Item Banking in Australia (Eye-tem) – Third Generation Questionnaire

Konrad Pesudovs

Professor & Chairman,

NH&MRC Centre for Clinical Eye Research,

Discipline of Optometry and Vision Science,

Flinders University, South Australia, Australia

Patient-reported outcome measures (PROM)

- Questionnaires/Instruments
- Not all the same - technology
  - First generation (summary scoring)
  - Second generation (Rasch / IRT scoring)
  - Third generation (Item Bank)

The trouble with 1st generation questionnaires is invalid scoring

- Summary scoring of ordinal values applied to response categories

Do you have difficulty driving…?

not at all (1), a little (2), quite a bit (3), a lot (4)

- Assumes the spacing between response categories are equidistant
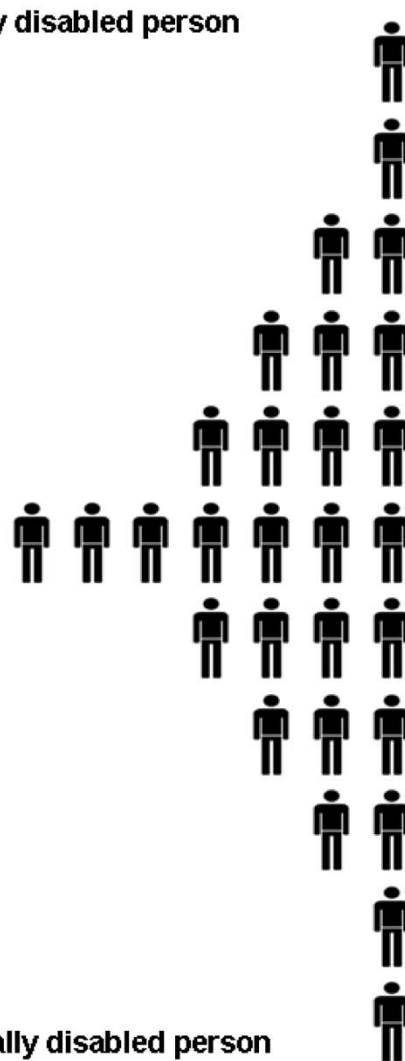
- Assumes all questions have the same "value"

- Not valid and cannot be used in statistical analysis of correlation or change
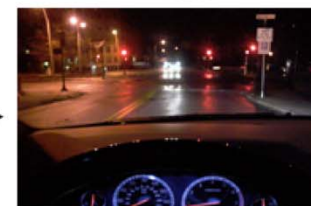
NHMRCf
Centre for Clinical
Eye Research

Rasch analysis solves this scoring problem in the second generation instruments
e.g. Catquest-9SF

Limitations of 2nd generation questionnaires

- Content of a questionnaire may not suit the population – e.g. too easy or too difficult
- Trade off of length versus applicability
- Not adaptable to change
- Paper-based format

3rd Generation - Item Banking

- Item bank - a very large collection of items
- Calibrate items on a single measurement scale using Rasch analysis
- All items connect to the underlying latent trait
- Key premise – responses to any item set provides a measure of the latent trait

What are we measuring?

- Latent trait e.g. ability
- The score an instrument produces is a score of the latent trait
- The score is not a sum of answers to questions
- The questions are not important, as long as they connect to the underlying latent trait – any questions will do

Questions

- Questions are just marks on a ruler
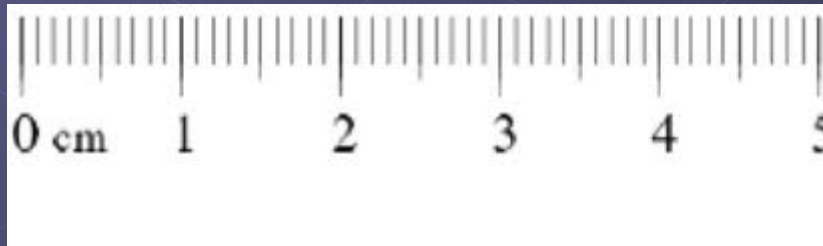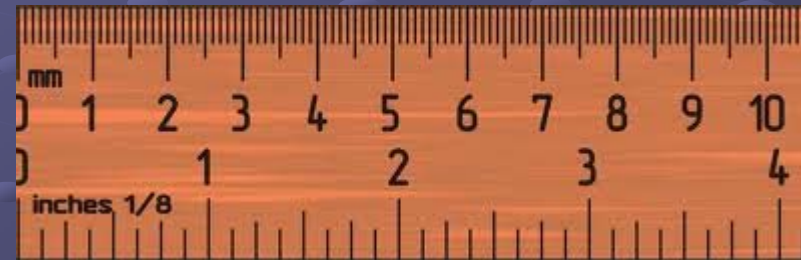
Item Banking

- A very large collection of items
- Many items suits all patient abilities
- Questionnaires (short ruler) have targeting problems
- Many items solves the targeting problems

- Important for use of instrument in different populations
  - Socio-economic
  - International

Item Banking

- Item banks can change and evolve
- It is possible to add new items to an item bank
- The implementation of an item bank is a computerised process which can include measurement items and calibration items
- Calibration of "new" items is done during measurement with "old" items

Item Banking

- Implement measurement in a more efficient format than paper questionnaires - computer adaptive testing (CAT)
- Staircase algorithm that selects questions based upon previous responses

Computer Adaptive Testing (CAT)

The Eye-tem Bank

Eye-tem Bank Project

The Eye-tem Bank project aims:

- To develop, validate, and implement an item banking and computer adaptive testing system to assess ophthalmic quality of life
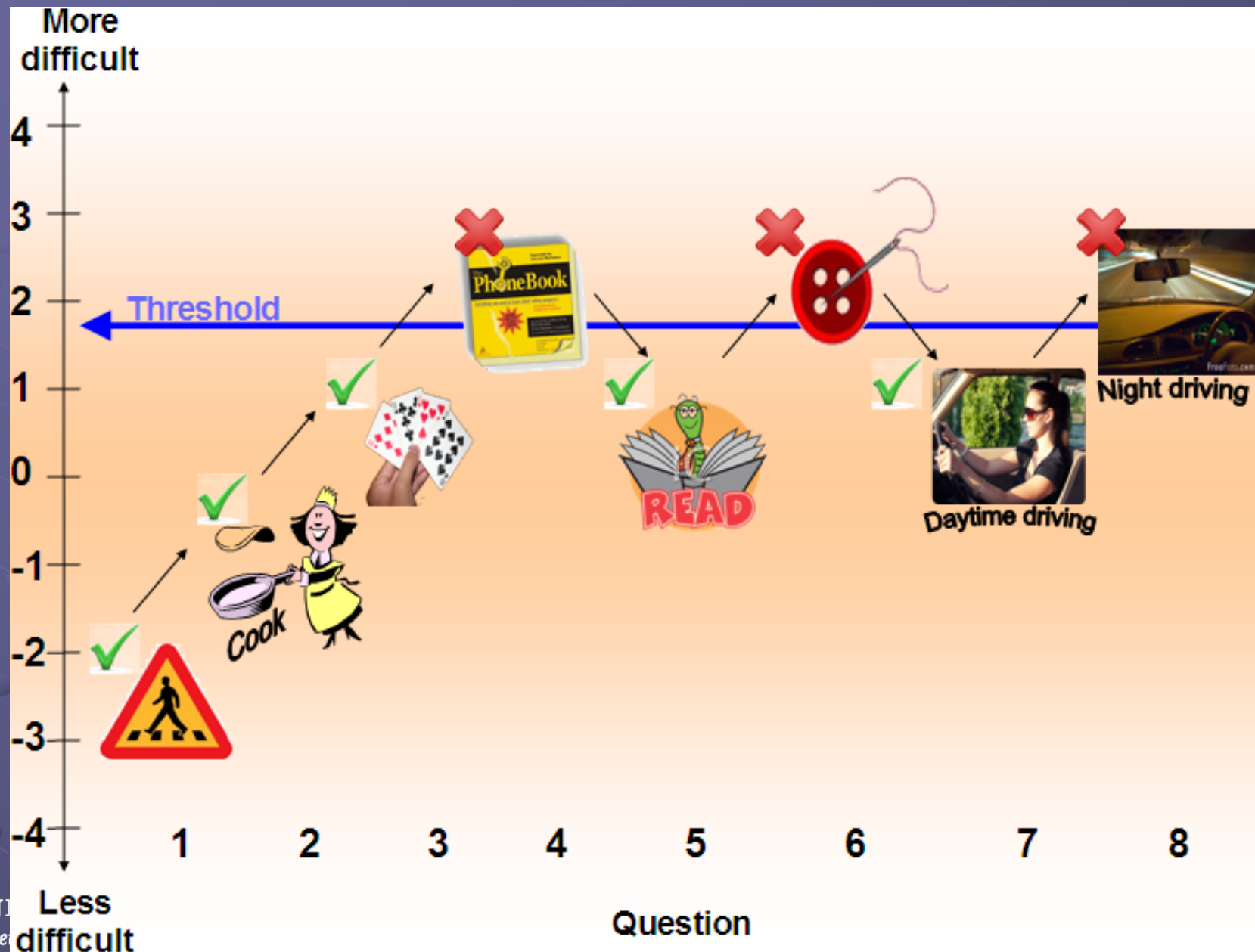
NHMRC
Centre for Clinical
Eye Research

Eye-tem Bank Project

Across 13 disease groups

Eye-tem Bank disease modules
Glaucoma
Diabetic retinopathy (DR)
Age-related macular degeneration (AMD)
Retinal Detachment (RD)
Other vitreo-retinal diseases/conditions
Refractive error
Cataract & corneal opacities
Corneal diseases
Strabismus & amblyopia
Uveitis spectrum of diseases
Inflammatory diseases other than uveitis
Neuro-ophthalmic
Lacrimal and Ocular surface diseases

Methods

- For each Eye-tem Bank module
  - Items are being developed and tested across 10 hypothesised QOL domains

Activity limitation
Mobility
Visual symptoms
Ocular surface symptoms
General symptoms
Convenience
Health concerns
Emotional well-being
Social well-being
Work/finance

Final system

- 10 latent traits x 13 disease groups
- 130 measures!
- Assumptions
  - Disease-specific
  - Latent traits stand alone
  - To be determined

Methods

- Four-phase development process (contd..)

Phase III → CAT system
Validity and reliability tests

↓

Validated disease-specific
Eye-tem Bank modules

Phase IV → Disease-specific population
normative QOL data

↓

Interactive internet based platform
- Desktops/ Tablets/ Smart phones Apps
Real time PRO assessment and reporting
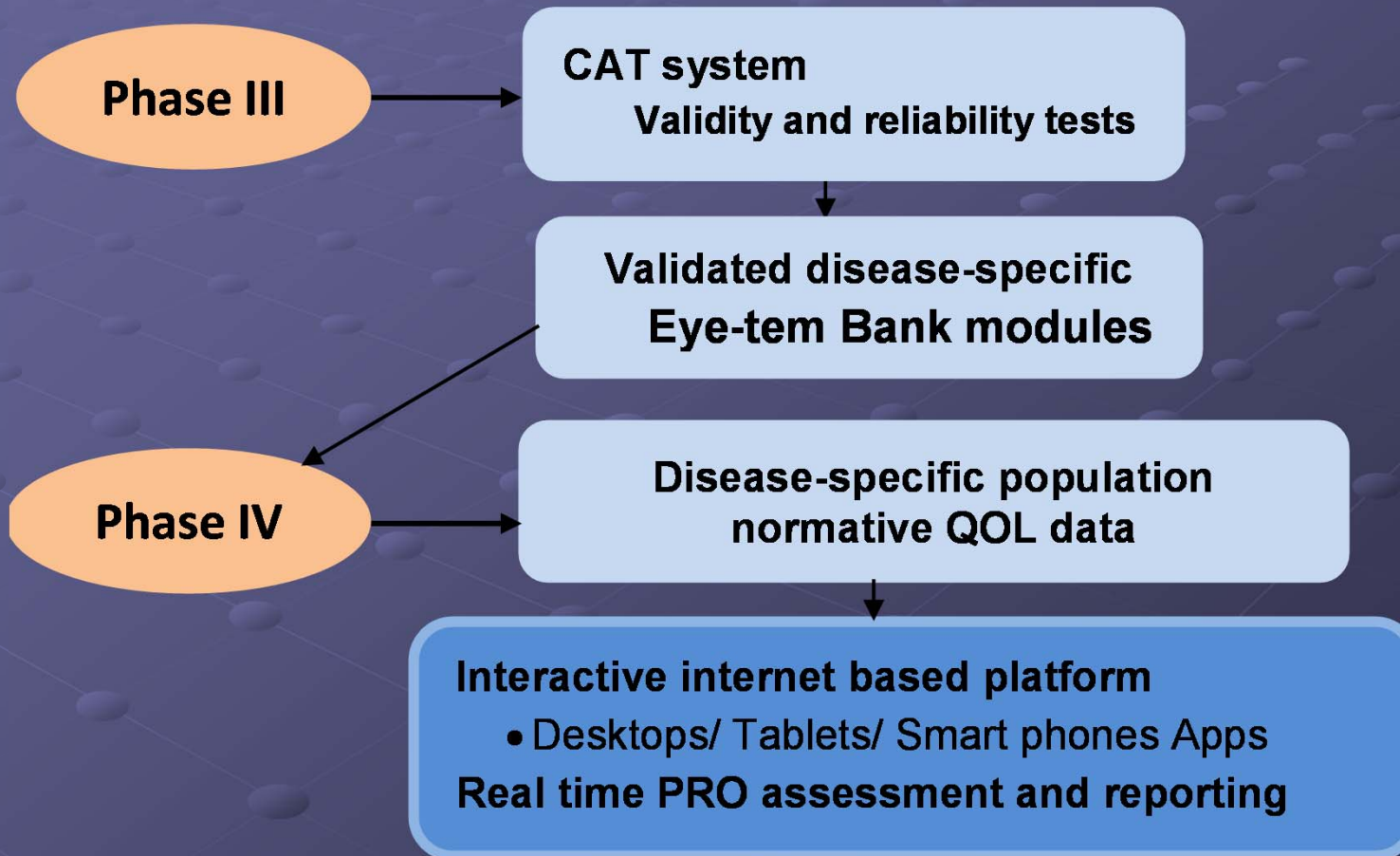
Phase 1

- Identical to questionnaire development methodology

Item Identification

- Literature review
  - Existing items in existing instruments
  - Extracting items from the qualitative literature
- Qualitative research
  - Patient focus groups / semi-structured interviews (sampling is very important)
  - Expert focus groups
  - Formal analysis framework

NHMRC
Centre for Clinical
Eye Research

Item Reduction

- Binning and winnowing
- Aggregate and count
- Reduce using experts
- Format into items
- Cognitive testing of items

Results - literature review

- 130 ophthalmic questionnaires were identified
- These were developed for:
  - 21 generic
  - 19 glaucoma
  - 19 dry eye
  - 17 cataract
  - 18 retinal disease
  - 15 paediatric
  - 9 refractive correction
  - 12 other

NHMRC
Centre for Clinical
Eye Research

Latent traits and items

- The review identified 3192 items in 8 traits
  - activity limitation (1629)
  - visual symptoms (560)
  - health (22)
  - ocular symptoms (432)
  - treatment (190)
  - emotions/feelings (274)
  - independence/coping (62)
  - work/finance (23)
- Many instruments incorporate similar items, so the total number of unique items is 1246 (39%)

Content Development Map
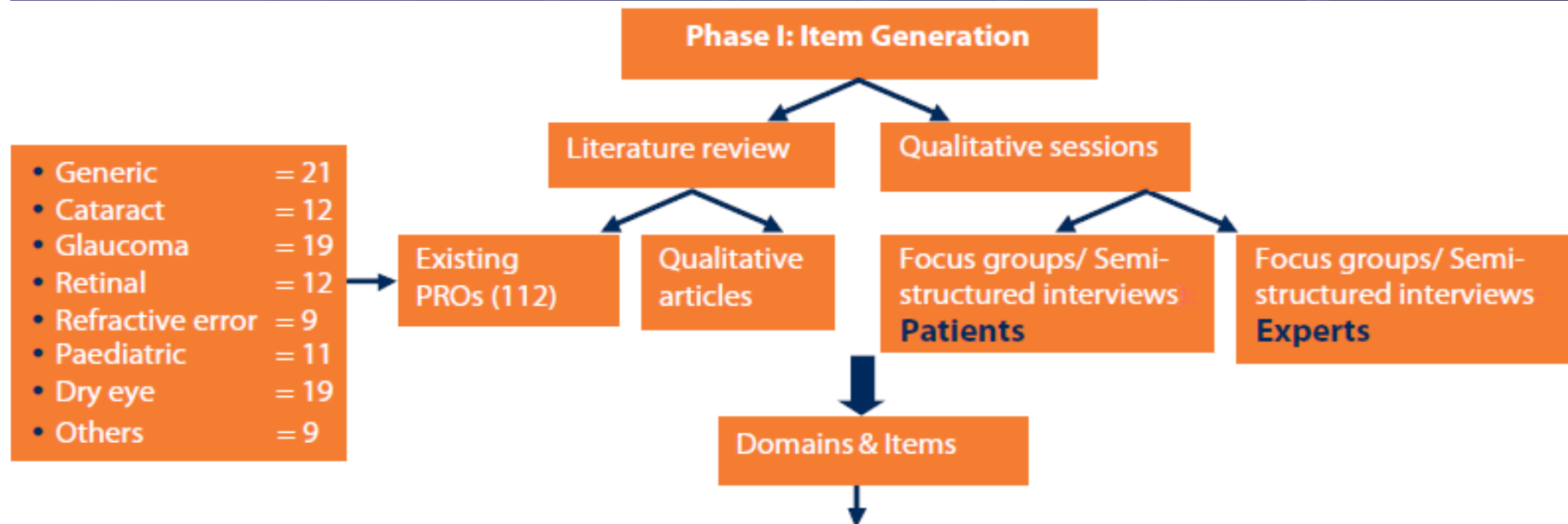
Results: Phase pilots

- Completed for 12 disease groups
- 6 Pilot instruments already developed
  - Diabetic Retinopathy (DR)
  - Glaucoma
  - Age-related macular degeneration (AMD)
  - Refractive error (RE)
  - Other vitreo-retinal (split into 2- only 45% QoL issues common)
    - Hereditary retinal diseases (HRD)
    - Acquired retinal diseases (ARD)
    - So now 14 disease groups!

NHMRC
Centre for Clinical
Eye Research

Results: Phase I focus groups

- Pilot instruments being developed for 6 modules
  - Amblyopia and Strabismus (n=48)
  - Retinal detachment, RD (n=42)
  - Uveitis spectrum of diseases (n=41)
  - Ocular inflammation other than uveitis (n= 40)
  - Cornea (n= 39)
  - Ocular surface and lacrimal (n= 39)

NHMRC
Centre for Clinical
Eye Research

Results: Phase I

- Now 14 disease groups
- Ongoing patient recruitment & data collection- 2 disease groups
  - Cataract and corneal opacities (n= 25)
  - Neuro-ophthalmic

Content development map

Phase I: Item Generation

- Generic = 21
- Cataract = 12
- Glaucoma = 19
- Retinal = 12
- Refractive error = 9
- Paediatric = 11
- Dry eye = 19
- Others = 9

Literature review

Qualitative sessions

Existing PROs (112)

Qualitative articles

Focus groups/ Semi-structured interviews **Patients**

Focus groups/ Semi-structured interviews **Experts**

Domains & Items

Glaucoma
Initial item pool = 737 *(unique items)*
- Literature review = 261
- Qualitative sessions = 476

Diabetic Retinopathy (DR)
Initial item pool = 1165 *(unique items)*
- Literature review = 756
- Qualitative sessions = 409

Other disease-specific item pools *(to follow)*

Phase II: Item refinement and revision
- Binning and winnowing
- Expert opinion
- Pre-testing testing with cognitive interviews

Final item pools = 349 items

Final item pools = 314 items

Phase I: Pilot Instruments

Domains	Glaucoma Total items
Activity limitation	88
Mobility	20
Ocular comfort symptoms	23
Visual symptoms	18
General symptoms	16
Convenience	40
Emotional well-being	54
Health concerns	45
Social	23
Economic	22
Total	349

Domains	DR Total items
Activity limitation	120
Mobility	19
Ocular comfort symptoms	10
Visual symptoms	18
General symptoms	0
Convenience	30
Emotional well-being	48
Health concerns	36
Social	21
Economic	12
Total	314

Domains	AMD Total items
Activity limitation	114
Mobility	23
Ocular comfort symptoms	25
Visual symptoms	14
General symptoms	0
Convenience	34
Emotional well-being	72
Health concerns	39
Social	21
Economic	17
Total	359

Khadka J, McAlinden C, Craig JE, Fenwick MA, Lamoureux EL, Pesudovs K. Identifying content for the glaucoma-specific item bank to measure quality of life parameters. *J Glaucoma* 2013; *[Epub ahead of print]*.

Fenwick E, Pesudovs K, Khadka J, Rees G, Wong TY, Lamoureux EL. Evaluation of item candidates for a diabetic retinopathy quality of life item bank. *Qual Life Res* 2012; **22**(7): 1851-8.

Phase I: Content overlap

Domain	AMD		Glaucoma		DR		HRD		ARD		RE		Common items	
	Total	Unique	Total	Unique	Total	Unique	Total	Unique	Total	Unique	Total	Unique	n	%
VS	18		19		18		20		22		28	8	6	28.8
OS	13		22		10		4		8		17	2	2	16.2
GS	0		15		0		6		4		10	3	0	0.0
AL	100		88		120		86		62		81	13	25	27.9
MB	19		20		19		23		10		20	3	2	10.8
HC	38		45		36		48		27		41	21	7	17.9
CV	29		39		30		16		21		45	37	2	6.7
EM	50		49		48		66		51		58	17	13	24.2
SC	20		23		21		28		18		17	2	1	4.7
EC	15		22		12		17		10		18	5	3	19.1
CP	0		0		0		29		24		13	5	0	0.0
Total	**302**	**0**	**342**	**0**	**314**	**0**	**343**	**0**	**257**	**0**	**348**	**116**	**61**	**19.2**

Note: AMD = age-related macular degeneration; DR= diabetic retinopathy; HRD= hereditary retinal diseases; ARD= acquired retinal diseases, RE= refractive error; VS= Visual symptoms, OS = Ocular comfort symptoms; GS = General symptoms; AL = Activity limitation; MB = Mobility; HC= Health Concern; CV= Convenience; EM= Emotional well-being; SC= Social, EC = Economic & CP= Coping

NHMRC
Centre for Clinical
Eye Research

Around 20% of items were common across the modules

Pilot instruments

- 20% of items were common between Glaucoma, DR, AMD, ARD, HRD and RE modules
- Our hypothesis
  - The final Eye-tem Bank will have a core item set plus disease-specific item sets
- Disease specificity is all important

Phase II: Item bank development

- **Completed**
  - **DR module**
    - Administered to 466 patients (median age, 62 yrs, range 22-88 yrs)

  - **Glaucoma module**
    - Administered to 293 patients (median age, 70 yrs, range 20-91 yrs)
      - Majority (80%) endorsed two lower end response categories signifying less impact on QOL

NHMRC
Centre for Clinical
Eye Research

Phase II: Rasch analysis

- **DR module**
  - Only 28 (9%) items were misfitting
  - Only 4 (1.3%) items showed differential item functioning (DIF) by age and gender

- **Glaucoma module**
  - Only 27 (7.9%) items were misfitting
  - Only 7 (2%) items showed DIF by age and gender

- **Both modules demonstrated**
  - Good psychometric properties against most Rasch based metrics across all the QOL domains

NHMRC
Centre for Clinical
Eye Research

Phase II: Psychometric properties of DR and Glaucoma

Rasch Parameters	Module	QOL domains												
		AL	MB	CV	HC	EM	SC	EC	VS			OS		
									Freq	Sev	Both	Freq	Sev	Both
PSI	Glaucoma	3.12	2.01	2.02	2.45	2.22	2.08	1.95	2.12	2.13	2.00	1.91	1.76	1.59
	DR	5.93	2.04	2.23	2.93	2.29	2.14	2.14	2.21	2.29	2.30	1.48	1.39	1.43
Misfitting items (n)	Glaucoma	5	1	2	5	5	2	0	2	0	1	2	1	1
	DR	4	2	5	1	5	4	2	0	1	1	0	0	0
PCA, variance explained by the measure (%)	Glaucoma	64.3	68.8	48.4	50.6	57.5	59.5	64.7	43.3	51.5	48.5	37.6	37.4	38.5
	DR	68.8	67.1	57	57.3	59.4	55.7	56.4	44.5	45.2	49.6	43	37.6	45.0
DIF by age (n)	Glaucoma	0	1	1	0	0	0	3	0	0	0	0	0	0
	DR	0	1	1	0	0	0	2	0	0	0	0	0	0
DIF by sex (n)	Glaucoma	0	0	0	0	0	0	0	0	0	0	0	0	0
	DR	0	0	0	0	0	0	0	0	0	0	0	0	0
Mean person location	Glaucoma	1.88	3.71	2.14	1.53	2.52	2.55	1.81	1.34	1.70	1.99	1.80	1.64	2.02
	DR	1.32	3.14	1.49	1.23	1.94	1.78	0.39	0.88	0.92	1.14	2.08	1.89	2.00

Note: AL = Activity Limitation, MB = Mobility, VS = Visual Symptoms, OS = Ocular Comfort Symptoms, CV = Convenience, HC = Health Concerns, EM = Emotional, SC = Social, EC = Economic, PCA = Principle Component Analysis, DIF = Differential item functioning, Freq = Frequency, Sev = Severity, Both = Bothersome, PSI = Person separation index

Both modules demonstrated
- Good psychometric properties against most Rasch based metrics across all the QOL domains

NHMRC
Centre for Clinical
Eye Research

Decision-making in analysis

- Removal of mis-fitting persons >2.00 or >1.5 fit statistic
  - Curtis DD. Person misfit in attitude surveys: influences, impacts and implications *Int Ed J* 2004;5(2):125-144.
- Removed items with fit >2.00 – more noise than signal!
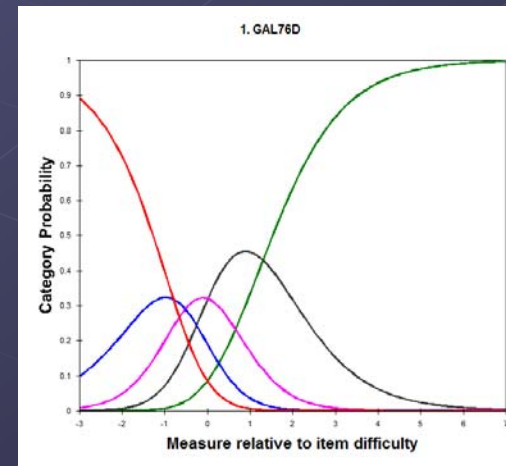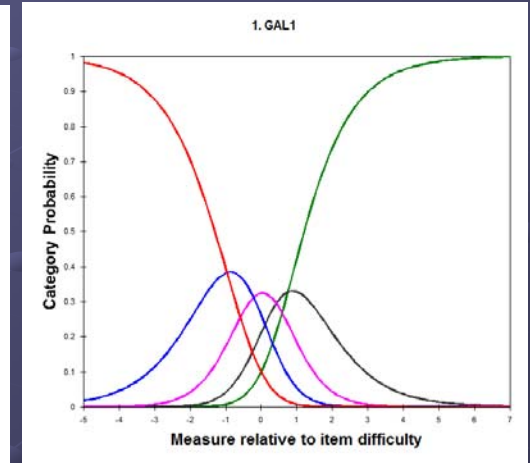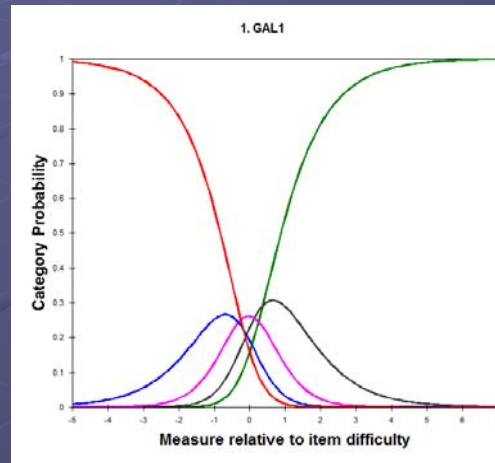- Dropping perfect response sets - glaucoma

NHMRC
Centre for Clinical Eye Research

Many items – detail challenges

- Activity limitation
  - All items
  - Remove driving
  - Driving alone
  - Reading
  - Lighting
- Scale or subscale?
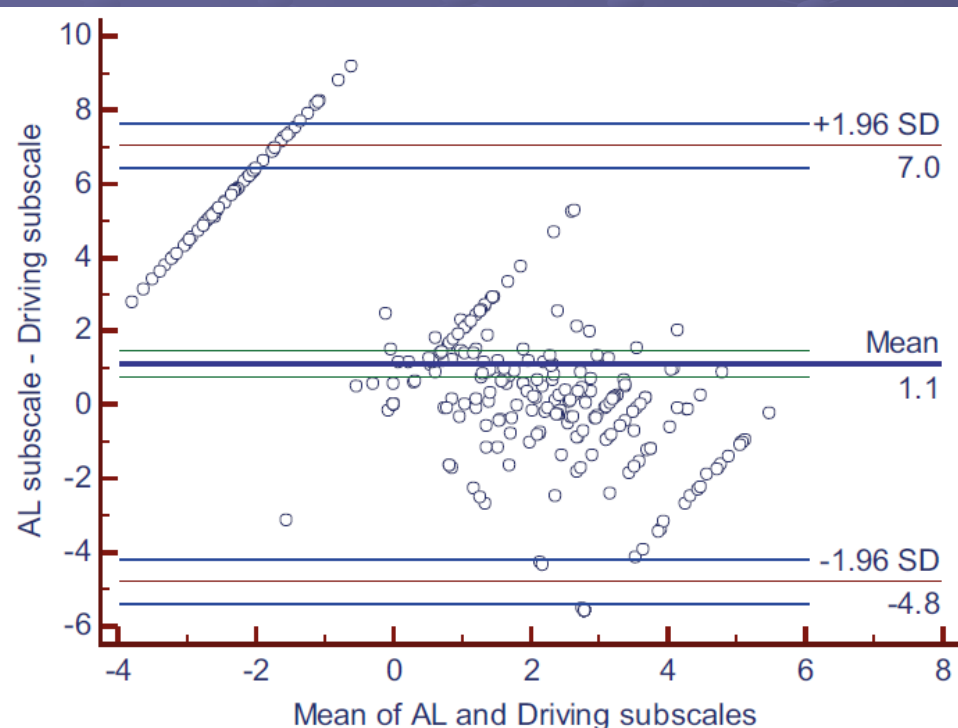- Unidimensional with secondary strands

NHMRC
Centre for Clinical
Eye Research

Phase II: Glaucoma module

- A new QoL domain identified
  - **Driving**

Bland and Altman plot shows the limit of agreement (Mean difference and 95% confidence interval) between the Activity limitation and the Driving
Note: AL = Activity limitation

Activity limitation domain of Glaucoma module			
Parameters	**All items**	**Driving**	**AL without driving**
Number of of items	88	13	75
Rating scale ordering	Disordered	Ordered	Ordered
Precision (PSI)	3.12	2.94	4.25
Item Infit MNSQ >1.5	5	0	6
Item Outfit MNSQ >1.5	9	0	7
PCA analysis % variance explained by measure	64.5	58.4	63.2
PCA analysis Eigen value 1st contrast	10.2	2.3	6.0

NHMRC
Centre for Clinical
Eye Research

10 Domains of QOL become 12!

- Activity limitation + 3 domains
- Driving a stand alone domain
- Coping a new domain in HRD
- Visual symptoms, Convenience, Mobility, Emotional, Health concerns, Social: OK
- In some modules domains not viable: Economic and Ocular surface symptoms in DR, General symptoms in glaucoma

Phase III: Computer adaptive testing

- Populate with calibrated items
- Algorithm development
- Trialled CAT in glaucoma and diabetic retinopathy clinical studies
- Diabetic retinopathy ready to roll out

Demonstration

Eye-tem Bank - Diabetic
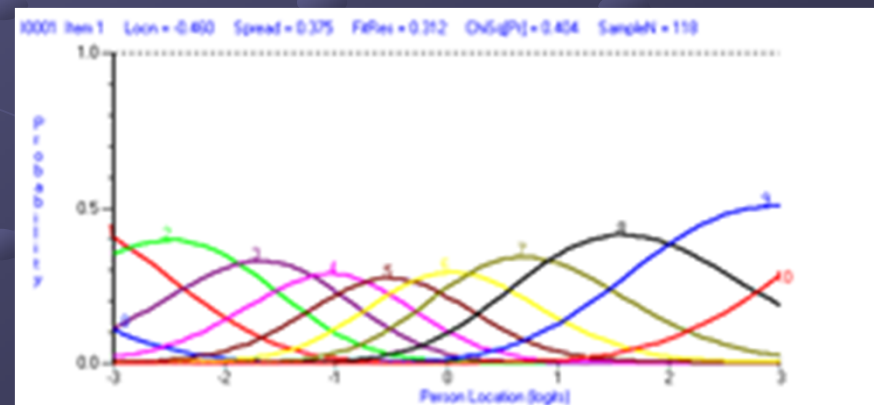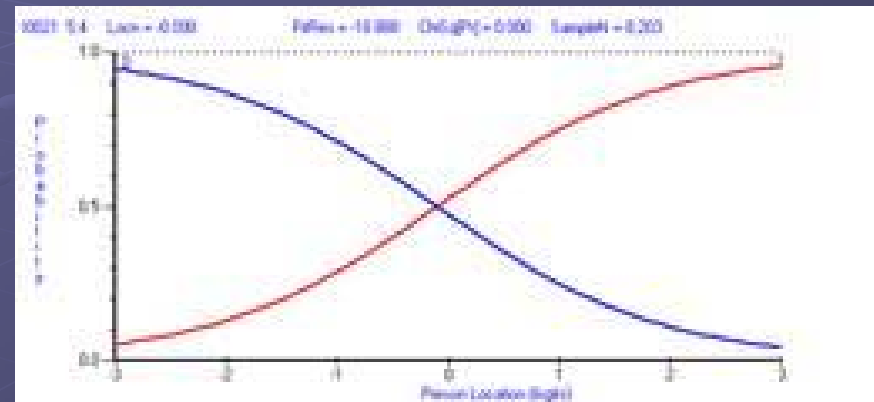Retinopathy - Activity Limitation

**BHB-906-273-215**

MUO-263-422-351

Estimating thresholds

- MCQ or binary question has a single threshold
- Rating scale has multiple thresholds
- Simple to select item if "no problem" or "can't do"
- Complicated for partial difficulty

Choosing items

- Fisher Information Functions
- Theta hat 0 – item 6 provides the most information
- "wrong"
- Theta hat -1 – item 4 provides the most information
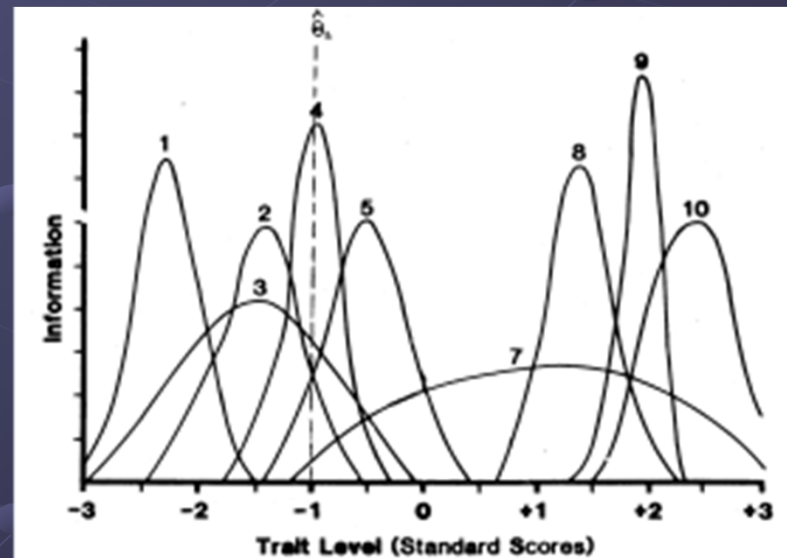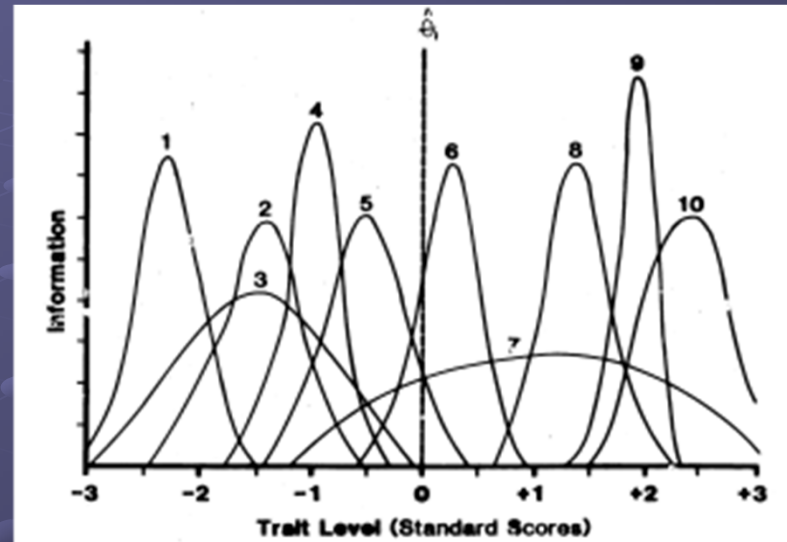
Finalising CAT

- Monte Carlo simulations for different distributions
- All known information functions
- Fixed length easier than fixed precision

Phase II: CAT simulation for DR

Standard error of measurement (SEM) set at 0.52			
Domain	No. of items available for CAT	Average no. of items used by CAT	Correlation between CAT and item bank theta
Visual symptoms	18	4	0.91
Activity Limitation	77	3	0.88
Mobility	17	5	0.90
Emotional	43	4	0.88
Health concerns	35	3	0.88
Social	20	3	0.90
Convenience	20	3	0.90
Economic	15	3	0.91
Driving	15	4	0.91
Luminance	10	4	0.93

An average of 3-5 items administered to obtain the desired SEM
Ran on "Firestar-D software (online software)"-n=1000

NHMRC
Centre for Clinical
Eye Research

Ongoing work

- Phase 1 for 2 incomplete groups
- Phase 2 for 4 groups, with 6 to follow
- Phase 3 testing of Glaucoma module, 4 soon to follow
- Conduct phase 4 validity studies for DR

NHMRC
Centre for Clinical
Eye Research

Vision for the future

- An internet-based CAT system will be made available to the international eye research community

- Rapid online testing, real-time scoring and data storage

- Available via all popular digital formats, e.g. iPhone, Android etc

Conclusion

- Item banking provides patient-reported measurement which
  - Has high quality psychometric measurement
  - Suits all populations
  - Is adaptable to change
  - Is implemented via modern technology
  - Is the future of PROMs

NHMRC
Centre for Clinical
Eye Research

Acknowledgements

- Prof Konrad Pesudovs
- Assoc Prof Ecosse L Lamoureux
- Dr Jyoti Khadka
- Dr Eva K Fenwick
- Ms Susan Aldhous
- Ms Prue Spencer
- Prof Robert Casson
- Assoc Prof Jagit Singh Ghilhotra
- Assoc Prof Michael Goggin
- Prof Justine Smith
- Assoc Prof Jamie Craig
- Dr Mallika Prem Senthil
- Ms Chris Lewis

Collaborators

Discipline of Optometry and Vision Science,

Flinders University **(Lead centre)**

Royal Victorian Eye and Ear Hospital, Melbourne

Centre for Eye Research Australia, Melbourne

Royal Adelaide Hospital, South Australia

Queen Elizabeth Hospital, South Australia

Ashford Advanced Eye Care, South Australia

NHMRC
Centre for Clinical
Eye Research