

Validation of the Behavioral Complexity Scale (BCS) 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 *Behavioral Complexity Scale (BCS)* using the Rasch measurement model. The *BCS* consists of 33 past-year yes/no items related to externalizing disorders, including attention deficit, hyperactivity/impulsivity, and conduct disorder.

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 *BCS* performs well as a measure of the construct of behavioral complexity. Both items and scales form the theoretically expected hierarchies with a person internal consistency reliability of .87 and an item reliability of 1.00. The persons' responses generally conformed to the expectations of the Rasch model. Of the 33 items in the *BCS*, significant DIF (i.e., $> .5 SD = .78$ logits) occurred in 4 items for males vs. females, 5 items for youth vs. adults, 10 items for race when using African American as the reference group, and 9 items for primary substances when using alcohol as the referent. No items misfit using the criterion of .75-1.33 MNSQ for both infit and outfit. The three atypical groups, Atypical Groups 1, 2, and 3, tend to have scores that underestimate their behavioral complexity symptoms.

The Atypical Type 1 group is a deceptive pattern because the person measures will tend to be lower than they should be based on the person's average endorsement of the lower severity inattentive disorder items but higher than expected endorsement of the more serious items. Atypical Group Types 2 and 3 will also tend to be deceptive because, even though they will tend to have high measures, they tend to endorse the most severe symptoms more than expected but endorse the lower severity items less than expected.

Conclusion. The results suggest the possibility of some important gender differences. It was somewhat more common for females than for males to endorse *Forced Sex*. Qualitative probing suggests that females endorsing this item were thinking of instances when they pressured or seduced males or females into having sex with them. Males, in contrast, were more likely to associate this with rape, and thus were less likely to endorse it. Further evaluation of alternative wording, e.g., using words such as "seduced or pressured against their will", may be useful to improve this item.

Other DIF analyses did not discern obvious unbalance that could result in biased measurement, but they did reveal some issues that might be of theoretical interest. We recommend flagging the Atypical Groups

1, 2, and 3 for clinicians because these groups tend to have scores that underestimate their behavioral complexity symptoms.

Citation. Conrad, K. J., Conrad, K. M., Dennis, M. L., Riley, B.B., & Funk, R. (2009). Validation of the Behavioral Complexity Scale (BCS) to the Rasch Measurement Model, GAIN Methods Report 1.1.

Chicago, IL: Chestnut Health Systems. Retrieved from

Retrieved from [http://www.chestnut.org/li/gain/psychometric_reports/Conrad et al 2009 BCS Rasch Report.pdf](http://www.chestnut.org/li/gain/psychometric_reports/Conrad_et_al_2009_BCS_Rasch_Report.pdf)

Purpose

The purpose of this report is to provide a brief report of a psychometric analysis of the *Behavioral Complexity Scale (BCS)* using the Rasch measurement model (Bond & Fox, 2007; Rasch, 1960). The Rasch analysis was conducted using *Winsteps* software (Linacre, 2007). The tables and figures are annotated to summarize the main points.

Background

The *BCS* 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 actually a series of related instruments that share the same general instructions, questions (and variable names in most cases), scoring, interpretation, and clinical decision trees. The *GAIN* is a biopsychosocial assessment that integrates research and clinical practice to do diagnosis, placement, individualized treatment planning, and program evaluation and meets major reporting requirements.

The *BCS* “is a count of past-year symptoms related to externalizing disorders, including attention deficit, hyperactivity/impulsivity, and conduct disorder; it is based on the *DSM_IV* (American Psychiatric Association, 2000) symptoms/disorders that are common in adolescents but still persist into adulthood and are correlated with substance use severity (Dennis, Chan, & Funk, 2006).

BCS Subscales and Items

The *BCS* consists of three subscales with a total of 33 items. Its subscales are the *Inattentive Disorder Scale (IDS; 9 items)*, *Hyperactivity-Impulsivity Scale (HIS; 9 items)*, and *Conduct Disorder Scale (CDS; 15 items)*. The item stem reads: “*During the past 12 months, have you done the following things two or more times?*” The 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.

Table 1. Scale and Item Information

Subscale Name and Item Stem	GAIN Item Number	Rasch Output Item Number	Item Label
<i>Inattentive Disorder Scale</i>			
1. Made mistakes because you were not paying attention.	M3A1	61.	MistakesAttn
2. Had a hard time paying attention at school, work or home	M3A2	62.	PayAttnSchool
3. Had a hard time listening to instructions at school, work or home.	M3A3	63	ListnInstructns
4. Not followed instructions or not finished your assignments	M3A4	64.	FollowInstruct
5. Had a hard time staying organized or getting everything done	M3A5	65.	StayOrg
6. Avoided things that took too much effort, like school work or paperwork.	M3A6	66.	AvoidEffort
7. Lost things that you needed for school, work or home.	M3A7	67.	LostThings
8. Been unable to pay attention when other things were going on.	M3A8	68.	UnablePayAttn
9. Been forgetful or absentminded	M3A9	69.	Absentminded
<i>Hyperactivity Impulsivity Scale</i>			
10. Fidgeted or had a hard time keeping your hands or feet still when you were supposed to	M3A10	70.	Fidget
11. Been unable to stay in a seat or where you were supposed to stay.	M3A11	71.	StaySeated
12. Felt restless or the need to run around or climb on things.	M3A12	72.	Restless
13. Gotten in trouble for being too loud when you were playing or relaxing.	M3A13	73.	LoudPlay
14. Felt like you were always on the go or driven by a motor.	M3A14	74.	FeltOnTheGo
15. Talked too much or had others complain that you talked too much.	M3A15	75.	TalkTooMuch
16. Gave answers before the other person finished asking the question.	M3A16	76.	AnswrB4Questn
17. Had a hard time waiting for your turn	M3A17	77.	Wait
18. Interrupted or butted into other people's conversations or games.	M3A18	78.	Interrupted
<i>Conduct Disorder Scale</i>			
19. Been a bully or threatened other people.	M3B1	79.	Bully
20. Started fights with other people	M3B2	80.	StartFights
21. Used a weapon in fights.	M3B3	81.	Weapon
22. Been physically cruel to other people.	M3B4	82.	PhysCrulPeopl

Subscale Name and Item Stem	GAIN Item Number	Rasch Output Item Number	Item Label
23. Been physically cruel to animals.	M3B5	83.	PhysCruAnmal
24. M3B6 Taken a purse, money or other things from another person by force.	M3B6	84.	TakMoneyForce
25. Forced someone to have sex with you when they did not want to.	M3B7	85.	ForcedSex
26. Set fires	M3B8	86.	SetFires
27. Broken windows or destroyed property	M3B9	87.	DstroydPrprty
28. Taken money or things from a house, building or car.	M3B10	88.	TakeMoneyHome
29. Lied or conned to get things you wanted or to avoid having to do something.	M3B11	89.	1LiedConned
30. Taken things from a store or written bad checks to buy things.	M3B12	90.	StolStorBadChks
31. Stayed out at night later than your parents or partner wanted.	M3B13	91.	StayOut2Late
32. Run away from home (partner) for at least one night.	M3B14	92.	RunAwayOvrnite
33. Skipped work or school.	M3B15	93.	SkipSchool

•
•
•
•
•

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 BCS 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 logits (log odd units) 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

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 most severe 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^a)

	Percent	Number
Age, Mean (sd) 19.9 (8.9)		
< 18 years	72.5	5388
≥18 years	27.5	2047
Gender		
Male	67.1	4992
Female	32.7	2437
Race		
African American	25.7	1913
Caucasian	45.2	3360
Hispanic	10.8	806
Mixed/other	17.7	1314
Drug, primary, most severe		
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
^a Numbers may not add up to 100% due to missing values		

Table 3. Person and Item Reliability

SUMMARY OF 5532 MEASURED (NON-EXTREME) PERSONS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	14.4	32.9	-.51	.46	.99	.0	1.02	.1
S.D.	6.8	1.1	1.35	.10	.24	1.2	.70	.9
MAX.	32.0	33.0	4.95	1.21	2.17	4.9	9.90	9.9
MIN.	1.0	12.0	-4.21	.41	.35	-3.6	.04	-1.6
REAL RMSE	.49	ADJ.SD	1.25	SEPARATION	2.54	PERSON RELIABILITY		.87
MODEL RMSE	.47	ADJ.SD	1.26	SEPARATION	2.67	PERSON RELIABILITY		.88
S.E. OF PERSON MEAN = .02								

MAXIMUM EXTREME SCORE: 1 PERSONS
MINIMUM EXTREME SCORE: 1856 PERSONS
LACKING RESPONSES: 46 PERSONS
VALID RESPONSES: 99.7%

SUMMARY OF 7389 MEASURED (EXTREME AND NON-EXTREME) PERSONS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	10.8	32.9	-1.76	.81				
S.D.	8.5	1.0	2.45	.61				
MAX.	32.0	33.0	4.95	1.85				
MIN.	.0	12.0	-5.47	.41				
REAL RMSE	1.02	ADJ.SD	2.22	SEPARATION	2.18	PERSON RELIABILITY		.83
MODEL RMSE	1.01	ADJ.SD	2.23	SEPARATION	2.20	PERSON RELIABILITY		.83
S.E. OF PERSON MEAN = .03								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .96 (approximate due to missing data)
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .94 (approximate due to missing data)

SUMMARY OF 33 MEASURED (NON-EXTREME) ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	2408.6	5516.3	.00	.04	1.00	-.5	1.03	.0
S.D.	1170.0	5.1	1.55	.02	.08	4.8	.15	4.2
MAX.	4583.0	5526.0	5.12	.15	1.21	9.9	1.38	9.9
MIN.	48.0	5506.0	-2.60	.03	.88	-9.1	.81	-7.2
REAL RMSE	.04	ADJ.SD	1.55	SEPARATION	35.06	ITEM RELIABILITY		1.00
MODEL RMSE	.04	ADJ.SD	1.55	SEPARATION	35.68	ITEM RELIABILITY		1.00
S.E. OF ITEM MEAN = .27								

DELETED: 91 ITEMS
UMEAN=.000 USCALE=1.000
ITEM RAW SCORE-TO-MEASURE CORRELATION = -.96 (approximate due to missing data)
182039 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 169509.46

- Person reliability is strong at .87
- Items are placed reliably on the ruler.
- Cronbach alpha =.94
- Cronbach's alpha is higher because it estimates extreme scores as measured perfectly, i.e., with no error.
- A separation value of 2.67 gives approximately three separation levels, thus splitting the persons into about 4 groups on the Rasch ruler (Table 3).

- Good item reliability of 1.00.
- Item separation is high at 35.06 meaning items are placed reliably on the ruler.

Table 4. Wright Item Map

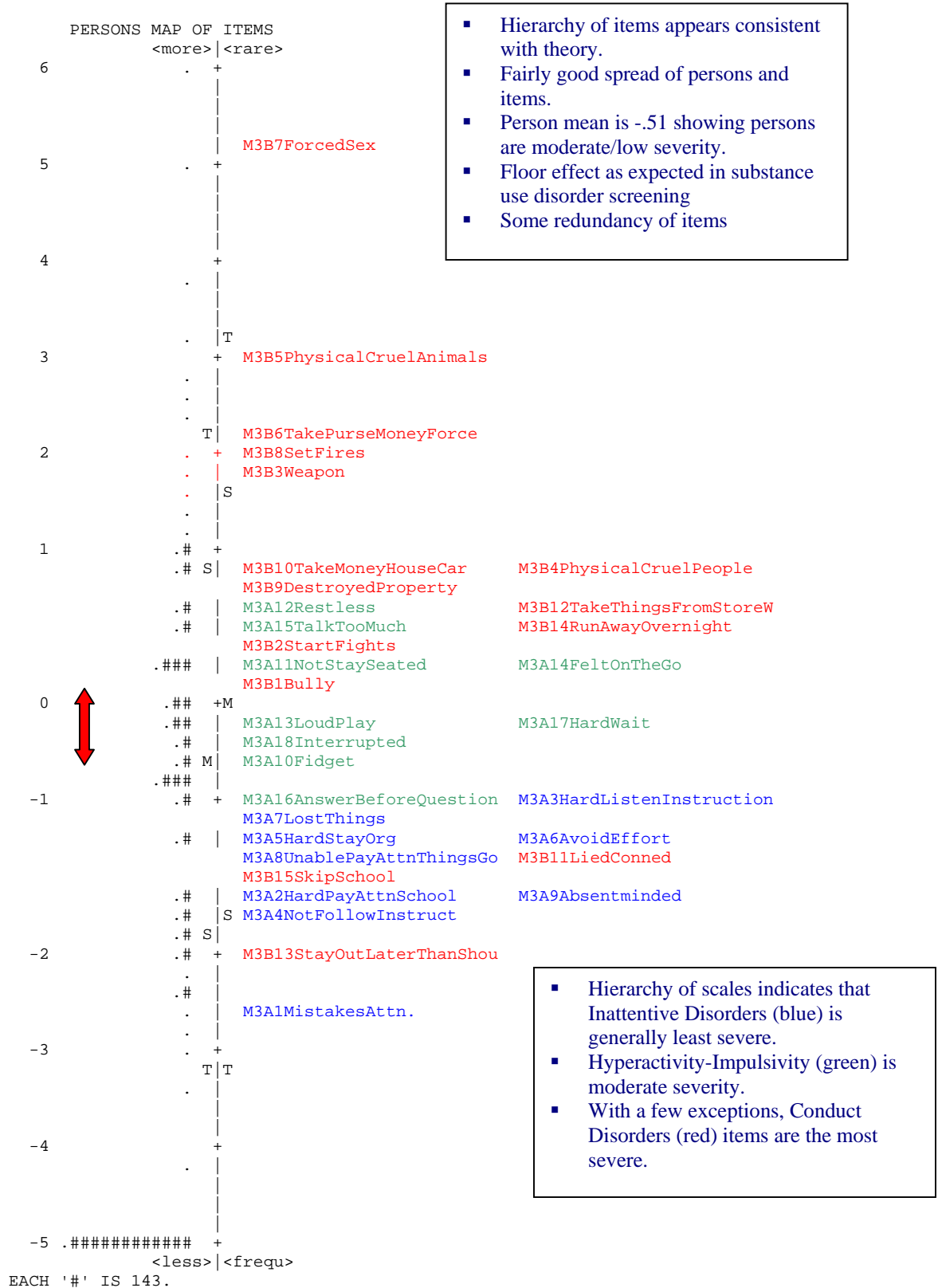


Table 5. Principal Component Analysis of Standardized Residual Correlations For Items

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

		-- Empirical --		Modeled
Total raw variance in observations	=	53.1	100.0%	100.0%
Raw variance explained by measures	=	20.1	37.9%	33.7%
Raw variance explained by persons	=	7.6	14.4%	12.8%
Raw Variance explained by items	=	12.5	23.5%	20.9%
Raw unexplained variance (total)	=	33.0	62.1%	100.0%
Unexplned variance in 1st contrast	=	3.0	5.6%	9.1%

STANDARDIZED RESIDUAL LOADINGS FOR ITEMS (SORTED BY LOADING)

CON-TRAST	LOADING	MEASURE	INFIT	OUTFIT	ENTRY	NUMBER	ITEM
1	.46	.79	.99	1.02	A	87	M3B9DstroydPrprty
1	.42	.13	1.01	1.02	B	79	M3B1Bully
1	.42	.36	1.05	1.08	C	80	M3B2StartFights
1	.42	.87	.98	1.01	D	82	M3B4PhysCrulPeopl
1	.40	.85	1.00	1.02	E	88	M3B10TakeMoneyHome
1	.39	1.75	1.05	1.33	F	81	M3B3Weapon
1	.37	2.30	.99	.94	G	84	M3B6TakMoneyForce
1	.32	.62	1.04	1.17	H	90	M3B12StolStorBadChks
1	.30	-1.17	1.21	1.38	I	93	M3B15SkipSchool
1	.30	-2.02	1.12	1.21	J	91	M3B13StayOut2Late
1	.27	.40	1.11	1.21	K	92	M3B14RunAwayOvrnite
1	.25	2.04	1.07	1.05	L	86	M3B8SetFires
1	.20	-1.19	.99	.94	M	89	M3B11LiedConned
1	.19	3.00	1.09	1.31	N	83	M3B5PhysCrulAnmal
1	.01	5.11	1.03	1.16	O	85	M3B7ForcedSex
1	-.41	-1.47	.90	.83	a	62	M3A2PayAttnSchool
1	-.39	-1.29	.87	.81	b	68	M3A8UnablePayAttn
1	-.37	-1.14	.97	.99	c	65	M3A5StayOrg
1	-.37	-.99	.88	.83	d	63	M3A3ListnInstructns
1	-.33	-1.47	.93	.90	e	69	M3A9Absentminded
1	-.33	-.53	.94	.90	f	70	M3A10Fidget
1	-.30	.18	.89	.83	g	71	M3A11StaySeated
1	-.26	.51	.94	.91	h	72	M3A12Restless
1	-.24	-.91	.99	.97	i	67	M3A7LostThings
1	-.24	-1.12	.92	.87	j	66	M3A6AvoidEffort
1	-.23	-1.65	.95	.91	k	64	M3A4FollowInstruct
1	-.21	-.30	.93	.90	l	77	M3A17Wait
1	-.20	-2.60	.99	1.00	m	61	M3A1MistakesAttn
1	-.19	.20	1.02	1.04	n	74	M3A14FeltOnTheGo
1	-.14	.47	1.08	1.21	o	75	M3A15TalkTooMuch
1	-.09	-.44	1.04	1.06	p	78	M3A18Interrupted
1	-.09	-1.06	1.11	1.17	Q	76	M3A16AnswrB4Questn
1	-.09	-.22	.97	.94	P	73	M3A13LoudPlay

- To judge the strength of the measurement dimension, we used the following internal guidelines for variance explained by the measure: $\geq 40\%$ is considered a strong measurement dimension, $\geq 30\%$ is considered a moderate measurement dimension, and $\geq 20\%$ is considered a minimal dimension. The 20% criterion is taken from Reckase (1979).
- The variance explained by the measure is 37.9%.
- The fact that only 9.1% of the variance is explained by the first factor of residuals supports unidimensionality
- Thus, the BCS is considered unidimensional using these guidelines.
- The PCA was conducted using *Winsteps v. 3.68*

- No infit problems
- “Skipped school” has highest outfit and infit, but the infit is less than 1.33 so there is no substantial problem.

Table 6. Most Misfitting Response Strings in Terms of Outfit Mean Square (OUTMNSQ)

ITEM	OUTMNSQ	PERSON
		6 5211 4414446 21515126211 663 6733366653332
		57448384185162578947191421639425599240114810842035
		76371508306847529872741997570189973313821000642662
		03933745026390747398582181293193205597266919472275
	high	-----
93 M3B15SkipScho	1.38	A 0...0.....11.1.....
81 M3B3Weapon	1.33	B 1..1.1.....1111.....
83 M3B5PhysCrulA	1.31	C 1.....11111111.11.1.....
91 M3B13StayOut2	1.21	D .0.....0.00.....
75 M3A15TalkTooM	1.21	E 1..1.....111
92 M3B14RunAwayO	1.21	F .0.....1..1.....1.....
76 M3A16AnswrB4Q	1.17	G 000.....
90 M3B12StolStor	1.17	H 1.....111.....
85 M3B7ForcedSex	1.16	I ...111111111111111111.....1.....
80 M3B2StartFigh	1.08	J 1.11.....
86 M3B8SetFires	1.05	K 1..1.....1..1.....
78 M3A18Interrup	1.06	L 1..1.....1.....
74 M3A14FeltOnTh	1.04	M 1.....
87 M3B9DstroydPr	1.02	N 1..1.....1.....
88 M3B10TakeMone	1.02	O 1.....1.....
79 M3B1Bully	1.02	P .00.....1..1..1.....
82 M3B4PhysCrulP	1.01	Q 1..1.....
61 M3A1MistakesA	1.00	p .00.....0.....
67 M3A7LostThing	.97	o ...00.0.....
84 M3B6TakMoneyF	.94	n ...1..11...1..11.....1..1.....
65 M3A5StayOrg	.99	m 0...0.0...0.....
89 M3B11LiedConn	.94	l 0.....
73 M3A13LoudPlay	.94	k 1.....
64 M3A4FollowIns	.91	j 0.....0.....
72 M3A12Restless	.91	i 1.....
70 M3A10Fidget	.90	h 1.....

This table shows the most misfitting items in terms of OUTMNSQ. To the right of the items, we see the persons who unexpectedly endorsed or did not endorse an item given their overall measure. **M3B15SkipScho** has the highest outfit (1.38) because a few people with high overall measures did not endorse it (the zeroes) even though it was a low severity item. A few other people misfit because they had low overall measures, but endorsed it (the ones). Also of note is **85 M3B7ForcedSex** which had several high outfitting persons who had low measures but misfit because they endorsed this high severity item.

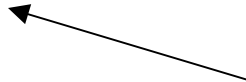
Table 8. Person Statistics: Misfit Order (table abridged due to space limitations)

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT OBS%	MATCH EXP%	PERSON
935	5	33	-2.31	.52	.99	.1	9.90	9.9	A .07	87.9	85.3	935 1036
1112	11	33	-1.05	.42	1.37	1.9	9.90	7.9	B .08	57.6	73.1	1112 2026
1175	12	33	-.88	.41	1.16	1.0	9.90	7.1	C .22	66.7	73.0	1175 2046
5187	1	33	-4.21	1.04	1.14	.5	9.90	3.2	D-.20	97.0	97.0	5187 2013
7256	11	33	-1.05	.42	1.08	.5	9.90	7.8	E .24	69.7	73.1	7256 1124
7036	14	33	-.54	.41	1.42	2.3	9.84	6.0	F .11	54.5	72.7	7036 1111
3866	5	33	-2.31	.52	1.40	1.3	9.18	4.4	G-.14	81.8	85.3	3866 1043
2879	15	33	-.38	.41	1.86	4.2	9.17	6.2	H-.12	45.5	72.6	2879 1013
1479	15	33	-.38	.41	1.60	3.1	8.93	6.1	I .01	51.5	72.6	1479 1022
897	16	33	-.21	.41	1.44	2.3	8.02	6.1	J .09	60.6	72.6	897 1026
4970	5	33	-2.31	.52	1.38	1.2	7.59	3.9	K-.08	81.8	85.3	4970 1023
235	5	33	-2.31	.52	1.10	.4	7.18	3.8	L .08	87.9	85.3	235 2013
5102	16	33	-.21	.41	1.11	.7	7.07	5.6	M .31	72.7	72.6	5102 1013
6235	17	32	.02	.42	1.72	3.3	6.67	5.7	N .02	56.3	73.3	6235 2114
3312	2	33	-3.45	.76	1.23	.6	6.16	2.5	O-.11	93.9	93.9	3312 2016
289	6	33	-2.05	.49	1.30	1.1	5.67	3.5	P .05	78.8	82.6	289 1013
5948	18	33	.12	.41	1.36	1.9	5.53	5.3	Q .22	60.6	72.9	5948 2114
6327	18	33	.12	.41	.97	-.1	5.08	4.9	R .43	72.7	72.9	6327 2115
4797	4	33	-2.61	.57	1.19	.6	5.02	2.6	S .03	87.9	87.9	4797 1036
5415	3	32	-2.95	.64	1.09	.4	4.98	2.4	T .08	90.6	90.6	5415 1023
1379	7	33	-1.82	.47	1.35	1.4	4.94	3.4	U .04	78.8	80.2	1379 1043
4676	7	33	-1.82	.47	1.35	1.4	4.94	3.4	V .04	78.8	80.2	4676 1043
2171	7	33	-1.82	.47	1.24	1.0	4.92	3.4	W .09	78.8	80.2	2171 1023
1183	19	33	.29	.42	1.38	1.9	4.83	4.9	X .23	60.6	73.3	1183 2026
3118	1	33	-4.21	1.04	1.13	.4	4.81	1.9	Y-.10	97.0	97.0	3118 1113
6480	19	33	.29	.42	1.24	1.3	4.74	4.8	Z .29	66.7	73.3	6480 1114
903	7	33	-1.82	.47	1.01	.1	4.62	3.3	.22	84.8	80.2	903 1036
1652	7	33	-1.82	.47	1.19	.8	4.60	3.3	.15	78.8	80.2	1652 1123
6059	11	33	-1.05	.42	1.87	4.0	4.51	3.6	-.21	51.5	73.1	6059 2115
5088	10	33	-1.23	.43	1.71	3.3	4.48	3.6	-.12	51.5	74.1	5088 1013
6244	19	33	.29	.42	1.05	.3	4.48	4.6	.40	72.7	73.3	6244 1114
3078	3	33	-2.97	.64	1.33	.8	4.42	2.2	-.08	90.9	90.9	3078 1013
411	7	33	-1.82	.47	1.12	.6	4.41	3.2	.20	78.8	80.2	411 1016
4904	7	33	-1.82	.47	1.50	1.9	4.41	3.2	-.11	78.8	80.2	4904 1034
130	20	33	.47	.42	1.41	2.0	4.33	4.6	.21	66.7	74.0	130 2026
1485	20	32	.58	.44	1.45	2.0	4.26	4.4	.20	65.6	75.2	1485 1045
473	27	33	2.01	.55	1.32	1.0	4.11	2.9	.28	81.8	86.5	473 1042
2752	4	33	-2.61	.57	1.39	1.1	4.09	2.3	-.09	87.9	87.9	2752 1013
843	10	33	-1.23	.43	1.52	2.5	4.09	3.3	-.01	63.6	74.1	843 1036
2287	10	33	-1.23	.43	1.91	4.0	3.93	3.2	-.24	57.6	74.1	2287 1043
5981	5	33	-2.31	.52	1.49	1.5	3.92	2.4	-.09	81.8	85.3	5981 2111
763	29	33	2.72	.65	1.05	.3	3.90	2.1	.39	93.9	90.8	763 1026
5199	6	33	-2.05	.49	1.40	1.4	3.86	2.6	-.02	78.8	82.6	5199 1016

There is minimal person misfit when using the criterion of less than 5% of the sample having both infit and outfit ≥ 1.33 logits. In this case, 4% of the sample misfit.

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

PERSON	OUTMNSQ	ITEM
		69666689667667777778977988888888
		114298935663708739140252078216435
		high-----
935	935 1036	9.90 A1
1112	1112 2026	9.90 B11.....1
1175	1175 2046	9.90 C1
5187	5187 2013	9.90 D1.....
7036	7036 1111	9.90 E 0.....1.....1
7256	7256 1124	9.90 F1.....1
2879	2879 1013	9.23 G .0.....1.1
3866	3866 1043	9.22 H11.....1.1
1479	1479 1022	8.99 I .0.....1.1
897	897 1026	8.07 J1.11
4970	4970 1023	7.62 K1.....1.1.1
235	235 2013	7.20 L1.....1
5102	5102 1013	7.12 M .0.....1
6235	6235 2114	6.71 N ...0.....1.1.1
3312	3312 2016	6.18 O1.....1.....
289	289 1013	5.69 P11.....1.....1
5948	5948 2114	5.57 Q ..0.....1
6327	6327 2115	5.12 R1
4797	4797 1036	5.05 S1.....1.....1..
5415	5415 1023	5.00 T1.....1.....1...
1379	1379 1043	4.96 U1.11.....1.
4676	4676 1043	4.96 V1.11.....1.
2171	2171 1023	4.94 W1.1.....1.
1183	1183 2026	4.86 X00.....1.1.1
3118	3118 1113	4.81 Y1.....
6480	6480 1114	4.77 Z00.0.....1
		-----low-
		69666689667667777778977988888888
		114298935663708739140252078216435



- This table shows the persons who unexpectedly endorsed an item given their overall measure. Most of the person misfit is being caused by items 81-86 when persons with low severity are unexpectedly endorsing the high severity items.
- Person #935 is one of the most misfitting persons with a low measure (-2.31) but an unexpected endorsement of item #85.

Table 10. Persons with Most Unexpected Responses in Terms of Measures

PERSON	MEASURE	ITEM
		69666689667667777778977988888888
		114298935663708739140252078216435
		high-----
570 570 2 1 0 4 570	3.86	←.....0.....
763 763 2 1 0 2 763	2.72	.0.....0.....
473 473 2 1 0 4 473	2.01	0.....0.....0.....
6439 4817 1 1 0 2 6439	2.01	0.....0.....
1804 1804 5 2 0 4 1804	.65000.0.....
2357 2357 3 2 0 1 2357	.65
5813 7121 1 1 1 4 5813	.6500...0.....
1485 1485 5 1 0 4 1485	.58	..0..0...0.....
130 130 6 2 0 2 130	.470..0.....
1183 1183 1 2 0 2 1183	.29 X00.....
4566 6244 4 1 1 1 4566	.29
4802 6480 4 1 1 1 4802	.29 Z00..0.....
4270 5948 4 2 1 1 4270	.12 Q	..0.....
4649 6327 5 2 1 1 4649	.12 R
4557 6235 4 2 1 1 4557	.02 N	..0.....1.1.1
897 897 4 1 0 2 897	-.21 J1..11
6724 5102 3 1 0 1 6724	-.21 M	..0.....1
1479 1479 6 1 0 2 1479	-.38 I	..0.....1.1
2983 2879 3 1 0 1 2983	-.38 H	..0.....1.1
5728 7036 1 1 1 1 5728	-.54 F	0.....1.....1
1175 1175 3 2 0 4 1175	-.88 C
1112 1112 2 2 0 2 1112	-1.05 B11.....1
5948 7256 4 1 1 2 5948	-1.05 D1.....1
2491 2491 3 1 0 2 2491	-1.611.....
411 411 3 1 0 1 411	-1.821.....1
903 903 3 1 0 3 903	-1.821.....1
1379 1379 3 1 0 4 1379	-1.82 U1..11.....1
1652 1652 3 1 1 2 1652	-1.821.....1
2171 2171 3 1 0 2 2171	-1.82 W1..1.....1
6298 4676 3 1 0 4 6298	-1.821..11.....1
289 289 3 1 0 1 289	-2.05 P11...1.....1
235 235 3 2 0 1 235	-2.31 L1.....1
935 935 3 1 0 3 935	-2.31 A
3970 3866 3 1 0 4 3970	-2.31 G11.....1.1
6592 4970 3 1 0 2 6592	-2.31 K1.....1..1
6593 4971 3 1 0 2 6593	-2.311.....1..
6419 4797 3 1 0 3 6419	-2.61 S1.....1.....1..
7037 5415 3 1 0 2 7037	-2.95 T1.....1.....
3126 3022 3 1 0 4 3126	-2.971.....1.....
3182 3078 3 1 0 1 3182	-2.9711.....1.....
3416 3312 3 2 0 1 3416	-3.45 O1.....1.....
525 525 1 1 0 4 525	-4.211.....
2367 2367 3 1 0 2 2367	-4.211.....
3062 2958 3 2 0 2 3062	-4.211.....
3222 3118 3 1 1 1 3222	-4.21 Y1.....
3447 3343 1 1 0 2 3447	-4.211.....
5864 7172 6 2 1 2 5864	-4.211.....
6009 7317 2 1 1 1 6009	-4.211.....
6101 7409 3 2 1 4 6101	-4.211.....
6809 5187 3 2 0 1 6809	-4.21 E1.....
		-----low-
		69666689667667777778977988888888
		114298935663708739140252078216435

We see some of the high scorers with unexpected non-endorsement of low severity items, i.e., zeros on the left. The other unexpected responses are the low scorers who are endorsing high severity items, especially 81-86.

Table 11. Summary of Category Structure

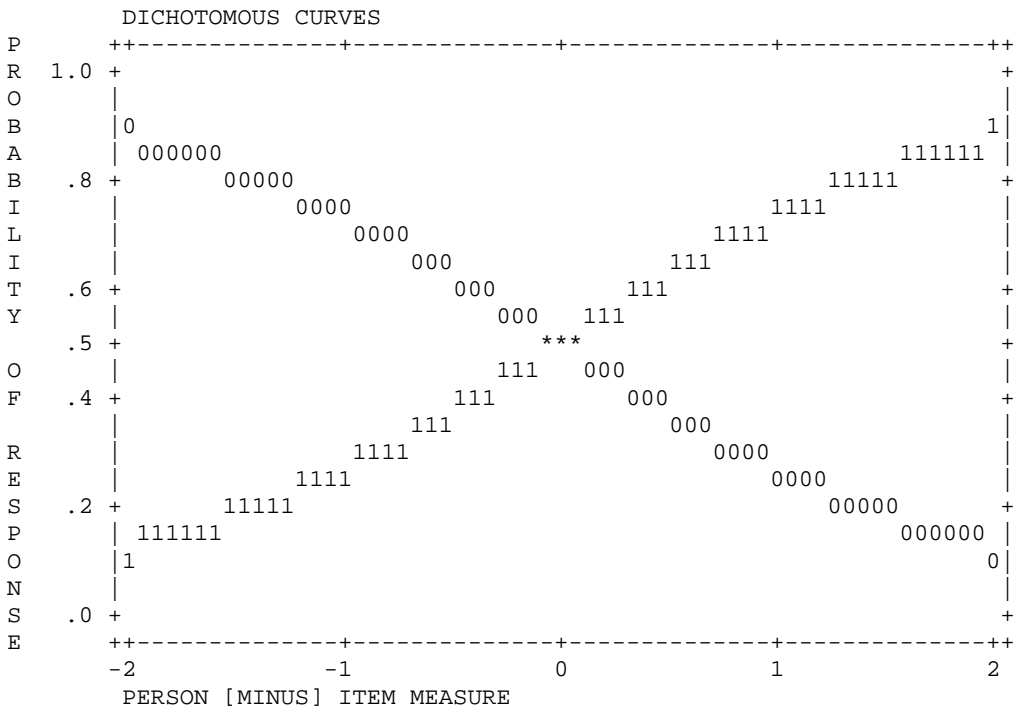
CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	COHERENCE	ESTIM
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ
				M->C	C->M	DISCR	
0	0	102018	56	-1.56	-1.56	.99	.98
1	1	79036	44	.84	.84	1.00	1.06
MISSING		512	0	-.17			

56% of all responses were zeros (did not endorse items).

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?



Differential Item Functioning for Age, Gender, Race, and Primary Drug Severity for the Behavioral Complexity Scale

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. They 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 ≥ 0.75 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 *SPS* items. Table 12 contains the data that formed these graphs, and provides the information to compute the exact differences between groups on each item. For example, to get the DIF contrast between males and females on M3A1MistakesAttn, subtract $-2.63 - (-2.55) = -.08$ which is a non-significant difference.

Gender DIF. In Figure 1, we can see that it was significantly easier for males to endorse *DestroyedProperty*, *UsedWeapon*, *SetFires*, and *PhysicallyCruelToAnimals*. However, the only item with a significant DIF contrast that was easier for females to endorse was *ForcedSex*. Qualitative probing suggests that females endorsing this item were thinking of instances when they pressured or seduced males or females into having sex with them. Males, in contrast, were more likely to associate this with rape, and thus were less likely to endorse it.

In any case, it appears that with four large DIF contrasts being easier for males, it is quite likely that the scale will be biased in the sense that females, who are only a third of the sample, will receive lower measures for these four items than they would if the items were calibrated for females alone. If there is no substantive reason for recalibrating these items for females, then there is really no problem.

Age DIF. In Figure 2, however, there are some much larger differences that tend to be balanced between youth and adults. Four items, i.e., *NotFollowedInstructions*, *SkippedSchool*, *DestroyedProperty*, and *SetFires*, are all much, i.e., $> .75$ logit easier for youth to endorse. Concurrently, there are three items that are $> .75$ logit easier for adults to endorse than it is for youth, i.e., *AbsentMinded*, *FeltOnTheGo*, and *ForcedSex*. While this equality in the numbers of significant DIF items will tend to balance the *BCS* so that the test will be unbiased for youth compared to adults, the differences are interesting theoretically. Specifically, youth will tend to endorse vandalism, while school-related items are not as relevant to adults.

Race DIF. In looking carefully at Figure 3, we observed that the high (harder to endorse) and low (easier to endorse) values usually involved either African Americans or Hispanics with Caucasians and Mixed tending to be in the middle. Therefore most of the significant differences were between African Americans and Hispanics. The items that were easier to endorse for African Americans compared to Hispanics were *Interrupted*, *FeltOnTheGo*, and *Forced Sex*. The items that were easier for Hispanics to endorse compared to African Americans were *SkipSchool* and *SetFires*. Additionally, it was easier for African Americans to endorse *UsedWeapon* and *TakeMoneyForce* compared to Caucasians. It was easier for African Americans to endorse *ForcedSex* when compared to all other groups. It was harder for Caucasians to endorse *PhysCruelAnimals* when compared to all other groups.

Primary Drug Severity DIF. With six different types of drugs, there were no obvious imbalances that would tend to bias the overall measures. However, the interactions here are quite complex and may deserve further study. A paper by Dennis, Conrad, & Chan, (2008) indicates that severity hierarchies do

differ by drug, but that most of the variation in severity is actually due to the drugs themselves rather than different hierarchies within drugs.

Figure 1. BCS Gender DIF

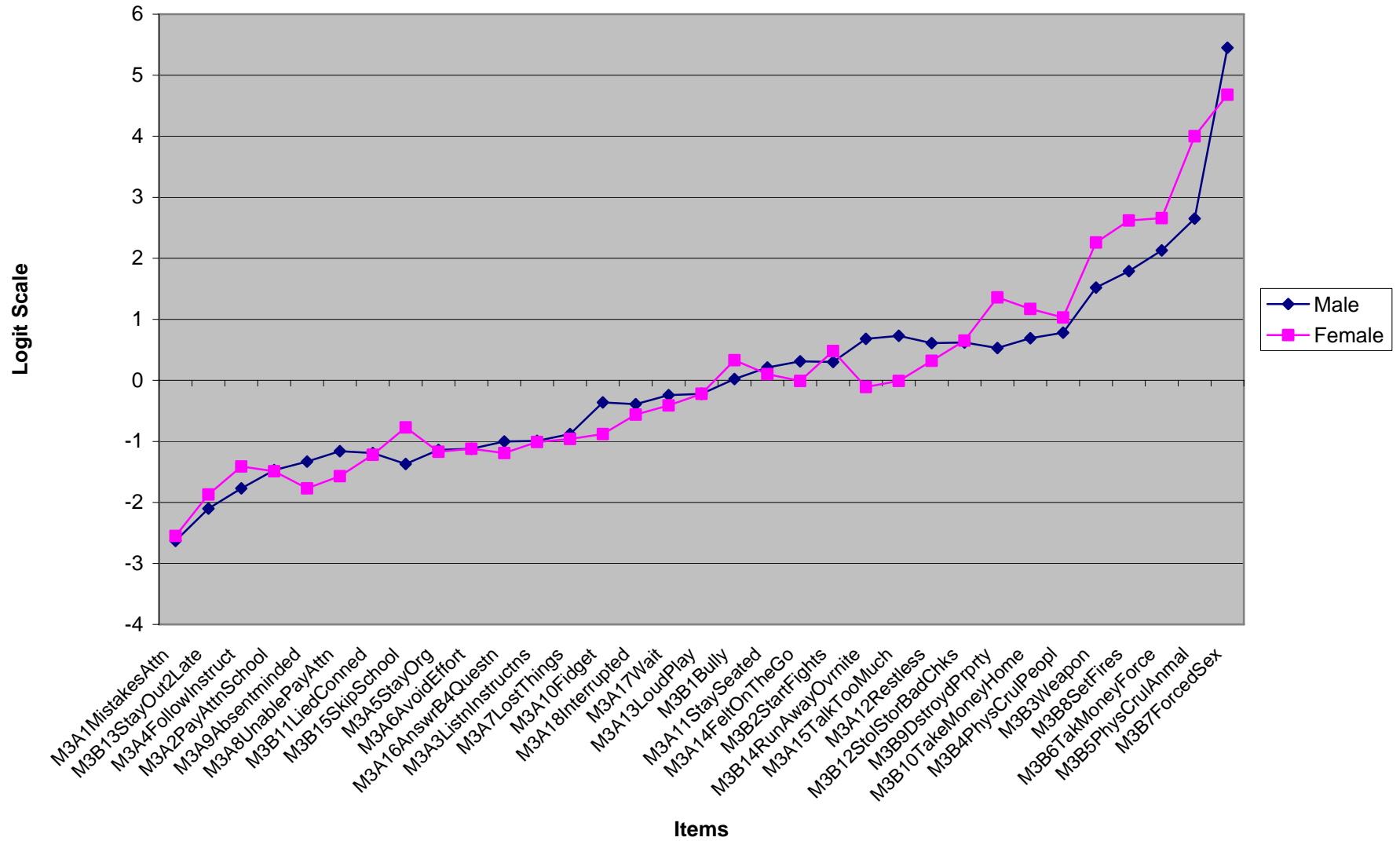


Figure 2. BCS Age DIF

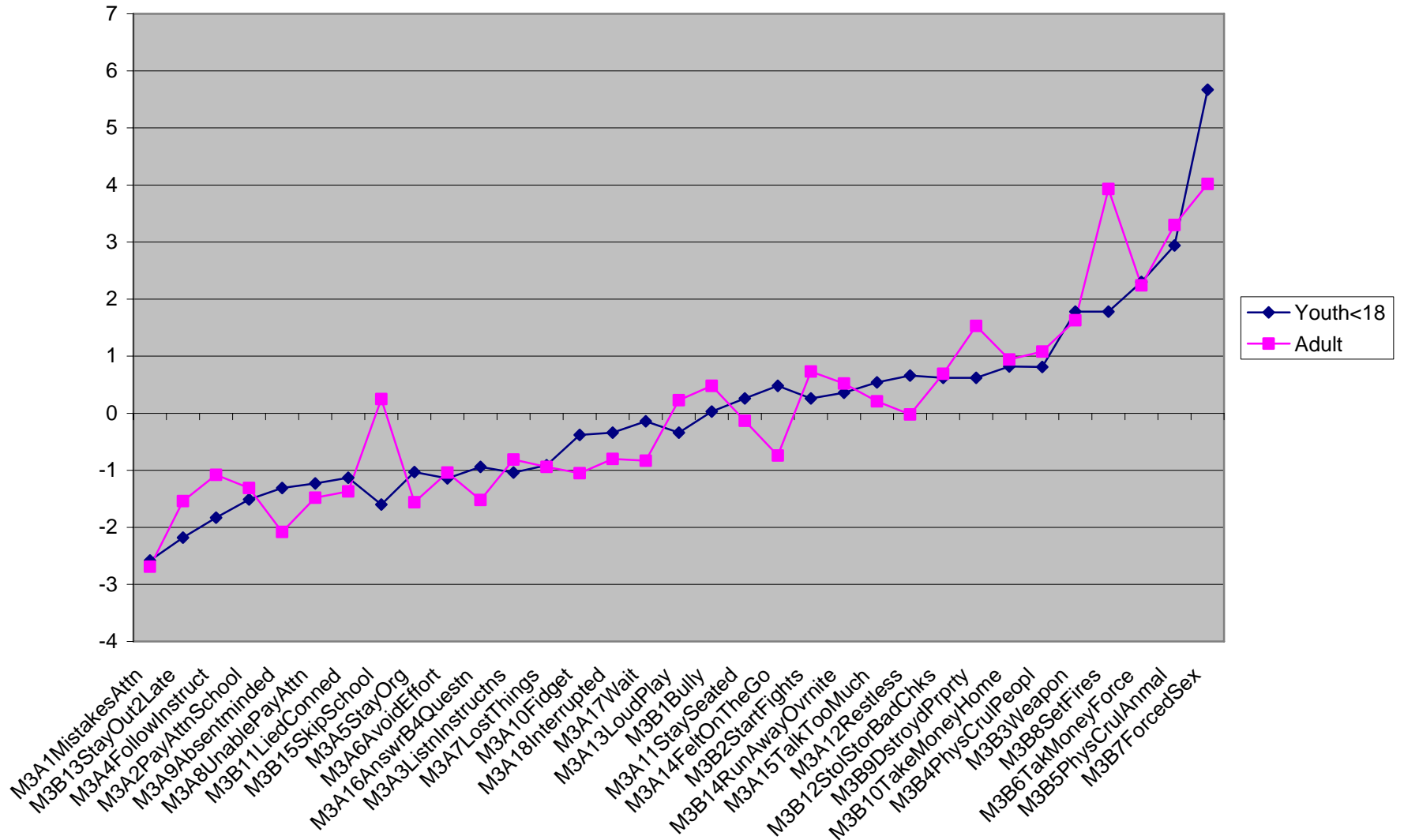


Figure 3. BCS Race DIF

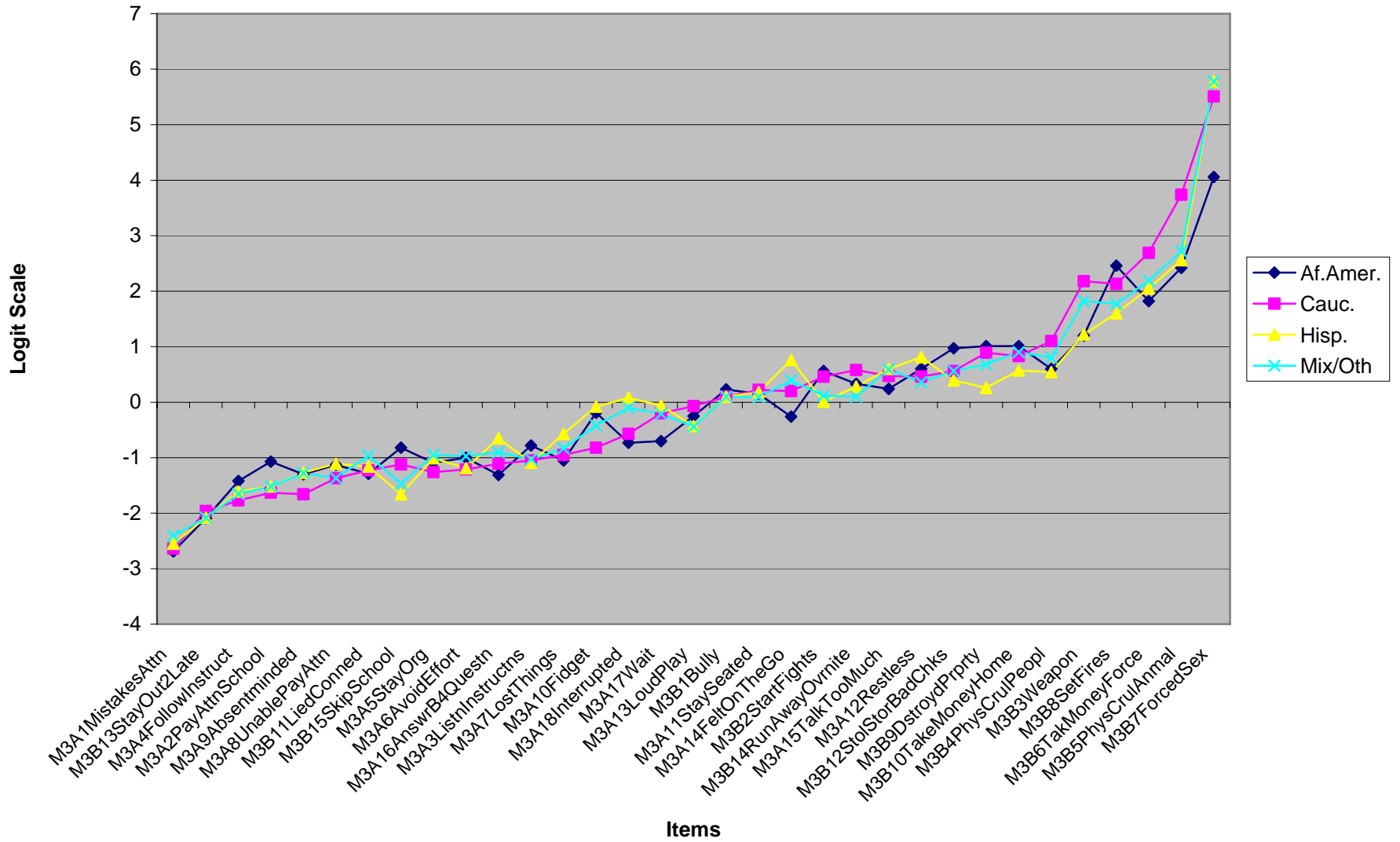


Figure 4. BCS Primary Drug DIF

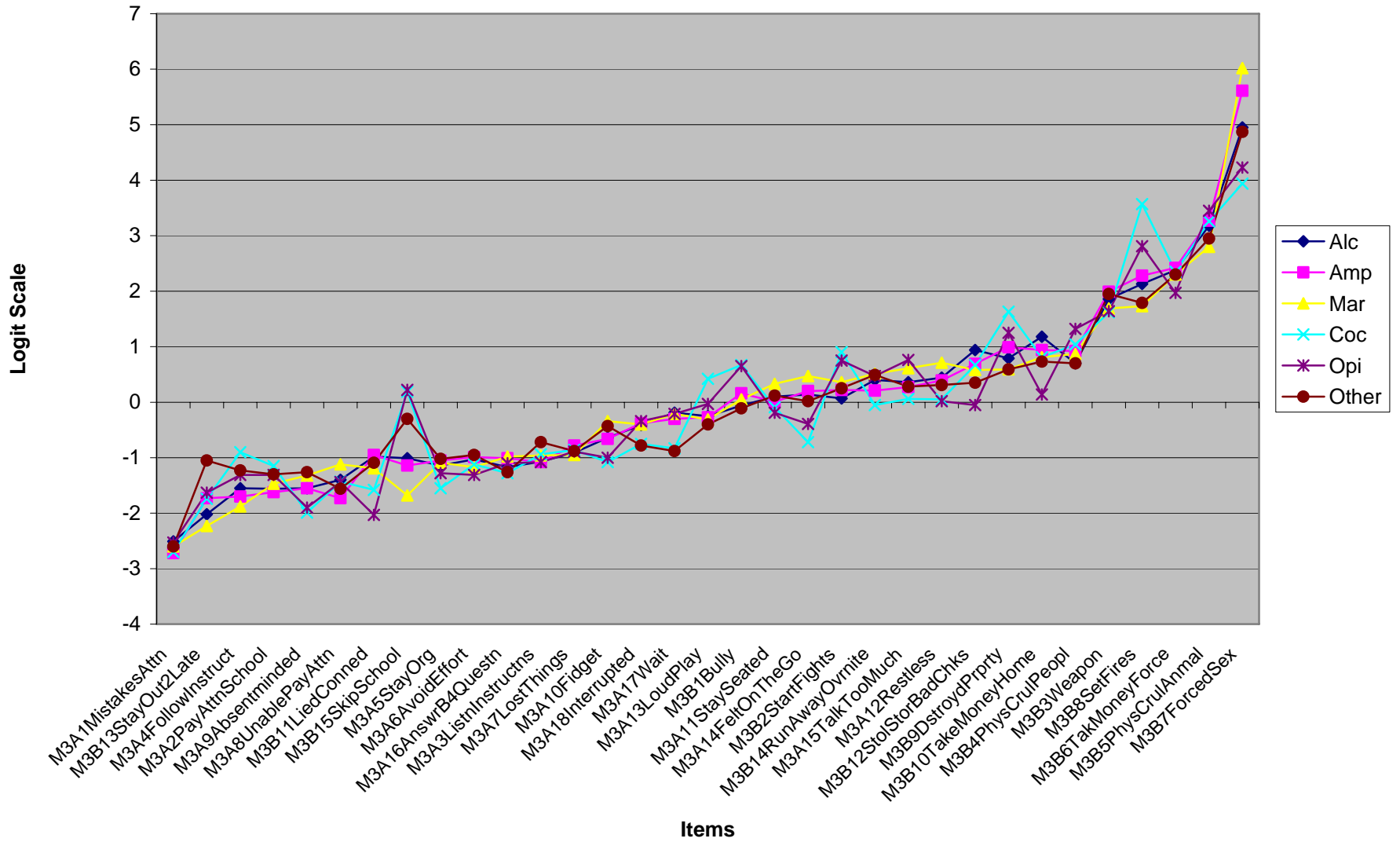


Table 12. BCS Item Measures by Demographic Groups (Items Listed in Severity Order)

Item	Gender		Age		Race				Primary Drug					
	Male	Female	Youth<18	Adult	AA	Cauc.	Hisp.	Mix/Oth	Alc	Amp	Mar	Coc	Opiate	Other
M3A1MistakesAttn	-2.63	-2.55	-2.58	-2.69	-2.69	-2.64	-2.55	-2.41	-2.51	-2.72	-2.6	-2.7	-2.53	-2.6
M3B13StayOut2Late	-2.1	-1.87	-2.18	-1.54	-2.09	-1.96	-2.08	-2.08	-2.02	-1.73	-2.23	-1.74	-1.63	-1.05
M3A4FollowInstruct	-1.77	-1.41	-1.83	-1.08	-1.42	-1.77	-1.61	-1.65	-1.55	-1.7	-1.88	-0.9	-1.31	-1.23
M3A2PayAttnSchool	-1.47	-1.49	-1.51	-1.31	-1.07	-1.63	-1.52	-1.52	-1.56	-1.62	-1.47	-1.15	-1.31	-1.3
M3A9Absentminded	-1.33	-1.77	-1.31	-2.08	-1.3	-1.66	-1.27	-1.28	-1.55	-1.55	-1.32	-1.99	-1.9	-1.26
M3A8UnablePayAttn	-1.16	-1.57	-1.23	-1.48	-1.13	-1.37	-1.11	-1.37	-1.4	-1.73	-1.12	-1.43	-1.44	-1.56
M3B11LiedConned	-1.19	-1.22	-1.13	-1.37	-1.29	-1.23	-1.16	-0.96	-0.98	-0.95	-1.19	-1.58	-2.03	-1.09
M3B15SkipSchool	-1.37	-0.77	-1.6	0.25	-0.82	-1.12	-1.66	-1.47	-1.01	-1.14	-1.68	0.19	0.22	-0.3
M3A5StayOrg	-1.14	-1.17	-1.03	-1.56	-1.09	-1.26	-1.01	-0.95	-1.14	-1.05	-1.09	-1.55	-1.28	-1.02
M3A6AvoidEffort	-1.12	-1.12	-1.14	-1.04	-1	-1.21	-1.19	-0.97	-1.03	-0.99	-1.17	-1.12	-1.31	-0.95
M3A16AnswrB4Questn	-1	-1.19	-0.94	-1.52	-1.31	-1.11	-0.65	-0.9	-1.17	-1.01	-0.98	-1.28	-1.1	-1.26
M3A3ListnInstructns	-0.99	-1.01	-1.04	-0.81	-0.78	-1.05	-1.09	-1.03	-1.07	-1.08	-0.96	-0.94	-1.08	-0.72
M3A7LostThings	-0.88	-0.96	-0.91	-0.94	-1.05	-0.95	-0.57	-0.83	-0.91	-0.78	-0.95	-0.86	-0.89	-0.88
M3A10Fidget	-0.36	-0.88	-0.38	-1.05	-0.19	-0.82	-0.08	-0.42	-0.64	-0.66	-0.34	-1.08	-1	-0.43
M3A18Interrupted	-0.39	-0.56	-0.34	-0.8	-0.73	-0.57	0.08	-0.1	-0.41	-0.38	-0.4	-0.75	-0.34	-0.78
M3A17Wait	-0.24	-0.41	-0.14	-0.83	-0.7	-0.2	-0.06	-0.21	-0.19	-0.3	-0.21	-0.83	-0.21	-0.88
M3A13LoudPlay	-0.22	-0.22	-0.34	0.23	-0.25	-0.07	-0.43	-0.44	-0.25	-0.27	-0.31	0.42	-0.03	-0.4
M3B1Bully	0.02	0.33	0.03	0.48	0.23	0.1	0.1	0.1	-0.06	0.16	0.05	0.67	0.65	-0.11
M3A11StaySeated	0.21	0.1	0.26	-0.13	0.15	0.22	0.18	0.09	0.1	0.01	0.33	-0.09	-0.19	0.12
M3A14FeltOnTheGo	0.31	-0.01	0.48	-0.74	-0.26	0.2	0.76	0.39	0.14	0.2	0.47	-0.72	-0.39	0.02
M3B2StartFights	0.3	0.48	0.26	0.73	0.56	0.46	0.01	0.12	0.07	0.22	0.36	0.9	0.75	0.25
M3B14RunAwayOvrnite	0.68	-0.11	0.36	0.52	0.33	0.58	0.28	0.1	0.4	0.21	0.51	-0.05	0.47	0.49
M3A15TalkTooMuch	0.73	-0.01	0.54	0.21	0.24	0.47	0.6	0.59	0.36	0.27	0.61	0.06	0.76	0.28
M3A12Restless	0.61	0.32	0.66	-0.02	0.6	0.46	0.81	0.36	0.44	0.39	0.71	0.05	0.02	0.31
M3B12StolStorBadChks	0.62	0.65	0.62	0.69	0.97	0.56	0.39	0.56	0.94	0.69	0.58	0.67	-0.05	0.35
M3B9DstroydPrprty	0.53	1.36	0.62	1.53	1.01	0.89	0.26	0.68	0.79	0.99	0.59	1.63	1.25	0.59
M3B10TakeMoneyHome	0.69	1.17	0.82	0.94	1.01	0.83	0.57	0.9	1.18	0.94	0.81	0.81	0.14	0.73
M3B4PhysCrulPeopl	0.78	1.03	0.81	1.08	0.6	1.1	0.54	0.8	0.7	0.92	0.87	1.04	1.32	0.7
M3B3Weapon	1.52	2.26	1.78	1.63	1.2	2.18	1.22	1.82	1.86	1.99	1.69	1.62	1.64	1.95

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

Item	Gender		Age		Race				Primary Drug					
	Male	Female	Youth<18	Adult	AA	Cauc.	Hisp.	Mix/Oth	Alc	Amp	Mar	Coc	Opiate	Other
M3B8SetFires	1.79	2.62	1.78	3.93	2.46	2.13	1.6	1.77	2.13	2.28	1.73	3.57	2.81	1.79
M3B6TakMoneyForce	2.13	2.66	2.3	2.24	1.82	2.69	2.05	2.19	2.38	2.42	2.3	2.35	1.97	2.3
M3B5PhysCrulAnmal	2.65	4	2.94	3.3	2.42	3.74	2.56	2.72	3.18	3.27	2.8	3.26	3.45	2.95
M3B7ForcedSex	5.45	4.68	5.67	4.02	4.06	5.51	5.78	5.78	4.95	5.61	6.02	3.94	4.23	4.87

Person Fit Group Analysis for the BCS

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. The response patterns for each fit group are discussed relative to the response pattern for the overall mean).

This information should enable us to interpret certain unusual patterns of scores more appropriately, e.g., low scorers who do not have attention difficulty, such as paying attention or staying organized, but who are actually at high risk because of their conduct disorders, such as destroying property and taking money by force. 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 ≤ 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 group mean.

In Figure 8, the L/M infit and high (H) outfit group consists of 508 people (7%) (termed Atypical Type 1). Relative to the overall mean, Atypical Type 1 response patterns on the *BCS* are as expected on inattentive disorder symptoms, slightly higher than expected for hyperactivity/impulsivity symptoms, and markedly higher than expected for conduct disorders with the exception of the most serious, *M3B7ForcedSex*, where the response pattern is slightly higher than expected. The overall score may underestimate the severity of this group's behavioral complexity.

In Figure 9, the H infit and L/M outfit group consists of only 60 people (1%) (termed Atypical Type 2). Relative to the overall mean, Atypical Type 2 response patterns on the *BCS* are lower than expected on several inattentive disorder symptoms but markedly higher than expected on both hyperactivity/impulsivity and on most of the conduct disorder symptoms, with the exception of the five most serious conduct disorder symptoms where their response pattern is slightly higher than expected for *M3B3Weapons*, *M3B8Set Fires*, *M3B6TakeMoneyForce*, and *M3B5 PhysCruAnimal* and as expected for *M3B7ForcedSex*. The overall score may underestimate the severity of this group's behavioral complexity.

In Figure 10, the H infit and H outfit group consists of 431 people (6%) (termed Atypical Type 3). Relative to the overall mean, Atypical Type 3 response patterns on the *BCS* are lower than expected for several inattentive disorder symptoms, with the exception of *M3A5StayOrg* where the response pattern indicates that this group has less trouble than expected staying organized. Their response patterns are markedly higher than expected on both hyperactivity and especially on conduct disorders, with the exception of *M3B7ForcedSex*, where this group's responses are slightly higher than expected. Compared to the Atypical Type 2 group, this group has more of a problem staying organized, are not as extreme in their hyperactivity/impulsivity but more

extreme in conduct disorder symptoms. The overall score may underestimate the severity of this group's behavioral complexity.

In summary, BCS underestimates severity in all three atypical groups. Relative to what was expected based on their BCS scores, people in Atypical Type 1 group will tend to be higher on hyperactivity and conduct disorder items. People in the Atypical Types 2 and 3 groups will tend to be lower than expected on the inattentive and higher than expected on hyperactivity and conduct disorder.

Figure 6. BCS Fit Group Pie Chart

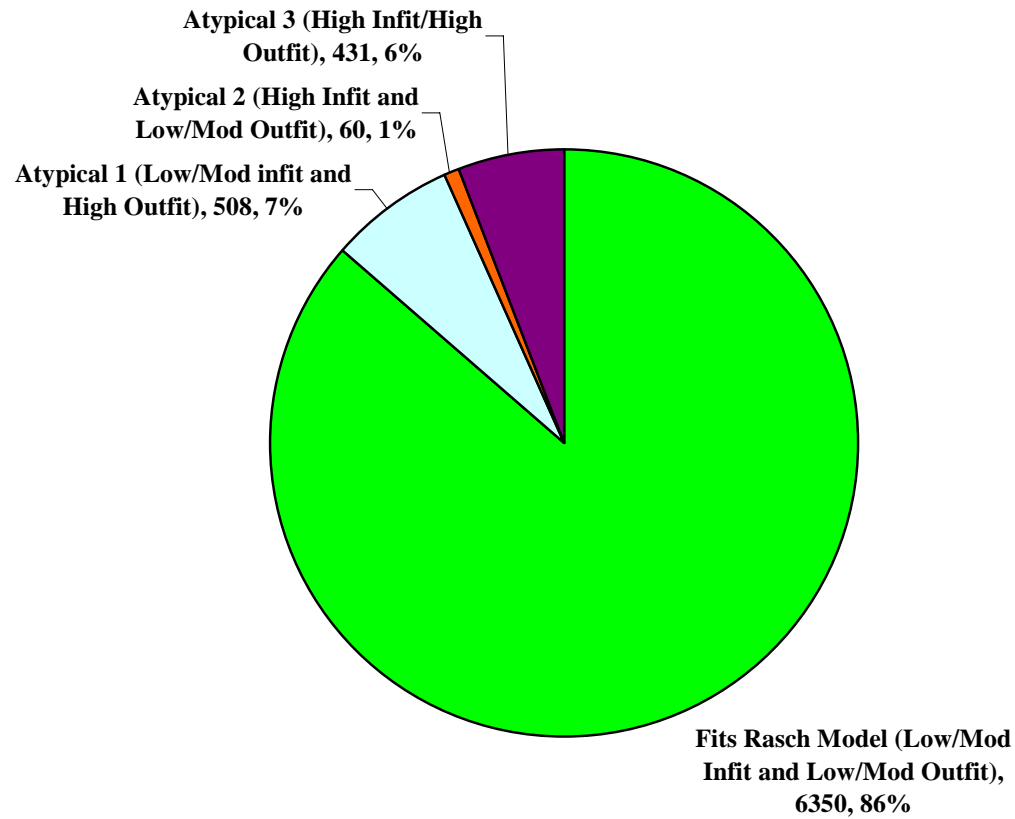


Figure 7. BCS Overall Item Mean vs. Fit Group Item Mean
Fits the Rasch Model
n=6350; 86%, Low/Mod Infit and Low/Mod Outfit
 (slightly lower than expected on conduct disorder)

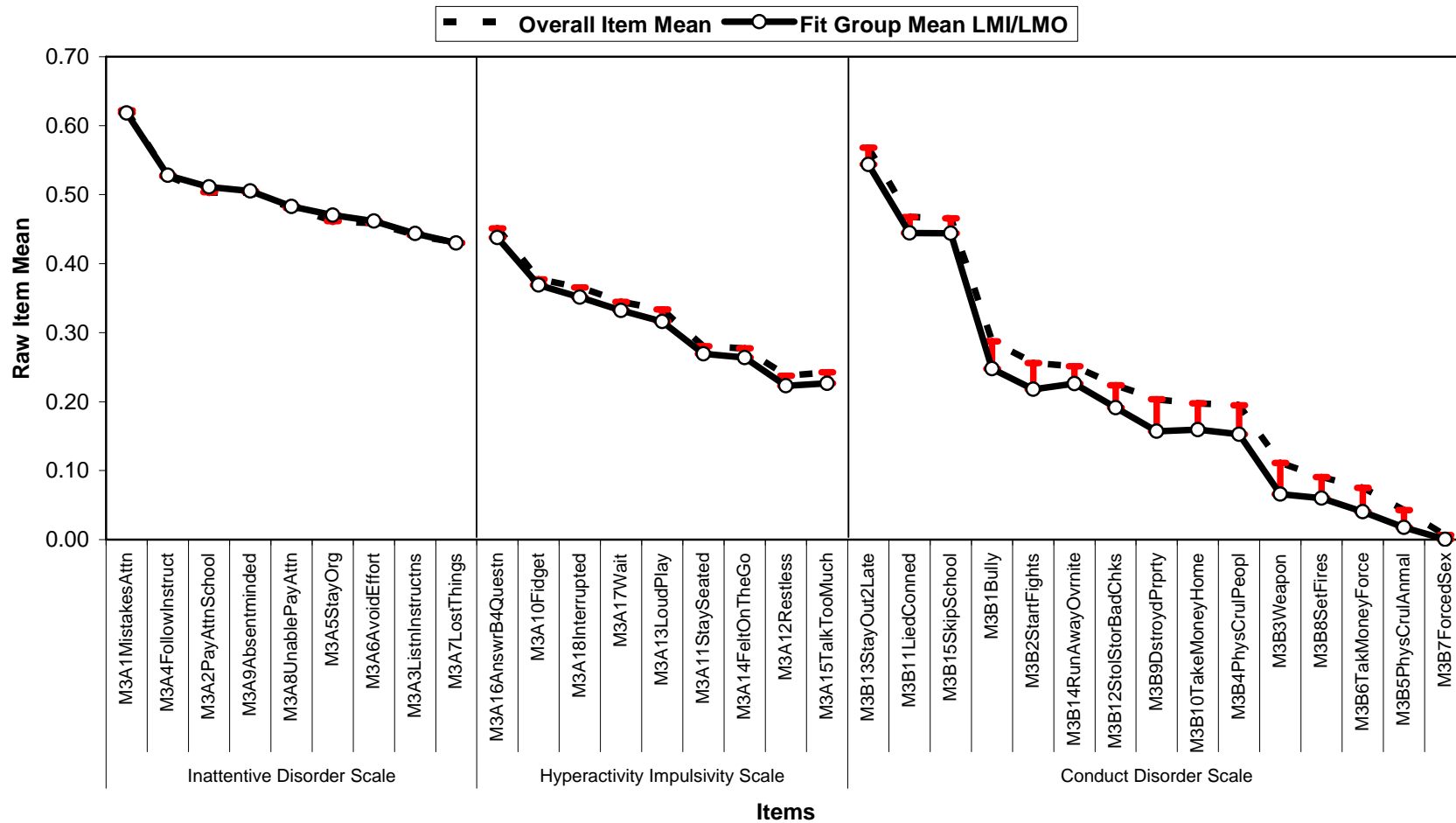


Figure 8. BCS Overall Item Mean vs. Fit Group Item Mean
Atypical Type 1
(n=508; 7%, Low/Mod Infit and High Outfit)
(slightly higher than expected on hyperactivity/impulsivity and markedly higher on conduct disorder)

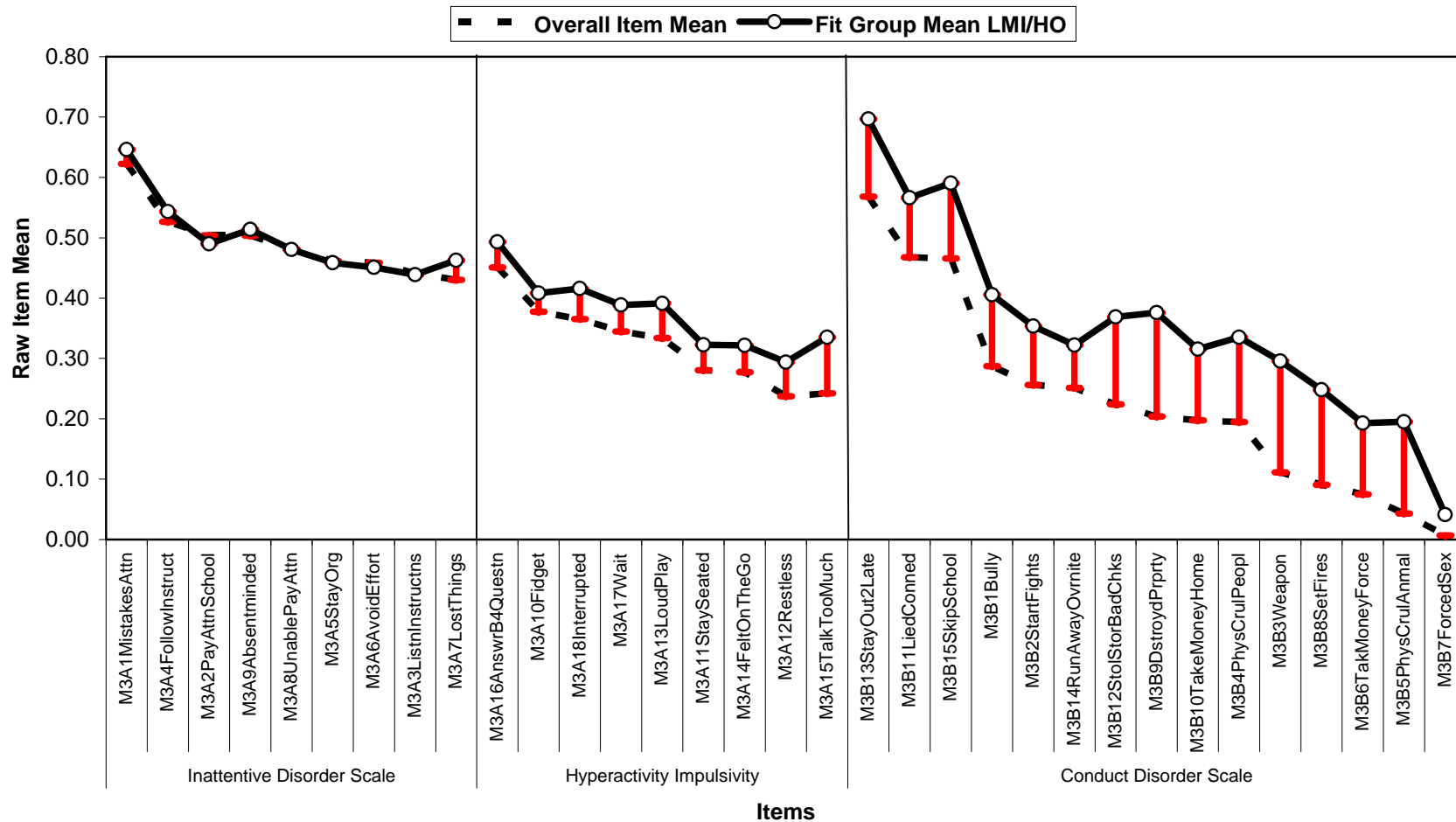


Figure 9. BCS Overall vs. Fit Group Item Mean
Atypical Type 2
(n=60; 1%, High Infit and Low/Mod Outfit)
(lower than expected on most inattentive/disorder items; markedly higher on hyperactivity/impulsivity and on all but most serious conduct disorders)

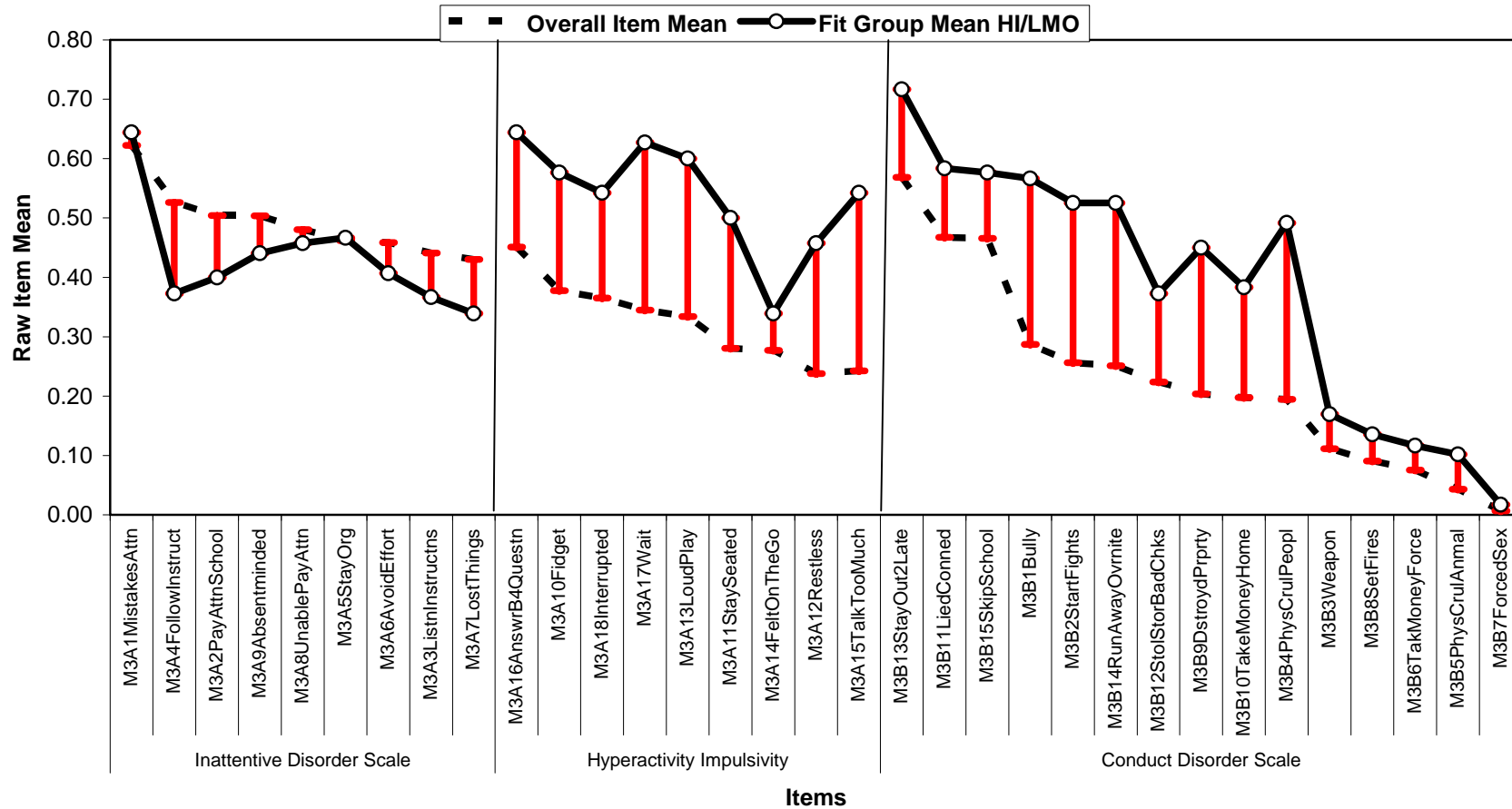
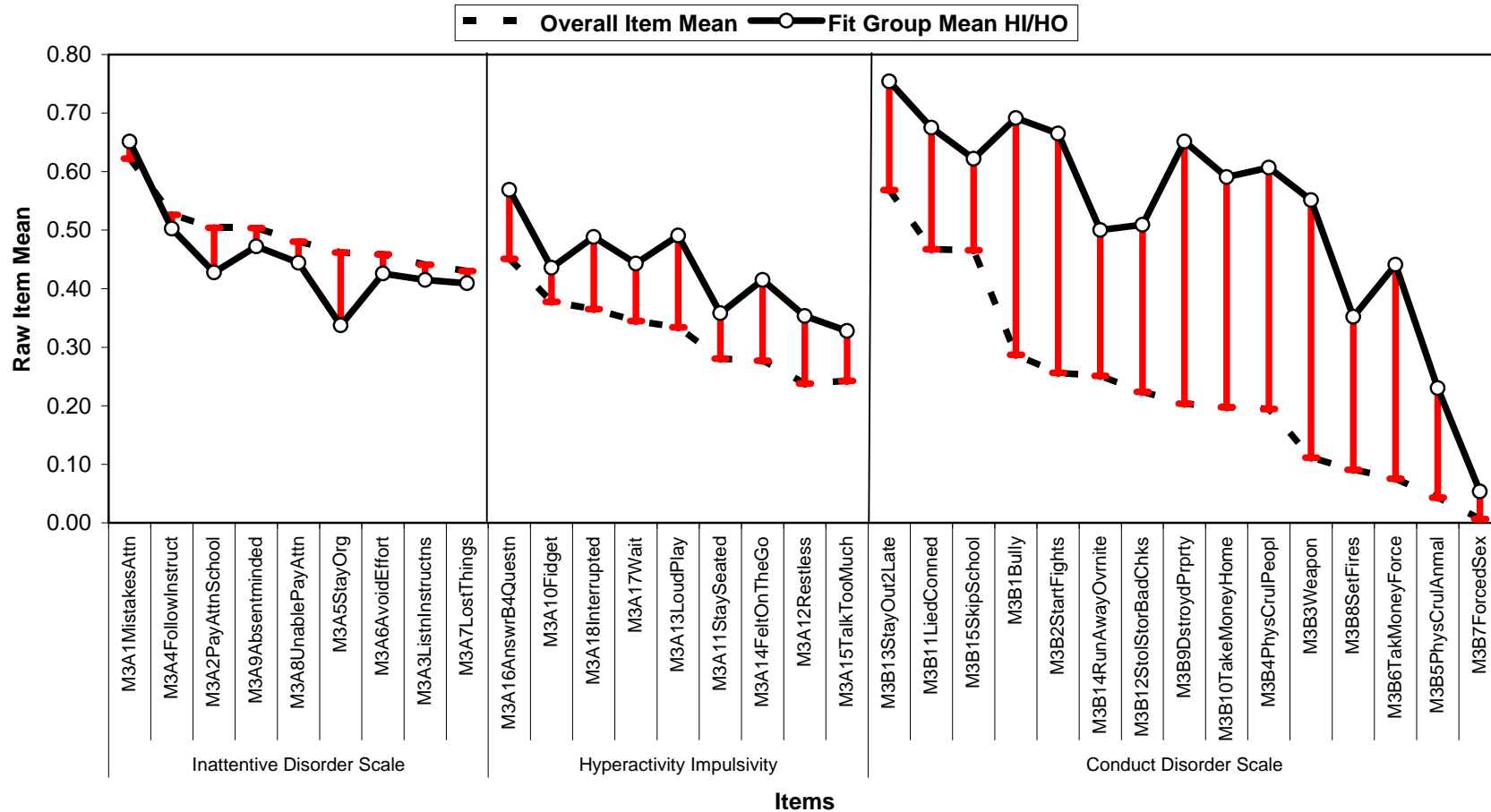


Figure 10. BCS Overall vs. Fit Group Item Mean
Atypical Type 3
(n=431; 6%, High Infit and High Outfit)
(lower than expected on most inattentive disorder items; higher than expected on
hyperactivity/impulsivity and markedly higher on conduct disorder)



Recommended Actions and Deliberations

The *BCS* 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.

- Significant DIF (i.e., $> .5$ SD = .78 logits) occurred in 4 items for males vs. females, 5 items for youth vs. adults, 10 items for race when using African American as the reference group, and 9 items for primary substances when using alcohol as the referent.
- No items misfit using the criterion of .75-1.33 MNSQ for both infit and outfit.
- The results suggest the possibility of some important gender differences. It was somewhat more common for females than for males to endorse *Forced Sex*. Qualitative probing suggests that females endorsing this item were thinking of instances when they pressured or seduced males or females into having sex with them. Males, in contrast, were more likely to associate this with rape, and thus were less likely to endorse it. Further evaluation of alternative wording, e.g., using words such as “seduced or pressured against their will”, may be useful to improve this item.
- Other DIF analyses did not discern obvious imbalance that could result in biased measurement, but they did reveal some issues that might be of theoretical interest.
- The three atypical groups, Atypical Groups 1, 2, and 3, tend to have scores that underestimate their behavioral complexity symptoms. We recommend flagging these three groups for clinicians in the evaluation/validity concerns section of the *GAIN*. Recommendation and Referral Summary (GRRS) as:
 - Atypical Type 1 response patterns on the *Behavioral Complexity Scale* (relative to the group mean) were higher than expected on hyperactivity/impulsivity and on conduct disorder (higher severity items).
 - Atypical Type 2 response patterns on the *BCS* (relative to the group mean) were lower than expected on inattentive disorder symptoms but markedly higher than expected on both hyperactivity/impulsivity and conduct disorder symptoms.
 - Atypical Type 3 response patterns on the *BCS* (relative to the group mean) were lower than expected on inattentive disorder symptoms but markedly higher than expected on both hyperactivity/impulsivity and conduct disorder symptoms (particularly the highest severity symptoms).
- More work on construct validity would be helpful to understand these fit groups better and to ensure proper interpretation of measures.

References

- American Psychiatric Association (2000). *Diagnostic and Statistical Manual (DSM-IV-TR, 4th Ed.* Washington, DC: American Psychiatric Association.
- Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences.* (2nd 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.
- 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. (2007). *Winsteps Rasch Measurement* (Version 3.64.2). www.winsteps.com. Author.
- Norman, G.R., Sloan, J.A., & Wywich, K.W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care, 41*(5), 582-592.
- 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.

June 2, 2009