Selected Findings
Adult Wellness Assessment

Second Pilot Study

Ann Doucette, Ph.D.

December 2008

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON DC

The Adult Wellness Assessment (AWA)

The value of behavioral healthcare treatment is characterized in terms of the beneficial change consumers/patients experience as a result of receiving care. As a consequence there is a growing demand to objectively monitor change through assessment; and, an increased emphasis given to the primacy of the consumer/patient perspective in monitoring their outcomes. Assessment approaches range from lengthy, multi-scale diagnostic instrumentation to brief measurement of global health and psychological status. While the lengthy instruments provide more precision in clinical application, such as diagnostic support, they are far too burdensome for consumers/patients to routinely complete during the course of their treatment.

The *Adult Wellness Assessment* (AWA) is purposefully designed to be brief, offering respondents (consumers/patients) an opportunity to provide feedback on their general emotional and psychological status. Items included in the AWA were intentionally developed and selected to capture broad areas (e.g., anxiety, depression, sleep, role function, etc.) in order to detect changes in global levels of emotional distress. Though the item content of the AWA reflects characteristics of major psychiatric disorders, it is not meant to be used as a substitute for more lengthy and comprehensive diagnostic self-report measures or clinician rating scales. In summary, the AWA measure is meant to be used as a barometer that is indicative of general improvement, stability, and in some cases increasing distress.

Financial Disclosure: This study was fully funded by United Behavioral Health. Dr. Ann Doucette was paid in full for conducting the analysis of the Adult Wellness Assessment under a contract as an academic consultant from George Washington University that included statements on confidential information and intellectual property rights, as well as academic freedom and clauses against publication censorship.

Sample Demographics – Descriptive Statistics

The following tables provide descriptive information on the clinical sample used in the psychometric analysis of the UBH *Wellness Assessment*. Baseline data were provided on 127,012 individuals. An examination of the data in terms of completeness yielded a sample of 99,319 adult records for individuals 18 years and older. The baseline assessment was conducted for the most part during the first or second session. However approximately 29% of the respondents competed the assessment during the third session of treatment or later.

Respondent Age	Baseline	Follow-UP
18-24 years	9.2%	6.0%
25-34 years	24.8%	18.9%
35-44 years	29.2%	26.3%
45-59 years	30.9%	38.1%
60-75 years	6.7%	10.1%
76 and older	0.3%	0.6%
$\chi^2(5, N = 134{,}243) = 5.540, p > .000$		

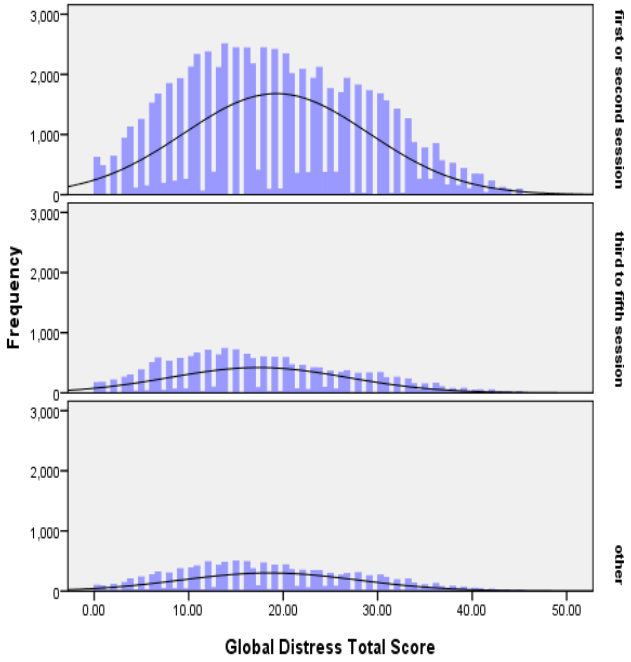
Older beneficiaries are more likely to respond to AWA Follow-up request.

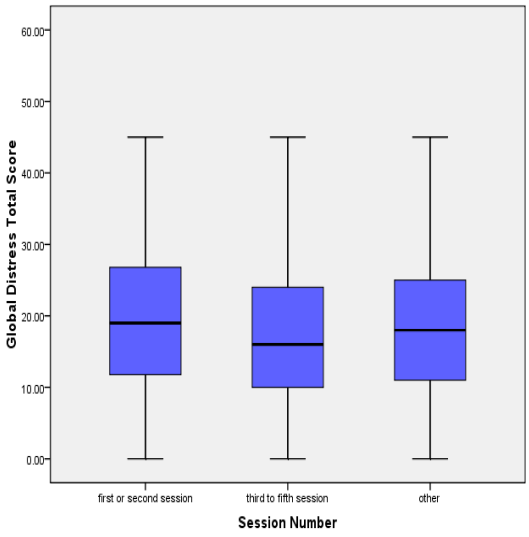
Respondent Gender	Baseline	Follow-UP
Female	66%	71%
Male	34%	29%
$\chi^2(1, N = 132{,}876) = 9.393, p > .000$		

Females are more likely to respond to AWA Follow-up requests.

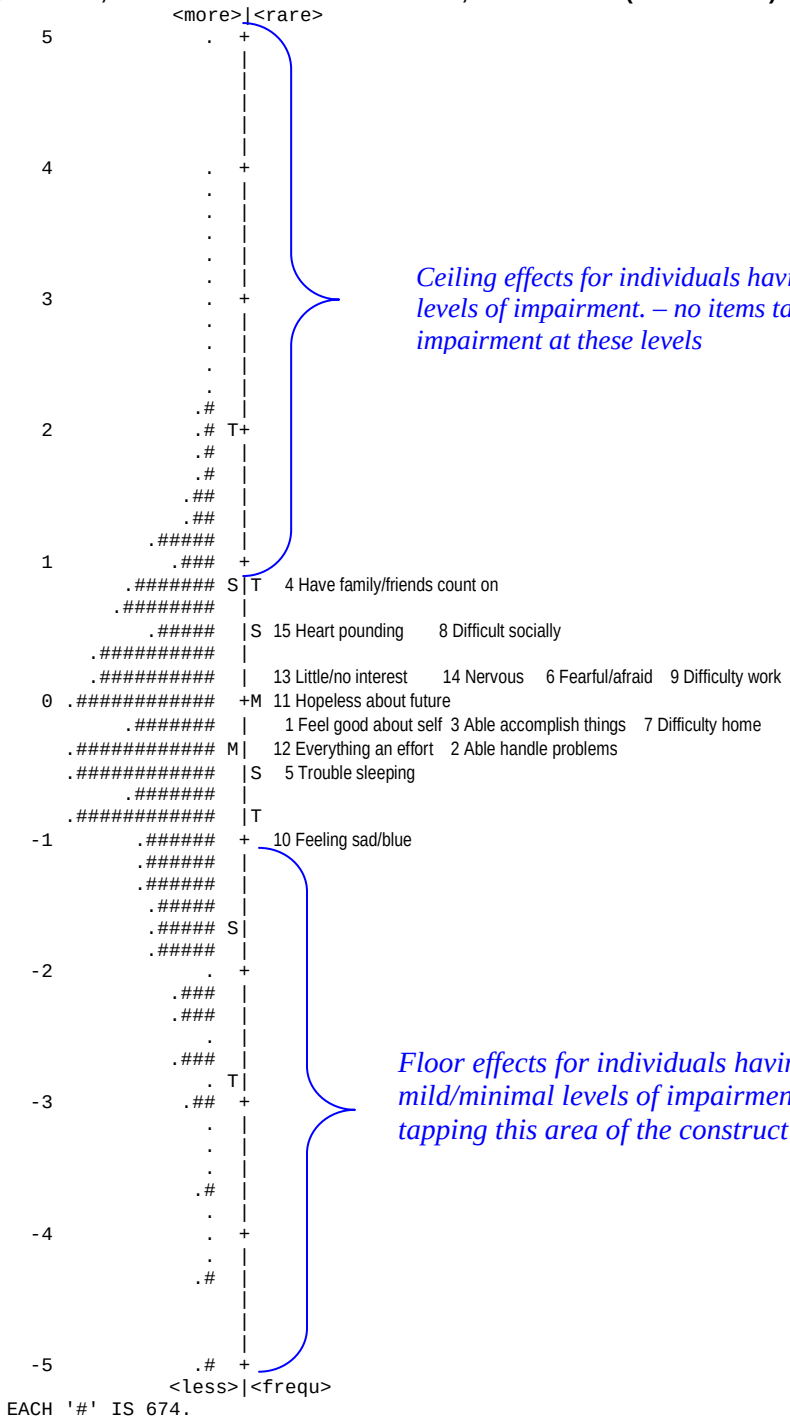
Session Completing Baseline Assessment

	Sample	Mean	SD	Skewness	Kurtosis
1st - 2nd session	70.9% (N = 70,266)	19.28	9.82	.218/.008	-.701/.018
3rd to 5th session	17.1% (N = 16,958)	17.39	9.45	.427/.019	-.457/.038
Other	12.1% (N = 12095)	18.53	9.41	.314/.022	-.536/.045





)15 ITEMS, 60 CATAGORIES MEASURED: 71,505 PERSONS (SESSION 1-2)

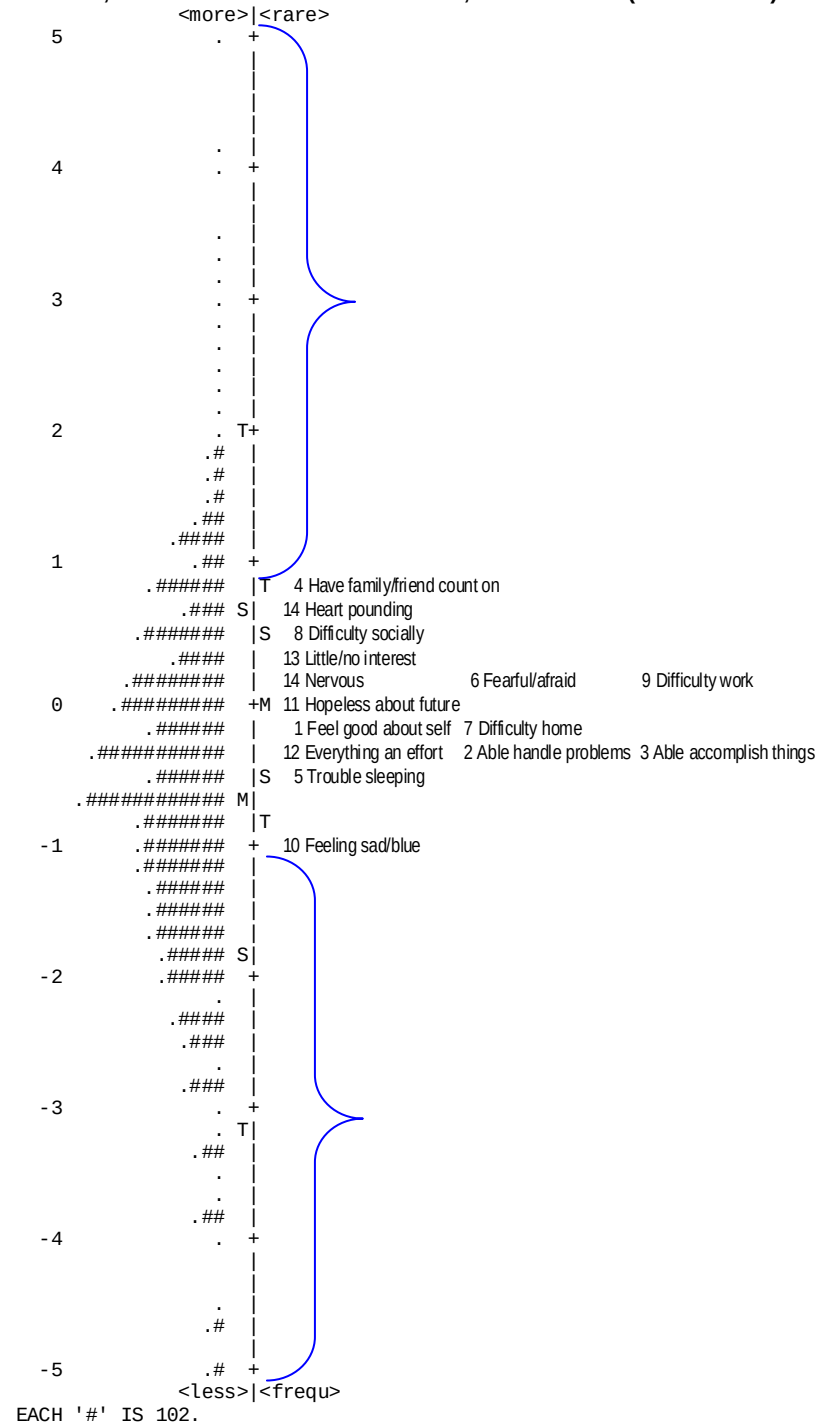


EACH '#' IS 674.

Ceiling effects for individuals having high levels of impairment. – no items tapping impairment at these levels

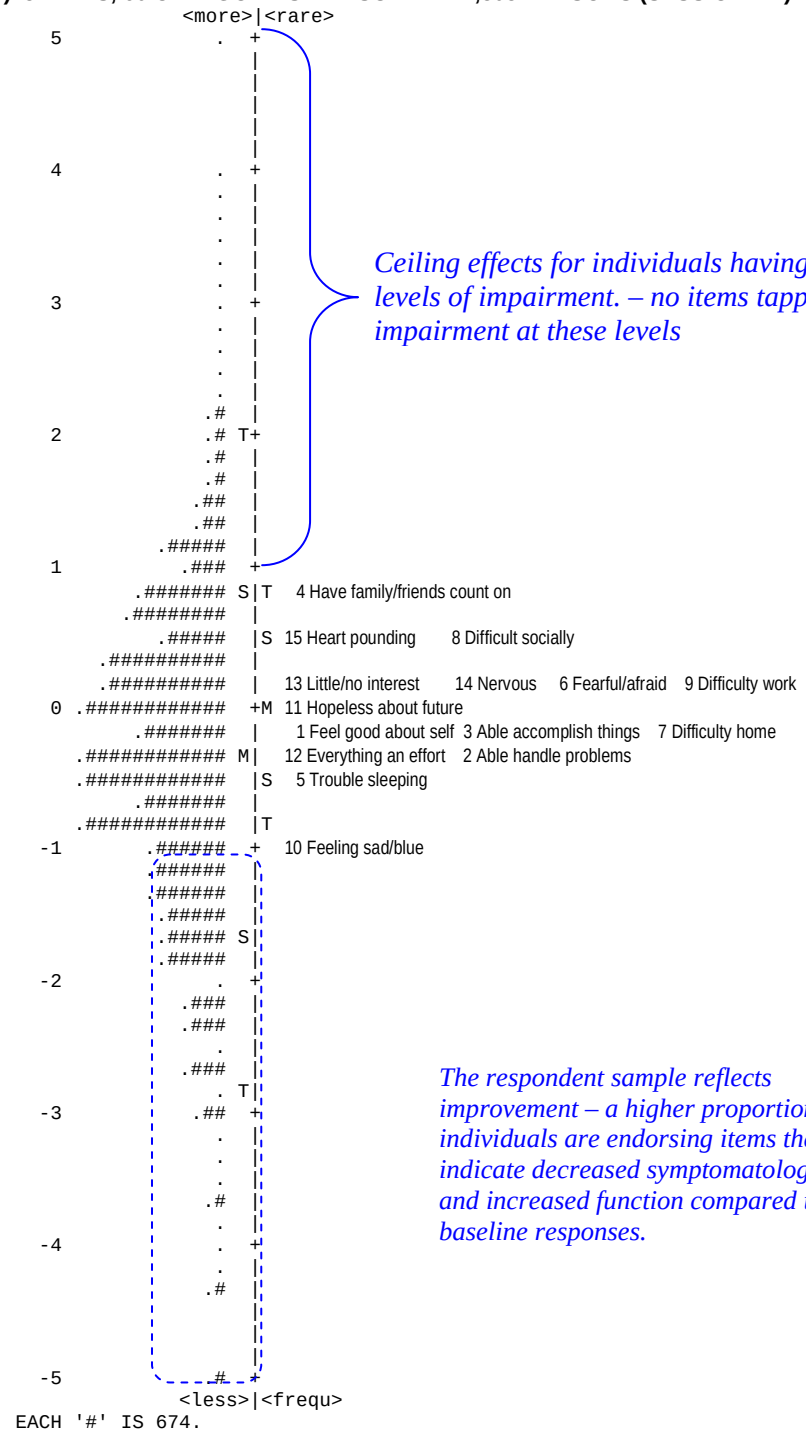
Floor effects for individuals having mild/minimal levels of impairment – no items tapping this area of the construct

15 ITEMS, 60 CATAGORIES MEASURED: 17,166 PERSONS (SESSION 3-5)



EACH '#' IS 102.

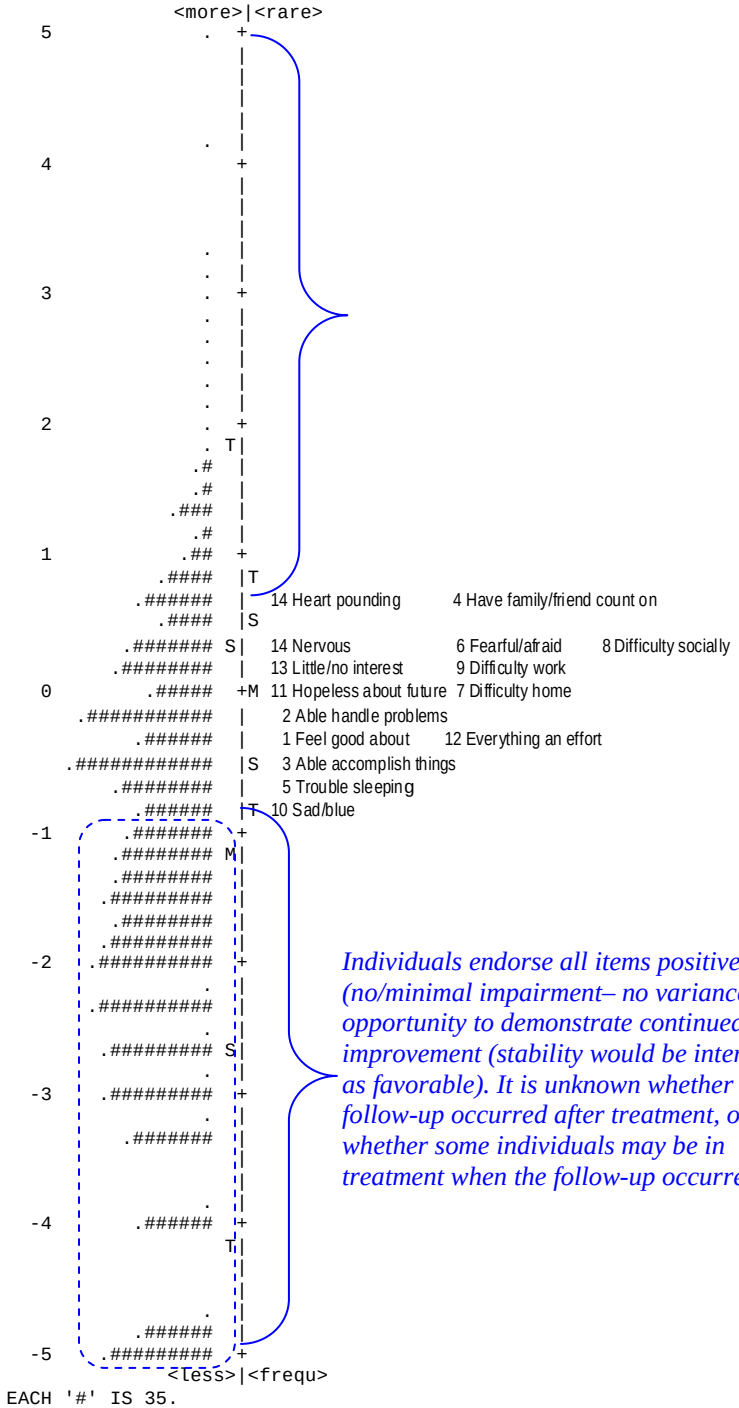
)15 ITEMS, 60 CATAGORIES MEASURED: 71,505 PERSONS (SESSION 1-2)



Ceiling effects for individuals having high levels of impairment. – no items tapping impairment at these levels

The respondent sample reflects improvement – a higher proportion of individuals are endorsing items that indicate decreased symptomatology and increased function compared to baseline responses.

15 ITEMS, 60 CATAGORIES MEASURED: 7,755 (FOLLOW-UP)



Individuals endorse all items positively (no/minimal impairment– no variance, no opportunity to demonstrate continued improvement (stability would be interpreted as favorable). It is unknown whether the follow-up occurred after treatment, or whether some individuals may be in treatment when the follow-up occurred.

TABLE 10.1 E: INPUT: 71,505 PERSONS (SESSION 1-2) -- 15 ITEMS MEASURED: 60 CATS [PERSON: REAL SEP: 2.57 REL: .87]
ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBSE CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	ESTIM DISCR	P-VALUE	ITEM	G
7	**102053**	**70002**	**-.27**	**.00**	**1.52**	**9.9**	**1.75**	**9.9**	**A .39**	**37.1**	**44.5**	**.18**	**1.46**	**7 Difficulty home**	**0**
4	**57006**	**68483**	**.93**	**.01**	**1.41**	**9.9**	**1.41**	**9.9**	**B .31**	**50.3**	**57.9**	**.55**	**.83**	**4 Have family/friends count on**	**0**
5	**109321**	**70363**	**-.43**	**.00**	**1.26**	**9.9**	**1.38**	**9.9**	**C .51**	**41.3**	**44.9**	**.60**	**1.55**	**5 Trouble sleeping**	**0**
15	67682	70234	.51	.00	1.12	9.9	1.17	9.9	D .54	47.7	49.3	.85	.96	15 Heart pounding	0
9	73403	67147	.24	.00	1.09	9.9	1.17	9.9	E .56	47.3	46.9	.88	1.09	9 Difficulty work	0
14	83194	69593	.14	.00	1.06	9.9	1.09	9.9	F .58	45.6	46.7	.92	1.20	14 Nervous anxious	0
6	81323	69982	.19	.00	.96	-8.4	.95	-7.2	G .63	48.5	46.6	1.07	1.16	6 Fearful afraid	0
3	95154	69870	-.12	.01	.92	-9.9	.91	-9.9	H .59	60.2	57.2	1.09	1.36	3 Able accomplish things	0
8	69701	70043	.49	.00	.89	-9.9	.86	-9.9	g .64	53.2	49.1	1.16	1.00	8 Difficulty socially	0
1	96505	70002	-.16	.01	.86	-9.9	.85	-9.9	f .63	61.4	55.9	1.17	1.38	1 Feel good about self	0
2	102121	69769	-.41	.01	.85	-9.9	.84	-9.9	e .63	63.0	58.1	1.17	1.46	2 Able deal w/ problems	0
12	101396	70020	-.27	.00	.80	-9.9	.79	-9.9	d .70	51.9	46.0	1.32	1.45	12 Everything an effort	0
10	129507	70415	-.98	.01	.79	-9.9	.77	-9.9	c .69	53.3	47.6	1.30	1.84	10 Feeling sad/blue	0
13	81258	70099	.21	.00	.78	-9.9	.75	-9.9	b .70	53.2	47.1	1.32	1.16	13 Little/no interest things	0
11	93403	70297	-.06	.00	.76	-9.9	.73	-9.9	a .72	51.3	45.3	1.38	1.33	11 Feeling hopeless future	0
MEAN	89535.1	69755	.00	.01	1.00	-1.9	1.03	-1.8		51.0	49.5				
S.D.	18084.3	826.5	.45	.00	.23	9.6	.29	9.6		6.8	4.9				

TABLE 10.1 E: INPUT: 17,079 PERSONS (SESSION 3-5) -- 15 ITEMS MEASURED: 60 CATS [PERSON: REAL SEP: 2.65 REL: .88]
ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBSE CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	ESTIM DISCR	P-VALUE	ITEM	G
7	**21610**	**16912**	**-.24**	**.01**	**1.43**	**9.9**	**1.62**	**9.9**	**A .45**	**41.7**	**48.3**	**.34**	**1.28**	**7 Difficulty at home**	**0**
5	**23591**	**16968**	**-.44**	**.01**	**1.31**	**9.9**	**1.45**	**9.9**	**B .50**	**42.3**	**47.8**	**.51**	**1.39**	**5 Trouble sleeping**	**0**
4	**13840**	**16565**	**.78**	**.01**	**1.40**	**9.9**	**1.39**	**9.9**	**C .35**	**52.6**	**60.0**	**.56**	**.84**	**4 Have family/friends to count on**	**0**
15	13792	16935	.62	.01	1.15	9.9	1.23	9.9	D .54	52.3	54.2	.83	.81	14 Heart pounding	0
9	15695	16134	.23	.01	1.11	9.5	1.18	9.9	E .57	51.2	51.1	.86	.97	9 Difficulty at work	0
14	18443	16871	.12	.01	1.05	4.6	1.07	5.3	F .60	49.2	50.0	.93	1.09	14 Nervous anxious	0
6	17602	16901	.19	.01	.94	-5.3	.93	-5.6	G .64	52.8	50.9	1.08	1.04	6 Fearful afraid	0
3	22272	16866	-.26	.01	.89	-9.9	.88	-9.9	H .62	64.3	60.7	1.12	1.32	3 Able to accomplish things	0
8	15257	16947	.48	.01	.89	-9.9	.88	-8.5	g .66	56.8	53.3	1.14	.90	8 Difficulty socially	0
2	22660	16867	-.37	.01	.84	-9.9	.82	-9.9	f .64	67.5	62.6	1.17	1.34	2 Able to deal with problems	0
1	21930	16896	-.24	.01	.83	-9.9	.82	-9.9	e .66	66.2	60.4	1.18	1.30	1 Feel good about self	0
12	22359	16927	-.31	.01	.80	-9.9	.80	-9.9	d .71	55.4	49.4	1.29	1.32	12 Everything is an effort	0
10	27559	16973	-.92	.01	.80	-9.9	.79	-9.9	c .70	55.5	50.1	1.28	1.62	10 Feeling sad or blue	0
13	16956	16905	.28	.01	.79	-9.9	.76	-9.9	b .71	56.8	51.2	1.29	1.00	13 Little or not interest things	0
11	18843	16960	.08	.01	.77	-9.9	.75	-9.9	a .72	55.4	49.6	1.32	1.11	11 Feeling hopeless future	0
MEAN	19493.9	16842	.00	.01	1.00	-2.1	1.02	-1.9		54.7	53.3				
S.D.	3886.8	211.3	.44	.00	.22	9.1	.28	9.1		7.2	4.9				

TABLE 10.1 E: INPUT: 7,732 PERSONS (FOLLOW-UP) -- 15 ITEMS MEASURED: 60 CATS [PERSON: REAL SEP: 2.67 REL: .88]
ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBSE CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	ESTIM DISCR	P-VALUE	ITEM	G
5	**9086**	**7705**	**-.65**	**.02**	**1.43**	**9.9**	**1.66**	**9.9**	**A .55**	**43.8**	**52.1**	**.36**	**1.18**	**5 Trouble sleeping**	**0**
7	**6859**	**7664**	**-.05**	**.02**	**1.34**	**9.9**	**1.59**	**9.9**	**B .56**	**51.8**	**56.5**	**.57**	**.89**	**7 Difficulty at home**	**0**
15	**4348**	**7695**	**.74**	**.02**	**1.22**	**9.9**	**1.46**	**9.9**	**C .55**	**61.1**	**64.7**	**.78**	**.57**	**14 Heart pounding**	**0**
4	**5418**	**7658**	**.62**	**.02**	**1.38**	**9.9**	**1.35**	**9.9**	**D .47**	**55.4**	**63.6**	**.61**	**.71**	**4 Have family/friends to count on**	**0**
9	5088	7170	.23	.02	1.13	6.8	1.22	6.7	E .61	59.4	60.1	.87	.71	9 Difficulty at work	0
14	5999	7655	.25	.02	1.10	5.3	1.16	5.9	F .63	57.3	58.7	.89	.78	14 Nervous anxious	0
6	5501	7664	.38	.02	.98	-.9	.96	-1.4	G .67	61.2	60.3	1.02	.72	6 Fearful afraid	0
8	5394	7668	.40	.02	.85	-8.6	.80	-7.4	H .72	64.3	60.4	1.17	.70	8 Difficulty socially	0
1	8434	7654	-.32	.02	.84	-9.7	.82	-9.9	g .70	67.8	62.5	1.17	1.10	1 Feel good about self	0
3	8935	7625	-.50	.02	.83	-9.9	.81	-9.9	f .70	68.0	62.6	1.19	1.17	3 Able to accomplish things	0
10	9733	7696	-.83	.02	.81	-9.9	.82	-9.9	e .75	58.6	54.0	1.22	1.26	10 Feeling sad or blue	0
2	8257	7644	-.22	.02	.80	-9.9	.77	-9.9	d .70	72.2	66.5	1.20	1.08	2 Able to deal with problems	0
12	7855	7678	-.31	.02	.75	-9.9	.76	-9.9	c .77	61.9	54.8	1.30	1.02	12 Everything is an effort	0
13	6139	7681	.18	.02	.76	-9.9	.72	-9.9	b .76	64.8	58.4	1.28	.80	13 Little or not interest things	0
11	6525	7681	.07	.02	.74	-9.9	.72	-9.9	a .77	63.1	56.8	1.30	.85	11 Feeling hopeless future	0
MEAN	6904.7	7635.9	.00	.02	1.00	-1.8	1.04	-1.7		60.7	59.5				
S.D.	1629.0	126.2	.45	.00	.24	8.9	.32	8.8		6.7	4.0				

INFIT – OUTFIT statistics should range between .7 and 1.3. Estimates below .7 indicate dependencies; while estimates above 1.3 indicate noise.

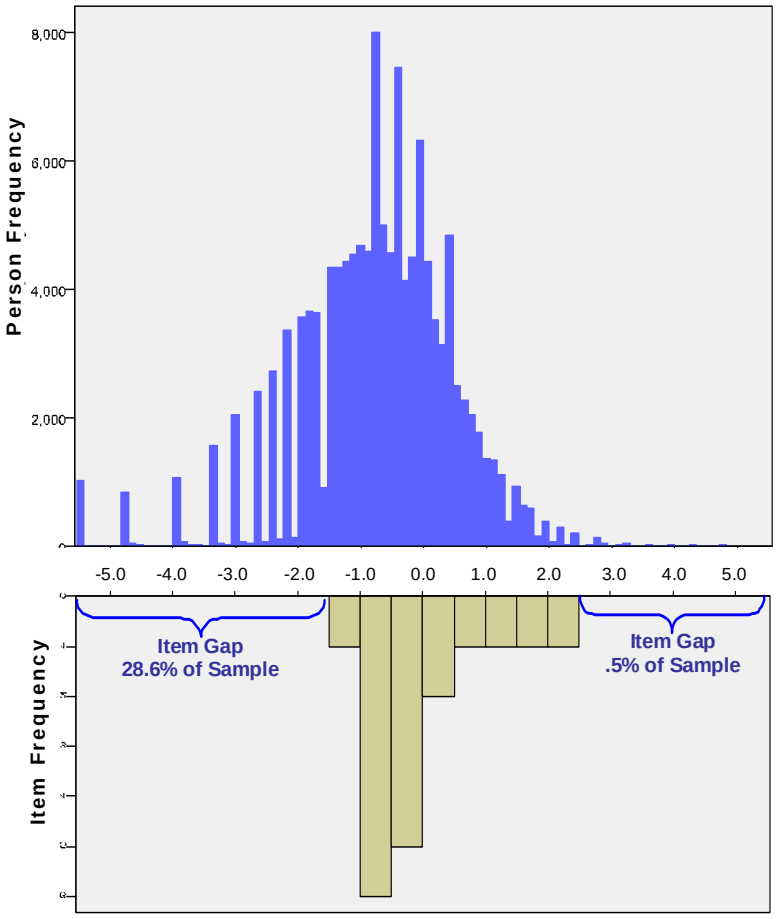
INFIT is an information-weighted fit statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level.

OUTFIT is an outlier-sensitive fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level.
MNSQ is the mean-square infit statistic with expectation 1. Values substantially below 1 indicate dependency in the data; values substantially above 1 indicate noise.

The item order is not significantly different between respondents completing the *Wellness Assessment* during session one and two, and those completing the assessment on the third session or later during the treatment period. Item order does not differ substantively between baseline and the follow-up assessment, administered via mail. The Wellness Assessment measurement model is stable and invariant across time.

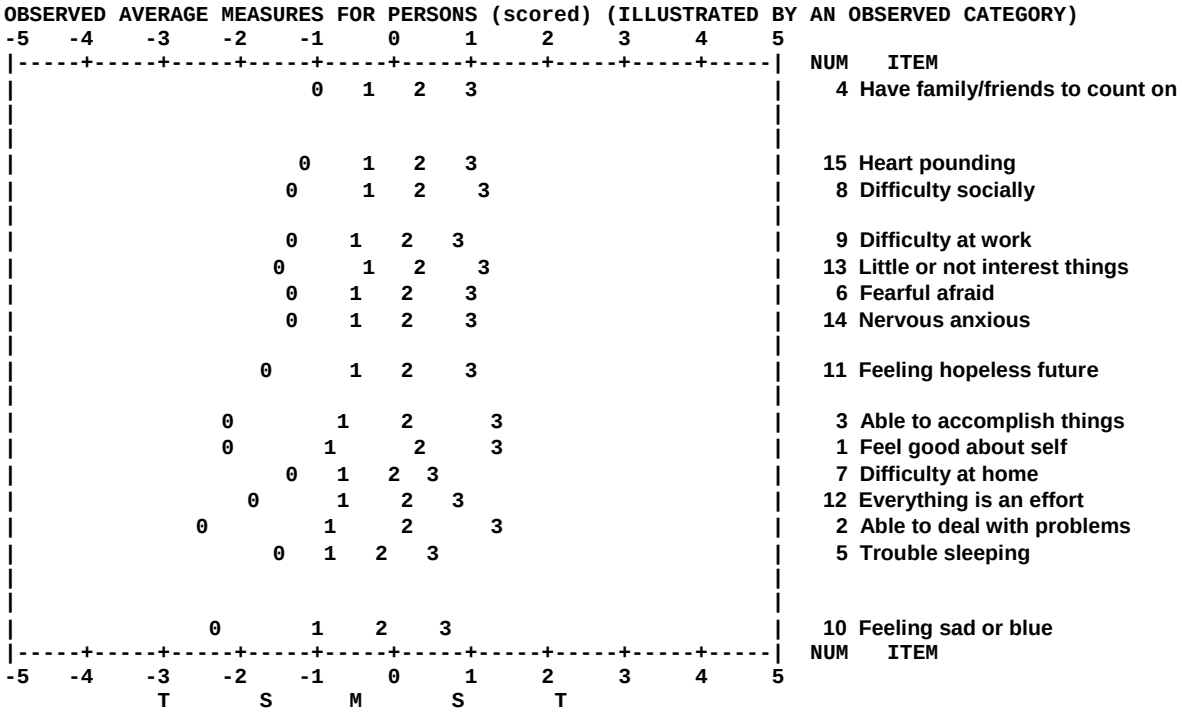
Three items reflected item misfit – *trouble sleeping*, *difficulty at home*, and *having friends/family to count on*. Estimates indicate noise, a lack of precision. The estimates, although greater the accepted ranges, do not exceed the ranges to the point of measurement degradation. The content of the items while logically related to emotional and psychological distress may not necessarily be confined to such distress. For example, difficulty sleeping may result from physical ailments or life stage status. The item asking about family and friends to count on may place a respondent in a quandary if they have friends but not family to count on for help. In summary, these items do no pose a serious problem in terms of inclusion in this assessment of *wellness*, an absence of emotional and/or psychological distress. Estimates for item 15 (Pounding/racing heart) at follow-up exceeded acceptable ranges. This is likely attributed to improvement and the diminishing experience of *heart pounding/racing* for individuals in treatment as they recover.

The person item maps illustrate a lack of items at the mild and the severe end of the assessed construct (wellness/distress). This phenomenon is characteristic of many if not most all behavioral healthcare assessment instruments. The absence of items addressing self-harm and suicidality restrict the assessment of the severe end of the measured construct, and mild items are notoriously difficult to write. The advantage of the using the Rasch measurement model is that the lack of change at some sections of the continuum can be identified as a measurement artifact as opposed to a lack of effective treatment. It is likely that individuals still continue to make progress as they improve towards the mild/moderate and mild end of the wellness construct, however, the *Wellness Assessment* is not sensitive to change/improvement at the milder end of the construct. Likewise, the *Wellness Assessment* can only assess deterioration up to a certain point, beyond which there are no items reflecting a higher level of severity. The graphic to the right illustrates the distribution of the respondent sample and the distribution of the items using the Rasch interval logit measure scores. Nearly 29% of the sample coming into treatment (sessions one and two) have score profiles indicating minimal if any distress using the UBH *Wellness Assessment*.



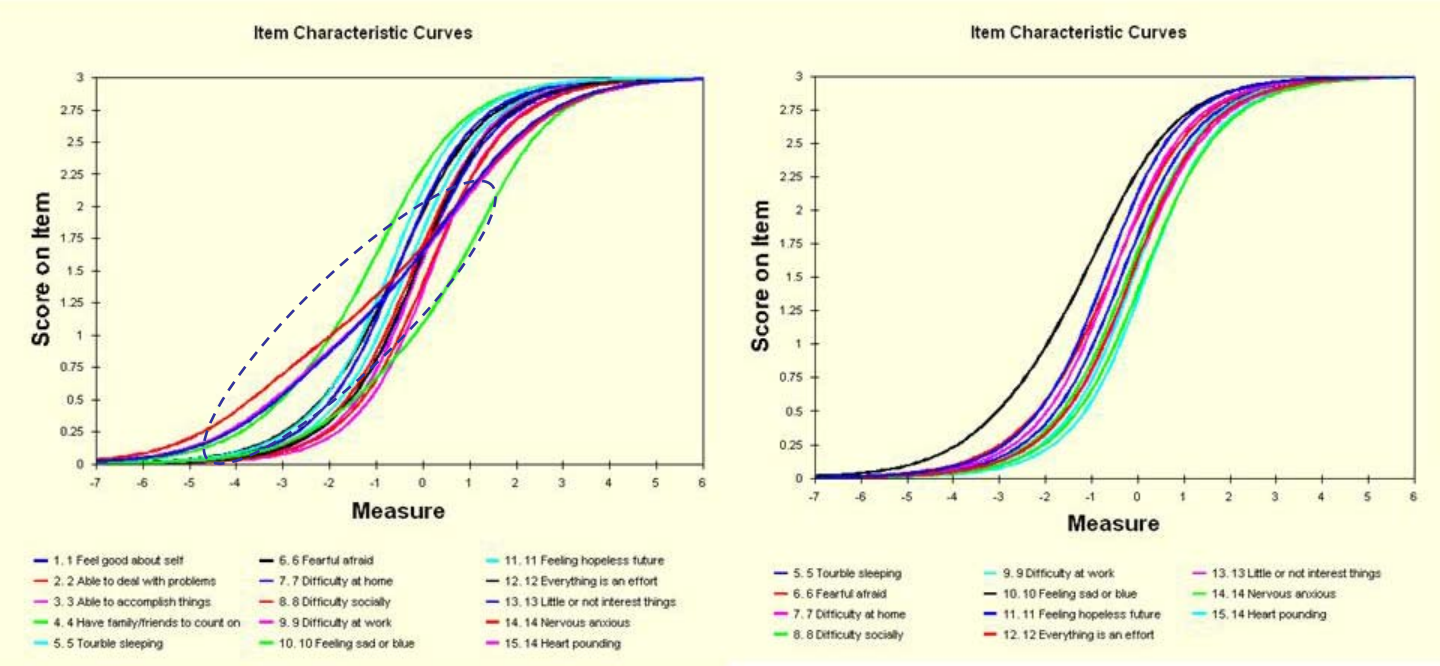
Response Scale Adequacy

The UBH Wellness Assessment uses a four-point Likert scale, ranging form 0 to 3. The scale functions within accepted parameters, as is indicated in the graphic below. This graphic arrays the Likert scale responses as interval (logit) data. The graphic illustrates that the response options for the most part are equally spaced. Some items have more restricted response ranges. For example, item 7 (Difficulty at home) has a logit range from -1.3 to .7, compared to the range for item 2 (Able to deal with problems) which ranges from -2.18 to 1.4. The items in a measure are intended to tap different areas of the assessment (wellness/distress) construct. The 15 items in this section of the *Wellness Assessment* does just that.



Dimensionality: Wellness Assessment

The Rasch model requires unidimensionality, however unidimensionality is never perfect – it is always approximate. The most important question is whether deviation from unidimensionality is substantive enough to warrant the construct of two or more measures from the item set, each dimension represented by separate measure. To test dimensionality, the item characteristic curves (ICCs) were initially examined. The Rasch 1PL model requires parallel ICCs, indicating that the items contribute additively to the overall assessment of the measured trait, emotional/psychological distress (*Wellness Assessment*). Four items (*feeling good about self*, *accomplishing things*, *handling problems* and *having friends/family to count on*) are characterized by crossed ICCs indicating that these items change in terms of difficulty relative to the placement of persons on the attribute level. For example, individuals with higher levels of impairment, these items become easier to endorse and contribute less in the response profile of an individual who is experiencing severe impairment and distress.



In addition to examining the ICCs, a principal components analysis (PCA) was conducted using residuals, as opposed to original observations (responses). As a first step, the first component (dimension) is removed, leaving secondary components to be examined in terms of whether the components are substantive enough to necessitate separating the items into separate measures.

In the case of the *Wellness Assessment*, the Rasch dimension explains 46% of the variance in the data. While the accepted *rule of thumb* is that a variance estimate of 60% or greater explained by the measure is considered good, it is important to remember that unidimensionality also depends on the size of the second dimension (contrast). In the case of the *Wellness Assessment*, the largest secondary dimension (first contrast of the residual data) explains 7.8% of the variance. The eigenvalue for this contrast is 2.2, indicating that it has the strength of approximately two items, which is the smallest number of items that would be considered in terms of a separate dimension. Eigenvalues for the remaining contrasts are 1.6, 1.4, 1.3, and 1.1. Given that random data can have eigenvalues of size 1.4, there is little to be gained in examining contrasts beyond the first contrast of the residual data. Most unexplained variance is hypothesized to be the random noise predicted by the Rasch model, rather than a degradation of the unidimensionality of the Rasch measurement model. The items and their respective loadings are presented in following Table.

Rasch Principal Component Analysis (Residuals)

Contrast	Loading	Measure	INFIT MNSQ	OUTFIT MNSQ	Item
1	.62	.51	1.12	1.17	15. Heart pounding
1	.54	.14	1.06	1.09	14. Nervous anxious
1	.42	.19	.96	.95	6. Fearful afraid
1	.41	-.43	1.26	1.38	5. Trouble sleeping
1	.13	-.27	1.52	1.75	7. Difficulty home
1	.11	.24	1.09	1.17	9. Difficulty work
1	-.53	-.16	.86	.85	1. Feel good about self *
1	-.48	-.12	.92	.91	3. Able accomplish things*
1	-.42	-.41	.85	.84	2. Able deal w/ problems*
1	-.40	-.06	.76	.73	11. Feeling hopeless future
1	-.38	.21	.78	.75	13. Little/no interest things
1	-.29	-.27	.80	.79	12. Everything an effort
1	-.24	-.98	.79	.77	11. Feeling sad/blue
1	-.11	.93	1.41	1.41	4. Have family/friends count on*
1	-.05	.49	.89	.86	5. Difficulty socially

INFIT – OUTFIT statistics should range between .7 and 1.3. Estimates below .7 indicate dependencies; while estimates above 1.3 indicate noise.

INFIT is an information-weighted fit statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level.
OUTFIT is an outlier-sensitive fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level.
MNSQ is the mean-square infit statistic with expectation 1. Values substantially below 1 indicate dependency in the data; values substantially above 1 indicate noise.

To further test the dimensionality of the *Wellness Assessment*, two separate scales were constructed based on the positive and negative loading on the first residual contrast. The correlation between these two scales using raw data is estimated at 0.71 (p =.01), and 0.693, using the Rasch measure scores. This indicates that these two separate sets of items are substantively related, as opposed to being distinct subscales. A latent (error-disattenuated) correlation was estimated using these data. This correlation approximates the correlation divided by the square root of the product of the reliabilities of the two sets of items in the first contrast. Reliability estimates for the two items sets are .767 (positive loading items) and .886 (negative loading items) [$C/\sqrt{(\alpha_1 * \alpha_2)} = .693/\sqrt{.767 * .886} = .693/.824 = .841$]. If this estimate approaches 1.0, the items sets are essentially telling the same statistical scenario.

From a practical stand point, in examining the content of the two sets of items, one should consider whether individuals high on one set of items would be treated differently than individuals high on the other set of items with regard to accessing treatment services, or demonstrating improvement resulting from receiving treatment services. In the case of the UBH Wellness Assessment, the 15 items reflect a substantively cohesive profile of emotional and psychological distress. The use of a summed score across the items is an empirically supported and parsimonious approach. In addition, the items also appear to sensitively reflect change over time.

Reliability

Scale Reliability Total Sample:	**0.87 (15 items, more precise estimate)**
Measurement Model Reliability:	**0.90**
Cronbach Alpha (Classical Test Theory):	**0.90**
Person Separation Index:	**2.93**

Rasch estimates of reliability incorporate item misfit estimates which are ignored by Cronbach Alpha estimates. As a result Cronbach Alpha is an overestimate of proportion of measurement variance that is true variance Separation is the number of statistically different performance strata that the test can identify in the sample. In this sample, there are about three measurably different levels of performance.

Clinical Cutoff Estimates

Approach

Clinical Cutoff Score: A cutoff score should balance both sensitivity (correctly identifying the proportion of individuals identified with clinical need who have high *UBH Wellness Assessment* scores within the clinical range) and specificity (correctly identifying the proportion of individuals without clinical distress, community sample – individuals with no indicated mental health problems) who have *UBH Wellness Assessment* scores within the nonclinical range). Most methods of calculating clinical cutoff scores incorporate scores from a nonclinical community sample. One accepted method of estimating cutoff scores using both treatment and nonclinical community samples is that proposed by Jacobsen and Truax (1991). The formula used is as follows:

$$Cutoff = \frac{(SD_{clinical})(Mean_{nonclinical}) + (SD_{nonclinical})(Mean_{clinical})}{SD_{clinical} + SD_{nonclinical}}$$

Data used to estimate a *UBH Wellness Assessment* clinical cutoff for adults consisted of two samples. The clinical sample included individuals completing the *UBH Wellness Assessment* survey on their *first* or *second visit*. The community sample consisted of individuals reporting that they had not received any psychotherapy services with a six month period prior to completing the survey, and were not taking psychotropic (e.g., anti-depressant) medication within the past 12 months.

Age differences were reported between the adult clinical and community samples. The clinical sample was proportionately older.

	Community (N=1,068)	**Clinical** (N=71, 480)
18 – 24	8.2%	9.9%
25 – 35	30.3%	26.3%
36 – 44	21.5%	29.3%
45 – 59	23.6%	29.1%
60 – 75	6.4%	5.0%
76 plus	—	0.3%
$\chi^2(5, N = 72{,}548) = 1.186, p > .000$		

Three community samples were used. In initial psychometric studies, health plan employees and/or affiliates were samples as part of a community sample. Several of these identified themselves as behavioral healthcare clinicians. There were mean *UBH Wellness Assessment* differences for behavioral healthcare clinicians respondents in the adult community sample as compared to non-behavioral healthcare clinicians adult community respondents (t = 3.373, df = 531, p = .001). Individuals with clinical background had significantly lower mean scores indicating minimal distress. A decision was made to retain the clinician respondents in the initial adult community sample, as the deletion of these data did not significantly change the estimates yielded from the Jacobsen & Truax algorithm establishing the cutoff scores (12.06 with clinician community respondents, 12.02 without clinician community respondents), Receiver

Operator Curve (ROC) or the Rasch measurement model analyses conducted with these data.

In addition to these samples, a third sample was collected using an Internet survey. An invitation was disseminated via professional Listservs (AEA, ASA, AERA, etc.) as well as circulated within several university academic departments (psychology departments were purposefully excluded). IP addresses were examined to ensure that respondents did not submit multiple surveys. An additional 537 individuals responded and completed the *UBH Wellness Assessment*.

Pilot Study: Adult Samples

	Total Community Sample	Internet Survey Community Sample	Community Sample No Clinicians	Community Sample Clinicians Only	Clinical Sample
Mean	7.33	7.29	7.96	5.91	19.28
SD	6.44	6.39	6.92	4.93	9.82
N	1,072	539	373	160	71,505

Possible score range: 0 – 45 (15 items, each scored 0 to 3)

Clinical Cutoff Score:

Jacobsen and Truax (1991) formula

Adult Wellness Assessment (UBH AWA): 12.06

Scale Reliability Total Sample: **0.87 (15 items, more precise estimate)**
Measurement Model Reliability: **0.90**
Cronbach Alpha (Classical Test Theory): **0.90**

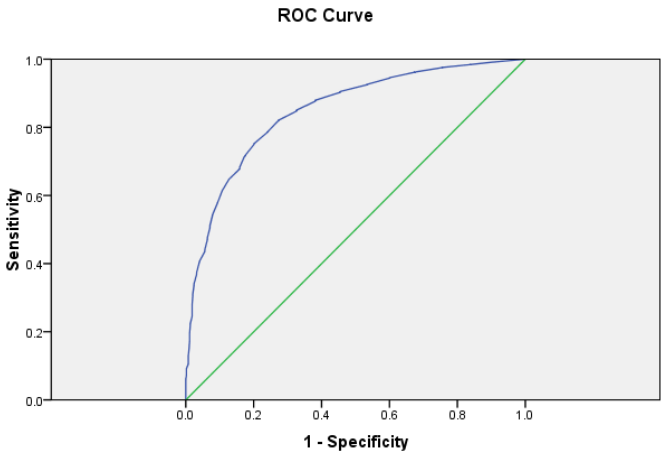
Receiver Operator Curve Analysis: Estimating Sensitivity and Specificity

AUC .848
Std. Error: .006
p .000

Estimates indicate a good fit in terms of separating clinical and community samples.

Cutoff Score = 12
- Sensitivity = ~.752
- Specificity = ~.798

Alternative Cutoff Score = 10
- Sensitivity = ~.816
- Specificity = ~.731



Wellness (cutoff score = 12.06)			
Test result	**Clinical Sample**	**Community Sample**	**Totals**
At or above Clinical Cutoff	**52,716 (true-positives – 75%)**	214 (false-positives – 20%)	52,930
Below Clinical Cutoff	17,572 (false-negatives – 25%)	**856 (true-negatives – 80%)**	18,428
Totals	70,288	1,070	71,358

Wellness (**Alternative** cutoff score = 10 / Sensitivity = ~.816, Specificity = ~.731			
Test result	**Clinical Sample**	**Community Sample**	**Totals**
At or above Clinical Cutoff	**55,528 (true-positives – 79%)**	257 (false-positives – 24%)	55,785
Below Clinical Cutoff	14,760 (false-negatives – 21%)	**813 (true-negatives – 76%)**	15,573
Totals	70,288	1,070	71,358

Coordinates of the Curve

Test Result Variable(s):UBH ADULT WELLNESS BASELINE SCORE

Positive if Greater Than or Equal To[a]	Sensitivity	Specificity
-1.0000	1.000	0.000
.5000	.991	0.099
1.0357	.985	0.155
1.1126	.984	0.161
1.2019	.984	0.161
1.6250	.984	0.161
2.0714	.976	0.243
2.2253	.975	0.245
2.4038	.975	0.245
2.7500	.975	0.245
3.1071	.963	0.320
3.3379	.962	0.325
3.6058	.962	0.325
3.8750	.962	0.325
4.1429	.946	0.398
4.4505	.945	0.400
4.8077	.945	0.400
5.1786	.927	0.464
5.5632	.925	0.467
5.8846	.925	0.467
6.1250	.905	0.544
6.3393	.905	0.544
6.6758	.903	0.546
6.9615	.903	0.546
7.2500	.879	0.619
7.7500	.877	0.619
8.0385	.851	0.673
8.3242	.850	0.673
8.6607	.847	0.675
8.8750	.847	0.675
9.1154	.821	0.727

Coordinates of the Curve

Test Result Variable(s):UBH ADULT WELLNESS BASELINE SCORE

Positive if Greater Than or Equal To[a]	Sensitivity	Specificity
9.4368	.820	0.727
9.8214	.816	0.731
10.1923	.787	0.759
10.5495	.786	0.759
10.8571	.783	0.763
11.1250	.753	0.798
11.3942	.753	0.798
11.6621	.752	0.798
11.8929	.749	0.800
12.2500	.718	0.824
12.5962	.718	0.824
12.7747	.717	0.824
12.9286	.713	0.828
13.3750	.683	0.841
13.7981	.683	0.841
13.8874	.682	0.841
13.9643	.678	0.841
14.5000	.647	0.873
15.5000	.613	0.893
16.0357	.583	0.905
16.1126	.579	0.907
16.2019	.578	0.907
16.6250	.578	0.907
17.0714	.547	0.920
17.2253	.543	0.921
17.4038	.542	0.921
17.7500	.541	0.921
18.1071	.511	0.929
18.3379	.507	0.929
18.6058	.506	0.929
18.8750	.506	0.929
19.1429	.476	0.936
19.4505	.471	0.936
19.8077	.470	0.936
20.1786	.441	0.944
20.5632	.437	0.944
20.8846	.436	0.944
21.1250	.408	0.959
21.3393	.408	0.959
21.6758	.403	0.961
21.9615	.402	0.961
22.2500	.373	0.968
22.7500	.368	0.968
23.0385	.342	0.976
23.3242	.341	0.976
23.6607	.336	0.976
23.8750	.336	0.976
24.1154	.311	0.979
24.4368	.310	0.979
24.8214	.306	0.979
25.1923	.281	0.981

Coordinates of the Curve

Test Result Variable(s):UBH ADULT WELLNESS BASELINE SCORE

Positive if Greater Than or Equal To[a]	Sensitivity	Specificity
25.5495	.280	0.981
25.8571	.275	0.981
26.1250	.251	0.981
26.3942	.251	0.981
26.6621	.250	0.981
26.8929	.246	0.981
27.2500	.224	0.987
27.5962	.223	0.987
27.7747	.223	0.987
27.9286	.218	0.987
28.3750	.197	0.989
28.7981	.197	0.989
28.8874	.196	0.989
28.9643	.192	0.989
29.5000	.173	0.989
30.5000	.149	0.991
31.0357	.132	0.993
31.1126	.128	0.993
31.2019	.127	0.993
31.6250	.126	0.993
32.0714	.111	0.993
32.2253	.107	0.993
32.4038	.106	0.993
32.7500	.106	0.993
33.1071	.092	0.998
33.3379	.089	0.998
33.6058	.088	0.998
33.8750	.088	0.998
34.1429	.076	0.998
34.4505	.073	0.998
34.8077	.072	0.998
35.1786	.061	1.000
35.5632	.058	1.000
35.8846	.058	1.000
36.1250	.048	1.000
36.3393	.048	1.000
36.6758	.046	1.000
36.9615	.045	1.000
37.2500	.038	1.000
37.7500	.035	1.000
38.0385	.028	1.000
38.3242	.028	1.000
38.6607	.026	1.000
38.8750	.026	1.000
39.1154	.020	1.000
39.4368	.020	1.000
39.8214	.018	1.000
40.1923	.014	1.000
40.5495	.014	1.000
40.8571	.012	1.000
41.1250	.009	1.000
41.3942	.009	1.000
41.6621	.008	1.000

Coordinates of the Curve

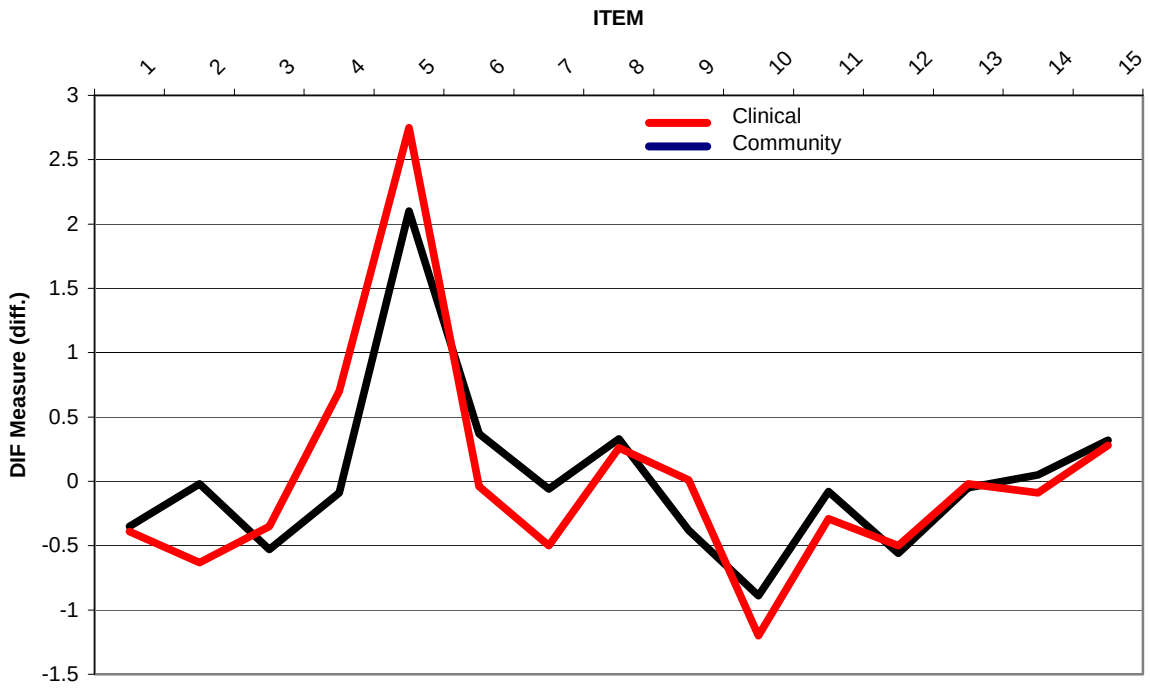
Test Result Variable(s):UBH ADULT WELLNESS BASELINE SCORE

Positive if Greater Than or Equal To[a]	Sensitivity	Specificity
41.8929	.008	1.000
42.2500	.005	1.000
42.5962	.005	1.000
42.7747	.005	1.000
42.9286	.004	1.000
43.4231	.002	1.000
43.8874	.002	1.000
43.9643	.002	1.000
44.5000	.001	1.000
46.0000	.000	1.000

The test result variable(s): UBH ADULT WELLNESS BASELINE SCORE has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

PERSON DIF Plot (DIF=CLINICAL-COMMUNITY)



1 Feel good about self
2 Able to deal w/ problems*
3 Able accomplish things*
4 Have family/friends count on*
5 Trouble sleeping*
6 Fearful afraid*
7 Difficulty home*
8 Difficulty socially
9 Difficulty work*
10 Feeling sad/blue*
11 Feeling hopeless future*
12 Everything an effort
13 Little/no interest things
14 Nervous anxious*
15 Heart pounding

**significant DIF*

Person Classes	χ^2	*df*	*p*	Item Name
2	.5631	1	.4530	1 Feel good about self
2	106.4990	1	.0000	2 Able deal w/ problems
2	10.9038	1	.0010	3 Able accomplish things
2	198.145	1	.0000	4 Have family/friends count on
2	248.274	1	.0000	5 Trouble sleeping
2	40.5501	1	.0000	6 Fearful afraid
2	64.2980	1	.0000	7 Difficulty home
2	1.3325	1	.2484	8 Difficulty socially
2	60.7260	1	.0000	9 Difficulty work
2	44.6183	1	.0000	10 Feeling sad/blue
2	15.0941	1	.0001	11 Feeling hopeless future
2	1.3329	1	.2483	12 Everything an effort
2	.2918	1	.5891	13 Little/no interest things
2	5.7569	1	.0164	14 Nervous anxious
2	.2851	1	.5934	15 Heart pounding

Differential item function (DIF) indicates that one group of respondents is scoring differently (better/worse) than another group of respondents on an item (after adjusting for the overall scores of the respondents. DIF analysis indicated that 10 of the 15 Wellness Assessment items functioned differentially for the clinical versus community respondents. Community, non-clinical respondent found it easier to endorse items such as able to handle problems, accomplishing things, having family and friends to count on; and more difficult to endorse such things as nervousness, feeling hopeless, or feeling sad and/or blue. DIF estimates support the ability of the Wellness Assessment to differentiate between clinically and non-clinical respondents.

DIF estimates are presented in the tables below for age groups and for female and male respondents. For the most part, female respondents have an easier time reporting distress. Females however had a more difficulty time endorsing problematic alcohol/drug use than their male respondent counterparts. A similar pattern was identified for older respondents.

ITEM	Female	Male
1 Feel good about self	-0.58	-0.42
2 Able to deal with problems	-0.83	-0.57
3 Able to accomplish things	-0.55	-0.45
4 Have family/friends to count on	0.56	0.38
5 Trouble sleeping	-0.77	-0.77
6 Fearful afraid	-0.20	-0.08
7 Difficulty at home	-0.53	-0.69
8 Difficulty socially	0.17	0.01
9 Difficulty at work	-0.04	-0.24
10 Feeling sad or blue	-1.38	-1.10
11 Feeling hopeless future	-0.37	-0.37
12 Everything is an effort	-0.65	-0.57
13 Little or not interest things	-0.12	-0.12
14 Nervous anxious	-0.26	-0.14
14 Heart pounding	0.14	0.27
C1 Ought to cut down	1.67	**0.66**
C2 People criticize drinking	2.65	**1.75**
C3 Guilty about drinking drug use	1.95	**1.08**

ITEM	18-24	25-34	35-44	45-59	60-75	76+
1 Feel good about self	-0.53	-0.55	-0.53	-0.50	-0.46	-0.45
2 Able to deal with problems	-0.70	-0.74	-0.72	-0.74	-0.90	-1.02
3 Able to accomplish things	-0.23	-0.36	-0.51	-0.67	-0.80	-0.90
4 Have family/friends to count on	0.87	0.64	0.44	0.37	0.34	0.50
5 Trouble sleeping	-0.71	-0.70	-0.79	-0.83	-0.77	-0.77
6 Fearful afraid	-0.09	-0.16	-0.16	-0.16	-0.24	-0.24
7 Difficulty at home	-0.18	-0.56	-0.78	-0.58	-0.33	-0.24
8 Difficulty socially	-0.05	0.07	0.15	0.17	0.21	0.23
9 Difficulty at work	-0.35	-0.18	-0.07	-0.07	0.22	1.15
10 Feeling sad or blue	-1.28	-1.28	-1.26	-1.28	-1.28	-1.28
11 Feeling hopeless future	-0.45	-0.37	-0.33	-0.37	-0.46	-0.48
12 Everything is an effort	-0.58	-0.63	-0.63	-0.63	-0.74	-0.82
13 Little or not interest things	-0.16	-0.09	-0.09	-0.15	-0.21	-0.31
14 Nervous anxious	-0.38	-0.24	-0.13	-0.22	-0.39	-0.67
14 Heart pounding	0.06	0.07	0.18	0.29	0.33	0.35
C1 Ought to cut down	0.85	1.19	**1.29**	**1.44**	**1.77**	**2.15**
C2 People criticize drinking	1.60	**2.21**	**2.34**	**2.57**	**2.85**	**3.45**
C3 Guilty about drinking drug use	1.14	1.48	**1.62**	**1.83**	**2.27**	**3.16**

The UBH Wellness Assessment asks several questions on health status – an item on general health status, as well as whether respondents are concerned about diabetes, asthma, heart conditions, and back pain. In addition, respondents are asked to estimate the number of days missed at work during the past month that were related to these health concerns. The following table provides the correlations among these variables. The experience of back and chronic pain has the strongest association with the self-reported number of medical visits in the past six months and the respondents' perception of general health status. There is substantive literature that links self-reported chronic pain, such as back pain and headaches to undiagnosed and treated behavioral health conditions. It would likely be worth while examining the distribution of medical visits before and after psychotherapy to determine whether there is an associated decrease in medical visits after successfully completing treatment.

	Asthma	Diabetes	Heart condition	Back pain	General health	Days missed
Asthma						
Diabetes	.077**					
Heart condition	.066**	.181**				
Back pain	.105**	.101**	.097**			
General health	.101**	.173**	.118**	.270**		
Days missed	.029**	.038**	.038**	.109**	.195**	
# Medical visits	.106**	.116**	.095**	.247**	.344**	.251**

**Correlation is significant at the 0.01 level (2-tailed).

In addition to items pertaining to physical health, three items adapted from the CAGE[1] questionnaire were included in the UBH *Wellness Assessment*. Problematic alcohol and drug use was associated with a general perception of poor health status. The UBH Wellness Assessment asks respondents to estimate the number of drinks they had in the past week. Responses ranged from 0 to more than 99 drinks of alcohol during the week. Responses to this item that exceeded 35 drinks (7 per day) were recoded as missing resulting in a loss of 13% of the sample. This item may need to be worded or perhaps scaled to give respondents reasonable choices, as opposed to asking for the number of alcoholic drinks consumed in the past week. The number of drinks, despite suspected error, was significantly related to perceived self-reported problems with alcohol and/or drug use. What is interesting is the fact that problematic alcohol and/or drug use is not associated with medical visits; however problematic alcohol and/or drug use is related to perceptions of poorer health status.

	Ought to cut down	Criticized drinking	Guilty alcohol/drug use	# Drinks/week
Ought to cut down				
Criticized drinking	.477**			
Guilty alcohol/drug use	.727**	.515**		
# Drinks/week	.398**	.212**	.308**	
General health	.056**	.070**	.067**	-.047**
Missed days	.055**	.049**	.058**	-.016**
# Medical visits/6 months	-.019**	.004	-.009	-.016**

[1] EWING, J. A. (1984). Detecting alcoholism: the CAGE questionnaire, *Journal of the American Medical Association*, 252, 1905-1907.