

Validation of the Internal Mental Distress Scale (IMDS) to the Rasch Measurement Model,  
GAIN Methods Report 1.1

Kendon J. Conrad  
University of Illinois at Chicago

Karen M. Conrad  
University of Illinois at Chicago

Michael L. Dennis  
Chestnut Health Systems

Barth B. Riley  
Chestnut Health Systems

Rod Funk  
Chestnut Health Systems

**Abstract**

**Purpose.** The purpose of this report is to provide a brief psychometric analysis of the *Internal Mental Distress Scale* using the Rasch measurement model. The 43-item *IMDS* is a count of past-year symptoms related to internalizing disorders, including somatic, anxiety, depression, traumatic stress and suicide thoughts. It is based on the *DSM-IV-TR*, the *Hopkins Symptom Checklist 90* (HSCL-90), the *Mississippi Scale of Post Traumatic Stress Disorder* (PTSD) and common screening items for suicide risk (homicidal thoughts, suicidal thoughts, plans, means, attempts).

**Methods.** Data were analyzed on 7,435 persons who presented for substance problem screening. Rasch analysis included an examination of: person and item reliabilities; construct validity including item and person fit statistics; and differential item functioning (DIF) across subgroups. DIF analysis allowed us to determine if the relative item estimates (i.e., item difficulty estimates) remained invariant across subgroups of persons.

**Results.** The *IMDS* performs well as a measure of the construct of internal mental distress. Both items and scales form the theoretically expected hierarchies with a person internal consistency reliability of .89 and an item reliability of 1.00. The persons' responses generally conformed to the expectations of the Rasch model. Of the 43 items in the *IMDS*, significant DIF (i.e.,  $> .5$  SD = .58 logits) occurred in 3 items for males vs. females, 11 items for youth vs. adults, 4 items for race when using African American as the reference group, and 17 items for primary substances when using alcohol as the referent. In terms of person fit, over 86% of the persons exhibited person infit and outfit that were low or moderate (LMI/LMO) and were thus regarded as fitting the Rasch model expectations well from a clinical perspective. The group termed, Atypical Type 1, (10 %) consisted of persons with low/moderate infit and high outfit (LMI/HO), where the overall score may underestimate severity since these tended to be people who were having suicidal thoughts (higher risk items) but were unexpectedly low on depression, about moderate on somatic, low to moderate on anxiety, except for the three highest anxiety items where they were higher, and low to moderate on trauma (which all tended to be lower risk than suicide). The HI /LMO group, Atypical Type 2, consisted of only 11 people (.1%). This group was characterized by low depression but was especially high on trauma, stress, and anxiety while being somewhat higher on somatic complaints and suicidal ideation. The HI/HO group, Atypical Type 3, (3.4%) tended to be valid high scorers who also had an overall measure that may underestimate severity somewhat and tended to have higher anxiety, higher trauma and higher suicidal ideation while their depression and somatic symptoms were moderate. These persons misfit because they strongly endorsed most of the higher risk items but only endorsed some of the lower risk items at moderate levels.

**Conclusion.** The *IMDS* functioned well as a unidimensional measure with good person and item reliability. In terms of item quality, there were no items with both infit and outfit values outside of our criterion of .75-1.33. While there were several items with high outfit values, these indicated that a few people endorsed these items

unexpectedly, but that, otherwise, the items performed well as indicators of the general construct of internal mental distress. Differential item functioning was not a major issue in terms of causing biased *IMDS* measures, but the DIF analyses did reveal some interesting group differences that may merit further study. This was especially true of the youth vs. adult groups and the primary drug severity groups. Regarding person fit, the two groups with the most misleading measures were Atypical Types 2 and 3. Both of these groups tended to have scores that underestimated the seriousness of their internal mental distress. Atypical Type 3 is of special concern because of the much higher endorsement of suicidal symptoms in this group compared to the overall group.

**Citation.** Conrad, K. J., Conrad, K. M., Dennis, M. L., Riley, B.B., & Funk (2009). Validation of the Internal Mental Distress Scale (IMDS) to the Rasch Measurement Model, GAIN Methods Report 1.1. Chicago, IL: Chestnut Health Systems. Retrieved from [http://www.chestnut.org/li/gain/psychometric\\_reports/Conrad et al 2009 IMDS Rasch Report.pdf](http://www.chestnut.org/li/gain/psychometric_reports/Conrad_et_al_2009_IMDS_Rasch_Report.pdf)

## Purpose of this Report

The purpose of this report is to provide a brief psychometric analysis of the *Internal Mental Distress Scales (IMDS)* using the Rasch measurement model (Rasch, 1960; Bond & Fox, 2007). The Rasch analysis was conducted using *Winsteps* software (Linacre, 2008). The report presents annotated tables and figures to summarize the main points.

## Background

The *IMDS* is a scale of the *General Individual Severity Scale (GISS)*. The *GISS* in turn is part of the larger *Global Appraisal of Individual Needs (GAIN)* which is a standardized biopsychosocial instrument that integrates research and clinical assessment for people presenting to substance abuse treatment or other behavioral health treatment (Dennis, Chan, & Funk, 2006).

The *IMDS* is a count of past-year symptoms related to internalizing disorders, including somatic, anxiety, depression, traumatic stress and suicide thoughts. It is based on the *DSM-IV-TR* (APA), the *Hopkins Symptom Checklist 90 (HSCL-90)*, the *Mississippi Scale of Post Traumatic Stress Disorder (PTSD)* and common screening items for suicide risk (homicidal thoughts, suicidal thoughts, plans, means, attempts) (Dennis, et al., 2006).

## IMDS Subscales and Items

The *IMDS* consists of five subscales with a total of 43 items. Its subscales are the: *Somatic Symptom Index (SSI; 4 items)*, *Depressive Symptom Scale (DSS; 10 items)*, *Homicidal Suicidal Thought Scale (HSTS; 5 items)*, *Anxiety/Fear Symptom Scale (AFSS; 12 items)*, and the *Traumatic Distress Scale (TDS; 13 items)*. The scale stem reads as follows: “The next questions are about common nerve, mental, or psychological problems that many people have. These problems are considered significant when you have them for two or more weeks, when they keep coming back, when they keep you from meeting your responsibilities or they make you feel like you cannot go on.” The item stem then reads: “During the past year, have you had significant problems with. . .” Response format is Yes/No (coded: no=0, yes=1). The scale names, item stems, *GAIN* item numbers, Rasch output item location codes, and item labels are shown in the table below. Note Item M1B2 is deleted from the analysis.

**Table 1. Scale and Item Information**

Subscale Name and Item Stem	GAIN Item Number	Rasch Output Item Number	Item Label
<i>Somatic Symptom Index</i>			
1. Headaches, faintness, dizziness, tingling, numbness, sweating or hot or cold spells?	M1A1	17	HeadachFaint
2. Sleep trouble, such as bad dreams, sleeping restlessly or falling asleep during the day?	M1A2	18	SlepTrbl
3. Having dry mouth, loose bowel movements, constipation, trouble controlling your bladder or related itching?	M1A3	19	DryMouth
4. Pain or a heavy feeling in your heart, chest, lower back, arms, legs or other muscles?	M1A4	20	PainChest
<i>Depressive Symptom Scale</i>			
5. Feeling very trapped, lonely, sad, blue, depressed, or hopeless about the future?	M1B1	21	Sadepressed
6. Remembering, concentrating, making decisions, or having your mind go blank?	M1B3	23	Remembering
7. Feeling very shy, self-conscious or uneasy about what people thought or were saying about you?	M1B4	24	4Shyness
8. Thoughts that other people did not understand you or appreciate your situation?	M1B5	25	DontUnderstand
9. Feeling easily annoyed, irritated, or having trouble controlling your temper?	M1B6	26	AnoyedIrritated
10. Feeling tired, having no energy, or feeling like you could not get things done?	M1B7	27	Tired
11. Losing interest or pleasure in work, school, friends, sex, or other things you cared about?	M1B8	28	LostInterest
12. Losing or gaining 10 or more pounds when you were not trying to?	M1B9	29	Wghtloss/gain
13. Moving or talking slower than usual?	M1B10	30	MovSlower
<i>Homicidal Suicidal Thought Scale</i>			
14. Thought about killing or hurting someone else?	M1C1	31	HrtSomeone
15. Thought about ending your life or committing suicide?	M1C2	32	Suicide
16. Had a plan to commit suicide?	M1C3	33	PlanSuicide
17. Gotten a gun, pills or other things to carry out your plan?	M1C4	34	GotGun
18. Attempted to commit suicide?	M1C5	35	AtemtSuicide
<i>Anxiety /Fear Symptom Scale</i>			
19. Feeling very anxious, nervous, tense, fearful, scared, panicked or like something bad was going to happen?	M1D1	36	Anxious
20. Having to repeat an action over and over, or having thoughts that kept running over in your mind?	M1D2	37	Repeatover
21. Trembling, having your heart race or feeling so restless that you could not sit still?	M1D3	38	Trembling
22. Getting into a lot of arguments and feeling the urge to shout, throw things, beat, injure or harm someone?	M1D4	39	Arguments
23. Feeling very afraid of open spaces, leaving your home, having to travel or being in a crowd?	M1D5	40	FearOpen

<b>Subscale Name and Item Stem</b>	<b>GAIN Item Number</b>	<b>Rasch Output Item Number</b>	<b>Item Label</b>
24. Avoiding snakes, the dark, being alone, elevators or other things because they frightened you?	M1D6	41	SnakesDark
25. Thoughts that other people were taking advantage of you, not giving you credit or causing you problems?	M1D7	42	TakAdvantage
26. Thoughts that someone was watching you, following you or out to get you?	M1D8	43	OthrsWatch
27. Seeing or hearing things that no one else could see or hear, or feeling that someone else could read or control your thoughts?	M1D9	44	OthsNoSee
28. Thoughts that you should be punished for thinking about sex or other things too much?	M1D10	45	ShudBpunished
29. Having a lot of tension or muscle aches because you were worried?	M1D11	46	MuslAches
30. Being unable or finding it difficult to control your worries?	M1D12	47	Worries
<b><i>Traumatic Stress Scale</i></b>			
31. When something reminds you of the past, you became very distressed and upset?	M2A	48	RemindDistressed
32. You had nightmares about things in your past that really happened?	M2B	49	Nitmares
33. When you think of things you have done, you wish you were dead?	M2C	50	WishDead
34. It seemed as if you have no feelings?	M2D	51	NoFeelings
35. Your dreams at night are so real that you awaken in a cold sweat and force yourself to stay awake?	M2E	52	WakColdSweat
36. You felt like you could not go on?	M2F	53	CantGoOn
37. You were frightened by your urges?	M2G	54	FearUrges
38. Sometimes you used alcohol or other drugs to help yourself sleep or forget about things that happened in the past?	M2H	55	ODHlpSleep
39. You lost your cool and exploded over minor, everyday things?	MF2J	56	LostCool
40. You were afraid to go to sleep at night?	M2K	57	AfradSleep
41. You had a hard time expressing your feelings, even to the people you cared about?	M2M	58	NoXpressFeel
42. You felt guilty about things that happened because you felt like you should have done something to prevent them?	M2N	59	FeltGuilty
43. Had any of the above problems for three or more months?	M2P	60	AnyAboveProbs3mos

Note: Item M1B2 (no energy) was deleted.

## Data Source

Data on the 7435 cases reported in this paper came from 12 projects/programs including 70 sites from around the country.

All interviews were conducted by interviewers with three to four days of training followed by rigorous field-based certification procedures. Field interviewers had ongoing supervision by local trainers who were trained and certified by Chestnut staff on the use of the GAIN.

Full details about the *IMDS* may be obtained at the following: <http://www.chestnut.org/LI/gain/index.html>

## Rasch Analysis

The Rasch measurement model (Rasch, 1960) was chosen for this analysis because it is the only item response theory model that has the desirable scaling properties of linear, interval measurement (Embretson & Reise, 2000). Therefore, Rasch measures are the most valid for mathematical operations, such as correlation and regression analysis, as well for assessing change. Rather than tailor models to fit the data, the Rasch one parameter model fulfills the requirements of fundamental measurement (i.e., linear interval scale (Bond & Fox, 2007), and examines the data, i.e., items and persons, for flaws or problems that are indicated by their failure to fit the model.

*Quality control with fit statistics.* Rasch analysis provides fit statistics to test assumptions of fundamental measurement (Wright & Stone, 1979). "Fitting the model" simply means meeting basic assumptions of measurement, e.g., high scorers should endorse or get right almost all of the easy items. Once identified, persons and items that "misfit" can then be examined qualitatively to determine the causes of the problems. Problems may include items with confusing wording or items that assess a construct that is different from the principal one being measured, i.e., multidimensionality. Understanding poor fit can lead to improving or dropping items.

The fit of the data to the model is evaluated by fit statistics that are calculated for both persons and items. The following link provides a handy guide to interpreting fit statistics: <http://www.rasch.org/rmt/rmt82a.htm>. The Rasch model provides two indicators of misfit: infit and outfit. The infit is sensitive to unexpected behavior affecting responses to items near the person ability level and the outfit is outlier sensitive. Mean square fit statistics are defined such that the model-specified uniform value of randomness is 1.0 (Wright & Stone, 1979). Person fit indicates the extent to which the person's performance is consistent with the way the items are used by the other respondents. Item fit indicates the extent to which the use of a particular item is consistent with the way the sample respondents have responded to the other items. For this type of analysis, values between .75 and 1.33 are considered acceptable (Wilson, 2005). In addition to fit statistics, principal component analysis of residuals is used to examine whether a substantial factor exists in the residuals after the primary measurement dimension has been estimated (Linacre, 1998; Smith, 2002).

## Construct Validation

In Rasch analysis the item hierarchy that is created by the item difficulty estimates provides an indication of construct validity (Smith, 2001). The items should form a ladder of low severity symptoms on the bottom to high severity symptoms on the top.

In summary, the advantages of Rasch analysis are that:

- Standard errors differ across scores of items and persons, e.g., improved estimation of error in extreme scores.
- Enables shorter measures that are more reliable, e.g., eliminate bad items, and via computerized adaptive testing.
- Facilitates analysis of construct validity
- Enables comparable scoring across different measures, i.e., item and test equating.
- Unbiased estimates of item difficulties can be obtained from non-representative samples.
- Interval scale properties are achieved. How? Probabilities, or log odds, are used.
- Analysis of response category usefulness is enhanced.
- Analysis of person and item characteristics is enhanced through fit statistics.
- Enables analysis of item bias, a.k.a., differential item functioning
- Facets beyond persons and items that affect the measures may be estimated

For references to articles that illustrate the applications noted above, we recommend Conrad & Smith (2004). For a complete treatment of Rasch analysis, we recommend Bond & Fox (2007) which includes a glossary of Rasch measurement terminology. Terminology may also be accessed online via *Rasch Measurement Transactions* located at <http://www.rasch.org/rmt/>. The tables below are output from *Winsteps* (Linacre, 2007) with annotated explanations and interpretations.

## Background Characteristics of the Sample

The data for this analysis come from 7435 respondents who completed the CVS. The respondents were being screened for substance use disorders. In the previous year, 86% had substance use disorders, 51% had internalizing disorders (e.g., somatic, depression, anxiety, trauma, suicide), 59% had externalizing disorders (i.e., attention-deficit or hyperactivity disorders, and conduct disorders), and 59% had problems with crime or violence. Approximately, 42% were entering residential treatment and 66% were involved in the criminal justice system.

As shown in the following table, the sample was predominately under 18 years of age (73%) and male (67%). Almost half were Caucasian (45%), a quarter were African American (26%), and the remainder Hispanic or mixed race. Of the top five primary drugs reported, marijuana was reported by 49% of the sample. The drug least often reported was opioids at 5%. Other drugs reported included amphetamines (11%), cocaine (11%), and alcohol (20.5%). Almost 3% percent of the sample reported other drugs.

**Table 2. Demographic Characteristics of the Sample (N=7435<sup>a</sup>)**

	Percent	Number
<b>Age, Mean (sd) 19.9 (8.9)</b>		
< 18 years	72.5	5388
≥18 years	27.5	2047
<b>Gender</b>		
Male	67.1	4992
Female	32.7	2437
<b>Race</b>		
African American	25.7	1913
Caucasian	45.2	3360
Hispanic	10.8	806
Mixed/other	17.7	1314
<b>Drug, primary, most severe</b>		
Alcohol	20.5	1527
Amphetamines	11.0	820
Marijuana	49.1	3654
Cocaine	10.9	808
Opiates	5.3	393
Other drug	2.9	214

<sup>a</sup> Numbers may not add up to 100% due to missing values

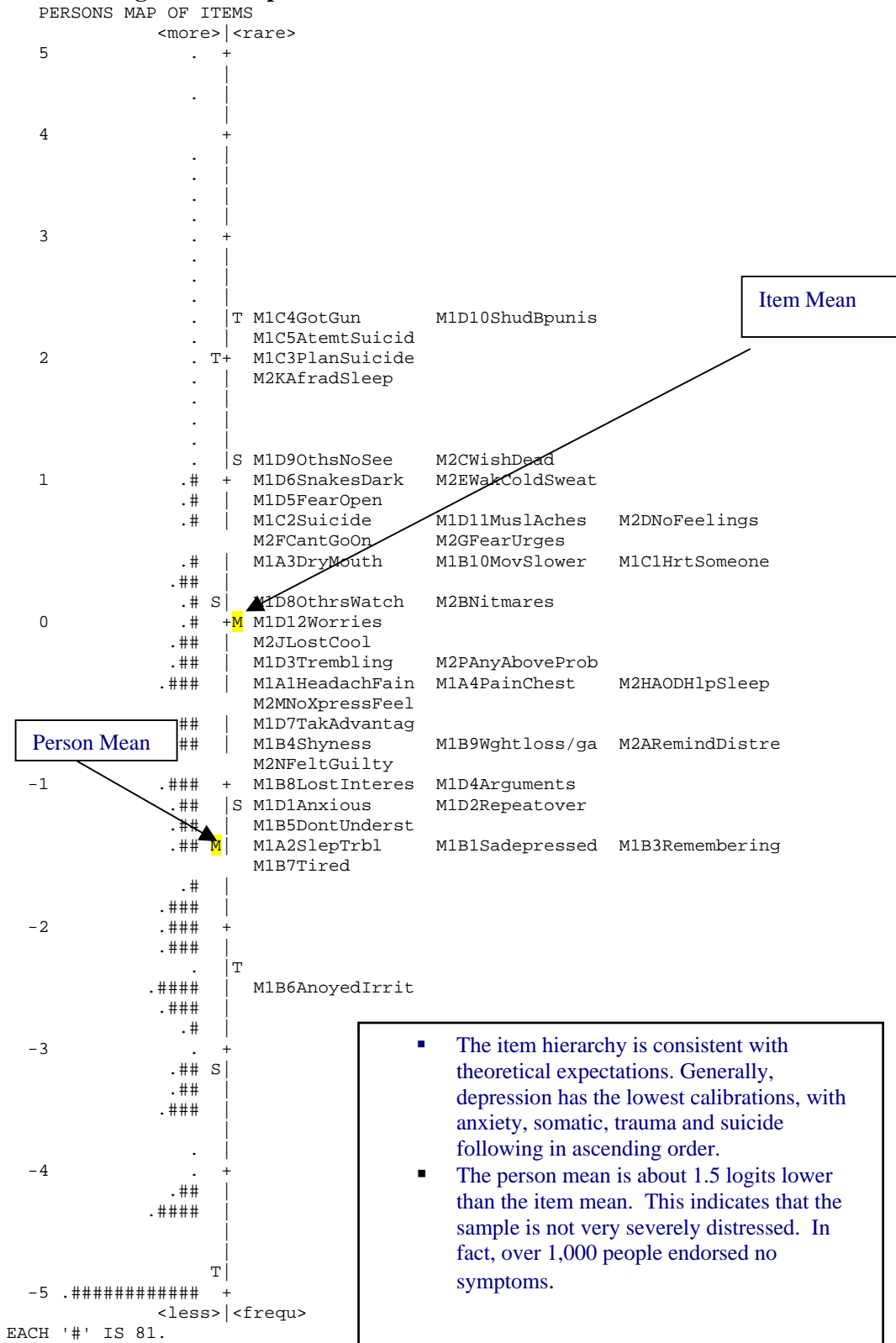
**TABLE 3. Person and Item Reliability**

SUMMARY OF 6353 MEASURED (NON-EXTREME) PERSONS									
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	11.8	40.5	-1.43	.51	1.00	.1	.98	.1	
S.D.	9.3	3.3	1.68	.20	.18	.9	.64	1.0	
MAX.	42.0	43.0	4.37	1.05	1.79	4.4	9.90	4.8	
MIN.	1.0	16.0	-4.29	.35	.41	-3.4	.16	-2.8	
REAL RMSE	.56	ADJ.SD	1.58	SEPARATION	2.83	PERSON RELIABILITY	.89		
MODEL RMSE	.55	ADJ.SD	1.58	SEPARATION	2.91	PERSON RELIABILITY	.89		
S.E. OF PERSON MEAN = .02									
MAXIMUM EXTREME SCORE: 2 PERSONS									
MINIMUM EXTREME SCORE: 1041 PERSONS									
LACKING RESPONSES: 30 PERSONS									
DELETED: 9 PERSONS									
VALID RESPONSES: 94.1%									
SUMMARY OF 7396 MEASURED (EXTREME AND NON-EXTREME) PERSONS									
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	10.2	40.5	-2.00	.70					
S.D.	9.6	3.4	2.09	.50					
MAX.	42.0	43.0	5.55	1.86					
MIN.	.0	13.0	-5.53	.35					
REAL RMSE	.86	ADJ.SD	1.91	SEPARATION	2.21	PERSON RELIABILITY	.83		
MODEL RMSE	.86	ADJ.SD	1.91	SEPARATION	2.23	PERSON RELIABILITY	.83		
S.E. OF PERSON MEAN = .02									
PERSON RAW SCORE-TO-MEASURE CORRELATION = .94 (approximate due to missing data)									
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .94 (approximate due to missing data)									
SUMMARY OF 43 MEASURED (NON-EXTREME) ITEMS									
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	1748.9	5979.4	.00	.04	1.00	-.2	.98	-1.1	
S.D.	914.7	861.8	1.16	.01	.14	6.7	.33	6.2	
MAX.	4172.0	6351.0	2.36	.06	1.31	9.9	2.03	9.9	
MIN.	396.0	3832.0	-2.46	.03	.77	-9.9	.49	-9.9	
REAL RMSE	.04	ADJ.SD	1.16	SEPARATION	28.93	ITEM RELIABILITY	1.00		
MODEL RMSE	.04	ADJ.SD	1.16	SEPARATION	29.80	ITEM RELIABILITY	1.00		
S.E. OF ITEM MEAN = .18									

- The 43 item scale has high person reliability of .89.
- Including extreme scores, the reliability is lower (.83) because Rasch estimates the error as higher for extreme scores.
- Cronbach's alpha is higher (.94) because it estimates extreme scores as measured perfectly, i.e., with no error.
- A separation value of 2.91 gives approximately two separation levels, thus splitting the persons into about 3 groups on the Rasch ruler (Table 3).

- Good item reliability of 1.00.
- Item separation is high at 29.80 meaning that items are placed reliably on the ruler.

**Table 4. Wright Item Map**



**Table 5. Principal Components Analysis of Standardized Residual Correlations for Items**

CONTRAST 1 FROM PRINCIPAL COMPONENT ANALYSIS OF  
 Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --	Modeled
Total raw variance in observations	=	70.6 100.0%	100.0%
Raw variance explained by measures	=	27.6 39.1%	37.8%
Raw variance explained by persons	=	18.2 25.8%	25.0%
Raw Variance explained by items	=	9.4 13.3%	12.8%
Raw unexplained variance (total)	=	43.0 60.9% 100.0%	62.2%
Unexplned variance in 1st contrast	=	4.0 5.7% 9.3%	

STANDARDIZED RESIDUAL LOADINGS FOR ITEMS (SORTED BY LOADING)

CON-TRAST	LOADING	MEASURE	INFIT MNSQ	OUTFIT MNSQ	ENTRY NUMBER	ENTRY ITEM	LOADING	MEASURE	INFIT MNSQ	OUTFIT MNSQ	ENTRY NUMBER	ENTRY ITEM
1	.69	-.82	.83	.73	A 59	M2NFeltGuilty	-.25	-1.47	1.06	1.10	a 23	M1B3Remembering
1	.66	-.82	.84	.74	B 48	M2AREmindDistressed	-.23	-.36	.93	.85	b 38	M1D3Trembling
1	.63	-.53	.82	.69	C 58	M2MNoXpressFeel	-.23	-.47	1.19	1.37	c 17	M1A1HeadachFaint
1	.59	-.48	.85	.72	D 55	M2HAODHlpSleep	-.22	-1.52	1.04	1.02	d 27	M1B7Tired
1	.57	-.25	.83	.67	E 60	M2PAnyAboveProbs3mos	-.22	-.50	1.17	1.31	e 20	M1A4PainChest
1	.49	-.12	.84	.68	F 56	M2JLostCool	-.21	.42	1.16	1.34	f 19	M1A3DryMouth
1	.42	.25	.93	.83	G 49	M2BNitmares	-.20	-.92	1.11	1.13	g 24	M1B4Shyness
1	.39	.74	.77	.49	H 53	M2FCantGoOn	-.20	-1.58	1.05	1.03	h 18	M1A2SlepTrbl
1	.38	.59	.92	.69	I 51	M2DNoFeelings	-.20	-1.08	1.16	1.20	i 39	M1D4Arguments
1	.37	.75	.78	.52	J 54	M2GFearUrges	-.20	-1.24	.91	.85	j 36	M1D1Anxious
1	.32	1.17	.86	.59	K 50	M2CWishDead	-.19	.44	1.06	1.12	k 30	M1B10MovSlower
1	.32	1.04	.87	.58	L 52	M2EWakColdSweat	-.19	-.65	1.04	1.04	l 42	M1D7TakAdvantage
1	.25	1.79	.87	.56	M 57	M2KAfradSleep	-.18	-1.10	.96	.89	m 37	M1D2Repeatover
							-.18	-1.01	1.08	1.09	n 28	M1B8LostInterest
							-.18	.85	1.10	1.11	p 40	M1D5FearOpen
							-.18	.12	1.10	1.18	o 43	M1D8OthrsWatch
							-.16	-1.32	1.00	.99	q 25	M1B5DontUnderstand
							-.15	-2.46	1.18	1.33	r 26	M1B6AnoyedIrritated
							-.15	1.25	1.05	1.00	s 44	M1D9OthsNoSee
							-.13	-.92	1.28	1.61	t 29	M1B9Wghtloss/gain
							-.13	.66	.95	.75	u 46	M1D11MuslAches
							-.13	-1.44	.90	.82	v 21	M1B1Sadepressed
							-.12	2.29	1.09	1.20	U 45	M1D10ShudBpunished
							-.11	1.05	1.22	2.04	T 41	M1D6SnakesDark
							-.11	.43	1.31	1.80	S 31	M1C1HrtSomeone
							-.11	2.14	1.08	1.04	R 35	M1C5AtemtSuicide
							-.11	.69	1.05	1.09	Q 32	M1C2Suicide
							-.09	1.99	1.06	.89	P 33	M1C3PlanSuicide
							-.09	2.37	1.03	.81	O 34	M1C4GotGun
							-.08	.04	.86	.68	N 47	M1D12Worries

- To judge the strength of the first dimension, we used the following guidelines for variance explained by a measurement dimension,  $\geq 40\%$  is considered a moderate measurement dimension, and  $\geq 20\%$  is considered a minimal dimension. The 2nd dimension taken from Reckase (1979) is  $\geq 20\%$ .
- The variance explained by the first factor is 39.1%.
- The fact that only 9.3% of the variance is explained by the first factor supports unidimensionality.
- Thus, the IMDS is considered unidimensional using these guidelines.
- The PCA was conducted using a Kaiser-Meyer-Olkin (KMO) of 0.368.

**Table 6. Most Misfitting Response Strings in Terms of OUTMNSQ**

ITEM	OUTMNSQ	PERSON
		6644332612541 732533777763221365544444444333333
		9561513436663 109588331090209336196444332299842273
		1976329334993 326680971437258988470877947031858790
		48122061234257252000476517181075907884906960174705
		-----high-----
41 M1D6SnakesDar	2.03	A .....11...11...11.11..11.11.....1.
31 M1C1HrtSomeon	1.80	B .11...1.....
17 M1A1HeadachFa	1.37	D .1.1.....1.....
20 M1A4PainChest	1.31	G .....1.....
39 M1D4Arguments	1.21	H ....1...1.....
45 M1D10ShudBpun	1.20	I 11.111...1...1.....1.....
43 M1D8OthrsWatc	1.18	J .....1.....1.....
24 M1B4Shyness	1.12	K .....1.....
30 M1B10MovSlowe	1.12	L ..... .. 11..1.....11..
40 M1D5FearOpen	1.11	M .....1.....1.....1.....
32 M1C2Suicide	1.10	O .1...1111.111.1...1.1...1...1...1...
28 M1B8LostInter	1.10	P 1..1. . . . .
35 M1C5AtemtSuic	1.04	Q .1...1.111.1.....
33 M1C3PlanSuici	.90	R .....111..11.....
18 M1A2SlepTrbl	1.03	S .....1.....
44 M1D9OthsNoSee	1.00	T .....11...1.....1.....
42 M1D7TakAdvant	1.03	U .....1.....
34 M1C4GotGun	.81	u .1...1.....
37 M1D2Repeatove	.89	s .....1.....
		-----low-----
		664433261254177325337777632213655444444433333373
		9561513436663 109588331090209336196444332299842290
		1976329334993 326680971437258988470877947031858705
		4812206123425 2520004765171810759078849069601747

This table shows the most misfitting items in terms of OUTMNSQ. Here we see the persons who unexpectedly endorsed an item given their overall measure. “Fear of snakes” has the highest outfit (2.03) because this item (item measure = 1.05 logits) was endorsed by persons with very low measures. For example, Person #790 on the far right endorsed the item “Fear of snakes” but had a very low overall measure of -4.29. (See Table 8). The box at the upper left of the table shows the three items with the high outfits.

**Table 7. Most Unexpected Person Responses in Terms of Measure**

ITEM	MEASURE	PERSON
		6644332612541 73253377763221365544444444333333
		9561513436663 109588331090209336196444332299842273
		1976329334993 326680971437258988470877947031858790
		48122061234257252000476517181075907884906960174705
	high-----	
18 M1A2SlepTrbl	-1.58 S	.....1.....
37 M1D2Repeatove	-1.10 s	.....1.....
39 M1D4Arguments	-1.08 H	...1...1.....
28 M1B8LostInter	-1.00 P	1..1. . . . .
24 M1B4Shyness	-.91 K	.....1.....
48 M2ARemindDist	-.82 f	.....1.....1.....
59 M2NFeltGuilty	-.82 d	.....1.....1.....
42 M1D7TakAdvant	-.65 U	.....1.....
20 M1A4PainChest	-.50 G	...1.....
17 M1A1HeadachFa	-.47 D	..1.1.....1.....
38 M1D3Trembling	-.37 p	11.....
43 M1D8OthrsWatc	.12 J	.....1.....1.....
49 M2BNitmares	.25 q	.....1.....
31 M1C1HrtSomeone	.43 B	..11...1.....
30 M1B10MovSlowe	.45 L	..... . . . . 11..1.....11..
32 M1C2Suicide	.69 O	..1...1111.111.1.....1.1... ..1...1...1...1...
40 M1D5FearOpen	.85 M	.....1.....1.....1.....1.....1.....1.....1.....1.....
41 M1D6SnakesDar	1.05 A	.....11...11...11.11.11.11.....1.
50 M2CWishDead	1.17 i	.....1.....
44 M1D9OthsNoSee	1.24 T	.....11...1.....1.....
57 M2KAfradSleep	1.79 k	.....1.1.....
33 M1C3PlanSuici	1.99 R	.....111.11 .....
35 M1C5AtemtSuic	2.14 Q	..1.....1.111.1.....
45 M1D10ShudBpun	2.30 I	11.111...1...1.....1.....
34 M1C4GotGun	2.36 u	..1...1..... . . . .
	low-----	
		664433261254177325337776322136554444444433333373
		9561513436663 109588331090209336196444332299842290
		1976329334993 326680971437258988470877947031858705
		4812206123425 2520004765171810759078849069601747

This table shows the item measures and the persons with the most unexpected responses.

- Person #790 has a low measure of -4.29 (see Table 8), but endorsed Item 41, "Snakes."
- The large number of 1's in the "Snakes" indicate that this atypical response has occurred several times.

**Table 8. Person Statistics: Misfit Order (The top listed persons)**

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEA CORR.	EXACT OBS%	MATCH EXP%	PERSON	DMOGRAPHIC CODES
1981	1	37	-4.11	1.03	1.13	.4	9.90	3.7	A-.31	97.3	97.3	1981 3	1 0 4 1981
1335	3	43	-3.09	.62	1.28	.7	9.15	3.7	B-.38	93.0	93.0	1335 3	2 0 4 1335
6314	3	43	-3.09	.62	1.28	.7	9.15	3.7	C-.38	93.0	93.0	4692 3	1 0 4 6314
7	2	34	-3.33	.75	1.24	.6	8.72	3.1	D-.33	94.1	94.1	7 3	2 0 2 7
3066	2	37	-3.35	.75	1.23	.6	8.15	2.8	E-.27	94.6	94.6	2962 3	1 0 2 3066
3129	2	37	-3.35	.75	.92	.1	7.76	2.8	F-.03	94.6	94.6	3025 3	1 1 1 3129
2396	4	37	-2.53	.56	1.44	1.1	7.44	3.8	G-.38	89.2	89.2	2396 3	1 0 1 2396
6293	5	43	-2.47	.50	1.40	1.2	6.57	3.9	H-.32	88.4	88.4	4671 3	1 0 4 6293
2511	6	37	-2.00	.48	1.26	1.0	5.92	4.2	I-.18	86.5	84.4	2511 3	1 0 2 2511
4014	1	43	-4.29	1.03	1.10	.4	5.92	2.2	J-.17	97.7	97.7	3910 3	1 0 1 4014
291	7	43	-2.03	.44	1.44	1.6	5.75	4.5	K-.30	81.4	84.2	291 1	2 0 2 291
5253	1	37	-4.11	1.03	1.12	.4	5.74	2.2	L-.16	97.3	97.3	6931 4	2 1 1 5253
2559	5	37	-2.24	.51	1.21	.7	5.70	3.6	M-.12	89.2	86.8	2559 3	1 0 2 2559
7316	3	43	-3.09	.62	1.07	.3	5.40	2.6	N-.03	93.0	93.0	5694 3	1 0 2 7316
2747	3	37	-2.88	.63	1.25	.7	5.35	2.6	O-.14	91.9	91.9	2643 3	1 0 3 2747
5808	1	37	-4.11	1.03	1.12	.4	5.33	2.1	P-.15	97.3	97.3	7116 2	1 1 1 5808
7182	2	43	-3.54	.74	1.13	.4	5.25	2.2	Q-.09	95.3	95.3	5560 2	2 0 3 7182
4605	5	36	-2.19	.52	1.28	.9	5.23	3.4	R-.11	83.3	86.5	6283 4	1 1 1 4605
5824	2	37	-3.35	.75	1.16	.5	5.04	2.1	S-.08	94.6	94.6	7132 1	1 1 2 5824
149	7	37	-1.78	.46	1.38	1.5	4.95	4.1	T-.17	78.4	81.9	149 1	2 0 2 149
2004	4	37	-2.53	.56	1.34	.9	4.91	2.8	U-.19	89.2	89.2	2004 1	1 0 4 2004
790	1	43	-4.29	1.03	1.09	.4	4.87	2.0	V-.14	97.7	97.7	790 3	1 1 4 790
4040	1	43	-4.29	1.03	1.09	.4	4.87	2.0	W-.14	97.7	97.7	3936 2	2 0 2 4040
5431	1	43	-4.29	1.03	1.09	.4	4.87	2.0	X-.14	97.7	97.7	4209 3	1 0 2 5431
5562	1	43	-4.29	1.03	1.09	.4	4.87	2.0	Y-.14	97.7	97.7	4340 1	1 0 2 5562
5621	1	43	-4.29	1.03	1.09	.4	4.87	2.0	Z-.14	97.7	97.7	4399 2	1 0 2 5621
1840	34	37	3.06	.64	1.34	.8	4.84	2.4	-.17	91.9	91.9	1840 1	2 0 3 1840
2171	6	37	-2.00	.48	1.32	1.2	4.66	3.5	-.10	81.1	84.4	2171 3	1 0 2 2171
2295	6	37	-2.00	.48	1.40	1.4	4.63	3.5	-.17	81.1	84.4	2295 3	1 0 2 2295
230	8	43	-1.84	.42	1.31	1.3	4.61	4.2	-.13	79.1	82.1	230 4	2 0 2 230
6475	8	43	-1.84	.42	1.30	1.3	4.59	4.1	-.12	79.1	82.1	4853 3	2 0 2 6475
6306	8	43	-1.84	.42	1.28	1.2	4.58	4.1	-.10	79.1	82.1	4684 2	2 0 4 6306
5482	7	43	-2.03	.44	1.48	1.7	4.37	3.6	-.23	81.4	84.2	4260 1	1 0 4 5482
7229	6	43	-2.24	.47	1.32	1.1	4.37	3.3	-.13	83.7	86.3	5607 3	2 0 3 7229
1559	8	37	-1.58	.44	1.45	1.9	4.35	4.1	-.18	75.7	79.5	1559 5	1 1 4 1559
1610	6	37	-2.00	.48	1.30	1.1	4.30	3.3	-.12	86.5	84.4	1610 3	2 0 1 1610
65	3	37	-2.88	.63	1.20	.6	4.13	2.2	-.08	91.9	91.9	65 1	1 0 2 65

**Table 9. Persons with Most Misfitting Response Strings in Terms of OUTMNSQ**

PERSON	OUTMNSQ	ITEM
		2122223332224545251365444133543554545453343
		6873156798498928057806739910162340210473554
	high-----	
1981 1981 1043	9.90 A	.....1.
1335 1335 2046	9.16 B	.....11..
4692 4692 1043	9.16 C	.....11..
7 7 2023	8.73 D	.....1.
2962 2962 1023	8.16 E	.....1..
3025 3025 1113	7.77 F	.....1.
2396 2396 1013	7.42 G	.....1.1
4671 4671 1043	6.55 H	.....1.1
2511 2511 1023	5.91 I	.....11.1
3910 3910 1013	5.90 J	.....1....
291 291 2021	5.74 K	.....11.1
6931 6931 2114	5.73 L	.....1....
2559 2559 1023	5.68 M	.....1.1
5694 5694 1026	5.40 N	.....1.
2643 2643 1033	5.35 O	.....1.
7116 7116 1112	5.33 P	.....1....
6283 6283 1114	5.22 Q	.....1.1
5560 5560 2032	5.21 R	.....1....
7132 7132 1121	5.01 S	.....1....
149 149 2021	4.94 T	.....11.1
2004 2004 1041	4.91 U	.....1.1.
790 790 1143	4.86 V	.....1.....
3936 3936 2022	4.86 W	.....1....
4209 4209 1023	4.86 X	.....1....
4340 4340 1021	4.86 Y	.....1....
4399 4399 1022	4.86 Z	.....1....
	low-----	
		2122223332224545251365444133543554545453343
		6873156798498928057806739910162340210473554

This table shows the unexpected responses of persons with the highest outfit mean squares (MNSQ)s. Note, Person #790 again.

We see that the most misfitting response strings for persons were usually caused by the suicidal ideation items: #32 Suicide, #33 Plan Suicide, #34 Got Gun, and #35 Attempt Suicide.



**Table 11. Summary of Category Structure**

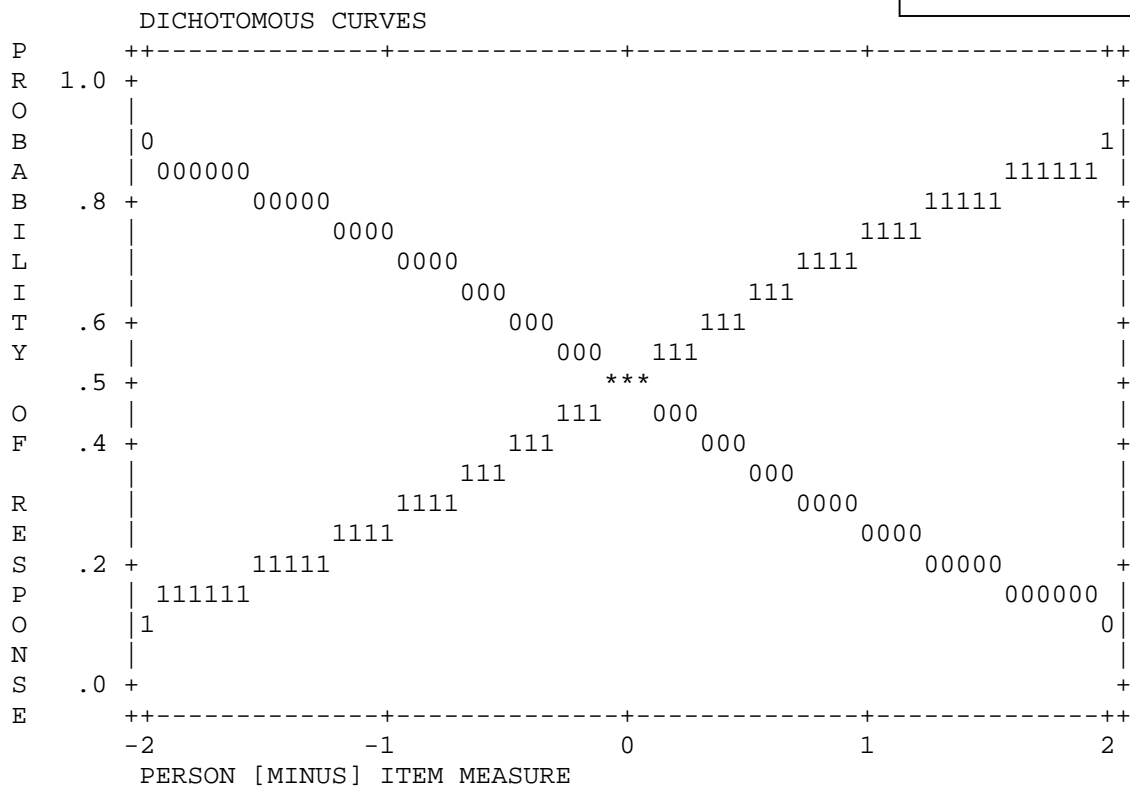
CATEGORY LABEL	OBSERVED SCORE	OBSERVED COUNT	OBSERVED %	SAMPLE AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	COHERENCE		ESTIM DISCR
								M->C	C->M	
0	0	180920	67	-2.24	-2.24	1.00	1.01	84%	90%	0
1	1	74785	28	.40	.40	1.00	.96	72%	60%	1
MISSING		16012	6	-.90						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

M->C = Does Measure imply Category?

C->M = Does Category imply Measure?

67% of all responses were zeroes (i.e., did not endorse items).



## Differential Item Functioning (DIF) for Age, Gender, Race, and Primary Drug Severity for the Internal Mental Distress Scales

As Bond and Fox (2007) note, the Rasch model requires that relative item estimates (i.e., item difficulty estimates) remain invariant across subgroups of persons (e.g., females and males). DIF allows us to examine whether items have significantly different meanings for different groups. The authors suggest that items that show DIF should be investigated to determine what may be inferred about the underlying construct and what that implies about the samples of persons detected. A significant DIF contrast is based on  $\geq 0.58$  logit difference for all comparisons which is approximately half a standard deviation (Table 3, bottom panel) for the items (Norman, Sloan, & Wyrwich, 2003; Conrad, Dennis, Bezruczko, Funk, & Riley, 2007).

The figures below present easily interpretable graphs of the relationships of the various groups on the *IMDS* items. Table 12 contains the data that formed these graphs, and it contains the information to compute the differences between groups on each item. For example, to get the DIF contrast between males and females on *MIB6AnoyedIrritated*, subtract  $-2.55 - (-2.26) = -.29$ .

*Gender DIF.* In Figure 1, we can see that it was significantly easier for females to endorse *Sadepressed*. Likewise, for males it was significantly easier to endorse *OthrsWatch*, a paranoia symptom, and *HrtSomeone*. While these were clearly different response patterns for these three items by gender, overall the DIF was not particularly strong since only three items were involved and they tended to balance out the overall differences between males and females. Therefore, there would not tend to be bias in the total *IMDS* measures, i.e., total scores.

*Age DIF.* In Figure 2, however, there are much larger differences, and they tend to be balanced. *Sadepressed*, *Worries*, *DryMouth*, *MuslAches*, and *FearUrges* are all much, i.e.,  $\geq .58$  logit easier for adults to endorse. Concurrently, there are six items, that are  $\geq .58$  logit easier for youth to endorse than it is for adults, i.e., *AnoyedIrritated*, *Arguments*, *HrtSomeone*, *PlanSuicide*, *AtemtSuicide*, and *GotGun*. While this equality in the numbers of significant DIF items will tend to balance the *IMDS* so that the test will be unbiased for youth compared to adults, the differences are interesting theoretically. Specifically, adults will tend to endorse neurotic and somatic symptoms more readily than will youth. Youth will tend to endorse suicidal ideation much more readily than will adults. Please see Conrad, Bezruczko, Park, Chan, Riley, Diamond, & Dennis (2008) for a complete treatment of this issue.

*Race DIF.* In Figure 3, we did not observe very much DIF among the four racial/ethnic groups. Of note was the difference for African Americans whereby items indicating suicidal ideation were more difficult to endorse.

*Primary Drug Severity DIF.* With six different types of drugs, it appears that opiates and cocaine are the easiest to endorse for the lower and mid severity items such as *Sadepressed*, *Wghtloss/gain*, and *Worries*. At the high severity end, we see amphetamines and other drugs being easiest to endorse, i.e., *AfradSleep* and *PlanSuicide*. In other words, persons with opiates and cocaine as the primary drug find the low and mid severity items to be somewhat easier to endorse while the users of amphetamines and other drugs find it easier to endorse the higher severity items. However, the interactions here are quite complex and are the subject of a paper in progress (Dennis, Conrad, & Chan, 2008). This paper indicates that severity hierarchies do differ by drug, but that most of the variation in severity is actually due to the drugs themselves.

Figure 1. IMDS Gender DIF

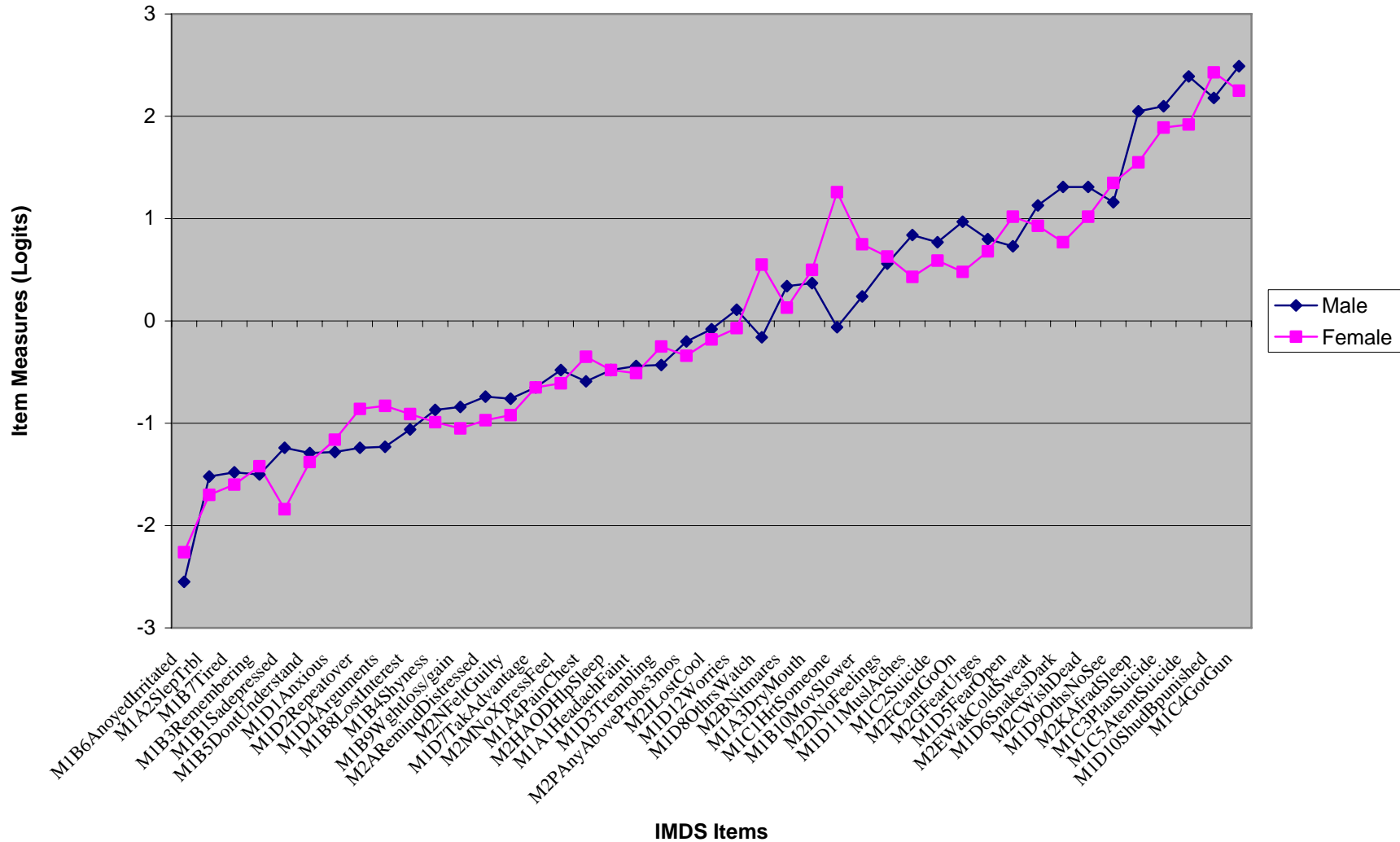


Figure 2. IMDS Youth vs. Adult DIF

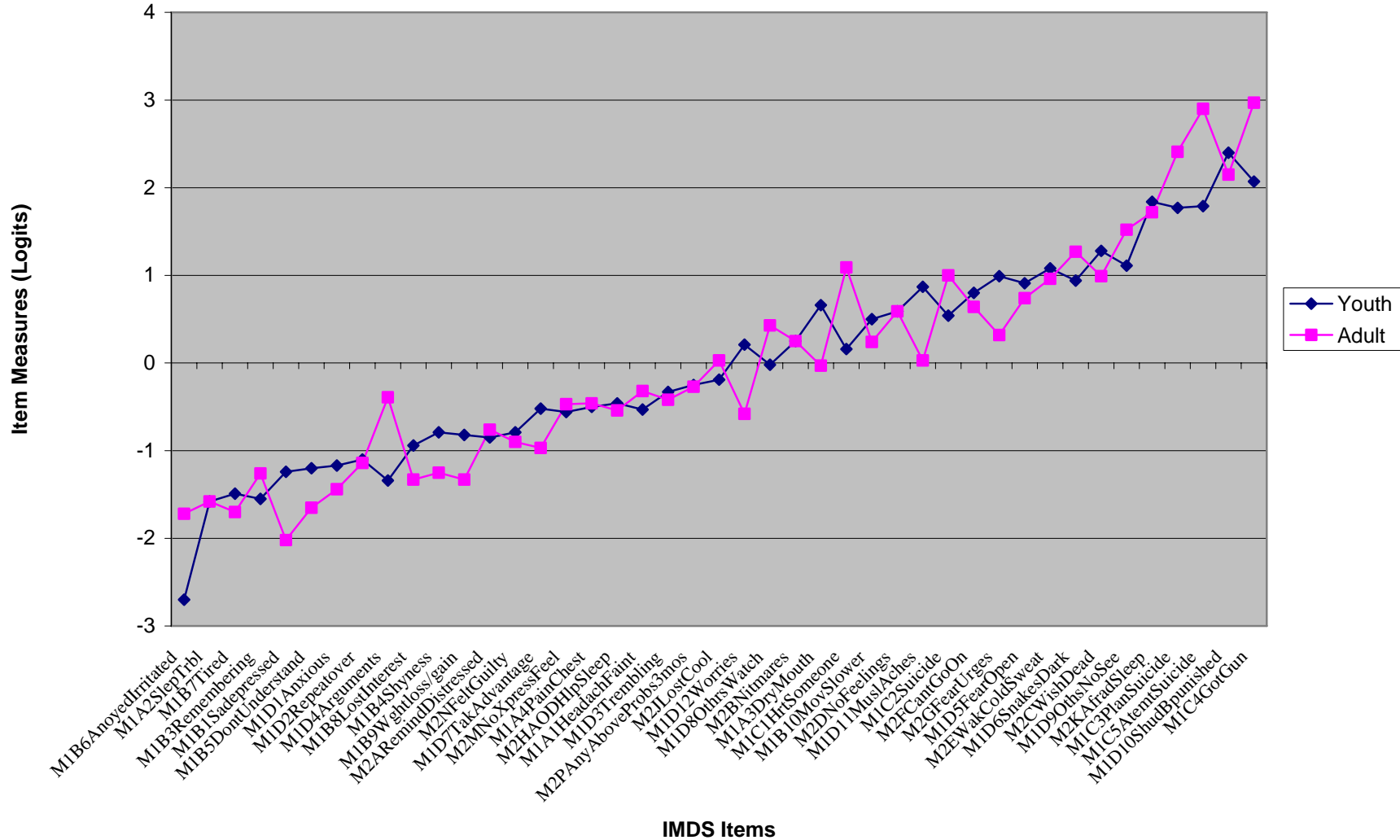


Figure 3. IMDS Race DIF

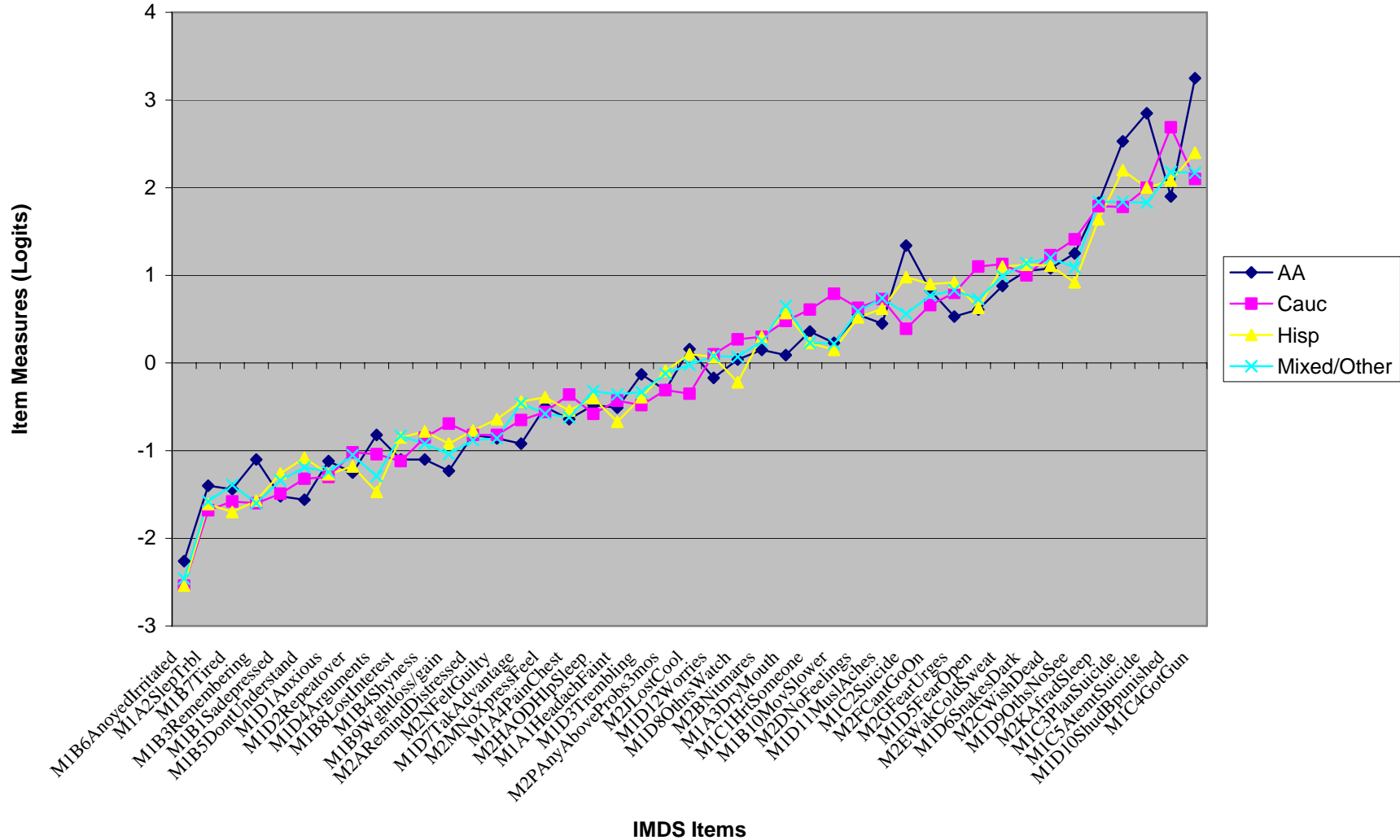
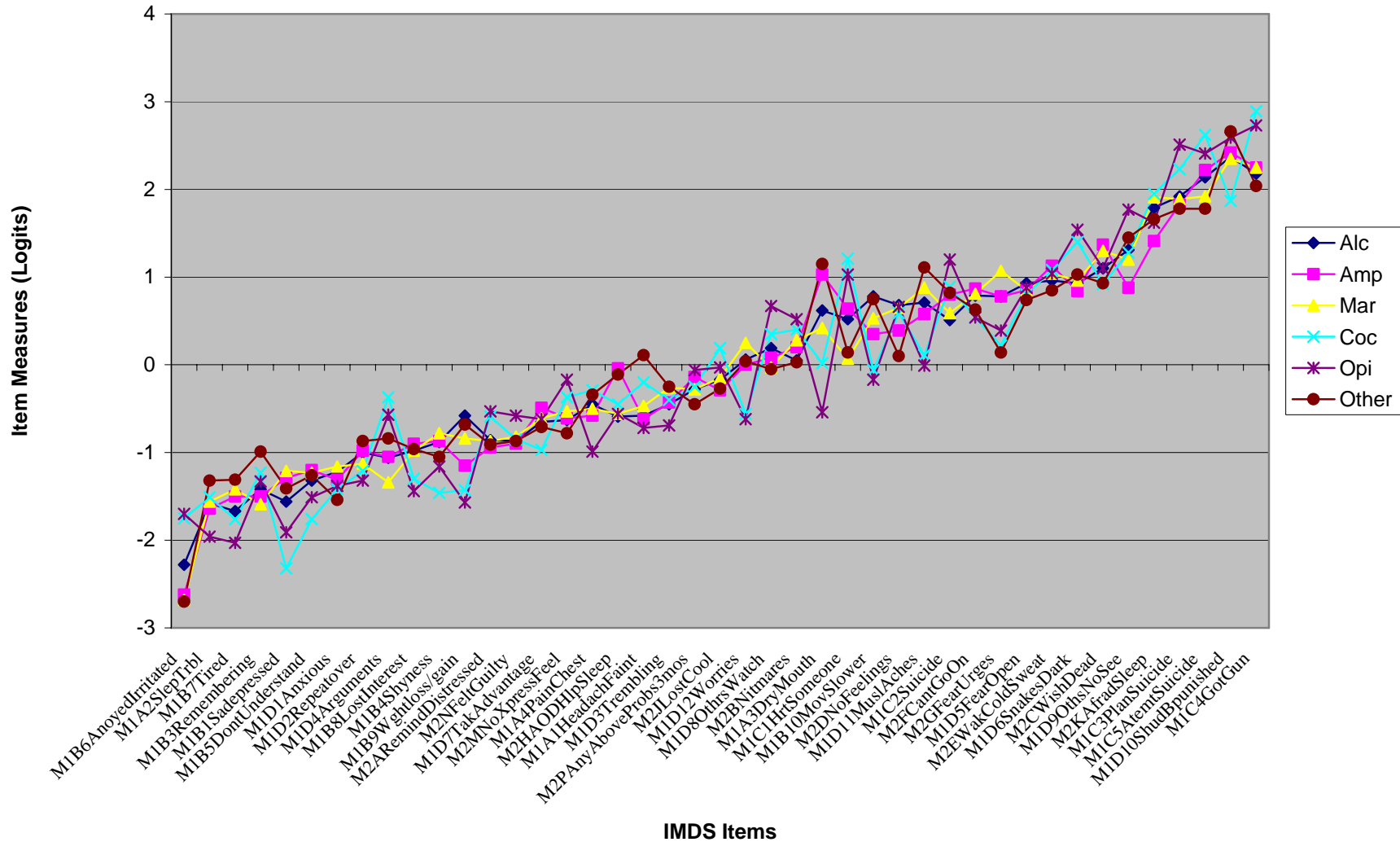


Figure 4. IMDS Primary Drug Severity



**Table 12. Item Measures by Demographic Groups (Items Listed in Severity Order)**

Entry	ITEM	Male	Female	Youth	Adult	AA	Cauc	Hisp	Mixed/ Other	Alc	Amp	Mar	Coc	Opi	Other
26	M1B6AnoyedIrritated	-2.55	-2.26	-2.7	-1.72	-2.26	-2.54	-2.54	-2.46	-2.28	-2.62	-2.68	-1.75	-1.7	-2.7
18	M1A2SlepTrbl	-1.52	-1.7	-1.58	-1.58	-1.4	-1.68	-1.61	-1.58	-1.58	-1.64	-1.56	-1.51	-1.96	-1.32
27	M1B7Tired	-1.48	-1.6	-1.49	-1.7	-1.44	-1.58	-1.7	-1.39	-1.67	-1.5	-1.42	-1.76	-2.03	-1.31
23	M1B3Remembering	-1.5	-1.42	-1.55	-1.26	-1.1	-1.6	-1.57	-1.6	-1.42	-1.5	-1.59	-1.23	-1.33	-0.99
21	M1B1Sadepressed	-1.24	-1.84	-1.24	-2.02	-1.52	-1.49	-1.26	-1.34	-1.56	-1.29	-1.21	-2.32	-1.91	-1.41
25	M1B5DontUnderstand	-1.29	-1.38	-1.2	-1.65	-1.56	-1.32	-1.08	-1.19	-1.32	-1.2	-1.23	-1.76	-1.51	-1.26
36	M1D1Anxious	-1.28	-1.16	-1.17	-1.44	-1.12	-1.3	-1.27	-1.24	-1.22	-1.31	-1.16	-1.42	-1.38	-1.54
37	M1D2Repeatover	-1.24	-0.86	-1.1	-1.14	-1.25	-1.02	-1.18	-1.05	-1	-0.98	-1.13	-1.21	-1.32	-0.87
39	M1D4Arguments	-1.23	-0.83	-1.34	-0.39	-0.82	-1.04	-1.47	-1.29	-1.06	-1.05	-1.34	-0.37	-0.57	-0.84
28	M1B8LostInterest	-1.06	-0.91	-0.94	-1.33	-1.1	-1.12	-0.85	-0.83	-0.98	-0.9	-0.98	-1.3	-1.44	-0.96
24	M1B4Shyness	-0.87	-0.99	-0.79	-1.25	-1.1	-0.85	-0.78	-0.92	-0.88	-0.87	-0.78	-1.46	-1.16	-1.05
29	M1B9Wghtloss/gain	-0.84	-1.05	-0.82	-1.33	-1.23	-0.69	-0.92	-1.04	-0.58	-1.15	-0.84	-1.43	-1.57	-0.68
48	M2ARemindDistressed	-0.74	-0.97	-0.85	-0.76	-0.82	-0.82	-0.77	-0.88	-0.86	-0.94	-0.87	-0.59	-0.53	-0.91
59	M2NFeltGuilty	-0.76	-0.92	-0.79	-0.9	-0.86	-0.82	-0.64	-0.86	-0.87	-0.9	-0.82	-0.85	-0.58	-0.87
42	M1D7TakAdvantage	-0.65	-0.65	-0.52	-0.97	-0.92	-0.65	-0.44	-0.46	-0.65	-0.49	-0.61	-0.97	-0.62	-0.71
58	M2MNoXpressFeel	-0.48	-0.61	-0.56	-0.47	-0.5	-0.56	-0.39	-0.58	-0.63	-0.61	-0.53	-0.37	-0.17	-0.78
20	M1A4PainChest	-0.59	-0.35	-0.5	-0.46	-0.64	-0.36	-0.54	-0.62	-0.45	-0.58	-0.5	-0.29	-0.99	-0.34
55	M2HAODHlpSleep	-0.48	-0.48	-0.46	-0.54	-0.48	-0.58	-0.4	-0.32	-0.59	-0.04	-0.56	-0.45	-0.56	-0.11
17	M1A1HeadachFaint	-0.44	-0.51	-0.53	-0.32	-0.51	-0.43	-0.67	-0.36	-0.58	-0.62	-0.47	-0.2	-0.72	0.11
38	M1D3Trembling	-0.43	-0.25	-0.33	-0.42	-0.13	-0.48	-0.39	-0.33	-0.45	-0.43	-0.26	-0.4	-0.69	-0.25
60	M2PAnyAboveProbs3mos	-0.2	-0.34	-0.25	-0.27	-0.31	-0.31	-0.09	-0.12	-0.28	-0.14	-0.28	-0.25	-0.06	-0.45
56	M2JLostCool	-0.08	-0.18	-0.19	0.03	0.16	-0.35	0.1	-0.02	-0.17	-0.29	-0.15	0.19	-0.03	-0.27

**Table 12. Item Measures by Demographic Groups (Items Listed in Severity Order)**

Entry	ITEM	Male	Female	Youth	Adult	AA	Cauc	Hisp	Mixed/ Other	Alc	Amp	Mar	Coc	Opi	Other
47	M1D12Worries	0.11	-0.07	0.21	-0.58	-0.17	0.1	0.07	0.08	0.06	0	0.25	-0.56	-0.62	0.04
43	M1D8OthrsWatch	-0.16	0.55	-0.02	0.43	0.04	0.27	-0.22	0.07	0.19	0.08	-0.03	0.35	0.67	-0.05
49	M2BNitmares	0.34	0.13	0.25	0.25	0.15	0.3	0.3	0.25	0.05	0.2	0.28	0.4	0.52	0.03
19	M1A3DryMouth	0.37	0.5	0.66	-0.03	0.09	0.48	0.57	0.65	0.62	1.03	0.42	0.02	-0.54	1.15
31	M1C1HrtSomeone	-0.06	1.26	0.16	1.09	0.36	0.61	0.23	0.23	0.52	0.64	0.07	1.21	1.03	0.14
30	M1B10MovSlower	0.24	0.75	0.5	0.24	0.23	0.79	0.15	0.22	0.78	0.35	0.53	-0.06	-0.17	0.75
51	M2DNoFeelings	0.56	0.63	0.59	0.59	0.55	0.63	0.52	0.59	0.68	0.39	0.65	0.59	0.66	0.1
46	M1D11MuslAches	0.84	0.43	0.87	0.03	0.45	0.73	0.62	0.74	0.71	0.58	0.88	0.11	-0.01	1.11
32	M1C2Suicide	0.77	0.59	0.54	1	1.34	0.39	0.98	0.56	0.51	0.8	0.59	0.89	1.2	0.82
53	M2FCantGoOn	0.97	0.48	0.8	0.64	0.83	0.66	0.9	0.77	0.79	0.87	0.81	0.54	0.54	0.63
54	M2GFearUrges	0.8	0.68	0.99	0.32	0.53	0.8	0.92	0.83	0.78	0.78	1.07	0.22	0.39	0.14
40	M1D5FearOpen	0.73	1.02	0.91	0.74	0.61	1.1	0.63	0.73	0.93	0.85	0.85	0.8	0.88	0.74
52	M2EWakColdSweat	1.13	0.93	1.08	0.96	0.88	1.13	1.1	0.99	0.96	1.13	1.04	1.08	1.04	0.85
41	M1D6SnakesDark	1.31	0.77	0.94	1.27	1.05	1	1.12	1.14	0.93	0.84	0.96	1.4	1.54	1.03
50	M2CWishDead	1.31	1.02	1.28	0.99	1.08	1.23	1.11	1.2	1.1	1.37	1.3	0.92	1.1	0.93
44	M1D9OthsNoSee	1.16	1.35	1.11	1.52	1.25	1.41	0.92	1.09	1.31	0.88	1.2	1.28	1.77	1.45
57	M2KAfradSleep	2.05	1.55	1.84	1.72	1.83	1.79	1.64	1.84	1.79	1.41	1.91	1.95	1.62	1.66
33	M1C3PlanSuicide	2.1	1.89	1.77	2.41	2.53	1.78	2.2	1.84	1.92	1.83	1.89	2.23	2.51	1.78
35	M1C5AtemtSuicide	2.39	1.92	1.79	2.9	2.85	2	2	1.83	2.14	2.22	1.92	2.62	2.41	1.78
45	M1D10ShudBpunished	2.18	2.43	2.4	2.15	1.9	2.69	2.08	2.18	2.37	2.42	2.35	1.87	2.59	2.66
34	M1C4GotGun	2.49	2.25	2.07	2.97	3.25	2.1	2.4	2.17	2.18	2.25	2.25	2.89	2.73	2.04

## Person Fit Group Analysis for the IMDS

The purpose of the person fit group analysis is to illustrate and interpret the expected and unexpected patterns of raw scores in terms of the expectations of the Rasch model. This information should enable us to interpret certain unusual patterns of scores more appropriately, e.g., low scorers without depression who are actually at high risk because of their suicidal thoughts. This type of analysis should inform the interpretation of Rasch measures and enable better treatment decisions. We alert the reader that these charts present raw score p-values (higher proportions endorsing the items are higher on the charts) so they are upside down from the typical Rasch charts where more rare is higher. We use the words lower, moderate and higher to express where the fit group means fall relative to the overall group mean. The items in the charts are arranged first in severity order by subscale and then arranged by severity order within subscales.

Figure 6 displays the numbers and percentages of persons in each fit group in a pie chart. In Figures 7-10 below, we present the four possible person fit patterns using Rasch person fit statistics (Wright & Stone, 1979) where  $\leq 1.33$  mean square on both infit and outfit is low or moderate (L/M) fit (Wilson, 2005). We are regarding this as good fit from a clinical perspective, though we recognize that some would say that very low values, e.g.,  $< .75$  would be over-fitting. Infit or outfit values above 1.33 are regarded as high (H) or poor fitting patterns.

Therefore, in Figure 7, L/M on infit and L/M on outfit would be a pattern that is consistent with Rasch model expectations, i.e., good fit, and over 86% of the persons were in this fit group. The solid lines represent the actual item (dichotomous 0/1 categories) response means for each fit group, and the dashed lines represent the item means over all persons. The red vertical lines indicate the difference between fit group mean and the overall mean.

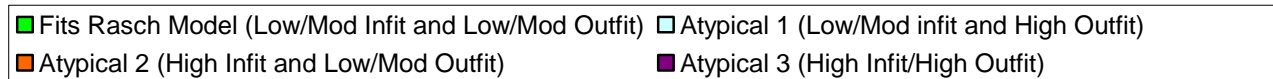
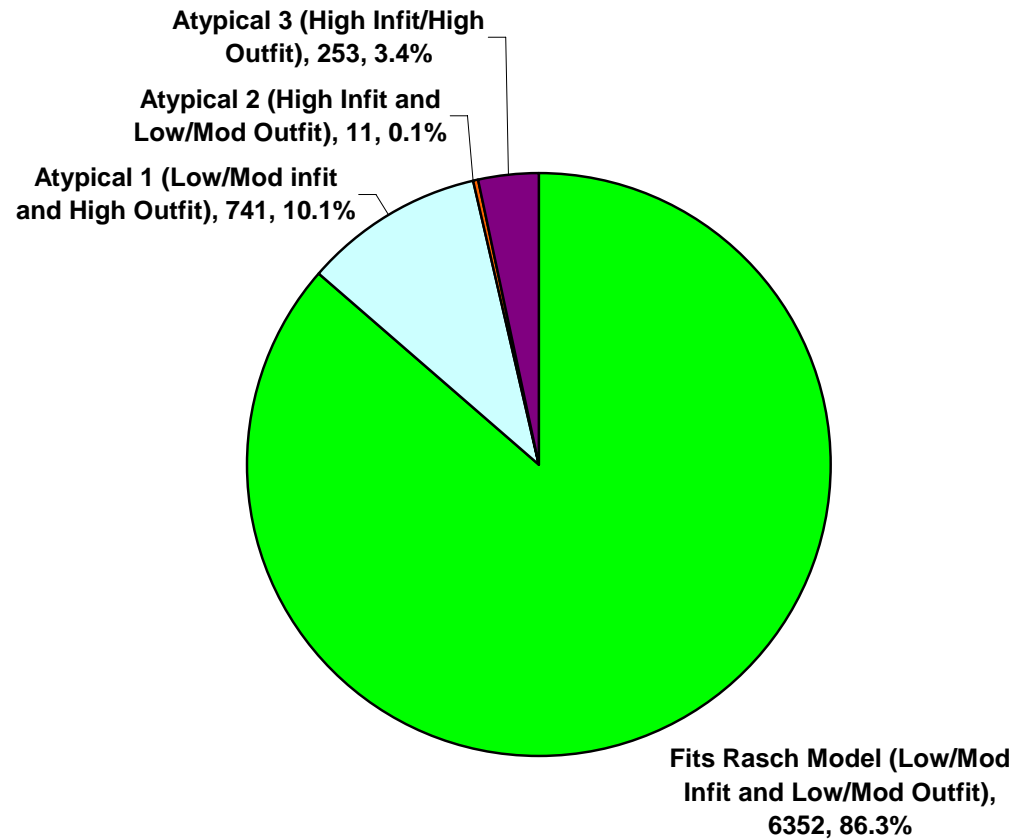
In Figure 8, the L/M infit and H outfit group (10.1%) is called Atypical Type 1, where the overall score may underestimate severity since these tend to be people who are having suicidal thoughts (higher risk items) but are unexpectedly low on depression, about normal on somatic, low to normal on anxiety except for the three highest anxiety items where they are higher, and low to normal on trauma (which all tend to be lower risk than suicide).

In Figure 9, the H infit and L/M outfit group consists of only 11 people (.1%) and is called Atypical Type 2. This group tends to be people who are lower on depression and the less severe anxiety items, but higher on somatic symptoms, high on the higher severity anxiety items, considerably higher on trauma, and higher on suicidal thoughts. Therefore, this group is characterized by low depression but is especially high on trauma, stress, and anxiety while being somewhat higher on somatic complaints and suicidal ideation.

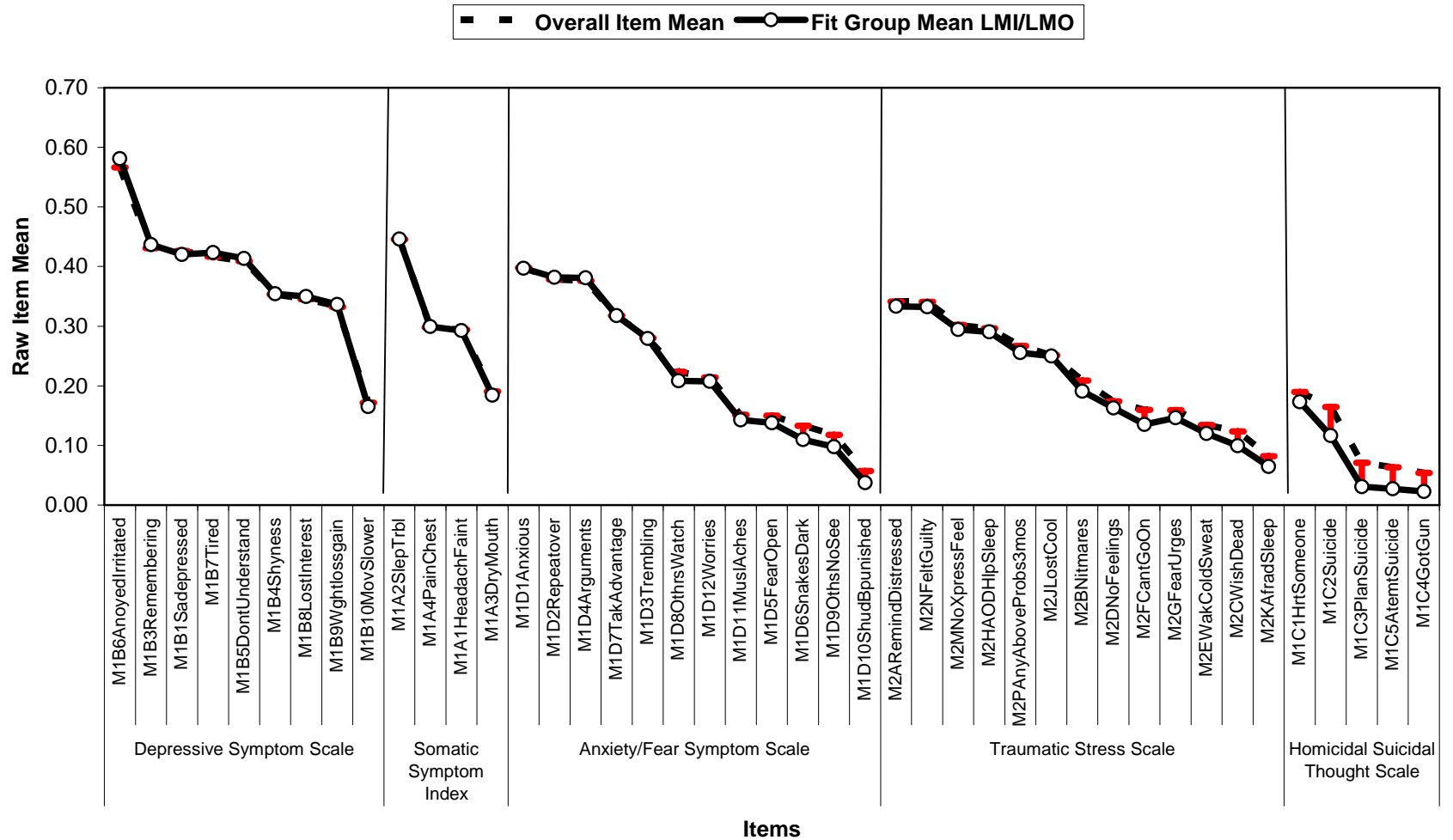
In Figure 10, the H infit and H outfit group consists of 253 people (3.4%) and is called Atypical Type 3. These tend to be valid high scorers who also have an overall measure that may underestimate severity somewhat and tend to have higher anxiety, higher trauma and higher suicidal ideation while their depression and somatic symptoms are moderate. These persons misfit because they strongly endorse most of the higher risk items but only endorse some of the lower risk items at moderate levels.

In summary, Atypical Type 1 is a deceptive pattern since the person measures will tend to be lower than they should be based on the person's low endorsement of the lower severity items such as depression. Atypical Types 2 and Type 3 will also tend to be deceptive since, even though they will tend to have high measures, their actual measures should be even higher because they tend to endorse the most severe symptoms but not many of the lower severity items. Both of these groups will tend to have measures that underestimate the seriousness of their internal mental distress. Atypical Type 3 is of special concern because of the much higher endorsement of suicidal symptoms. (Conrad, Bezruczko, Park, Chan, Riley, Diamond, & Dennis (2009) discuss this issue in detail).

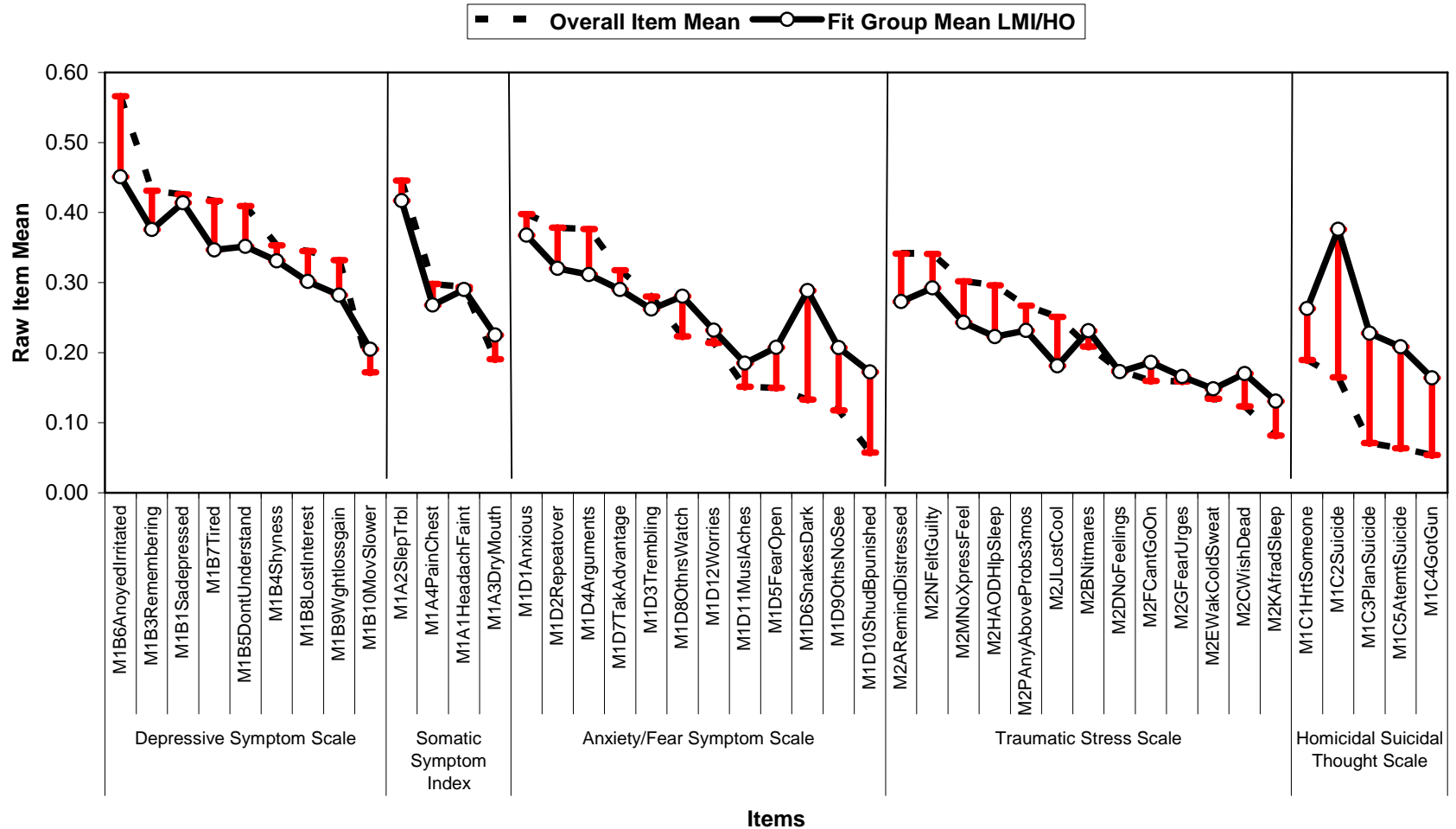
Figure 6. IMDS Fit Group Pie Chart



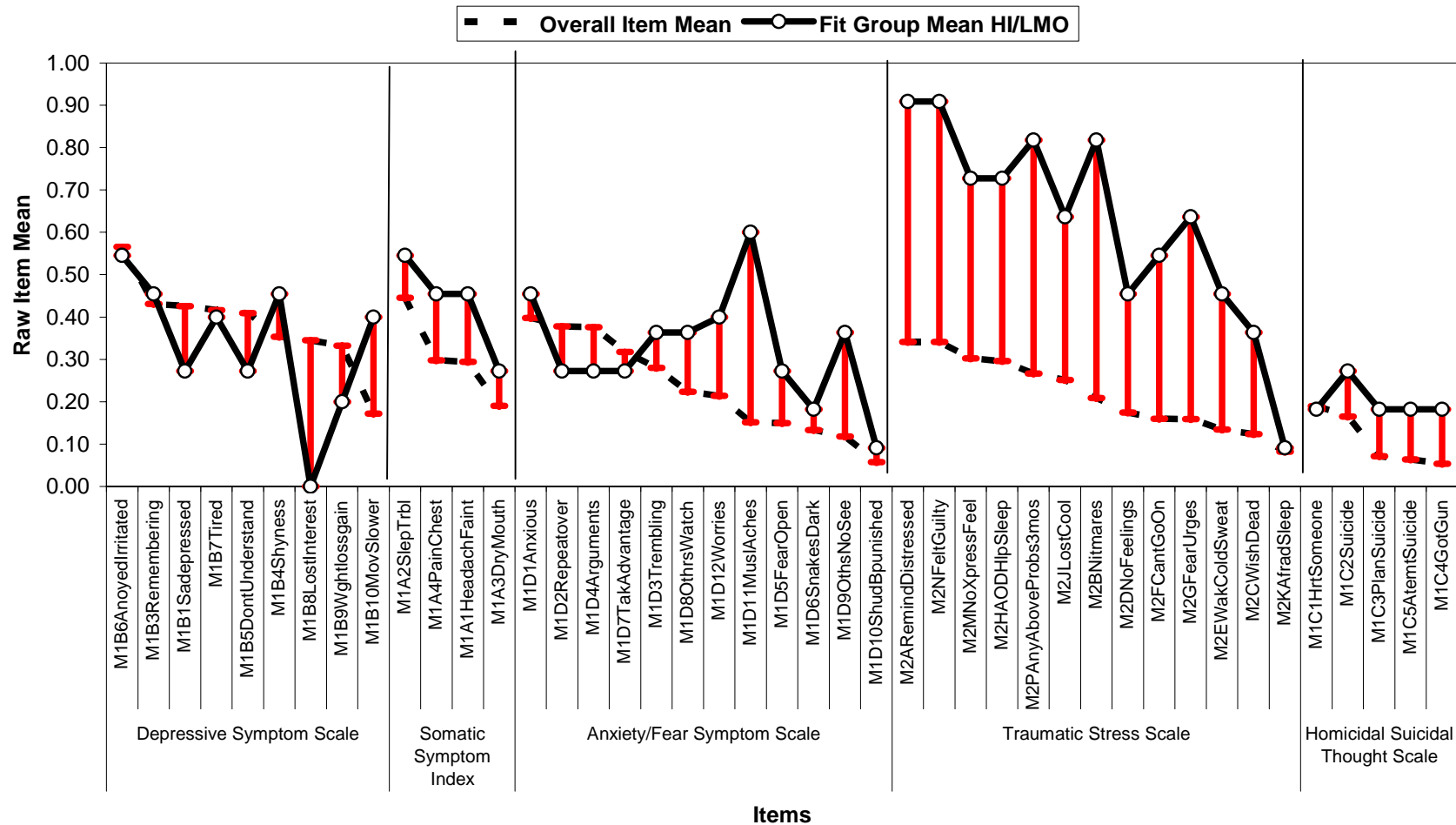
**Figure 7. IMDS Overall Item Mean vs. Fit Group Item Mean  
 Fits the Rasch Model (LMLM)  
 (n=6352; 86.3%, Low/Mod Infit and Low/Mod Outfit)**



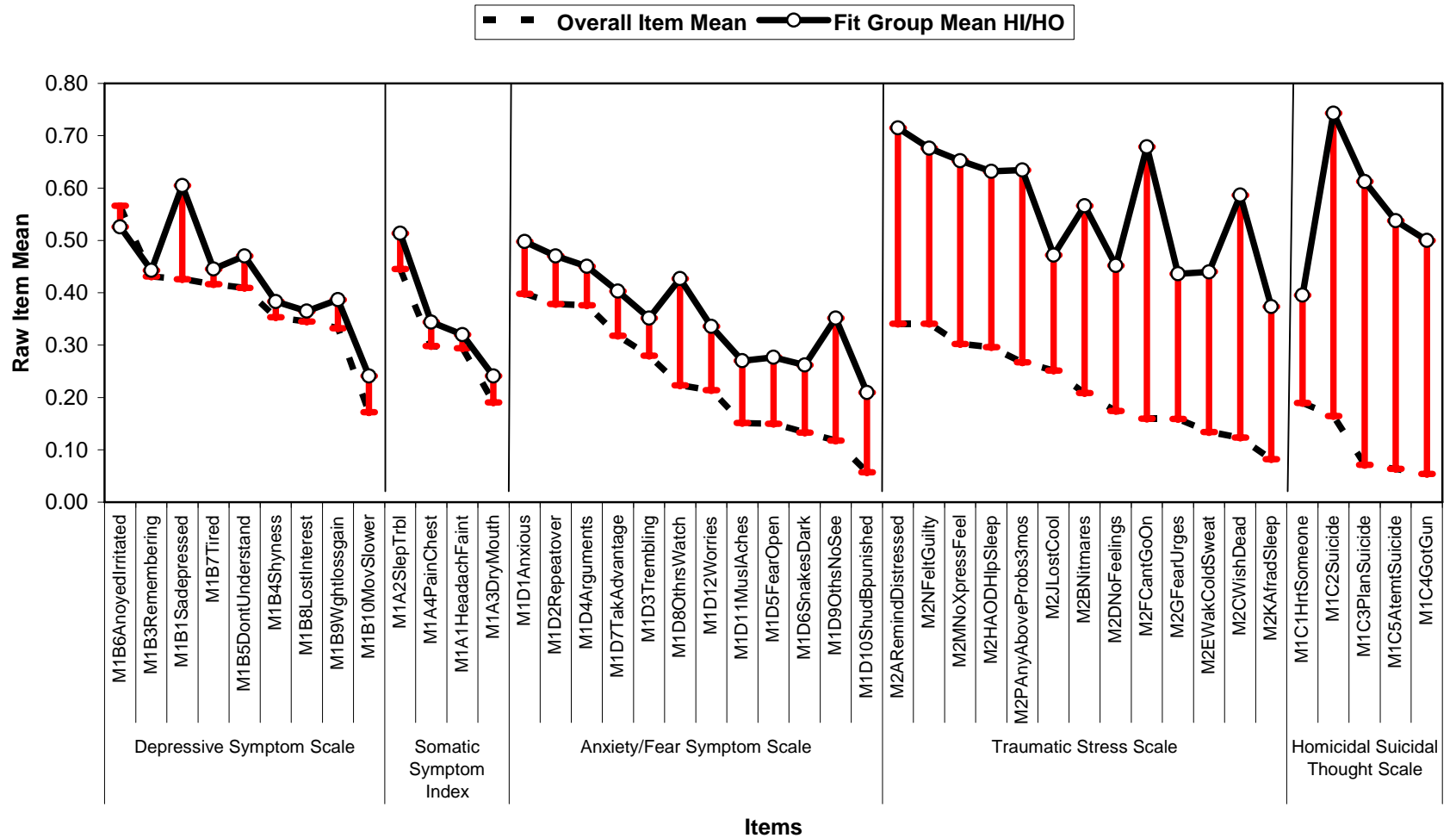
**Figure 8. IMDS Overall Item Mean vs. Fit Group Item Mean**  
**Atypical Type 1 Higher on Suicide and Some Anxiety Items, but Lower/Moderate on the Other**  
**Subscales**  
**(n=741; 10.1%, Low/Mod Infit and High Outfit)**



**Figure 9. IMDS Overall Item Mean vs. Fit Group Item Mean**  
**Atypical Type 2 Especially Higher on Trauma, Also Higher on Somatic, Suicide and Some**  
**Anxiety Items**  
**(n=11; 0.1%, High Infit and Low/Mod Outfit)**



**Figure 10. IMDS Overall Item Mean vs. Fit Group Item Mean**  
**Atypical Type 3 Especially Higher on Suicide and Trauma, Also Higher on Anxiety**  
**(n=253; 3.4%, High Infit and High Outfit)**



### Recommended Actions and Deliberations

The *IMDS* is useful in assessing its target construct, and it has strong psychometric properties (see abstract and text). However, there are several recommended actions and deliberations that might improve it.

- Differential item functioning was not a major issue in terms of the bias in the overall IMDS measures, but the DIF analyses did reveal some interesting group differences that may merit further study because of their substantive implications for tailoring treatment to needs. This is especially true of the primary drug severity group and the youth vs. adult contrasts.
  - Of the 43 items in the *IMDS*, significant DIF (i.e.,  $> .5$  SD = .58 logits) occurred in 3 items for males vs. females, 11 items for youth vs. adults, 4 items for race when using African American as the reference group, and 17 items for primary substances when using alcohol as the referent.
  - In terms of the proportion of contrasts with significant DIF, while there were more significant contrasts (n=17 out of 215 contrasts for primary drugs - where the contrast was alcohol vs. 5 other drugs; about 8%), there was a higher proportion for youth vs. adults (25%; 11 significant contrasts out of 43).
- Further examination of DIF between youth and adults could be helpful in improving appropriateness for one or both of these groups
- Two groups, Atypical Types 2 and 3, tend to have measures that underestimate the seriousness of their internal mental distress. Atypical Type 3 is of special concern because of the much higher endorsement of suicidal symptoms in this group compared to the overall group. We recommend flagging these two groups for clinicians in the evaluation/validity concerns section of the GAIN Recommendation and Referral Summary (GRRS) as:
  - Atypical 2 response pattern on Internalizing Mental Distress Scale (Relative to total score, lower than expected on depression and common anxiety symptoms, but higher than expected on somatic complaints, trauma and suicidal symptoms)
  - Atypical 3 response pattern on Internalizing Mental Distress Scale (Relative to total score, higher than expected on anxiety, trauma and suicidal symptoms)
- More work on construct validity would be helpful to understand these fit groups better and to ensure proper interpretation of measures.

## References

- Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences*. (2<sup>nd</sup> Ed.) Mahwah, NJ: Erlbaum Associates.
- Conrad, K.J. & Smith, E.V. (2004). International conference on objective measurement: Applications of Rasch analysis in health care. *Medical Care*, 2004; 42 (suppl I) 1-6.
- Conrad, K.J., Bezruczko, N., Park, H., Chan, Y., Riley, B., Diamond, G., & Dennis, M.L. (2009). Atypical suicide risk: Using Rasch person fit statistics to improve screening and understanding. Manuscript submitted for publication.
- Conrad, K.J., Dennis, M.L., Bezruczko, N., Funk, R., & Riley, B. (2007). Substance use disorder symptoms: Evidence of differential item functioning by age. *Journal of Applied Measurement*, 8(4), 373-387.
- Dennis, M. L., Chan, Y-F., & Funk, R.R. (2006). Development and validation of the GAIN short screener (GSS) for internalizing, externalizing, and substance use disorders and crime/violence problems among adolescents and adults. *The American Journal on Addictions*, 15, 80-91.
- Dennis, M. L. Conrad, K. J. & Chan, Y. (2008). Variation in *DSM-IV* symptom severity depending on type of drug. Manuscript in progress.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J.: Erlbaum.
- Linacre, J.M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2(3), 266-283.
- Linacre, J.M. (2008). *Winsteps Rasch Measurement* (Version 3.64.2). [www.winsteps.com](http://www.winsteps.com). Author.
- Lipman, R.S., Covi, L., & Shapiro, A.K.. (1979). The Hopkins Symptom Checklist (HSCL): Factors derived from the HSCL-90. *Journal of Affective Disorders*, 1, 9-24.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut. (Republished Chicago: The University of Chicago Press: 1980).
- Reckase, M. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.
- Smith, E.V. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2(3), 281-311.
- Smith, E.V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3(2), 205-231.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum Associates.
- Wright, B.D., & Stone, M.H. (1979). *Best test design*. Chicago: University of Chicago, MESA Press.

**ACKNOWLEDGEMENT:** This development of this paper was supported by the Center for Substance Abuse Treatment (CSAT), Substance Abuse and Mental Health Services Administration (SAMHSA) via Westat under contract 270-07-0191 to Dr. Dennis at Chestnut Health Systems in Bloomington, Illinois using data provided by the following grants and contracts from CSAT (TI-11320, TI-11317, TI-11321, TI-11323, TI-11324, TI-11422, TI-11424, TI-11423, TI-11894, TI-11874, TI-11888, TI-11892, TI-11871, TI-13309, TI-13356, TI-13305, TI-13340, TI-13344, TI-13322, TI-13323, TI-13345, TI-13308, TI-13354, TI-13313, TI-14254, TI-14376, TI-14311, TI-14196, TI-14214, TI-14261, TI-14090, TI-14189, TI-14252, TI-14283, TI-14355, TI-14272, TI-14103, TI-14267, TI-14315, TI-14188, TI-14271, TI-15686, TI-15671, TI-15486, TI-15545, TI-15672, TI-15475, TI-15678, TI-15447, TI-15461, TI-15433, TI-15481, TI-15514, TI-15478, TI-15413, TI-15483, TI-15670, TI-15674, TI-15479, TI-15682, TI-15467, TI-15511, TI-15562, TI-13601, TI-13190, TI-12541, TI-00567; Contract 207-98-7047, Contract 277-00-6500, Contract 270-2003-00006), the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (R01 AA 10368), the National Institute on Drug Abuse (NIDA) (R37 DA11323; R01 DA 018183), the Illinois Criminal Justice Information Authority (95-DB-VX-0017), and the Illinois Office of Alcoholism and Substance Abuse (PI 00567). The opinions are those of the author and do not reflect official positions of the contributing project directors or government.

March 20, 2009