

## Abstract

- The RNL has an acceptable reliability
  - Meaning that if we measured a consumer twice with the RNL, they would get the same score each time
- All the items in the RNL do seem to fit a service continuum
  - Meaning that all the questions on the RNL go together very well and are measuring the same thing
- The items cover the full range of services for our consumers, though this could be improved
  - We will explain in more detail, but our questions are measuring our consumer's level of services, but we could still have improvements

## The Recovery Needs Level

- The RNL is designed to evaluate the level of services received by the consumer based on clinical variables
- 15 items with varying response sets from 2 to 5
- Levels of service determined by algorithm
  - Level 1 – High service needs (HITT)
  - Level 2 – High service minus residential needs (HITT)
  - Level 3 – Moderate service needs (CTT)
  - Level 4 – Low service needs (OP)

## Advantages of Item Response Theory (IRT)

- The hallmark feature: **Item-person separation** (a.k.a. invariance)
  - Can discriminate the characteristics of the item ("hard item" vs. "easy items") separately from the characteristics of the consumers ("high service needs" vs. "low service needs")
- Through examination of **infit** and **outfit**, it can be determined how well items are fitting the continuum of service needs
- Among many other features, refer to Hambleton, Swaminathan, Roger 1991

## Methodology

- The information was gathered between March and September 2002 from 625 consumers, with complete information from 575 consumers.
- Data cleaning was conducted in SPSS
- Analysis was conducted using a Rasch Partial Credit Model (Master, 1982) conducted using Winsteps

## Results

- Initially the top 5% of misfitting persons were dropped, however this did not affect reliability or fit statistics significantly.
- After item calibration, the 5% was reintroduced for ability calculations and group means.
- Reliability Estimates:

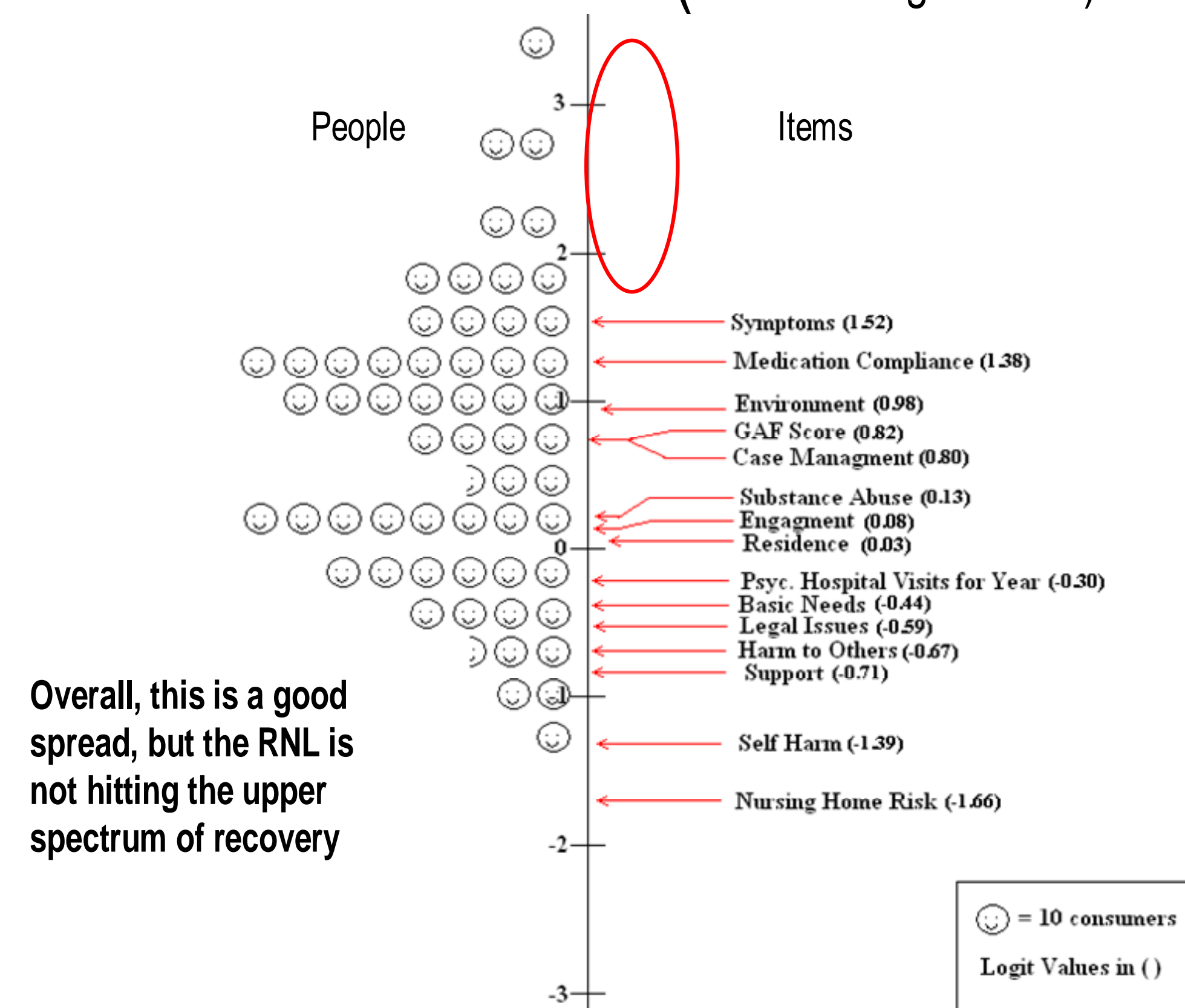
CTT (Cronbach's $\alpha$ )	IRT
.78	.75

## Fit Statistics

Item	Infit (z)	Outfit (z)
GAF Score	1.04 (0.7)	1.05 (0.9)
Hospital Visits	1.07 (1.3)	1.08 (0.8)
Basic Needs	0.87 (-3.1)	0.76 (-3.2)
Nurse H. Risk	1.07 (0.8)	1.13 (0.8)
Legal Issues	1.04 (0.6)	1.45 (2.5)
Substance Abuse	1.12 (1.9)	1.49 (3.5)
Residence Issues	1.01 (0.2)	0.94 (-0.5)
Harm to Others	1.03 (0.5)	1.16 (0.8)
Engagement	0.95 (-0.9)	1.02 (0.4)
Self Harm	1.05 (0.6)	0.89 (-0.7)

Med. Compliance	0.98 (-0.5)	1.00 (0)
Case Management	0.92 (-1.6)	0.91 (-1.8)
Environment	0.88 (-2.5)	0.86 (-2.8)
Symptoms	1.19 (3.0)	1.34 (4.9)
Support	0.92 (-1.6)	0.79 (-2.4)

## IRT Discrimination (aka Ordering of Items)

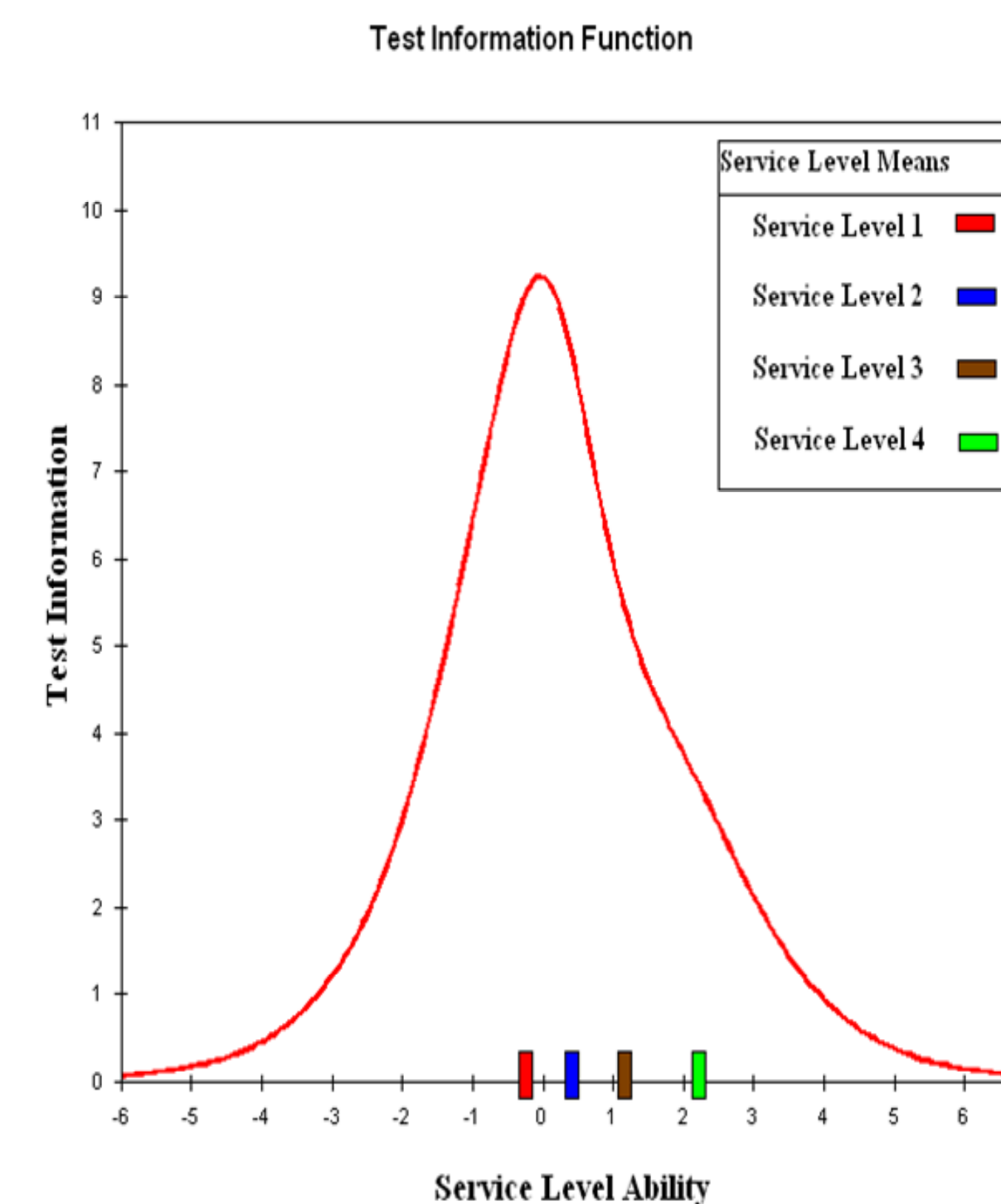


Overall, this is a good spread, but the RNL is not hitting the upper spectrum of recovery

\*These are mean "difficulties" of items as a whole, the responses are more spread.

## Can RNL Differentiate Among Consumers

- In order to examine if the RNL could differentiate among consumers, we ran a one-way ANOVA with IRT scores as the DV and predicted Service Level as the IV.



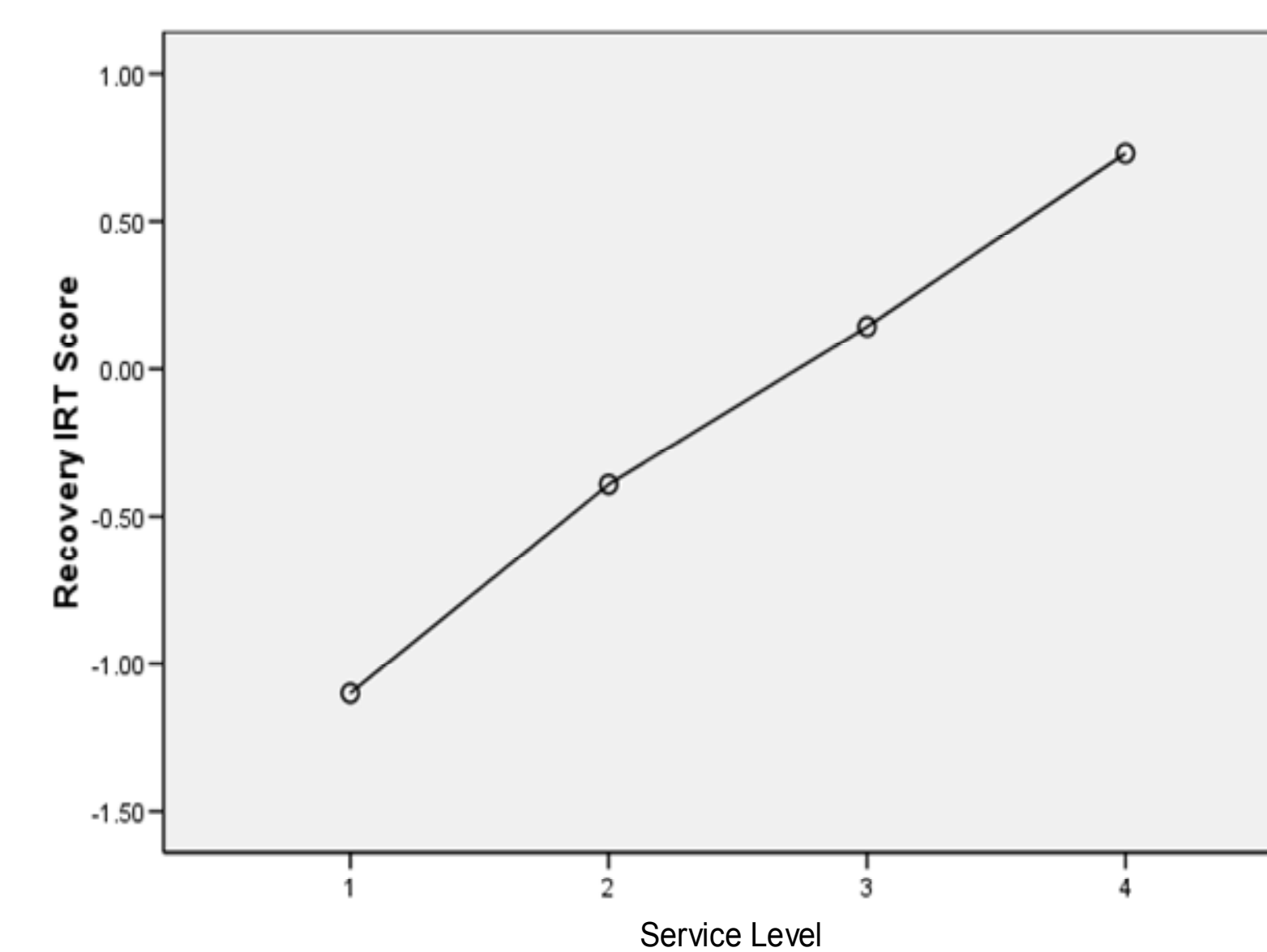
The lowest error of measurement (a.k.a. highest Test Information) are found for Service Levels 1 and 2.

## Service Level Comparison of IRT Ability Scoring

	Level 1	Level 2	Level 3	Level 4
Mean	-0.228	0.379	1.013	2.166
(SD)	(0.457)	(0.355)	(0.483)	(1.101)
	N = 184	N = 77	N = 250	N = 114

An ANOVA revealed that levels of services, as computed through the algorithm, were significantly different,  $F(3,621) = 368.268$  ( $p > 0.001$ )

## ANOVA Plot



As the Rasch Calibrated scores increase, the consumers' service needs decrease.

## Post Hoc Group Comparisons Using a Bonferroni Adjustment

Dependent Variable: measure

	(I) RecLevel	(J) RecLevel	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Bonferroni	1	2	-.7077*	.05988	.000	-.8662	-.5492
		3	-1.2432*	.04285	.000	-1.3566	-1.1298
		4	-1.8298*	.05258	.000	-1.9690	-1.6906
2	1	3	.7077*	.05988	.000	.5492	.8662
		4	-.5355*	.05750	.000	-.6877	-.3834
		3	-1.1221*	.06508	.000	-1.2944	-.9499
3	1	2	1.2432*	.04285	.000	1.1298	1.3566
		4	.5355*	.05750	.000	.3834	.6877
		2	-.5866*	.04986	.000	-.7185	-.4546
4	1	2	1.8298*	.05258	.000	1.6906	1.9690
		3	1.1221*	.06508	.000	.9499	1.2944
		2	-.5866*	.04986	.000	-.4546	-.7185

Based on observed means.

\*. The mean difference is significant at the .05 level.

This shows that the different level of services were significantly different from each other in terms of their IRT recovery level scores.

## Discussion

- Many items are yes/no questions (dichotomous). Unless you have a lot of questions (~50), yes/no items don't give us a lot of information.
- A possible change might be to rewrite items to construct a larger response set for selected items (i.e. to produce more variance)
  - ex. "At risk of nursing home placement" to "How at risk for nursing home placement" – No risk, little risk, some risk, etc.
- Using IRT to validate service levels (i.e. Level 1-4) is not ideal, but we have validated the scoring in which the algorithm is based, displayed from the ANOVA results
  - Since the RNL scoring is psychometrically valid, it would be feasible for centers to adopt the RNL and create alternative service levels that better suit their consumers.

## Conclusions

- Combination of issues to solve:
  - Increasing the precision and accuracy of the RNL (with some item changes), a more precise RNL would give a more interpretable score
  - Double checking the accuracy of the algorithm, how many times are clinicians overriding the service level.
- Overall the RNL seems to work very well. The items are at the same "level" with the majority of the consumers.
- Reliability and other statistics are "OK", but not great.
- Test information is high, but mostly around levels 2 and 3 (meaning standard errors are lowest at levels 2 and 3). It would be helpful if we had more items or response options that appear in the lower range of the continuum of services (meaning higher levels of services)

Presented at the OPEC Conference, Lyons, CO May 17, 2007