccer player's jersey). The numbers on the thresholds represent the order in which they are counted, not necessarily the order of the threshold magnitudes.

Mathematically, to satisfy expression (5.2) it is not necessary for thresholds to define boundaries of intervals – each threshold is tracked from trial to trial on the  $\varphi$  scale as an independent entity. When the person responds with rating  $x$ , it is assumed by the model that  $\varphi_{n,j}$  is greater than all thresholds less than  $\tau_{x+1} + \zeta_{n,j,x+1}$  and less than all thresholds greater than  $\tau_x + \zeta_{n,j,x}$ . Trials that do not satisfy expression (5.2) are ignored (*i.e.*, estimated probabilities are conditioned on the requirement that dichotomous scores assigned to each threshold satisfy a Guttman scale on each trial) [6]. Unlike the requirement for expression (5.1), on each eligible trial the two sets of thresholds plus error segregated by  $\varphi_{n,j}$  can have magnitudes in any order on their respective sides of the inequalities. For this reason, Rasch models derived from expression (5.2) can estimate disordered average thresholds, whereas Rasch models derived from expression (5.1) always estimate ordered thresholds [5, 7, 8].

Some Rasch models permit expected values of thresholds for response categories to vary across items (*i.e.*,  $\tau_x$  is redefined as  $\tau_{j, x}$ ) [9]. If expression (5.1) must be satisfied by  $\tau_{j, x}$  (*i.e.*, thresholds for each item must be ordered), expression (5.1) can be rewritten as

$$\dots < \tau_{j,x-1} + \rho_j + \zeta_{n,j,x-1} < \tau_{j,x} + \rho_j + \zeta_{n,j,x} \leq \alpha_n < \tau_{j,x+1} + \rho_j + \zeta_{n,j,x+1} < \tau_{j,x+2} + \rho_j + \zeta_{n,j,x+2} < \dots \tag{5.3}$$

by adding  $\rho_j$  to every threshold and to functional reserve, which in effect, creates a different Likert scale for each item ( $j$ ). Expression (5.3) is consistent with Samejima’s [10] graded response model if the variance of  $\zeta_{n, j, x}$  depends on the item ( $j$ ) – a common assumption of item response theory models. If the variance of  $\zeta_{n, j, x}$  is constant with respect to  $n, j$ , and  $x$ , as required by Rasch models [8], then expression (5.3) reduces to a dichotomous Rasch model with  $\tau_{j, x} + \rho_j$  defining the measure for “item”  $x$  in domain  $j$  [6]. Using rating scale questionnaire terminology, the  $\tau_{j, x}$  values are playing the role of  $x = 1$  to  $m_j$  dichotomously scored items ( $j, x$ ) and the  $\rho_j$  values are playing the role of domain-dependent ( $j$ ) constants that offset the item measures in domain  $j$  so that all items in the instrument share the same origin on a common scale. The scoring rule for expressions (5.1, 5.2 and 5.3) is the same, which can be stated explicitly as: if the respondent chooses category  $j, x$ , then response category  $j, x$  and all categories less than  $j, x$  are given a score of 1 and all categories greater than  $j, x$  are given a score of 0 [6, 11, 12]. This rule is the consequence of requiring normative measurement models to conform to a Guttman scale.

Unlike Rasch’s original application of his model to reading tests [12] and most current applications in educational testing, in which ordinal scores represent error counts, ordinal scores for rating scales are assigned to subjective magnitude estimates, which are represented by a distance between two points on a number line [13]. The only counts involved in a distance measure are counts of unit distances that are concatenated to span the distance being measured, not counts of events. Expression (5.1) (and algebraic variations) represents a class of Rasch models for which measurement units correspond to a distance on a number line. Expression (5.2) (and algebraic variations) represents a class of Rasch models for which measurement units correspond to counts of sequentially ordered events that are positioned independently at points on the number line [6].

If  $r_{j, x}$  is the correlation between  $\eta_{n, x}$  and  $\varepsilon_{n, j}$ , then the between-person variance in  $\zeta_{n, j, x}$  is  $\sigma_{j,x}^2 = \sigma_j^2 + \sigma_x^2 + 2r_{j,x}\sigma_j\sigma_x$ . Invoking the central limit theorem, we assume the probability density function for  $\zeta_{n, j, x}$  is approximately normal, which in turn is further approximated with the logistic density function:  $f(\zeta_{n,j,x}|0, \sigma_{j,x}) = \frac{e^{-\zeta_{n,j,x}/\sigma_{j,x}}}{\sigma_{j,x} \left(1 + e^{-\zeta_{n,j,x}/\sigma_{j,x}}\right)^2}$ .

From expression (5.1), the probability person  $n$  will rate item  $j$  with rating category

$x$  is the probability  $\tau_x + \zeta_{n,j,x} < \alpha_n - \rho_j$  and  $\alpha_n - \rho_j < \tau_{x+1} + \zeta_{n,j,x+1}$ . To satisfy this requirement in expression (5.1), it must be true that  $\tau_{x+1} + \zeta_{n,j,x+1} > \tau_x + \zeta_{n,j,x}$  for every person/item combination and for every value of  $x$ .

For dichotomous response categories, there is only one threshold, which can be added to  $\rho_j$  and its variance can be added to  $\sigma_j^2$ . The major difference between dichotomous Rasch models and most dichotomous item response theory (IRT) models is that dichotomous Rasch models assume  $\sigma_j = 1$  for all items, whereas IRT models typically estimate  $\sigma_j$  for each item. This difference means that Rasch models are normative measurement models (*i.e.*, modeling the measurement, not the data) and IRT models are descriptive statistical models (modeling the data, not the measurement). This difference carries through to polytomous rating scale models also – Rasch models assume  $\sigma_{j,x} = 1$  for all items and thresholds and most IRT models assume  $\sigma_{j,x} = \sigma_j$  for all items and thresholds – with the same consequences for how the models are characterized. Our aim is to estimate *measures* of visual ability from item difficulty ratings, so we employ a Rasch model, which means that the probability density function for measurement errors is  $f(\zeta_{n,j,x} | 0, \sigma_{j,x}) = f(\zeta | 0, 1)$ . A result of this assumption is that all rating scale responses estimated by the model must adhere to a Guttman scale, a *sine qua non* of measurement. In effect, Rasch models estimate person and item measures that generate the most likely Guttman scale underlying the observations, whereas IRT models estimate parameters that generate responses that best fit the data [8].

Again returning to expression (5.1) with the assumption that  $\sigma_{j,x} = 1$ , we see that the probability person  $n$  responds to item  $j$  with **category  $x$  or greater** is  $p(\tau_x + \zeta_{n,j,x} < \alpha_n - \rho_j)$ , which is equal to  $p(\zeta_{n,j,x} < \alpha_n - \rho_j - \tau_x) = \int_{-\infty}^{\alpha_n - \rho_j - \tau_x} f(\zeta | 0, 1) d\zeta$ , a dichotomous Rasch model. Similarly, we see from expression (5.1) that the probability person  $n$  responds to item  $j$  with **less than category  $x + 1$**  is  $p(\alpha_n - \rho_j < \tau_{x+1} + \zeta_{n,j,x+1})$ , which also is in the form of a dichotomous Rasch model,  $p(\zeta_{n,j,x+1} > \alpha_n - \rho_j - \tau_{x+1}) = \int_{\alpha_n - \rho_j - \tau_{x+1}}^{\infty} f(\zeta | 0, 1) d\zeta = 1 - \int_{-\infty}^{\alpha_n - \rho_j - \tau_{x+1}} f(\zeta | 0, 1) d\zeta$ .

The probability person  $n$  responds with category  $x$  to item  $j$  is

$$p(x | \alpha_n - \rho_j) = 1 - (p(\zeta_{n,j,x} < \alpha_n - \rho_j - \tau_x) + p(\zeta_{n,j,x+1} > \alpha_n - \rho_j - \tau_{x+1})) \quad (5.4)$$

which is in the Rasch model form ( $\sigma_{j,x} = 1$ ) of Muraki's modification of Samejima's graded response model [10, 14]. The three polytomous Rasch model parameters –  $\alpha_n$  for each person,  $\rho_j$  for each item, and  $\tau_x$  for each threshold – are estimated for the logistic difference model in Eq. (5.4) [15] using the *method of successive dichotomizations* [5].



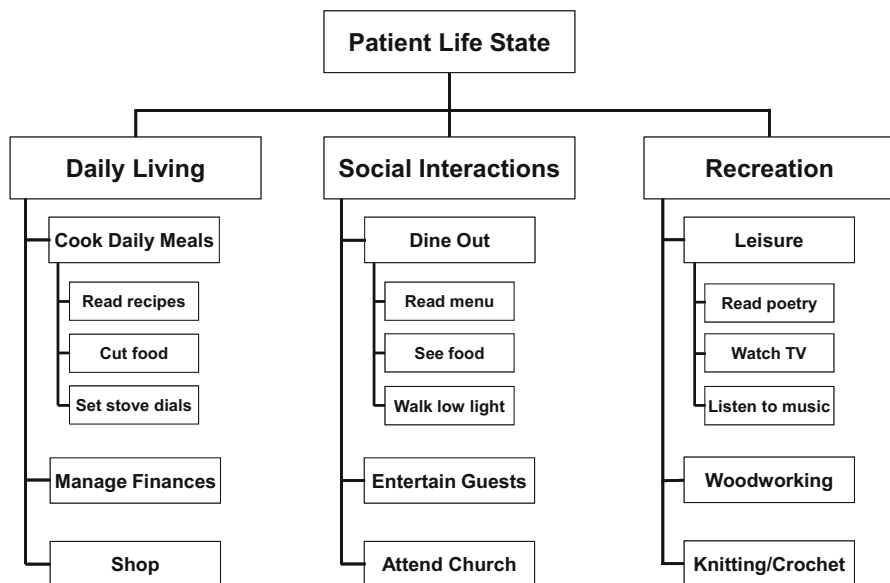
### 5.2.2 *Defining and Organizing the Activity Inventory Item Content*

Because of the demographics of low vision [16], vision rehabilitation primarily targets older visually impaired adults [17]. The low vision patient evaluation typically begins with an intake interview that consists of a health history, visual impairment history, psychosocial history, and functional history [18]. The functional history itemizes how the patient's visual impairment limits her/his ability to live independently, ability to engage in social activities, and ability to engage in favored leisure activities and avocations. For younger patients, the functional history might also cover limitations on employment-related and/or school-related activities, however because most low vision is due to age-related eye diseases, those patients are rare and tend to be referred out of the health care system for vocational rehabilitation or special education services.

The Activity Inventory (AI) was developed both to structure the intake history, so as to facilitate the development of individualized rehabilitation plans, and to provide an adaptive visual function rating scale questionnaire for measuring the low vision patient's functional ability and outcomes of vision rehabilitation [19, 20]. A retrospective chart review of low vision patient intake histories identified 460 common cognitive and motor activities that frequently were mentioned by patients as important for them to be able to perform but were made unusually difficult or precluded by their visual impairments. These very specific activities, called "Tasks" in the AI, were grouped according to the activity "Goals" they serve [19]. Goals in the AI refer to the reason for performing a coordinated set of Tasks (the first letter in the terms "Goals" and "Tasks" is capitalized when referring to items in the AI). For example, "prepare daily meals" is a Goal. There are many ways to prepare daily meals ranging from performing a suite of customary Tasks required to prepare a meal from scratch (*e.g.*, read recipes, measure ingredients, cut food, adjust appliance controls, time cooking, judge doneness of food, etc.) to heating up prepared food in a microwave or conventional oven (*e.g.*, read package instructions and adjust appliance controls).

Two rehabilitation strategies are employed to achieve activity Goals: (1) make usual and customary Tasks less difficult to perform by using vision assistive equipment to enhance vision (*e.g.*, magnifier) or using sensory substitution technology to obviate vision (*e.g.*, bump dots on appliance controls, talking timer, electronic liquid level indicator), or (2) employ adaptive strategies to make it possible to achieve the Goal without having to perform the patient's usual and customary Tasks (*e.g.*, heat prepared food instead of cooking from scratch). Overall outcomes of vision rehabilitation generally are judged in terms of goal attainment, whereas the effectiveness of vision assistive equipment and visual skills instruction tends to be judged in terms of improvements in the performance of tasks.

The difference between education and rehabilitation can be captured by the difference between learning to cook and regaining the ability to cook. Rehabilitation goals are defined by lost functional ability and by patient preferences. However, how much value the patient places on an activity depends on the Objective of the Goal.



**Fig. 5.1** Schematic of the hierarchical Activity Breakdown Structure (ABS) for the AI. Activities can be grouped into three Objectives that define the patient’s life state: Daily Living, Social Interactions, and Recreation (vocation and education are not included as objectives that are addressed by vision rehabilitation services provided within the health care system). Under each Objective are specific activity Goals (*e.g.*, Cook Daily Meals under Daily Living) that often are identified by low vision patients as needed to meet the parent Objective (the AI has 50 Goals: 18 under Daily Living, 11 under Social Interactions, and 21 under Recreation – Goals rated “not important” or “not difficult” are not included in the ABS). The AI has a total 460 Tasks in its calibrated item bank that are nested under the Goals they serve. Tasks rated by the patient as “not relevant” or “not difficult” are not included in the ABS. Besides being nested under Goals, each Task is also assigned to one of 4 functional domains: Reading; Mobility; Visual Motor (*i.e.*, eye-hand coordination); or Visual Information (*i.e.*, perception)

For example, cooking to prepare daily meals might be an extremely important Goal to the patient because it serves the Objective of daily living (*i.e.*, necessary for the patient to live independently). Cooking meals also could serve a social interaction Objective that is important to the patient because the patient places high value on entertaining guests. And cooking could serve a recreation Objective because the patient is rewarded by the joy of cooking. Although heating prepared food in a microwave oven might be an acceptable adaptation if cooking is serving the Daily Living Objective, it might not be acceptable if it is serving the Social Interaction or Recreation Objectives. As schematized by the example in Fig. 5.1, there are 50 common Goals in the AI that are nested under the three Objectives. Borrowing terminology from project management, this hierarchical structure of AI Objectives, Goals, and Tasks is called the *Activity Breakdown Structure (ABS)* [19]. The ABS organizes the items in the AI (Goals and Tasks) in a way that parallels the functional history.

The AI Tasks also are organized by grouping them into four sets according to the type of rehabilitation intervention that might be used. Each set of Tasks is identified as a functional domain: (1) reading function, (2) mobility function, (3) visual motor function, and (4) visual information processing function. Interventions for reduced reading function include various forms of magnification, speech output optical character recognition apps, synthetic speech devices, audio recordings, and braille. Interventions for reduced mobility function include orientation and mobility training (e.g., white cane use), dog guide, navigation apps (e.g., GPS systems), ride share services, talking signs, and remote sighted guide services. Interventions for reduced visual motor function include signature guides, raised line guide for writing, needle threader, syringe fillers, vegetable cutters, oversize keyboards, bump dots, oversize nail clippers with magnifier, and organizational skill instruction. Interventions for reduced visual information processing function include face and object recognition smartphone apps, descriptive audio for movies and events, color identification smartphone apps, and environmental modifications.

It is important to note, as illustrated above, most interventions that constitute vision rehabilitation consist of tools and methods that are specific to a narrow set of activities. The aim of each intervention is to reduce the difficulty of performing the troublesome activity (e.g., liquid level indicators sound an alarm when liquid in a glass or cup reaches a criterion level). Theoretically, this piecemeal approach to vision rehabilitation translates to increasing functional reserve for each item by reducing the visual ability required by the item, *i.e.*, reducing  $\rho_j$  in the model [21]. Selective reductions in item measures because of piecemeal intervention translate to *intervention-specific DIF*.

### 5.2.3 Adaptive Administration of the AI

The present AI item bank consists of 50 Goals distributed under 3 Objectives (Daily Living, Social Interactions, Recreation) and 460 Tasks nested under the Goals. There is redundancy of item content within the list of 460 Tasks, however item measures may vary between different Tasks that have the same content depending on the Goal and Objective the Task is serving. Similarly, there is some redundancy in Goal content because Goals could serve more than one Objective. Returning to the cooking example, “prepare your daily meals” is a Goal under Daily Living, “prepare food for guests” is a Goal under Social Interactions, and “cook or bake for recreation” is a Goal under Recreation. Although the list of Tasks for each of these Goals is very similar with respect to content, they differ in their item measures due to varying performance criteria that must be met to satisfy the Objective served by the parent Goal.

Because the respondent is visually impaired, the AI is administered by interview, usually over the telephone with the assistance of a secure computer-assisted telephone interview (CATI) system. For the baseline interview (before vision rehabilitation services are rendered), the patient is asked to rate the importance of being able

to attain without the assistance of another person one of the 50 Goals. The response choices are “not important”, “slightly important”, “moderately important”, or “very important”. If the patient chooses “not important”, the response is recorded and the interviewer moves on to the next Goal. If the patient chooses any of the other levels of importance, the response is recorded and the interviewer asks the patient how difficult it is to attain the Goal without the assistance of another person. The response choices are “not difficult”, “somewhat difficult”, “moderately difficult”, “very difficult”, or “impossible to do”. The patient’s response is recorded. If the patient responded “not difficult”, then the interviewer advances to the next Goal. If the patient responded with any of the other difficulty rating categories, then the interviewer asks the patient to rate the difficulty of performing each of that Goal’s subsidiary Tasks using the same five difficulty rating categories as used with the Goal, or to respond the Task is “not applicable” to the respondent’s customary way of achieving the Goal.

After completing the rating of Tasks under the Goal, the interviewer moves to the next Goal and repeats all the same steps. This approach is considered adaptive because the patient’s preferences (importance ratings of Goals and applicability responses for Tasks) determine which items will have difficulty ratings elicited. Goals also must be rated at least “slightly difficult” to have difficulty ratings of its subsidiary tasks elicited. The rationale for this adaptive approach is that if a Goal is not important or not difficult at baseline, it will not be included in the individualized rehabilitation plan.

#### ***5.2.4 Properties of Estimated AI Item and Person Measures at Baseline***

Generically, we refer to the estimated functional reserve,  $\varphi_{nj}$ , as *functional ability* of person  $n$  with respect to the content of item  $j$ . Functional ability is a multidimensional construct (*e.g.*, cognitive ability, motor ability, sensory ability, psychological ability, etc. are components of functional ability). Each functional ability dimension has its own multidimensional structure (*e.g.*, sensory ability can be expressed as visual ability, hearing ability, taste ability, etc.), and each component ability subdimension can be expressed further with its own dimensional structures (*e.g.*, visual ability can have independent components differentially affected by ophthalmic diseases: night vision, visual acuity, peripheral vision, color vision, etc.), all of which are properties of the person expressed at different levels of detail that must be inferred from observations by way of theory.

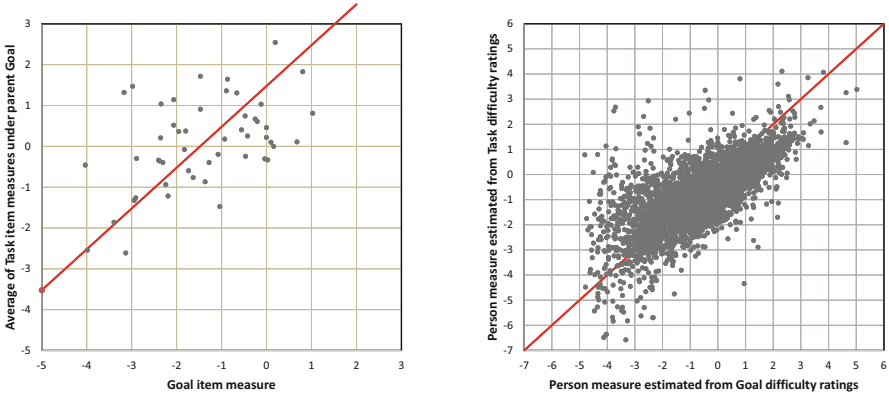
The person and item measures estimated from Rasch analysis of AI difficulty ratings correspond to the magnitude of an origin-bound functional ability vector in the multidimensional functional ability space. The direction of the vector is determined by item content and by traits of the people rating the items. For example, some vision-dependent items might describe activities, like reading medication instructions, that depend more heavily on cognitive ability than on motor ability and the

reverse might be true for other items, such as signing a check. Also, some low vision patients might have cognitive impairments, whereas other low vision patients might have physical impairments that differentially influence their ratings of different items.

These differences combined with variations between items in the demand placed on vision and variations between people in vision impairment severity will give rise to variations in the direction and magnitude of the resultant vector. To the extent that variations in visual ability is the common denominator for all patients and items, we presume the magnitude of the average vector across items and persons represents *visual ability*. All else being equal, or at least randomly distributed, variations of vision impairment severity between persons and variations of demand on vision between items will give rise to variations in functional reserve, which can be characterized as variations in the magnitude of the operationally-defined “visual ability” vector in functional ability space. Variations between items in the direction of this visual ability vector depend on how much demand the items place on other functional ability dimensions (e.g., cognitive demand, physical motor demand, psychological demand). Variations between persons in the direction of this visual ability vector depend on the types and magnitudes of other functional impairments the person may have (e.g., cognitive, physical motor, psychological disorders). Variations in measurements of “visual ability” depend on both the amount of deviation in vector direction and the magnitude of the deviated vector, which projects onto the defined visual ability vector, the magnitude of which corresponds to the estimated measures. We use this vector representation to guide our analyses of sources of variance and covariance in estimated person and item measures from AI difficulty ratings by people with low vision.

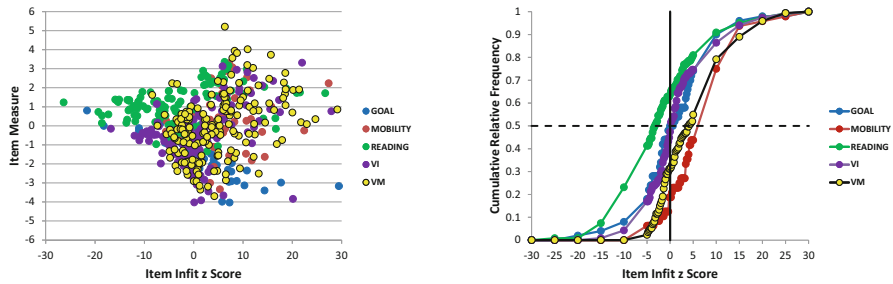
Item measures for the 510 Goals and Tasks in the AI were estimated using the method of successive dichotomizations [5] from the difficulty ratings of about 3600 low vision patients at pre-rehabilitation baseline [22]. Because activity Goals are attained by successfully performing some subset of their subsidiary Tasks, the difficulties of attaining Goals, and therefore their item measures, are expected to be a monotonically increasing function of the difficulty of performing their subsidiary Tasks. The left panel of Fig. 5.2 illustrates a scatterplot of mean Task item measures (on the ordinate) vs parent Goal item measures (on the abscissa), both specified as a difference from the mean of all Goal item measures and the mean of all average Task item measures respectively, along with the expected relationship given the respective means (red line) [23]. The Pearson correlation is 0.48. These results are consistent with the hypothesis that the difficulty of an activity Goal is inherited from the difficulties of the more specific activity Tasks that serve the Goal.

The Goal and Task item measures are estimated simultaneously on the same scale, therefore person measures estimated from Goal difficulty ratings should agree with person measures estimated from Task difficulty ratings. The right panel of Fig. 5.2 illustrates a scatterplot of person measures estimated from Task difficulty ratings (on the ordinate) versus person measures estimated from Goal difficulty ratings (on the abscissa) along with the expected identity relationship (red line) for approximately 3600 low vision patients at pre-rehabilitation baseline. The Pearson correlation is 0.71.



**Fig. 5.2** *Left panel:* Scatter plot of the average of all AI item measures of Tasks serving each Goal versus the corresponding Goal item measure. The red line is the identity line with respect to deviations of Goal measures from the mean Goal measure and deviation of the average Task measures from the mean of the average Task measures. Each data point corresponds to one of the 50 Goals. *Right panel:* Scatter plot of person measures estimated from all Tasks rated by the person at baseline versus person measures estimated from all Goals rated by the same person at baseline. The red line is the identity line. Each data point corresponds to a separate person

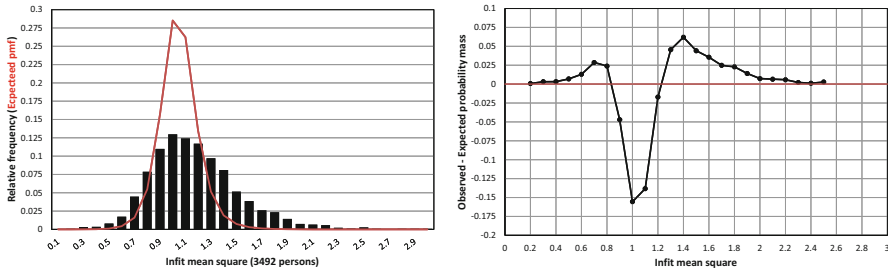
Rasch models are normative measurement models. As opposed to most IRT models, which are descriptive statistical models of observations, Rasch models assume that the density function for random deviates,  $f(\zeta|0, 1)$ , is the same for every combination of persons and items. The validity of using a Rasch model to estimate measures from observations is tested by determining if the observations conform to the premises of the model (*i.e.*, determining if “the data fit the model” rather than fitting the model to the data). The *information-weighted mean square fit statistic* (*infit*) is used to test the validity of item and person measure estimates. Equation (5.4) is the probability that person  $n$  would respond with category  $x$  to item  $j$ . Thus, if  $x_{n,j}$  is the observed response of person  $n$  to item  $j$  and  $\mathbb{E}\{x_{n,j}|\alpha_n, \rho_j\} = \sum_{x=0}^4 xp(x|\alpha_n - \rho_j)$  is the response of person  $n$  to item  $j$  expected by the model, then the sums of squares of observed minus expected responses across persons for each item is  $SS_j = \sum_{n=1}^{N_j} (x_{n,j} - \mathbb{E}\{x_{n,j}|\alpha_n, \rho_j\})^2$ , where  $N_j$  is the number of persons who rated item  $j$ . The expected sums of squares for item  $j$  is  $\mathbb{E}\{SS_j\} = \mathbb{E}\{x_{n,j}^2|\alpha_n, \rho_j\} - \mathbb{E}\{x_{n,j}|\alpha_n, \rho_j\}^2$ , for which  $\mathbb{E}\{x_{n,j}^2|\alpha_n, \rho_j\} = \sum_{x=0}^4 x^2 p(x|\alpha_n - \rho_j)$ . Assuming the source of variance is normally distributed, we expect  $SS_j$  to have a chi-square distribution with  $N_j - 1$  degrees of freedom. The expected value of a chi-square distribution is its degrees of freedom (*df*). Thus, the ratio of  $SS_j$  to the expected value of  $SS_j$ , given the assumption of an underlying normal density function as the source of variance, should be  $\frac{SS_j}{\mathbb{E}\{SS_j\}} = \frac{\chi^2}{df}$ . This ratio is called the *infit*. If  $df > 25$ , a cube-root transformation of  $\chi^2$  is a good approximation to a normal distribution (*i.e.*, Wilson-Hilferty transform [24]).



**Fig. 5.3** *Left panel:* Scatter plot of item measures for each of the 50 Goals and 460 Tasks as a function of the z-score for their respective item measure infit mean square. The Tasks are color coded by the functional domain to which they are assigned: Goals (blue), Mobility (red), Reading (green), Visual Information Processing (purple), and Visual Motor (yellow). The expected value for the infit z-score is 0. *Right panel:* Cumulative frequencies of the infit z-scores for each functional domain as identified by color in the left panel

We estimated the infit for each of the 50 Goals and 460 Tasks in the AI from the responses of our large sample of low vision patients rating the difficulty of AI items and transformed the infits to standard normal deviates (z-scores) for each item. The results for all Goals and Tasks are illustrated in the left panel of Fig. 5.3 as the covariance of infit z-scores (abscissa) and item measures (ordinate) for each item. There is a very weak correlation between infit mean squares and item measures ( $r = 0.12$ ). If the estimated measures conformed to the unidimensional assumption of the Rasch model (*i.e.*, magnitude of a single visual ability vector,  $\varphi_{n,j}$ ) contaminated by a single source of normally distributed random error ( $\zeta$ ), 97% of the points would fall symmetrically about zero within a z-score range of  $\pm 2$ . Clearly, these assumptions are violated [25]. [N.B. Because the AI is adaptive, the number of persons who rated each item varied between items, so  $df$  must vary between items. The mean of a chi-square distribution is  $df$ , the variance is  $2df$ , the skewness is  $\sqrt{8/df}$ , and the kurtosis is  $12/df$ , thus variations in  $df$  across items result in variations in the shape of the composite infit distribution. However, because  $df$  is very large for all items, skewness and kurtosis are approximately zero. Although the Wilson-Hilferty transform can be used for each item infit, it is necessary to employ each item's respective  $df_j$  when transforming to z-scores.]

As described above, each Task in the AI is assigned to one of four functional domains: (1) reading, (2) mobility, (3) visual information processing, or (4) visual motor. The AI Goals and the functional domains to which the AI Tasks are assigned are color-coded in Fig. 5.3. It can be seen that the colors form different clusters of infit z-scores. The right panel of Fig. 5.3 shows cumulative frequencies of infit z-scores for Goals and for Tasks in each of the four functional domains. The infit z-score cumulative frequency functions are similar for each of the five groups of items, but the median z-score values (z-score corresponding to 0.5 cumulative frequency) vary between functional domains (the expected value of the median is



**Fig. 5.4** *Left panel:* Histogram (black bars) of the person measure infit mean square values. The red curve is the expected probability mass function (same bin width as the histogram) for the person measure infit mean square given the degrees of freedom (1 less than the number of items rated) for each patient. *Right panel:* Difference between the observed (black bars in the left panel) and expected (red curve in the left panel) probability mass values for each infit mean square value. Differences greater than 0 indicate more variance observed in the distribution of person measure deviates than expected and differences less than 0 indicate less variance than expected

a z-score of 0 – solid vertical line). The median of the infit z-scores for Goals (0.2) and visual information processing Tasks (0.1) are very close to the expected value of zero. The median infit z-score for reading Tasks (−3.2) is much lower than the expected value of zero (indicative of error variance less than the expected amount of variance) and the median infit z-score for mobility Tasks (6) and visual motor Tasks (4.1) are much higher than the expected value of zero (indicative of error variance more than the expected amount of variance). These results suggest a strong dimensional structure to the estimated visual ability measure.

An infit also can be estimated for each person, but since the AI is administered adaptively, the number of items rated varies between people. The infit for person  $n$  is  $\frac{SS_n}{E\{SS_n\}} = \frac{\chi^2}{df_n}$ , for which  $df_n = J_n - 1$ , where  $J_n$  is the number of items rated by person  $n$ . Thus, for adaptive testing, the expected infit frequency distribution is a sum of weighted chi-square distributions. The left panel of Fig. 5.4 illustrates a histogram of person infits (black bars). The red curve is the expected chi-square mixture probability mass function [26] (pmf – same bin width as the histogram) estimated for the 3600 respondents to the AI. The chi-square mixture pmf is the sum of weighted chi-square pmfs for different values of  $df_n$  with the weight equal to the fraction of persons who rated  $df_n + 1$  items in the AI.

The right panel of Fig. 5.4 illustrates the difference between the observed and expected pmfs in the left panel of Fig. 5.4. About 27% of the persons had excess error variance in the observed responses (positive differences for infits greater than 1) and 8% had less error variance than expected (positive differences for infits less than 1). As discussed more formally in a later section, much of the excess variance may be due to functional limitations caused by comorbidities.

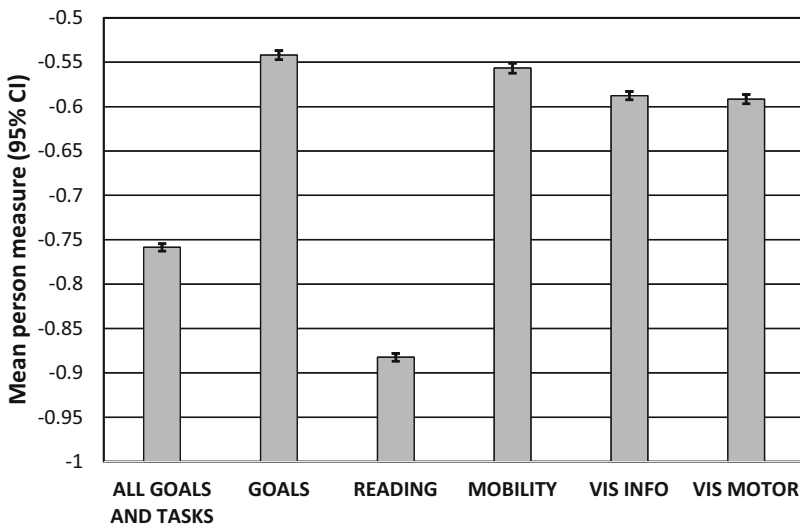


### 5.3 Functional Domains and Differential Person Functioning (DPF)

The analysis of item infit statistics by functional domains, as summarized in Fig. 5.3, suggests a strong dimensional structure to the estimated visual ability variable. Such a dimensional structure would correspond to DPF. Figure 5.5 shows that person measure distributions are significantly different when estimated from difficulty ratings of Goals and of Tasks in each of the four functional domains (ANOVA:  $F = 30.95$ ,  $df_B = 5$ ,  $df_W = 20,502$ ,  $p = 1.7 \times 10^{-31}$ ). Post hoc paired t-tests with Bonferroni adjustment for multiple comparisons showed that differences between all pairs are highly significant, except for visual information processing and visual motor functions ( $p = 0.37$ ).

In a vector representation of covariances, vector magnitude corresponds to the square root of the variance (*i.e.*, standard deviation in units of the measure) explained by orthogonal factors and the cosine of the angle between any pair of vectors corresponds to the correlation between the variables those vectors represent. Such factors could represent visual components of the visual ability variable and/or contributions to estimated measures from other functional ability dimensions, such as cognitive disorders, physical motor limitations, and depressed psychological state.

Therefore, we employed factor analysis with principal axis factoring and varimax rotation (*i.e.*, optimizing the rotation of the orthogonal factors to maximize the



**Fig. 5.5** Mean estimated person measure for each functional domain (Reading, Mobility, Visual Information Processing, and Visual Motor); mean of person measures estimated from all difficulty ratings of Goals and Tasks combined and from difficulty ratings of only Goals. Error bars represent 95% confidence intervals. All differences are statistically significant except the difference between Visual Information Processing person measures and Visual Motor person measures

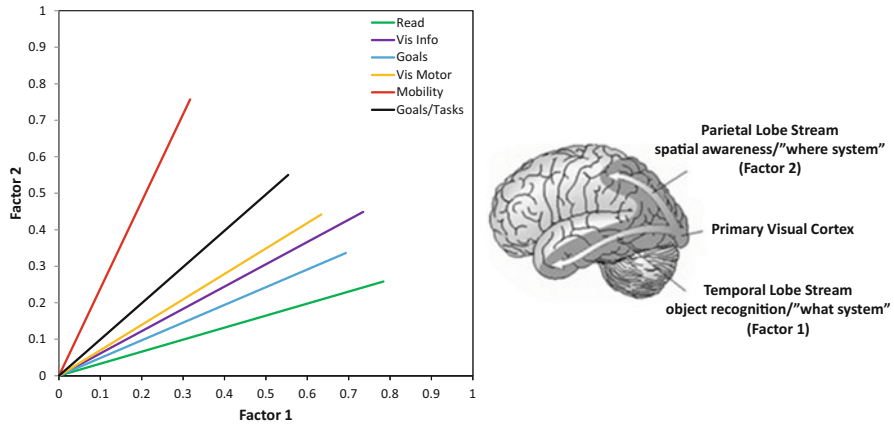
**Table 5.1** Correlation matrix for person measures estimated from AI difficulty ratings (gray cells) and from Rasch analysis of responses to more general health questionnaires (orange cells) – the GDS (depression), SF-36-PFS (physical ability), and TICS (cognitive disorders). The blue cells in the upper right are normalized linear regression model weights ( $\beta$ ) on person measures estimated from GDS, SF-36-PFS, and TICS ratings to predict person measures estimated from AI difficulty ratings for each of the functional domains

<i>r</i>	READING	MOBILITY	VIS INFO	VIS MOTOR	GOALS	$\beta$		
						GDS	SF-36-PMS	TICS
READING	1.00	*	*	*	*	0.16	0.06	0.13
MOBILITY	0.48	1.00	*	*	*	0.11	0.13	0.02
VIS INFO	0.69	0.58	1.00	*	*	0.11	0.06	0.01
VIS MOTOR	0.75	0.62	0.77	1.00	*	0.13	0.11	0.02
GOALS	0.75	0.60	0.72	0.81	1.00	0.29	0.24	0.01
GDS	0.21	0.30	0.24	0.32	0.39	1.00	*	*
SF-36-PMS	0.18	0.37	0.22	0.35	0.36	0.43	1.00	*
TICS	0.19	0.26	0.15	0.23	0.15	0.21	0.29	1.00

variance in each measure explained by each factor), on the five sets of person measures estimated from difficulty ratings of Goals and of the four subsets of Tasks representing the functional domains.

We learned that two factors explain the correlation matrix (gray cells in Table 5.1) and account for 70% of the variance in the five sets of person measures [22, 23, 27]. The left panel of Fig. 5.6 illustrates the two orthogonal factors and the vectors for the five sets of person measures plus the vector for the principal axis (black). Reading (green) loads most heavily on factor 1 and Mobility (red) loads most heavily on factor 2. Visual information processing (yellow) and visual motor (violet) vectors, and the vector for Goals (blue) are close to the principal axis (black vector, which corresponds to all Goals and Tasks combined).

We speculate that factor 1 represents central vision (*e.g.*, altered by visual acuity and contrast sensitivity losses) and factor 2 represents peripheral vision (*e.g.*, altered by visual field loss and blind areas in vision called scotomas). Not only can these two types of visual impairment occur independently, they also have different effects on visual perception. There is neuroanatomical and neurophysiological evidence of two visual pathways in the brain following visual processing in the primary visual cortex (V1), one in the parietal cortex and the other in the temporal cortex. The parietal pathway, sometimes called the “where” system, receives most of its input from the peripheral retina and appears to be responsible for visual perceptual processing related to spatial awareness and visual control of actions, whereas the temporal pathway, sometimes called the “what” system, receives most of its visual input from the central retina and appears to be responsible for object identification and interpretation of patterns [28]. Changes in visual acuity results in changes in the reading threshold (minimum size of print that can be read at all), whereas central scotomas reduce the maximum (asymptotic) reading rate that can be achieved with enlarged print [29], which suggests how the two factors can contribute independently to reduced reading function in low vision.



**Fig. 5.6** *Left panel:* Results of exploratory factor analysis with varimax rotation on the correlation matrix of 6 sets of person measures estimated from Task difficulty ratings in each of the 4 functional domains (green, red, purple and yellow vectors), Goal difficulty ratings (blue vector), and difficulty ratings of all Goals and Tasks combined (black vector). Vector magnitude corresponds to the square root of the variance in the person measures that is explained by the two factors and the cosine of the angle between any two vectors is equal to the correlation between the respective person measures. *Right panel:* The Mobility vector loads most heavily on factor 2, the Reading vector loads most heavily on factor 1, and the Goals/Tasks vector loads equally on the two factors. We hypothesize that these factors are independent components of visual ability that reflect the two visual pathways observed in the parietal and temporal lobes of the cortex. To the extent that mobility relies on spatial awareness and control of actions, which are attributed to the parietal lobe that receives most of its input from peripheral retina, and reading relies on object recognition, which is attributed to the temporal lobe that receives most of its input from the central retina, we expect visual field loss and scotomas to have their largest effect on functions that depend on factor 2 and visual acuity loss to have its largest effect on functions that depend on factor 1

These two independent vision-related factors have been observed repeatedly over the past several years contributing to person measures estimated for different samples of visually impaired patients from their responses to different visual function questionnaires [30, 31]. There also have been reports of significant contributions to person measure estimates from physical functioning (as measured by the SF-36 physical functioning scale [SF-36-PFS] [32]), cognitive functioning (as measured by the Telephone Interview for Cognitive Status [TICS] [33]), and psychological state (as measured by the Geriatric Depression Scale [GDS] [34], Center for Epidemiologic Studies – Depression Scale [34], or Patient Health Questionnaire-9 [35]).

However, as itemized in the last three rows of cells below the diagonal in Table 5.1 (pink highlight), the correlations ( $r$ ) between these health state measures and AI measures for Goals and each of the four functional domains are weak [36]. These low correlations suggest that additional independent factors are required to explain the covariances that are added to visual ability estimates by co-morbidities [27, 37]. Even though individual correlations are weak, the physical, psychological, and cognitive health states are statistically significant predictors of the Goal and

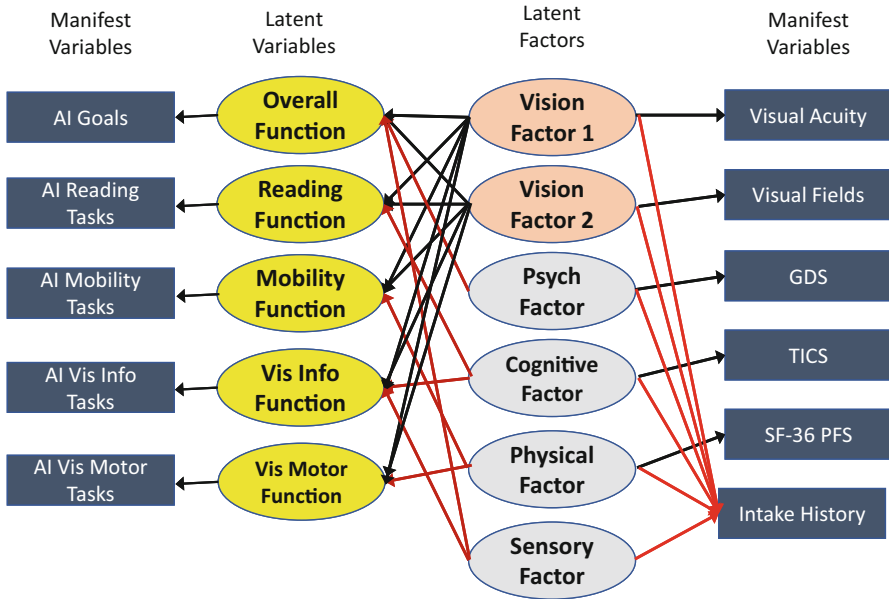
functional domain measures in a multivariate linear model. Table 5.1 lists above the diagonal (blue highlight) the normalized weights in the linear model ( $\beta$ ) [36]. Although small in all cases, the GDS weight is significantly different from zero for all five functional domains; the SF-36-PFS weight is significantly different from zero for all domains except reading; and the TICS weight is significantly different from zero for all domains except visual information processing.

### 5.3.1 Latent Variable Model for Sources of Variance in AI Visual Ability Measures

To better understand how co-morbidities contribute to estimates of visual ability measures from AI Goal and Task difficulty ratings, we constructed the conceptual path diagram schematized in Fig. 5.7. Estimated latent intervening variables are symbolized with yellow ellipses; latent factors are symbolized with salmon and gray ellipses; and observed manifest variables are symbolized with blue rectangles. The arrows identify paths by which the inferred latent factors give rise to the observed indicators (manifest variables, which are Goal and Task difficulty ratings), both directly and by way of intervening latent variables (estimated person measures). Vision Factor 1 corresponds to Factor 1 (“what” visual processing) in Fig. 5.7 and Vision Factor 2 corresponds to Factor 2 (“where” visual processing) in Fig. 5.7.

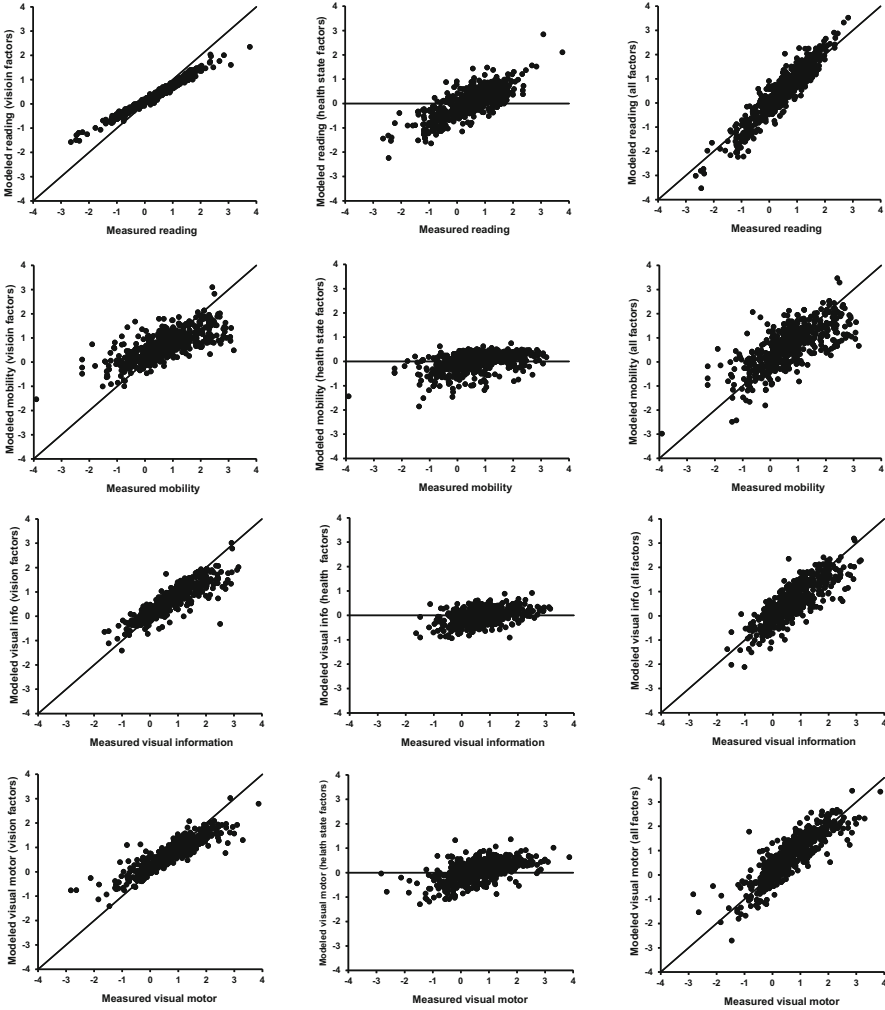
The black arrows from the vision factors (salmon ellipses) to each intervening latent variable has a fixed weight that is estimated from the factor analysis summarized in Fig. 5.6. Each weight corresponds to the projection of the respective Goals or functional domain vector onto that factor. Each red arrow from the independent latent systemic health state factors (gray ellipses) to each intervening latent variable has a weight estimated from structural equation modeling. From the perspective of visual ability, the latent systemic health state factors are acting in the role of effect modifiers on the intervening latent variables (*i.e.*, person measures estimated from Rasch analysis). The latent visual factors also are predictors of visual impairment measures (visual acuity and visual fields by way of regression models), continuous latent variables (incorporated in Psych, Cognitive, and Physical latent intervening variables that are not shown in the paths to the manifest variables on the right in Fig. 5.7) estimated from Rasch analysis of rating scale responses to depression (GDS), cognitive functioning (TICS), and physical functioning (SF-36 PFS) questionnaires and categorical responses (scored as a dichotomous grouping variable, *i.e.*, 0,1) to a detailed health/functional/psychosocial intake history questionnaire [36].

Each observation also has sources of random variance (not shown in Fig. 5.7) and covariances (also not shown). The weights on the unfixed paths were estimated by structural equation modeling.



**Fig. 5.7** Conceptual path model of how latent vision and non-visual factors contribute to explaining observed difficulty ratings of items in the AI (blue rectangles on the left); to observed psychophysical measures of visual impairments (blue rectangles for visual acuity and visual fields on the right); to observed psychometric measures of depression severity (GDS), cognitive impairment (TICS), and physical limitations (SF-36-PFS); and to self-reported co-morbidities (intake history). The black arrows identify paths for which weights are estimated from Rasch analysis and regression models and the red arrows identify paths for which weights are estimated from structural equation modeling

The structural equation model constructed from the conceptual design in Fig. 5.7 anchored the intervening latent visual function domain variables to person measures estimated from Rasch analysis of AI Goal and Task difficulty ratings; the latent vision factor value for each person to values estimated from principal axis factoring of the visual function domains plus visual acuity and contrast sensitivity covariance matrix; and the systemic health state factor values for each person from Rasch analysis of item responses to the GDS, TICS, and SF-36 PFS and from regression models of intake self-reported indicators in the Intake History [27]. The first row of Fig. 5.8 illustrates predicted reading function from the two vision factors vs measured reading function (left panel), predicted reading function from the 4 health state factors vs measured reading function (center panel), and predicted reading function from all 6 latent factors combined vs measured reading function (right panel). The second row makes the same comparisons for mobility function, the third for visual information function, and the fourth for visual motor function. The predicted person measures from the vision factors for each functional domain (first panel in each row) are linearly related to the measured values, but they are not accurate (i.e., they do not fall on the identity line). If health state factors make a consistent contribution to



**Fig. 5.8** *Left column:* Scatter plots of person measures modeled from a general linear model (GLM) using only loadings from exploratory factor analysis (vision factor 1 and vision factor 2 in Fig. 5.7) versus the person measure estimated from Rasch analysis of difficulty ratings of AI items for each functional domain: Reading (row 1), Mobility (row 2), Visual Information Processing (row 3), and Visual Motor (row 4). The identity line is the expected relationship if only vision factors contributed to the estimated measures. *Middle column:* Scatter plots of person measures modeled from a GLM using only loadings from the systemic factors (Psych, Cognitive, Physical, and Sensory factors in Fig. 5.7) versus person measures estimated from Rasch analysis of difficulty ratings of AI items in each functional domain. The horizontal line is the expected relationship if non-vision factors did not contribute to the measures estimated from Rasch analysis. *Right column:* Scatter plots of person measures modeled from a GLM using loadings from all 6 factors in Fig. 5.7 versus person measures estimated from Rasch analysis of difficulty ratings of AI items in each functional domain. The identity line is the expected relationship if the 6 independent factors accounted for all contributions to the observed person measures

visual function measures, then the predicted values in the second panel of each row should correlate with the measured values. The Pearson correlations are 0.77 for Reading (first row), 0.44 for Mobility (second row), 0.44 for Visual Information (third row), 0.56 for Visual Motor (fourth row), and 0.61 for overall visual ability (estimated from Goal difficulty ratings and not shown). The predictions of measured values for the full model (2 vision factors and 4 systemic health state factors combined) are shown in the third column of Fig. 5.8. The addition of contributions from the systemic health state factors not only improve the accuracy of the predicted person measures, it also increases variance about the identity line which provides explanations of previously unexplained variance in the person measures.

## 5.4 Intervention-Specific Differential Item Functioning (DIF)

Upon completion of rehabilitation services, the AI is re-administered adaptively with a follow-up CATI using a slightly different algorithm. Patient ratings are elicited the same way they were at baseline, except that any Goals rated not important or not difficult at baseline are not re-administered. If at baseline a Goal was given a difficulty rating greater than not difficult, then the difficulties of that Goal’s subsidiary Tasks are rated irrespective of the Goal difficulty rating at follow-up. Also, any Tasks rated not difficult or not applicable at baseline are not re-administered. The rationale for this “item-filtering” approach is that rehabilitation of Goals and Tasks rated not important or not applicable or not difficult at baseline is of no utility to the patient and those activities would not be included in the individualized rehabilitation plan.

### 5.4.1 Increasing Functional Reserve

In the case of vision rehabilitation, the aim of intervention is to increase the patient’s ability to function, *i.e.*, increase the patient’s functional reserve for activities targeted in the individualized rehabilitation plan. We typically think of increasing functional reserve by increasing the patient’s visual ability,  $\varphi_{nj} + \Delta\varphi_{nj} = (\alpha_n + \Delta\alpha_n) - \rho_j$ , *e.g.*, improving the patient’s vision by correcting refractive error with new glasses. However, we also can increase the patient’s functional reserve by decreasing the visual ability required to perform the activity,  $\varphi_{nj} + \Delta\varphi_{nj} = \alpha_n - (\rho_j + \Delta\rho_j)$ , *e.g.*, by equipping the patient with a magnifier. Most generally, a change in the patient’s functional reserve reflects either a change in the person measure and/or a change in the item measure,  $\Delta\varphi_{nj} = \Delta\alpha_n - \Delta\rho_j$ . Thus, the overall outcome of vision rehabilitation for person  $n$  is the average change in functional reserve over the  $J_n$  activities identified in that person’s individualized rehabilitation plan:

$$\Delta\varphi_n = \sum_{j=1}^{J_n} \frac{\Delta\varphi_{nj}}{J_n} = \sum_{j=1}^{J_n} \frac{\Delta\alpha_n - \Delta\rho_j}{J_n} = \Delta\alpha_n - \sum_{j=1}^{J_n} \frac{\Delta\rho_j}{J_n} \quad (5.5)$$

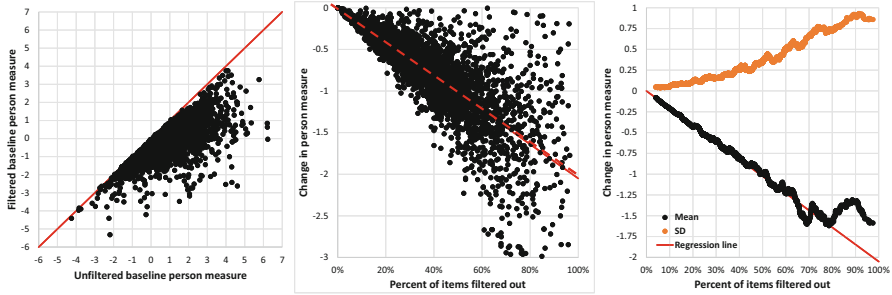
If  $\Delta\rho_j \neq 0$  in Eq. (5.5), we must conclude that the intervention resulted in intervention-specific DIF [21]. In this case, calibrating the AI item bank at baseline and anchoring the item measures,  $\rho_j$ , to their baseline values effectively defines  $\Delta\rho_j = 0$  for all items. Thus, anchoring item measures to calibrated values forces the average change in functional reserve for person  $n$ ,  $\sum_{j=1}^{J_n} \frac{\Delta\varphi_{nj}}{J_n}$ , into  $\Delta\alpha_n$  when estimating outcome measures [38]. Although mathematically equivalent for a single patient ( $n$ ), from the patient's perspective an average of changes in the difficulty experienced performing selected individual activities might not be equivalent to an equal size change in visual ability. Indeed, most low vision patients simply want to be able to "see better", not have to learn new behavior and function with the assistance of an array of activity-specific and often costly devices. The limitation of Eq. (5.5) is that  $\Delta\rho_j$  has the same weight for every item even though the same size change in item difficulty for different items might have different utilities for a given person depending on their difficulty at pre-rehabilitation baseline.

#### 5.4.2 Rehabilitation Demand and Item Filtering

If a person reported that none of the activities sampled by the AI were both important (or relevant) and difficult, then it is likely that person would have no need for rehabilitation. The visually impaired consumer's demand for rehabilitation is driven by that person's desire to be able to perform activities that he or she deems necessary to regain lost quality of life. If we use the criterion that an activity rated "not difficult" or "not important" (or not applicable) at baseline is not worthy of being included in an individualized rehabilitation plan, then that activity has no *rehabilitation demand* and the item should be dropped from the analysis [19, 20]. We refer to this selective removal of items from the analysis based on the patient's responses as "item filtering".

Filtering items by estimating the person measure at baseline from responses only to items rated at least "somewhat difficult" biases the person measure toward more negative values. The left panel of Fig. 5.9 illustrates this effect of item filtering on the person measure estimate in a scatterplot of  $\alpha_{n(\text{filtered})}$  versus  $\alpha_{n(\text{unfiltered})}$  (points) compared to the identity line (red line) for 3600 low vision patients. The person measures estimated from responses to filtered items (remaining items after removing those rated "not difficult") are more negative than or equal to person measures estimated from responses to unfiltered items (all rated items). The center panel of Fig. 5.9 shows that the difference between filtered and unfiltered person measures ( $\alpha_{n(\text{filtered})} - \alpha_{n(\text{unfiltered})}$ ) is linear with a slope of  $-2$  and an intercept of 0 as a





**Fig. 5.9** *Left panel:* Scatter plot of filtered person measures (i.e., items rated “not difficult” excluded from the person measure estimates) versus unfiltered person measures (estimated from all item ratings). Each data point is a different person and all points are on or below the red identity line indicating that item filtering biases the person measure estimate toward lessor ability. *Center panel:* Difference between the filtered and unfiltered person measure for each person versus the percent of items that were filtered out before the estimate was made. The red-dashed regression line fit to the data is plotted for comparison. *Right panel:* Moving average change in person measure with item filtering (black points) is plotted along with the red regression line fit to the data in the center panel. As the average change in person measure decreases with increasing percentage of the items filtered out, the moving standard deviation of the change in person measure increases (orange points)

function of the percent of items filtered out. Written another way, which would describe the trend in the data in the left panel of Fig. 5.9,  $\alpha_{n(\text{filtered})}$  is equal to  $\alpha_{n(\text{unfiltered})}$  plus a negative bias that is twice the percentage of items that have been filtered out for that person,  $\alpha_{n(\text{filtered})} = \alpha_{n(\text{unfiltered})} - \frac{2(J_{n(\text{unfiltered})} - J_{n(\text{filtered})})}{J_{n(\text{unfiltered})}}$ , for which  $J_{n(\text{filtered})}$  and  $J_{n(\text{unfiltered})}$  refer to the number of filtered and unfiltered items, respectively, that were rated by person  $n$ . However, note in the right panel of Fig. 5.9 that the moving standard deviation of the change in person measure increases with increasing percentage of the items filtered out (orange points) as the moving average change in person measure (black points) decreases linearly along the regression line (red line from the center panel of Fig. 5.9). The increase in variance with increases in the percentage of items filtered out is likely linked to differences between persons in the distributions of item measures among remaining items.

### 5.4.3 Utility to the Patient of Increasing Functional Reserve

In economics, the term *utility* refers to a quantity of how useful or desirable something is to a person. Utilities are preference values specified in relative units, often on a continuous ratio scale ranging from zero (no value) to 1.0 (maximum value of things being compared). Currently, the closest the AI comes to eliciting information from the patient about their preferences for being able to perform an activity is the elicitation of ordinal importance ratings of Goals and dichotomous

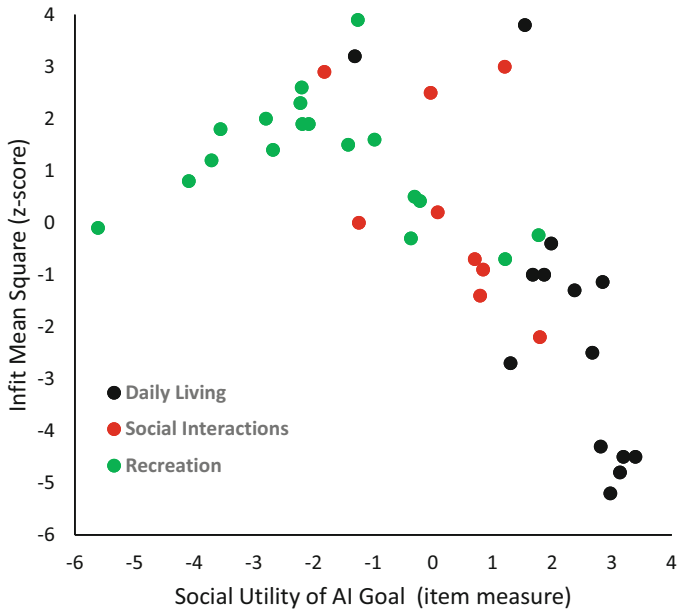
responses of relevance to the individual of Tasks, both scored as 0 or 1. These binary scores are used to weight the items for the purpose of determining whether the item is filtered out or retained when estimating the person measure. Even though a polytomous scale is used to rate the importance of Goals, the importance ratings are dichotomized for the purpose of item filtering.

For an outcome measure to be truly person-centered, it must factor in the individual's preferences for specific outcomes. The term *disutility* refers to a negative utility value, *i.e.*, the utility of something a person is willing to trade to be rid of something else that is undesirable. The greater the disutility of the patient's functional state, the greater is the rehabilitation demand. In the case of vision rehabilitation, the disutility for an individual of a particular functional state should be estimated in terms of the amount of time, effort, and resources the person is willing to expend to achieve a specific less disabled state. Although we can define functional ability by how difficult it is for an individual to perform activities that are important to her or him, the utility of the specific functional ability outcome would be determined not only by the importance of the activity to the individual, but also by the level of effort required to satisfy the objective of the activity. For example, cooking daily meals may have high utility to an individual even if it is difficult to do because of its contribution to the objective of independent daily living, whereas cooking for the objective of recreation (*i.e.*, joy of cooking) may have high utility to the person only if it is easy enough to be enjoyable. Looked at another way, for the person to realize a net gain from vision rehabilitation, the utility of a functional ability outcome must be equal to or greater than the utility of the person's time, effort, and resources that must be expended to achieve that outcome.

#### 5.4.4 *Social Utility of AI Goals*

The question we are raising is one of how best to measure the utility of a functional outcome. So, why not simply have the patient rate the importance of being able to perform with ease activities described by items in the AI and then use Rasch analysis of those ratings to estimate the utility of the outcome of vision rehabilitation? As part of its adaptive design, the patient already rates the importance of all the Goals in the AI. Rasch analysis assumes agreement within the population of the ordering of items, which might not be true of item importance. Applying Rasch analysis to importance ratings of AI Goals implies the existence of a consensus, which would translate to a latent variable that might best be described as the "social utility" of performing the itemized activities independently [39]. We explored this idea by using Rasch analysis to construct social utility measurement scales from importance ratings by 600 low vision patients [39].

As illustrated in Fig. 5.10, we observed that the putative social utility of performing personal hygiene activities independently (item measure = 3.39) is greater than the social utility placed on shopping independently (item measure = 1.86), but this is not necessarily true for an individual. The level of consensus



**Fig. 5.10** Scatter plot of the z-score for the item infit mean square versus the item measure estimated from Rasch analysis of importance ratings of AI Goals. Each point is 1 of the 50 AI Goals color-coded by the Objective in the Activity Breakdown Structure (Fig. 5.1) to which the Goal is assigned: Daily Living (black), Social Interactions (red), or Recreation (green). The infit z-score is 0 if variance of the distribution of deviates from the item measure estimate is at the expected value, negative if the variance is less than expected (high agreement between persons), or positive if the variance is more than expected (high disagreement between persons)

(dispersion between people of importance ratings relative to the expected ratings for each Goal) is captured in the Rasch model infit statistic and, as shown in Fig. 5.10, consensus is strong (low dispersion) for AI Goals with high social utility (Daily Living – black circles) and Goals with low social utility (Recreation – red circles) and is weakest (high dispersion) for AI Goals with medium social utility (Social Interactions – green circles), giving rise to a parabolic relationship between consensus and estimated social utility (*i.e.*, people agree most on how to order the importance of Goals on high valued activities and on low valued activities). Overall, AI Goals under the Daily Living Objective have the highest social utility (mean item measure = 2.17; SD = 1.21), Goals under the Social Interaction Objective have intermediate social utility (mean = 0.26; SD = 1.16), and Goals under the Recreation Objective have low social utility (mean = -1.82; SD = 1.16).

The objective of vision rehabilitation is to make activity goals that are important to the individual and difficult or impossible to achieve easier by way of vision enhancement methods and technology, adaptations (*i.e.*, adopting new strategies and using assistive technology that obviate performance of customary vision-dependent tasks), and modified independence (limited use of human assistance to overcome

intractable or high safety risk barriers that prevent attainment of a specific end goal). In terms of the theory behind the AI, the amount of difficulty a person experiences when attempting a specific activity is determined by functional reserve ( $\varphi_{nj}$ ). The objective of vision rehabilitation is to increase functional reserve, thereby reducing the difficulty of performing the activity, by modifying the way activities that are important to the person are performed.

Although the concept of social utility might be useful for policy making, it is not a good starting point for developing a truly person-centered rehabilitation plan and a person-centered measure of visual function outcomes in terms of net gain to the individual. To achieve this aim, we must start with a model of utility measures of vision rehabilitation functional outcomes in terms of reductions in rehabilitation demand and verify the internal validity of the model and analytic methods by applying them to simulations of the approach the model implies. The change in difficulty of performing different activities that are important and difficult to the patient can be estimated by the Rasch model from the change in rehabilitation demand, which is equal to the change in functional reserve,  $\Delta\varphi_n = \Delta\alpha_n$  in Eq. (5.3) when item measures are anchored to baseline values so  $\Delta\rho_j = 0$  for all items and all changes are forced into the person measure. However, the utility of the functional outcome for each specific activity is likely to be idiosyncratic to an individual patient. For example, the utility of a small realized change in functional reserve for cooking might have high value when serving the Daily Living Objective (social utility of preparing daily meals is 1.30), less value when serving the Recreation Objective (social value of recreational cooking and baking is  $-0.98$ ), and even less value when serving the Social Interaction Objective (social utility of cooking to entertain guess is  $-1.24$ ).

#### 5.4.5 Utility of Vision Rehabilitation Outcomes

The utility of reducing rehabilitation demand of item  $j$  for person  $n$  to zero is  $v_{nj}$ . Since we are referring to a single item,  $v_{nj}$  denotes a marginal utility, which is a person-specific function of both the item's difficulty to the person ( $D_{nj}$ ) and the importance to the person of being able to perform the activity without difficulty ( $I_{nj}$ ). While  $\varphi_{nj}$ , which determines the ordinal value of  $D_{nj}$ , is a continuous latent variable, no equivalent continuous variable has been modeled for determining the ordinal importance rating,  $I_{nj}$ . Therefore, to illustrate the derivation of a utility weighting model for the AI, we will employ discrete ordinal variables assigned by person  $n$  to the difficulty rating ( $D_{nj}$ ) and importance rating ( $I_{nj}$ ) of AI Goal  $j$ . The marginal utility of totally successful vision rehabilitation (*i.e.*, achieve a non-disabled state) of Goal  $j$  for person  $n$  is  $v_{nj} = U_n(I_{nj}, D_{nj})$ .

To paraphrase Gertrude Stein, the utility of a utility is a utility. By definition, we assume that the utility of vision rehabilitation on Goal  $j$  for the average person,  $U(I_j, D_j)$ , is a function of the respective part worth utilities associated with the importance and difficulty ratings, conditioned on the  $k$ th Objective,  $O_{k(j)}$ , so  $U(I_j, D_j) = f(U_i(I_j, I_j, O_{k(j)}), U_i(D_j, D_j, O_{k(j)}))$ .

$O_{k(j)}, U_d(D_j | O_{k(j)})$ ). Individuals will randomly deviate from this average relationship, so we model the mapping of  $I_{nj}$  and  $D_{nj}$  to the utility of vision rehabilitation for Goal  $j$  for person  $n$  as  $U_n(I_{nj}, D_{nj}) = U(I_{nj}, D_{nj}) + \epsilon_{nj}$ , where  $(I_{nj}, D_{nj})$  are the ratings of Goal  $j$  by person  $n$  and  $\epsilon_{nj}$  is the randomly distributed deviate. If either the importance or difficulty of Goal  $j$  is 0 for person  $n$ , then the utility of vision rehabilitation for Goal  $j$  will be zero, i.e.,  $U(I_{nj}, D_{nj}) = f(U_i(0 | O_{k(j)}), U_d(D_{nj} | O_{k(j)})) = f(U_i(I_{nj} | O_{k(j)}), U_d(0 | O_{k(j)})) = 0$ , and by definition  $\epsilon_{nj} = 0$  when  $I_{nj} = 0$  or  $D_{nj} = 0$ .

We are conditioning the utility function for each Goal on the Objective it serves because of the observed segregation by Objective of social values (estimated from Goal importance ratings) and level of consensus in the LV population, which is summarized in Fig. 5.9. Conditioning by Objective also is motivated by the possibility that the relation of Goal utility to difficulty may be determined by the reason for doing the activity (e.g., cooking to maintain independent living vs joy of cooking). If there are three different utility functions, one for each of the three Objectives, then we expect differences in utilities between Goals to be represented by the distance between them in a 3-dimensional utility space.

The marginal utility of Goal  $j$  is likely to be characterized by a nonlinear function (unique to Objective  $k$ ) of the part worth utilities associated with the observed ordinal importance and difficulty ratings of the Goal. For purposes of estimation, we can approximate this nonlinear function with a Taylor series:

$$U(I_{nj}, D_{nj}) \approx b_0 + b_1 I_{nj} + b_2 D_{nj} + b_3 I_{nj}^2 + b_4 D_{nj}^2 + b_5 I_{nj} D_{nj} + \dots$$

in which  $b_0$  is a constant and the other coefficients correspond to factorial-weighted first, second, and higher order partial derivatives (ellipsis denotes the higher order terms in the infinite series). To illustrate how this model can be applied to estimate utilities of vision rehabilitation outcomes, we simulated the method. The simplifying assumptions made within the model to create the simulation are: (1) the utility function is the product of the part worth utilities for importance and difficulty, i.e.,  $U(I_{nj}, D_{nj}) = U_i(I_{nj}) \times U_d(D_{nj})$ , which yields an overall utility of zero if either part worth utility is zero, an overall utility of 1 if both part worth utilities are 1, and an overall utility less than or equal to the lowest part worth utility; (2) the utility function is the same for all objectives, i.e.,  $U_{k(j)}(I_{nj}, D_{nj}) = U(I_{nj}, D_{nj})$  for all  $k$ ; and (3) the relation of utility to importance/difficulty rating combinations is fixed for the population, i.e.,  $\epsilon_{nj} = 0$ . These assumptions enable us to estimate a variable that is equivalent to utility by constructing a dissimilarity matrix comparing all possible deterministic importance/difficulty pairs  $(I, D)$ .

In the AI, ordinal values for importance ratings range from 0 to 3 (where 0 is “not important” and 3 is “very important”) and ordinal values for difficulty ratings range from 0 to 4 (where 0 is “not difficult” and 4 is “impossible”). As schematized in Fig. 5.11, we already have defined the utility of LVR for Goals with ratings  $(I, 0)$  or  $(0, D)$  to be zero for all values of  $I$  and  $D$ , leaving 12  $(I, D)$  combinations for which the partial utility associated with Goal importance and decrease of difficulty are greater than zero. For the simulation we simply filled in arbitrarily-chosen ascending values

Utilities		Reducing Difficulty to 0					Difficulty
		0	0.1	0.3	0.7	1	
Importance	Rating	Not	Slight	Moderate	Very	Impossible	
0	Not	0	0	0	0	0	
0.2	Slight	0	0.02	0.06	0.14	0.2	
0.5	Moderate	0	0.05	0.15	0.35	0.5	
1	Very	0	0.1	0.3	0.7	1	
	Importance						Products

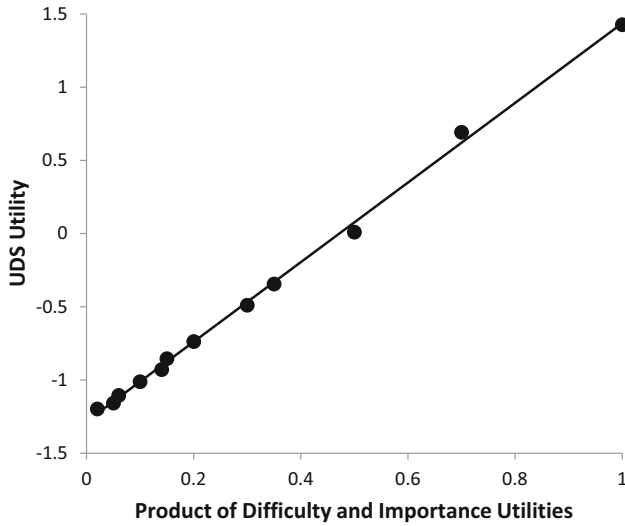
**Fig. 5.11** Relation of utilities to AI Goal importance and difficulty ratings as used in the simulation to test estimation of the utility to the patient of vision rehabilitation outcomes. The 4 Goal importance ratings range from “Not Important” to “Very Important” and are assigned the part worth utilities in the left most column. The 5 Goal difficulty ratings range from “Not difficult” to “Impossible” and are assigned the part worth utilities (of reducing difficulty to 0) in the top row. The marginal utility for each Goal is the product of the part-worth utilities corresponding to Importance and Difficulty ratings assigned by the patient to the Goal (entries in the purple cells)

for the marginal utilities (light green areas) and products of the marginal values for each  $(I,D)$  combination (purple area). Thus, a triangular matrix for the AI has 66 unique paired comparisons of different non-zero  $(I,D)$  combinations.

We can think of the marginal vision rehabilitation utility of Goal  $x$  with importance and difficulty ratings  $(I_x, D_x)$  as mapping to a point  $U_x$  on a number line that represents the utility of time, effort, and resources that would have to be expended on vision rehabilitation. Similarly the partial vision rehabilitation utilities of Goals  $y$  and  $z$  with respective importance and difficulty ratings  $(I_y, D_y)$  and  $(I_z, D_z)$ , would map to different points,  $U_y$  and  $U_z$ , on the same number line for the expenditure of time, effort, and resources. Individuals then can be asked to judge the relative distances between each pair of points.

For example, the person would be asked, “In terms of allocating your time, effort and resources to rehabilitation, which of the three Goals would you give the highest priority?” That question would be followed by, “Which of the remaining two Goals would you give the lowest priority?” The final question would be, “Is the priority you give to the left-over Goal closer to the highest priority Goal or the lowest priority Goal?” Let’s imagine the person gave the highest priority to Goal  $x$ , the lowest priority to Goal  $y$ , and said that the priority of Goal  $z$  is closer to the priority of Goal  $y$ . Thus, on the utility number line, the distance between  $U_x$  and  $U_y$  is the largest of the three comparisons and the pairing of  $(I_x, D_x)$  and  $(I_y, D_y)$  in the matrix would be assigned a dissimilarity rank of 3; the distance between  $U_y$  and  $U_z$  is the smallest of the three comparisons and the pairing of  $(I_y, D_y)$  and  $(I_z, D_z)$  in the matrix would be assigned a dissimilarity rank of 1; and the remaining pairing of  $(I_x, D_x)$  and  $(I_z, D_z)$  corresponds to an intermediate distance and would be assigned a dissimilarity rank of 2 in the matrix.

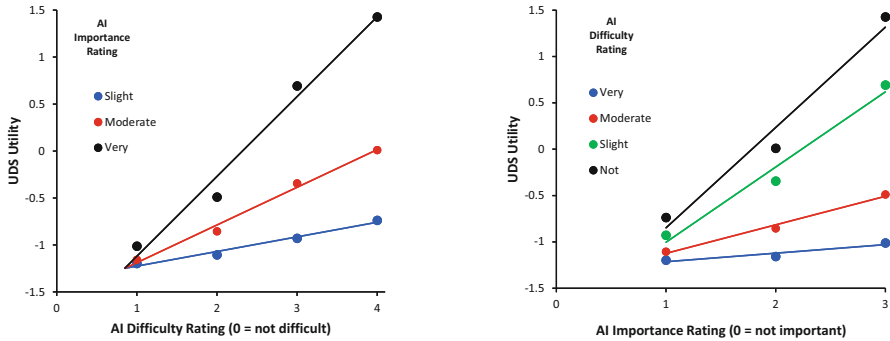
Repeating these judgments and assignment of dissimilarity rank scores for all feasible triadic comparisons across all persons in the sample, and averaging relative distance rank scores in each cell of the triangular dissimilarity matrix, we can employ non-metric unidimensional scaling [40] (UDS) to map each  $(I,D)$  rating pair to a variable that is monotonic with vision rehabilitation utility.



**Fig. 5.12** Utilities estimated from unidimensional scaling (UDS) of sums of ordinal-ranked differences in utility of 3 unique combinations of Goal difficulty and importance ratings (*i.e.*, triadic comparisons) as a function of the product of corresponding part worth utilities for the ratings. In this case, because the monotonic function was exponential, the log of the estimated distance metric is plotted as a function of the 12 marginal utilities defined in Fig. 5.11 and the linear relationship defines log distance as the UDS-estimated utility

Using the table in Fig. 5.11, we simulated the triadic comparison judgments by assigning rank scores to  $(I,D)$  pairings in all 1320 possible triads based on the products of the marginal utilities assigned to importance ( $I$ ) and difficulty ( $D$ ) resulting in 20 dissimilarity ranks contributing to the average in each of the 66 cells of the triangular matrix. [*N.B.* This scaling method effectively made  $\epsilon_{nj} = 0$  for all simulated judgments]. Figure 5.12 illustrates a scatter plot comparing the estimated utility (in arbitrary units) for each of the 12  $(I,D)$  pairs from unidimensional scaling (UDS) versus the products of the partial utilities assigned to importance and difficulty ratings in each pair (values in purple area of Fig. 5.11), which were used to make the ranked distance judgments in each triadic comparison.

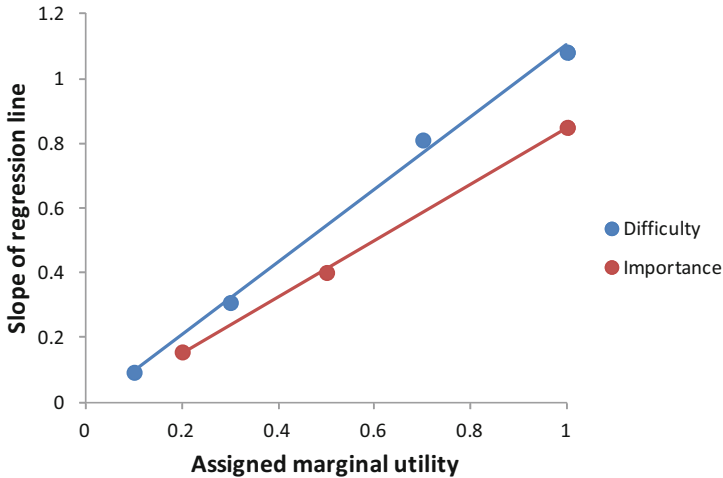
When applied to the real world, we will not know the true utilities associated with the importance and difficulty rating categories as we do for this simulation. The UDS (and more generally multidimensional scaling [MDS], which would be used if the Objectives represent different utility dimensions) estimates distances using the ordinal data in the dissimilarity matrix (non-metric scaling). The UDS (and MDS) approach enables us to perform statistical tests of the goodness of fit of the estimated distances in the dissimilarity matrix to the average ranks of dissimilarity scores from the triadic comparisons (*e.g.*, Shepard plots and estimates of “stress”).



**Fig. 5.13** *Left panel:* Estimated utilities as a function of the difficulty ratings of AI Goals for each rating of importance. The slope of the line fit to the results is a function of the importance ratings. *Right panel:* Estimated utilities as a function of the importance ratings of AI Goals for each rating of difficulty. The slope of the line fit to the results is a function of the difficulty ratings

The left panel of Fig. 5.13 illustrates a scatterplot of marginal utility estimates of  $U(I,D)$  from UDS coordinates vs ordinal ranks of different non-zero difficulty rating categories for the 3 different levels of non-zero importance ratings (color-coded). These linear relationships in the left panel of Fig. 5.13 imply  $U(I,D) = m_I D$ , for which the slope  $m_I$  is dependent on  $I$ . The right panel of Fig. 5.13 similarly illustrates a scatterplot of UDS estimates of  $U(I,D)$  vs ordinal ranks for different non-zero importance ratings for the 4 different levels of non-zero difficulty ratings. The linear relationships in the right panel of Fig. 5.13 imply  $U(I,D) = m_D I$ , for which the slope  $m_D$  is dependent on  $D$ . The abscissa value where the lines converge corresponds to the location of zero difficulty or zero importance, respectively (ordinal rank scores arbitrarily are equally spaced on the abscissa). The horizontal deviations of the points from the line must be construed as errors resulting from the assumption that the ordinal ratings represent equal intervals. From these linear relationships, we conclude that  $\frac{\partial U(I,D)}{\partial I} = m_D$  and  $\frac{\partial U(I,D)}{\partial D} = m_I$ . In Fig. 5.11 we defined the overall utilities estimated from UDS to be the product of the respective importance and difficulty partial utilities,  $U(I,D) = U_i(I) \times U_d(D)$ . This definition means that the Taylor series approximation is first-order,  $\frac{\partial U(I,D)}{\partial I} = \frac{\partial U_i(I)}{\partial I} U_d(D)$  and  $\frac{\partial U(I,D)}{\partial D} = \frac{\partial U_d(D)}{\partial D} U_i(I)$ . Combined with the conclusions we drew above from the linear relationships in Fig. 5.13, we can see that the estimated slopes must be linear with the respective partial utilities,  $m_D = \frac{\partial U_i(I)}{\partial I} U_d(D)$  (blue points in Fig. 5.14) and  $m_I = \frac{\partial U_d(D)}{\partial D} U_i(I)$  (red points in Fig. 5.14). Figure 5.14 confirms the tautology, which validates the analysis – we can estimate overall utilities from UDS (or more generally, MDS) on a matrix of average dissimilarity rank scores obtained from triadic comparisons using the pre-assigned overall utilities to rank distances.





**Fig. 5.14** Slopes of the regression lines in Fig. 5.13 as a function of the marginal utility corresponding to the assigned part-worth utilities for each combination of non-zero ordinal importance (red) and difficulty (blue) ratings of AI Goals

#### 5.4.6 Extension of the Utility Model to Estimation of Net Gain from Vision Rehabilitation

Consider the functional outcome for person *A* who has 3 Goals in her individualized rehabilitation plan and the functional outcome for person *B* who has 10 Goals in his individualized rehabilitation plan, for which 3 of them are the same as the Goals in person *A*'s plan with the same combinations of importance and difficulty ratings. If vision rehabilitation results in the difficulty of all 3 of *A*'s Goals being reduced to zero and the same 3 of *B*'s Goals also reduced to zero, but less than full reduction in difficulty for some of *B*'s other 7 Goals, how would the net gain from vision rehabilitation for patient *A* compare to the net gain for patient *B*? Using the current method of measuring functional outcomes from vision rehabilitation, which is equivalent to the average change in functional reserve for Goals in the individualized rehabilitation plan, the effect size for *B* could be larger than the effect size for *A*, or vice versa, depending on the magnitudes of changes in functional reserve for *B*'s other 7 Goals relative to the change in the 3 that are the same as *A*'s.

However, from another perspective, *A* would have no need for additional rehabilitation after completion of vision rehabilitation, whereas *B* would. So, for every scenario short of all 10 of *B*'s Goals being reduced to zero difficulty, in terms of remaining *rehabilitation demand* [19], *A*'s functional outcomes would have greater utility than *B*'s. If the difficulty of all 10 of *B*'s Goals were reduced to zero, then neither *A* nor *B* would have any remaining rehabilitation demand and the utility of additional vision rehabilitation would be zero for both. The conundrums raised by

this example suggest that to be truly person-centered, we should define the utility of the vision rehabilitation function outcome as the difference between the disutilities (rehabilitation demands) of post-rehabilitation and pre-rehabilitation. The question we must answer is, how do we combine the marginal utilities of reducing rehabilitation demand for different Goals to estimate the multi-attribute (all Goals combined) utility of reducing an individual's overall disability?

In a manifestation of dynamics related to the law of diminishing marginal utility, *B* might temporarily be euphoric with the reduction of difficulty of the 3 Goals having high rehabilitation demand, only to have the emotional high of successful vision rehabilitation dissipate and the disutility of the remaining rehabilitation demand emerge. The multi-attribute disutility function we seek, which combines disutilities of rehabilitation demand for all Goals in the patient's individualized rehabilitation plan, could be the linear sum of disutilities for individual Goals, or a nonlinear combination that ranges from diminishing rate of change in marginal utility (attenuation effect) to increasing rate of change in marginal utility (amplification effect) with increasing numbers of Goals having non-zero rehabilitation demand. This range of options can be expressed with a Minkowski distance,

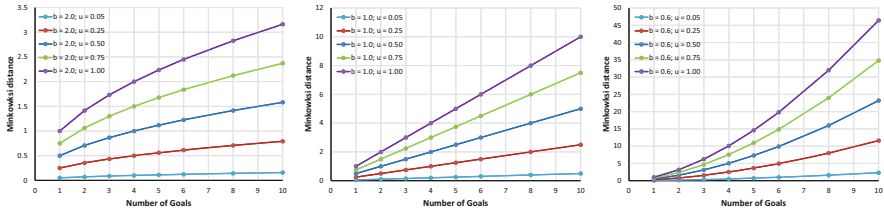
$$v_{nJ_n} = \left( \sum_j^{J_n} u_{nj}^b \right)^{1/b} \quad \text{where } v_{nJ_n} \text{ is the multi-attribute utility of vision rehabilitation}$$

(or multi-attribute disutility of rehabilitation demand) for an individualized rehabilitation plan for patient *n* with  $J_n$  Goals after filtering.

Multi-attribute utility is a linear sum for  $b = 1$ ; attenuation corresponds to  $b > 1$ ; and amplification corresponds to  $b < 1$ . The Minkowski distance (ordinate) across Goals with the same marginal utility ranging from 0.05 to 1.0 (colored functions) are shown in Fig. 5.15 as a function of the number of Goals ranging from 1 to 10 (abscissa) in the patient's plan. The left panel of Fig. 5.15 depicts attenuation in the growth of multi-attribute utility with  $b = 2$ ; the middle panel of Fig. 5.15 shows the linear sum in the growth with  $b = 1$ ; and the right panel of Fig. 5.15 shows amplification in the growth with  $b = 0.6$ .

## 5.5 Visual Ability Outcomes of Vision Rehabilitation

In the preceding section we reviewed a strategy for measuring visual ability outcomes in terms of net gain to the patient by way of a multi-attribute utility model. This model entails summed changes in the utilities of reducing the difficulty (and/or reducing the importance) of attaining individual AI Goals. However, at the current stage of development, operations in the multi-attribute utility model employ ordinal rank scores assigned to the person's importance and difficulty ratings of being able to perform activities. Ideally, we would define utility to be a function of continuous latent variables for importance and difficulty,  $v_{nj} = u_n(t_{nj}, \delta_{nj})$ . Difficulty is the inverse of functional reserve,  $\delta_{nj} = -\varphi_{nj} = \rho_j - \alpha_n$ , a continuous latent variable



**Fig. 5.15** *Left panel:* Multi-attribute utility as a function of the number of Goals, estimated from marginal utilities using a Minkowski distance formula with the exponent variable  $b = 2$ , which results in an attenuation in the rate of growth with an increasing number of Goals. The marginal utility is the same for every Goal with a different value for each curve ranging from  $u = 0.05$  to 1.0 (see legend). *Center panel:* Same as the left panel but for a Minkowski distance formula in which the exponent variable  $b = 1$ , which results in a constant rate of growth with an increasing number of Goals. *Right panel:* Same as the left and center panels but for a Minkowski distance formula in which the exponent variable  $b = 0.6$ , which results in an acceleration of the rate of growth with an increasing number of Goals

we already know how to estimate. However, estimating  $t_{nj}$ , which is the strength of the personal preference assigned to item  $j$  by person  $n$ , is a thornier problem because of the lack of consensus between people in the ordering of items by personal preferences (*cf.*, “social utility”).

We will return to the issue of how the multi-attribute utility model can be used to estimate the net gain from vision rehabilitation. But first we must review two methods of measuring patient-centered visual ability outcomes of vision rehabilitation with the AI in terms of (1) the distribution of changes in a continuous visual ability outcome variable and (2) the likelihood of attaining a change in a visual ability clinical endpoint. In both cases we employ the average change in functional reserve as the measure of the patient’s self-reported visual ability before and after rehabilitation.

### 5.5.1 Continuous Visual Ability Outcome Measure: Average Change in Functional Reserve

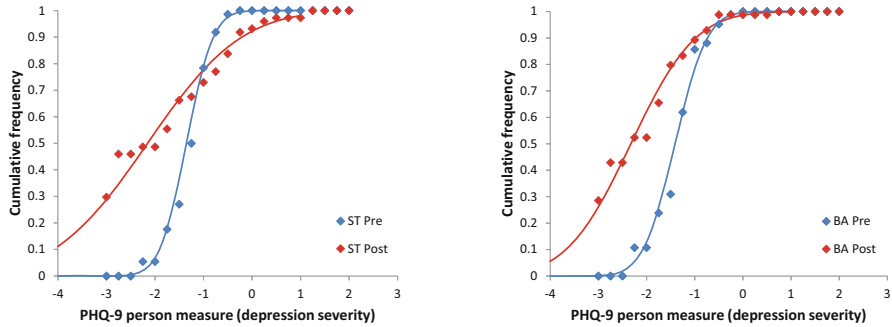
The Low Vision Depression Prevention Trial in Age-Related Macular Degeneration (VITAL) was a randomized, attention-controlled, clinical trial to determine the effectiveness of behavior activation therapy as a supplement to in-home vision rehabilitation with an occupational therapist in preventing the development of major or minor depression in low vision patients with subsyndromal depressive symptoms [41]. Low vision patients were randomized to six weekly sessions of vision rehabilitation provided in the home by an occupational therapist who also provided behavior activation therapy (BA – the treatment being tested) or to six weekly sessions of supportive therapy (ST – a placebo attention control) provided in

the home by a clinical social worker. The ST control group received no additional low vision services. The primary outcome measure, which was administered prior to any low vision services (PRE) and again at 2 months after the completion of services and assigned therapy (POST), was the PHQ-9, which was used to determine if the patient exhibited depressive symptoms consistent with DSM-IV criteria for major or minor depression. The AI also was administered PRE and POST low vision services and psychotherapy. Prior to randomization, all study participants received standard optometric low vision consultations; required vision assistive equipment was dispensed to all participants at study expense; and all participants were trained at the low vision clinic in how to use the equipment.

Rasch analysis of the PHQ-9 responses was used to estimate person measures of depression severity PRE and POST low vision services [35]. Rasch analysis also was used to estimate overall visual ability from AI Goal difficulty ratings with anchored item measures and thresholds for both the BA treatment group and ST control group (with item filtering) prior to receiving low vision services and again 2 months after completion of low vision services. The BA treatment group exhibited a statistically significant improvement in visual ability (Cohen's  $d = 0.71$ ;  $p < 0.001$ ). The ST control group also exhibited a statistically significant improvement in visual ability (Cohen's  $d = 0.55$ ;  $p = 0.003$ ). However, the distributions of change in visual ability for the BA treatment group was not significantly different from the change for the ST control group (Cohen's  $d = 0.10$ ;  $p = 0.39$ ) [35]. The significant medium size effect (POST-PRE) seen for both groups most likely can be attributed to the low vision devices and services that were provided in the clinic after the PRE measures of visual ability, but before the in-home vision rehabilitation supplemented by BA psychotherapy for the treatment group or the sham psychotherapy with no additional vision rehabilitation for the control group.

Although there was no difference between groups for the effects of study intervention on visual ability, the primary outcome for the VITAL study was a psychiatric clinical diagnosis in low vision patients of major or minor depression as defined by a criterion PHQ-9 score. As a study eligibility criterion, none of the participants in the study had a PHQ-9 score at PRE that exceeded the threshold for a clinical depression diagnosis. Thus, the PHQ-9 threshold for depression defined a clinical endpoint, which was exceeded at POST by significantly more patients in the ST control group than in the BA treatment group. The take-home conclusion of the VITAL study was that in-home vision rehabilitation supplemented with BA psychotherapy prevented the development of clinical depression in at-risk low vision patients.

Applying Rasch analysis to PHQ-9 item responses results in valid estimates of interval-scaled continuous person measures that can be interpreted as *depression severity* [35]. The left panel of Fig. 5.16 shows PRE (blue curve) and POST (red curve) cumulative frequency functions of PHQ-9 person measures for patients assigned to the ST control group. Negative person measures correspond to low depression severity and positive person measures correspond to high depression severity.



**Fig. 5.16** *Left panel:* Relative cumulative frequency of PHQ-9 person measure values of depression severity at baseline (blue curve) and at post-intervention follow-up (red curve) for the supportive therapy control group in the VITAL study. The shift of the median value to the left on the person measure axis at follow-up indicates a decrease in depression severity post-intervention. However, the curves cross, which means that about 80% of patients in the ST group had a decrease in depression severity after intervention, whereas about 20% of patients had an increase – the main effect reported for the study. *Right panel:* Same results for the BA group as shown for the ST group at baseline, but at follow-up the cumulative functions do not cross, which indicates that nearly all patients in the BA group had a decrease or no change in depression severity at follow-up

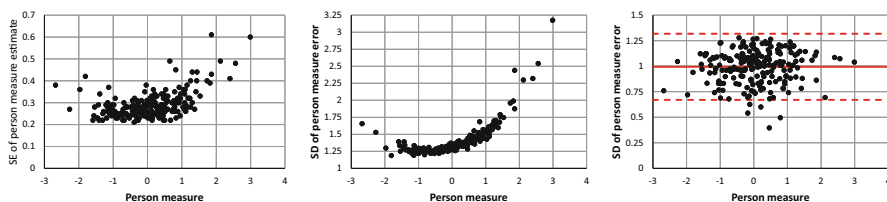
Similarly, the right panel of Fig. 5.16 shows Pre and Post cumulative frequency functions of PHQ-9 person measures for patients assigned to the BA treatment group. The decrease in depression severity from PRE to POST (leftward shift of the red curve relative to the blue curve) demonstrates a large significant effect of intervention for both the BA treatment group (Cohen's  $d = 2.12$ ;  $p < 0.001$ ) and for the ST control group (Cohen's  $d = 2.02$ ;  $p < 0.001$ ). There is no difference between depression severity distributions for the two groups at baseline (t-test;  $p = 0.28$  for PRE).

There also is no difference between the means of the depression severity distributions for the two groups at follow-up (t-test;  $p = 0.25$  for POST), but the slope of the depression severity cumulative distribution is shallower for the ST group than for the BA group. This change of slope that causes the ST PRE and POST curves to cross in the right panel of Fig. 5.16 underlies the main effect of BA psychotherapy supplementing in-home vision rehabilitation preventing the development of major or minor depression, which was reported as the VITAL study primary outcome. But this study also shows that the low vision devices and services provided in the clinic, weekly in-home sessions with a professional therapist or counselor, and whatever else the two groups have in common result in a large reduction of severity in depression symptoms. Since the two groups in the study had equivalent improvements in visual ability, we explored changes in that variable as a possible explanation.

### 5.5.2 Minimum Clinically Important Difference in Visual Ability as a Clinical Endpoint

With reference to Eq. (5.5), anchoring AI item measures and thresholds to baseline values forces all intervention-specific DIF into changes in the person measure,  $\Delta\alpha_n$ , with all randomly distributed measurement error incorporated in the person measure estimates. The lower bound on the standard error of the person measure estimate for person  $n$  ( $SE_n$ ) is proportional to the standard deviation on  $\zeta$  in expression (5.1) and inversely proportional to the square root of the number of items rated by person  $n$ . The left panel of Fig. 5.17 displays the distribution of standard errors of the person measure estimate at baseline versus the person measure for VITAL study participants. To estimate the standard deviation of the person measure error distribution, we multiplied each standard error of the person measure estimate by the square root of the number of items rated by the person. The standard deviations of the person measure error distributions, so estimated for each patient, are plotted as a function of the patient's estimated visual ability (person measure) in the middle panel of Fig. 5.17.

However, the conventional logistic Rasch model normalizes the estimated measures to the standard deviation of  $\zeta$ , so the expected value of the standard deviation of the error distribution should be 1 for each person. The U-shaped functional relationship between the estimate of  $SD_n$  and  $\alpha_n$  in the middle panel of Fig. 5.17 (which shows that all  $SD_n$  values are greater than 1) can be attributed to the increased uncertainty at the extremes of the person measure distribution due to the progressive change in frequency of responding with the half-open categories (“not difficult” or “impossible” that extend from  $\tau_4$  to  $\infty$  and from  $\tau_1$  to  $-\infty$ , respectively). The right



**Fig. 5.17** *Left panel:* Scatter plot of the standard error (SE) of the person measure estimates from difficulty ratings of AI Goals versus the estimated person measure at baseline in the VITAL study. *Center panel:* The same results shown in the left panel but with each SE value multiplied by the square root of the number of items rated by that person resulting in an estimate of the standard deviation of the distribution of deviates ( $\zeta$ ). Notice the U-shape in the plot of the data which can be attributed to greater numbers of difficulty ratings corresponding to half-open intervals contributing to the estimate as person measures become more extreme. Also note that all values are greater than the expected value of 1. *Right panel:* Person measure standard errors were re-estimated after omitting items for which the response represented a half-open interval. The revised standard error estimates were multiplied by the square root of the number of items retained in the estimate (points). These revised estimates of the standard deviation of the deviates,  $\zeta$ , have an average value of 1 (red line) – the expected value. The red dashed lines define the 95% confidence interval

panel of Fig. 5.17 displays the distribution of  $SD_n$  when the value is estimated by multiplying  $SE_n$  by the square root of the number of items the person rated with non-extreme difficulty categories (i.e., “somewhat difficult”, “moderately difficult”, and/or “very difficult”). The average estimated standard deviation of  $\zeta$  is 1 (solid red line), which agrees with the measurement scale normalization built into the model. The dashed lines in the right panel of Fig. 5.17 bound plus and minus two standard deviations of the between person distribution of  $SD_n$  estimates.

The variance of each person’s error distribution,  $\sigma_{\zeta_{n,j,x}}^2$ , is the sum of within and between person squared deviations from the expected value of zero,  $\mathbb{E}\left\{\zeta_{n,j,x}^2\right\}$ , whereas the variance of the between person error distribution is  $\sigma_{j,x}^2 = \sigma_j^2 + \sigma_x^2 + 2r_{j,x}\sigma_j\sigma_x$ , as introduced earlier in this chapter, in which  $\sigma_j^2$  refers to between person variance in the item measure for the  $j$ th item and  $\sigma_x^2$  refers to between person variance in the threshold for the  $x$ th response category. All within person variance can be assigned to visual ability,  $\sigma_{\alpha_n}^2$ , so the total variance of each person’s measurement error distribution is  $\sigma_{\zeta_{n,j,x}}^2 = \sigma_{\alpha_n}^2 + \sum_{j=1}^{J_n} \sigma_{j,x}^2$ . But, in the case of comparing PRE to POST intervention measures for each person, the item measures,  $\rho_j$ , and response category thresholds,  $\tau_x$ , are fixed to calibrated values that are the same for both measures (the deviates due to between person differences are fixed and manifest as the same person-dependent bias for PRE and POST measures, so  $\sum_{j=1}^{J_n} \sigma_{j,x}^2 = 0$ ) and the error variance on the estimate of  $\Delta\alpha_n$  is determined entirely by within person variance,  $2 \times \sigma_{\alpha_n}^2$ . In most cases, the number of items rated ( $J_n$ ) is the same at PRE and POST, however, that is not a requirement. Also the proportion of items rated with extreme response categories is likely to be different between PRE and POST, which will differentially affect the standard error of the estimate, even when there is no change between PRE and POST in within person variance. Thus, the standard error of the estimate of  $\Delta\alpha_n$  is  $SE_{\Delta\alpha_n} = \sqrt{SE_n^2(Pre) + SE_n^2(Post)}$ .

The smallest change in the person measure of an individual that we can say with confidence represents a real change in response to an intervention is called the *minimum clinically important difference* (MCID). The MCID is a clinical endpoint. We transform the clinical outcome for person  $n$  to a  $t$ -statistic,  $t(\Delta\alpha_n, df_n) = \frac{\Delta\alpha_n}{SE_{\Delta\alpha_n}}$  with  $df_n = J_n(Pre) + J_n(Post) - 2$ , and the MCID for person  $n$  as the  $t$  value that corresponds to a criterion probability of making a type I error (e.g.,  $p = 0.05$ ). If  $t(\Delta\alpha_n, df_n)$  exceeds the criterion corresponding to the chosen  $p$  value, then  $MCID = 1$  for person  $n$ , otherwise  $MCID = 0$ .

The odds of  $MCID = 1$  is 0.45 for the BA treatment group and 0.395 for the ST control group, resulting in an odds ratio of 1.14, which is significantly different from 1.00 ( $p < 0.05$ ). In other words, a significantly greater number of patients in the BA treatment group had a change in visual ability that exceeded the MCID clinical endpoint than occurred in the ST control group.

After combining the BA and ST groups, we compared the change in depression severity estimated from Rasch analysis of PHQ-9 responses of patients with MCID = 1 to the change in depression severity of patients with MCID = 0. There was no significant effect in the VITAL study of MCID in visual ability on changes in depression severity (t-test,  $p = 0.22$ ).

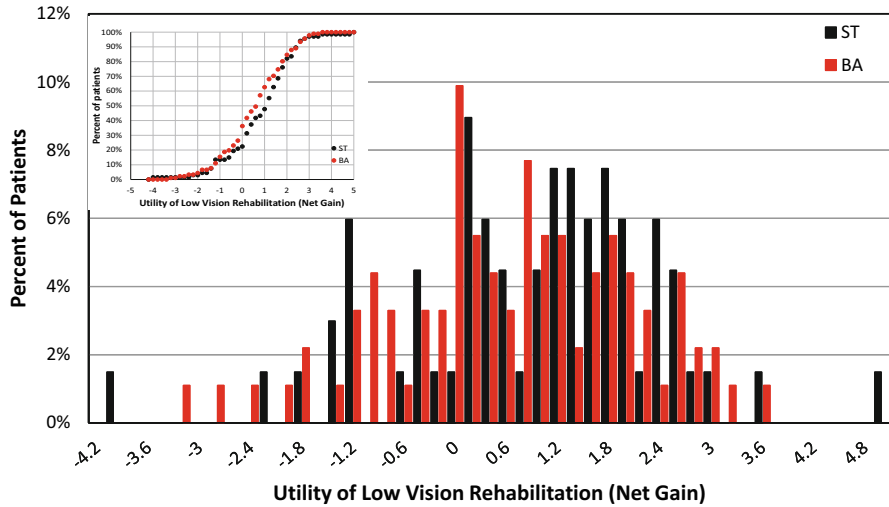
### 5.5.3 *Reducing Rehabilitation Demand: Net Gain from Vision Rehabilitation*

The VITAL study and other vision rehabilitation outcome studies that employed the AI [42, 43] agree that on average vision rehabilitation results in a moderate to large size effect of intervention (Cohen's effect size in the range of 0.7 and 1.1). However, as described above for the VITAL study, a recent Cochrane review of randomized controlled trials that compared the effectiveness of different levels or components of vision rehabilitation concluded that additional services beyond the initial low vision consultation produce no or very small incremental effects [44]. In other words, based on current practices there appears to be a diminishing return on investment with increasing amounts of rehabilitation. Thus, to be truly patient-centered we not only want to measure improvements in functional ability, but also measure the utility of those improvements to the patient. To demonstrate how this can be done, even though we still have an incomplete model of a continuous latent variable for the utility of reducing rehabilitation demand, we apply the fabricated parameters that we used for the simulation (listed in Fig. 5.11) to VITAL study outcome data obtained with the AI.

Both importance ( $I_{nj}$ ) and difficulty ( $D_{nj}$ ) ratings were obtained on AI Goals at PRE and POST intervention in the VITAL study. Goal items were filtered out (no difficulty rating elicited) if the Goal was rated "not important" (*i.e.*, if  $I_{nj} = 0$ ). Using the simulated "as if" model specified in Fig. 5.11, importance and difficulty ratings of each Goal for each patient at PRE and at POST were replaced with their corresponding part worth utilities (numbers created for the simulation in the green margins of Fig. 5.11). Next, as done for the simulation, the marginal utility of each Goal for each patient at PRE and at POST was computed by taking the product of the assigned part worth utilities. Finally, multi-attribute utilities of totally successful rehabilitation (*i.e.*, utility of reducing rehabilitation demand to zero) were estimated for both PRE and POST intervention Goals for each patient using the Minkowski distance with  $b = 2$  (re. left panel of Fig. 5.2), an arbitrarily chosen value that results in attenuation of utility growth with increasing numbers of Goals (the number of Goals with non-zero utilities varied across patients from 2 to 40 at PRE [mean = 16 and SD = 7] and from 2 to 44 at POST [mean = 15 and SD = 9]).

In this hybrid simulation, there is no significant difference in rehabilitation demand (multi-attribute utility) estimated between the BA and ST groups at PRE ( $p = 0.344$ ) or at POST ( $p = 0.405$ ). However, this "as if" model does result in a significant reduction in rehabilitation demand from PRE to POST for both groups





**Fig. 5.18** Histogram of the distribution of outcomes of intervention in the VITAL study when the outcome measure is estimated as the utility of rehabilitation demand reduction, estimated with the hybrid (simulation and data) model, from intervention for the ST (black bars) and BA (red bars) groups. Both groups exhibited a significant increase in the utility of visual ability outcomes (decrease in rehabilitation demand). As shown by the cumulative distribution of the outcome measures in the inset, the ST group (control intervention) had slightly better outcomes than did the BA group (experimental intervention), but that difference is not statistically significant

( $p = 1.27 \times 10^{-5}$  for ST and  $p = 0.00055$  for BA). Figure 5.18 displays histograms of net gains in the utility of vision rehabilitation outcomes (*i.e.*, reductions in rehabilitation demand) for the BA (red) and ST (black) groups. The inset in Fig. 5.18 displays the two distributions as relative cumulative frequency functions. These results would be interpreted as an average reduction in rehabilitation demand of 0.53 (SD = 1.42) for the BA group and 0.82 (SD = 1.5) for the ST group. This difference between groups, however, is not statistically significant ( $p = 0.11$ ).

### 5.5.4 Next Steps in the Development of Preference-Based Patient-Centered Outcome Measures for Vision Rehabilitation

The above estimates from AI Goal importance and difficulty ratings of multi-attribute utilities representing rehabilitation demand are premature. They were presented here as a demonstration of the next aim in the development of patient-centered outcome measures that incorporate patient preferences. To achieve this aim we ultimately must develop a valid method of estimating the importance of each AI Goal on a continuous interval scale for each respondent ( $t_{nj}$ ) that incorporates the stochastic error distributions ( $\epsilon_{nj}$ ). We then must collect sufficient triadic comparison

data on a large sample representing the low vision patient population to map continuous importance ( $t_{nj}$ ) and difficulty ( $\delta_{nj}$ ) latent variables onto marginal utilities and to define the utility function that maps the part worth utilities onto the total utility for the Goal,  $v_{nj} = u_n(t_{nj}, \delta_{nj})$ .

Rasch analysis, or some variant of traditional Rasch analysis in the case of importance ratings, must be used to measure the continuous latent variables ( $t_{nj}$  and  $\delta_{nj}$ ) estimated from ordinal ratings of individual patients ( $I_{nj}$  and  $D_{nj}$ ). It then will be necessary to build a large database for a sample of the target low vision patient population to estimate, validate, and anchor model algorithms and parameters for the part worth utility, marginal utility for each Goal, and multi-attribute (rehabilitation demand) utility functions. This theory-driven approach also can give us the tools to identify, estimate, and ultimately understand stochastic and systematic deviations of individual patients from the trends for the targeted population.

A theory-driven approach to the development of patient-centered outcome measurements also promises to provide the tools needed for principled cost-benefit analyses of specific interventions. The ultimate concern to the clinician when assessing risks, costs, and benefits of intervention is the clinical outcome, including adverse events, at a physiological (*e.g.*, ocular pathology) and/or behavioral (*e.g.*, visual impairment) level. The ultimate concern to the patient when assessing risks and benefits of the same intervention is net gains and losses in her or his quality of life, a multi-dimensional construct that ultimately is quantified as a personal multi-attribute utility of the intervention. To facilitate communication between the patient and clinician and thereby facilitate meaningful and ethical shared decision-making, it is necessary to model the relationships between manifest variables observed by the clinician (*e.g.*, visual impairment measures) and latent variables observed by the patient (*e.g.*, visual ability), which we attempt to do with the conceptual (and preliminary computational) model schematized in Fig. 5.7. The clinician has a myriad of sophisticated tools to make objective measurements of publicly observable variables. Rasch models provide us with the tools needed to make objective measurements of latent variables that are observed privately by the patient. What we need now is a rigorous *psychophysics* to build a crosswalk between the two worlds of measurement by way of testable theories.

## References

1. R.W. Massof, The measurement of vision disability. *Optom. Vis. Sci.* **79**, 516–552 (2002)
2. T.G. Bond, C.M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 3rd edn. (Routledge, New York, 2015)
3. R.W. Massof, G.S. Rubin, Visual function assessment questionnaires. *Surv. Ophthalmol.* **45**, 531–548 (2001)
4. R.L. Kirby, The nature of disability and handicap, in *Medical Rehabilitation*, ed. by J. V. Basmajian, R. L. Kirby, (Williams & Wilkins, Baltimore, 1984), pp. 14–18
5. C. Bradley, R.W. Massof, Method of successive dichotomizations: An improved method for estimating measures of latent variables from rating scale data. *PLoS One* **13**, e0206106 (2018)

6. R.W. Massof, Is the partial credit model a Rasch model? *J. Appl. Meas.* **13**, 1–18 (2012)
7. P.G.W. Jansen, E.E. Roskam, Latent trait models and dichotomization of graded responses. *Psychometrika* **51**, 69–91 (1986)
8. R.W. Massof, Understanding Rasch and item response theory models: Applications to the estimation and validation of interval latent trait measures from responses to rating scale questionnaires. *Ophthalmic Epidemiol.* **18**, 1–19 (2011)
9. G.N. Masters, A Rasch model for partial credit scoring. *Psychometrika* **47**, 149–174 (1982)
10. F. Samejima, Estimation of latent ability using a response pattern of graded scores. *Psychometrika Suppl* **17** (1969)
11. D. Andrich, Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika* (2010)
12. G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests* (Danish Institute for Educational Research, Copenhagen, 1960)
13. W.S. Torgerson, *Theory and Methods of Scaling* (Wiley, New York, 1958)
14. E. Muraki, Fitting a polytomous item response model to Likert-type data. *Appl. Psychol. Measur.* **14**, 59–71 (1990)
15. E.E. Roskam, P.G.W. Jansen, Conditions for Rasch-dichotomizability of the unidimensional polytomous Rasch model. *Psychometrika* **54**, 317–333 (1989)
16. T. Chan, D.S. Friedman, C. Bradley, R. Massof, Updated estimates of incidence and prevalence of visual impairment, low vision, and blindness in the U.S. *JAMA Ophthalmol.* **136**, 12–19 (2018)
17. J.E. Goldstein, R.W. Massof, J.T. Deremeik, S. Braudway, M.L. Jackson, K.B. Kehler, S.A. Primo, J.S. Sunness, LOVRNET Study Group, Baseline traits of low vision patients served by private outpatient clinical centers in the United States. *Arch. Ophthalmol.* **130**, 1028–1037 (2012)
18. R.W. Massof, G. Dagnelie, J.T. Deremeik, J.L. DeRose, S.S. Alibhai, N.M. Glasner, Low vision rehabilitation in the U.S. health care system. *J. Vis. Rehabil.* **9**, 3–31 (1995)
19. R.W. Massof, A systems model for low vision rehabilitation. I. Basic concepts. *Optom. Vis. Sci.* **72**, 725–736 (1995)
20. R.W. Massof, A systems model for low vision rehabilitation. II. Measurement of vision disabilities. *Optom. Vis. Sci.* **75**, 949–973 (1998)
21. J.A. Stelmack, T.R. Stelmack, R.W. Massof, Measuring low vision rehabilitation outcomes with the NEI VFQ-25. *Invest. Ophthalmol. Vis. Sci.* **43**, 2859–2868 (2002)
22. M. Gobeille, C. Bradley, J.E. Goldstein, R.W. Massof, Calibration of the Activity Inventory item bank: A patient-reported outcome measurement instrument for low vision rehabilitation. *Trans. Vis. Sci. Technol.* **10**, 12 (2021)
23. R.W. Massof, C.T. Hsu, F.H. Baker, G.D. Barnett, W.L. Park, J.T. Deremeik, C. Rainey, C. Epstein, Visual disability variables. II: The difficulty of tasks for a sample of low-vision patients. *Arch. Phys. Med. Rehabil.* **86**, 954–967 (2005)
24. E. Wilson, M. Hilferty, The distribution of chi-square. *Proc. Natl. Acad. Sci. U. S. A.* **17**(12), 684–688 (1931)
25. R.W. Massof, L. Ahmadian, L.L. Grover, J.T. Deremeik, J.E. Goldstein, C. Rainey, C. Epstein, G.D. Barnett, The Activity Inventory: An adaptive visual function questionnaire. *Optom. Vis. Sci.* **84**, 763–774 (2007)
26. P.G. Moschopoulos, W.B. Canada, The distribution function of a linear combination of chi-squares. *Comput. Math. Appl.* **10**, 383–386 (1984)
27. R.W. Massof, A clinically meaningful theory of outcome measures in rehabilitation medicine. *J. Appl. Meas.* **11**, 253–270 (2010)
28. M.A. Goodale, A.D. Milner, Separate visual pathways for perception and action. *Trends Neurosci.* **15**, 20–25 (1992)
29. G.E. Legge, G.S. Rubin, D.G. Pelli, M.M. Schleske, Psychophysics of reading – II. Low vision. *Vis. Res.* **25**, 253–265 (1985)
30. R.W. Massof, An interval-scaled scoring algorithm for visual function questionnaires. *Optom. Vis. Sci.* **84**, 690–705 (2007)

31. J.A. Stelmack, X.C. Tang, D.J. Reda, D. Moran, S. Rinne, R.M. Mancil, R. Cummings, G. Mancil, K. Stroupe, N. Ellis, R.W. Massof, The Veterans Affairs Low Vision Intervention Trial (LOVIT): Design and methodology. *Clin. Trials* **4**, 650–660 (2007)
32. J.K. Cooper, T. Kohlmann, J.A. Michael, S.C. Haffer, M. Stevic, Health outcomes: New quality measure for Medicare. *Int. J. Qual. Health Care* **13**, 9–16 (2001)
33. J.J. Manly, N. Schupf, Y. Stern, A.M. Brickman, M.X. Tang, T. Mayeux, Telephone-based identification of mild cognitive impairment and dementia in a multicultural cohort. *Arch. Neurol.* **68**, 607–614 (2011)
34. J.M. Lyness, T.K. Noel, C. Cox, D.A. King, Y. Conwell, E.D. Caine, Screening for depression in elderly primary care patients: A comparison of the Center for Epidemiologic Studies–Depression Scale and the Geriatric Depression Scale. *Arch. Intern. Med.* **157**, 449–454 (1997)
35. A.D. Deemer, R.W. Massof, B.W. Rovner, R.J. Casten, C.V. Piersol, Functional outcomes of the low vision depression trial in age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* **58**, 1514–1520 (2017)
36. J.E. Goldstein, M.W. Chun, D.C. Fletcher, J.T. Deremeik, R.W. Massof, Low Vision Research Network Study Group, Visual ability of patients seeking outpatient low vision services in the United States. *JAMA Ophthalmol.* **132**, 1169–1177 (2014)
37. R.W. Massof, L. Ahmadian, What do different visual function questionnaires measure? *Ophthalmic Epidemiol.* **14**, 198–204 (2007)
38. R.W. Massof, A general theoretical framework for interpreting patient-reported outcomes estimated from ordinally scaled item responses. *Stat. Methods Med. Res.* **23**, 409–429 (2014)
39. R.W. Massof, C.T. Hsu, F.H. Baker, G.D. Barnett, W.L. Park, J.T. Deremeik, C. Rainey, C. Epstein, Visual disability variables. I. The importance and difficulty of activity goals for a sample of low vision patients. *Arch. Phys. Med. Rehabil.* **86**, 946–953 (2005)
40. T.F. Cox, M.A.A. Cox, *Multidimensional Scaling*, 2nd edn. (CRC Press LLC, Boca Raton, 2001)
41. B.W. Rovner, R.J. Casten, M.T. Hegel, R.W. Massof, B.E. Leiby, A.C. Ho, W.S. Tasman, Low vision depression prevention trial in age-related macular degeneration: A randomized clinical trial. *Ophthalmology* **121**, 2204–2211 (2014)
42. J. Goldstein, M. Jackson, S. Fox, J.T. Deremeik, R.W. Massof, Low Vision Research Network Study Group, Clinically meaningful rehabilitation outcomes of low vision patients served by outpatient clinical centers. *JAMA Ophthalmol.* **133**(7), E1–E8 (2015)
43. M. Gobeille, A. Malkin, R. Jamara, N.C. Ross, Clinical outcomes of low vision rehabilitation delivered by a mobile clinic. *Ophthalmic Physiol. Opt.* **38**(2), 193–202 (2018)
44. R.M.A. Van Nispen, G. Virgili, M. Hoeben, M. Langelaan, J. Klevering, J.E.E. Keunen, G.H.M.B. van Rens, Low vision rehabilitation for better quality of life in visually impaired adults (review). *Cochrane Database Syst. Rev.* **Issue1**, CD006543 (2020). <https://doi.org/10.1002/14651858>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

