



Self Discovery Assessment

Statistical Analysis

August, 2002

An Independent Study By:
Jerry MacDaid
2724 SW 14th Drive
Gainesville, Florida 32608

Report on the Do What You Are assessment

This report summarizes the methods and the findings of research conducted on the psychometric properties of the Do What You Are assessment. Analysis began with a data set of 11,028 (54%) females and 9,385 (56%) males, combining into a total group of 20,413. The sample is an international sample of young people from thirty five states and four countries. The states comprising the largest portion of the sample are California 17.5%, Connecticut 15.8%, Washington 12.7%, and Texas 10.2%; 14 states comprise between 9.0% and 1.0%, with the remaining 17 states under 1.0%. The four countries comprised 1.5%.

The other major variables of this sample are described below:

GRADE

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 9	5550	27.2	27.2	27.2
10	4085	20.0	20.0	47.2
11	6064	29.7	29.7	76.9
12	4714	23.1	23.1	100.0
Total	20413	100.0	100.0	

SYSTEM

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Private	1906	9.3	9.3	9.3
Public	18507	90.7	90.7	100.0
Total	20413	100.0	100.0	

FIT

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Very	11557	56.6	56.6	56.6
Mostly	7099	34.8	34.8	91.4
Somewhat	1489	7.3	7.3	98.7
Not	268	1.3	1.3	100.0
Total	20413	100.0	100.0	

Three different samples were drawn from this super ordinate sample for analysis. Two are ideal samples and one is a maximized sample.

The two ideal samples were created from this larger sample using three common criteria. These criteria were US residents, a reported accuracy of fit of mostly or very accurate, and the choice of type made by people with low scores matching the calculated type. This group was then used to build two samples, one male and one female, each with an equal number of each of the sixteen types. The sample sizes are limited by the sample size of the least frequently occurring type, For females that was ISTP (n=91) and for males it was ENFJ (n=119). The purpose of an equal

number of each type is to eliminate any bias in the item analysis or test of independence that can be caused by an uneven type distribution. Males and females were run separately to look for gender differences in item responses.

The descriptive statistics for the females:

GRADE

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 9	369	25.3	25.3	25.3
10	263	18.1	18.1	43.4
11	476	32.7	32.7	76.1
12	348	23.9	23.9	100.0
Total	1456	100.0	100.0	

SYSTEM

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Private	178	12.2	12.2	12.2
Public	1278	87.8	87.8	100.0
Total	1456	100.0	100.0	

FIT

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Very	942	64.7	64.7	64.7
Mostly	514	35.3	35.3	100.0
Total	1456	100.0	100.0	

The descriptive statistics for the males:

GRADE

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 9	543	29.0	29.0	29.0
10	335	17.9	17.9	46.9
11	553	29.5	29.5	76.4
12	441	23.6	23.6	100.0
Total	1872	100.0	100.0	

SYSTEM

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Private	146	7.8	7.8	7.8
Public	1726	92.2	92.2	100.0
Total	1872	100.0	100.0	

FIT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very	1154	61.6	61.6	61.6
	Mostly	718	38.4	38.4	100.0
	Total	1872	100.0	100.0	

Parallel analyses were then conducted on these two samples: an item analysis, internal consistency using coefficient alpha, and independence of the scales.

The item analysis conducted was a Bayesian procedure which shows the predictive power in the form a probability for each choice of a question. Items are deemed predictive if both choices have at least a 0.66 probability of predicting to the preference. This threshold is more rigorous than what was used for the MBTI. Items listed in the weak category are above or close to the standard used with the MBTI. Items are also examined to make sure they do not have predictive power on any index other than the one they are assigned.

The results from the item analysis showed that the vast majority of items have predictive power only for the intended index. The specific probabilities for each item are attached. Gender differences averaged 0.02 for EI and SN and averaged 0.03 for T and J and 0.04 for F and P. The following item failed:

TF Item 32 totally fails for males and females.

The following items are flawed, but useable:

EI Item 09 is weak for I females.

SN Item 14 also predicts to EI above .60 on males and females.

SN Item 19 also predicts to JP above .60 on males and females.

SN Item 19 is weak for N males.

TF Item 26 is weak for F for males and females.

TF Item 33 is weak for F for females.

JP Item 34 is weak for P for males.

JP Item 39 is weak for P for females.

JP Item 40 is weak for J for females.

Next, an internal consistency analysis was performed to evaluate the reliability of each of the indices. The method employed was coefficient alpha. The minimum level of correlation for acceptable reliability is 0.60. The following tables show the coefficient alphas for females and males separately.

	Females	Males
EI	0.81	0.83
SN	0.73	0.74
TF	0.78	0.74

JP 0.82 0.81

These reliability coefficients fall into the good and very good ranges. The MBTI form G reliabilities are consistently above 0.85 with samples of this same size, while form M levels cluster in the low 0.90's. More specifically, this is a measure designed for young people of high school age. Reliability coefficients are not reported in the third edition of the MBTI Manual for form M for subjects under 18 years old.. Split half reliabilities calculated on the MBTI form F and G for high school aged groups range consistently in the low 0.80's for large samples and between the low 0.80's and high 0.70's for smaller samples.

The Do What You Are assessment is a much shorter instrument than the currently published versions of the MBTI. An appropriate comparison would be the reliabilities for MBTI form AV, a short lived shortened version of the MBTI with it's fifty most predictive items, designed to be a surrogate measure of type. Reliabilities were not originally calculated for form AV due to the lack of logical split half assignments for the items. A sample of 739 form G responses was used to calculate coefficient alpha for each scale of form AV. Those values ranged from 0.71 to 0.78, except JP which had only 0.52.

While the MBTI is the most frequently used Jungian Type instrument, other assessments have been developed to identify type. The Murphy-Meisgeier Type Indicator for Children (Murphy, E and Meisgeier, C, 1987) reports reliabilities in their Manual that range from 0.42 to 0.78 for seventh and eighth graders. The reliabilities for The Personality Profiler (Johnson, DA, 1995) range from 0.73 to 0.89 for students.

The last analysis is the independence of the scales. This procedure correlates each index to the others. The indices of a Jungian instrument should not correlate since the constructs are independent and separate ideas. Thus a correlation of the indices should not yield coefficients greater 0.30 in magnitude. Below are the results of the test of independence for females and males. These tables show that all correlations are below the 0.30 threshold. The highest is SN/JP, reaching above 0.20. Results of these analyses on the MBTI form G for adult and high school samples are very similar when also controlling for the type distribution by using an equal number of each type. Studies conducted with form M show similar finding on all intercorrelations except SN/JP which consistently exceed 0.40.

Correlations

		EICONT	SNCONT	TFCONT	JPCONT
EICONT	Pearson Correlation	1.000	-.044	.049	-.068**
	Sig. (2-tailed)	.	.096	.062	.009
	N	1456	1456	1456	1456
SNCONT	Pearson Correlation	-.044	1.000	.070**	.218**
	Sig. (2-tailed)	.096	.	.007	.000
	N	1456	1456	1456	1456
TFCONT	Pearson Correlation	.049	.070**	1.000	.035
	Sig. (2-tailed)	.062	.007	.	.181
	N	1456	1456	1456	1456
JPCONT	Pearson Correlation	-.068**	.218**	.035	1.000
	Sig. (2-tailed)	.009	.000	.181	.
	N	1456	1456	1456	1456

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

Correlations

		EICONT	SNCONT	TFCONT	JPCONT
EICONT	Pearson Correlation	1.000	-.064**	.014	-.043
	Sig. (2-tailed)	.	.006	.539	.060
	N	1872	1872	1872	1872
SNCONT	Pearson Correlation	-.064**	1.000	.080**	.223**
	Sig. (2-tailed)	.006	.	.001	.000
	N	1872	1872	1872	1872
TFCONT	Pearson Correlation	.014	.080**	1.000	.092**
	Sig. (2-tailed)	.539	.001	.	.000
	N	1872	1872	1872	1872
JPCONT	Pearson Correlation	-.043	.223**	.092**	1.000
	Sig. (2-tailed)	.060	.000	.000	.
	N	1872	1872	1872	1872

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

The third sample drawn, called maximal, has the goal to provide the maximum sample size to examine these sample psychometric properties. Since the gender differences found in the last item analysis were minimal, this sample was randomly drawn from the entire sample. The type distribution again was the limiting factor since the sample needs to have an equal number of each of the sixteen types. The table shows the frequencies for each of the types. INTJ was the least frequently occurring type so a sample of 550 subjects was randomly drawn from each of the other types.

type table order

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ISTJ	1201	5.9	5.9	5.9
	ISFJ	1344	6.6	6.6	12.5
	INFJ	841	4.1	4.1	16.6
	INTJ	551	2.7	2.7	19.3
	ISTP	799	3.9	3.9	23.2
	ISFP	1000	4.9	4.9	28.1
	INFP	2012	9.9	9.9	38.0
	INTP	1021	5.0	5.0	43.0
	ESTP	1178	5.8	5.8	48.7
	ESFP	1560	7.6	7.6	56.4
	ENFP	3196	15.7	15.7	72.0
	ENTP	1326	6.5	6.5	78.5
	ESTJ	1592	7.8	7.8	86.3
	ESFJ	1152	5.6	5.6	92.0
	ENFJ	992	4.9	4.9	96.8
	ENTJ	648	3.2	3.2	100.0
	Total	20413	100.0	100.0	

The descriptive statistic for the other variables follow:

GENDER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	4589	52.1	52.1	52.1
	Male	4211	47.9	47.9	100.0
	Total	8800	100.0	100.0	

GRADE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	9	2326	26.4	26.4	26.4
	10	1733	19.7	19.7	46.1
	11	2681	30.5	30.5	76.6
	12	2060	23.4	23.4	100.0
	Total	8800	100.0	100.0	

SYSTEM

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Private	856	9.7	9.7	9.7
	Public	7944	90.3	90.3	100.0
	Total	8800	100.0	100.0	

FIT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very	4959	56.4	56.4	56.4
	Mostly	3073	34.9	34.9	91.3
	Somewhat	645	7.3	7.3	98.6
	Not	123	1.4	1.4	100.0
	Total	8800	100.0	100.0	

The item analysis procedure was repeated for this maximal sample with males and females combined. The item weights replicated the first analysis, with the absolute value of the differences averaging from 0.02 to 0.04 across the four indices.

This sample was also used to conduct an exploratory factor analysis. This analysis showed four unambiguous factors: JP, EI, TF, and SN. The first two factors had ten of the eleven items at the top of the unrotated factor with correlations above 0.35. The last two factors had all eleven items at the top but correlations fell as low as 0.20.

The internal consistency analysis was again performed using coefficient alpha to evaluate the reliability of each of the indices. The following table show the previously reported coefficient alphas for females and males separately compared to the results for this maximal sample.

	Females	Males	Maximal
EI	0.81	0.83	0.78
SN	0.73	0.74	0.68
TF	0.78	0.74	0.71
JP	0.82	0.81	0.78

The test for independence of the indices was repeated for this maximal sample. The table below shows that while the SN/JP correlation has increased by about 0.50, all correlations are below the 0.30 threshold.

Correlations

		EICONT	SNCONT	TFCONT	JPCONT
EICONT	Pearson Correlation	1.000	-.051**	.032**	-.067**
	Sig. (2-tailed)	.	.000	.003	.000
	N	8800	8800	8800	8800
SNCONT	Pearson Correlation	-.051**	1.000	.096**	.282**
	Sig. (2-tailed)	.000	.	.000	.000
	N	8800	8800	8800	8800
TFCONT	Pearson Correlation	.032**	.096**	1.000	.078**
	Sig. (2-tailed)	.003	.000	.	.000
	N	8800	8800	8800	8800
JPCONT	Pearson Correlation	-.067**	.282**	.078**	1.000
	Sig. (2-tailed)	.000	.000	.000	.
	N	8800	8800	8800	8800

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

Goodness of fit data has been reported so far as one of the demographics showing the similarity of the samples. Here the subjects report the accuracy of the results. Thus fit can also be an approach to measure criterion validity. The table below shows that 91% of the sample report the psychological comfort of the fit of the type descriptions for their results to mostly or very comfortable.

The scoring of the Do What You Are assessment attempts to increase the accuracy by asking students to read different type descriptions and pick the better fitting description if they had very close scores on an index. The fit variable was examined by a variable call typesame, that is whether the subject picked the type as scored (four the same) or picked a type different than scored (three or only two the same).

The statistically significant findings below show that people who picked the same type as scored by the assessment had the highest percentage of very good fit. Analysis of fit by each preference index and controlling for close scores shows that only the SN index had a significant difference with 55% of people picking the same preference on SN assessing finding the fit very good versus only 52% for the people picking a different preference on SN.

FIT * TYPESAME Crosstabulation

% within TYPESAME

		TYPESAME			Total
		2	3	4	
FIT	Very	47.8%	51.6%	59.5%	56.6%
	Mostly	37.6%	38.2%	33.1%	34.8%
	Somewhat	11.9%	8.7%	6.3%	7.3%
	Not	2.6%	1.5%	1.1%	1.3%
Total		100.0%	100.0%	100.0%	100.0%

Studies of a parallel nature done with the MBTI measured percentage of agreement with MBTI results and a type chosen as the subjects best fit after a feedback session with a trained professional. and parallel to the Do What You Are assessment's approach to clarifying close scores. Studies show that on average 75% of people pick a type that matches their MBTI results.

In studies where participants were asked to pick a description blind to their results and without professional feedback, an average of 50% of subjects picked the same type description as their results predicted.

The final analysis is a comparison type distribution of the entire sample drawn to a Pennsylvania High School sample reported by Myers. This high school sample is one of many drawn to attempt to estimate the distribution of the types in the United States population. While it is the oldest of such estimates it is the most well suited as a comparison because it directly parallels the high school population being examined in this study.

The test performed is called selection ratio type table (SRTT) analysis. It is a contingency table analysis using chi-square and was specifically designed for the comparison of the type distributions of two groups.

This analysis shows this sample is significantly more NF than the Pennsylvania High School sample with the ratio showing twice as many NF types. Where the ES types were significantly underrepresented with almost half the ES.

The entire sample was analyzed one last time by applying the item weights calculated above and using them to rescore the Do What You Are assessment. The purpose of this approach is to give the questions with more predictive power more weight. The reliabilities coefficients were again calculated for the rescored sample and are in the table that follows.

	Maximal	Rescored
EI	0.78	0.79
SN	0.68	0.71
TF	0.71	0.72
JP	0.78	0.80

Conclusions

These results show that the psychometric properties of the Do What You Are assessment instrument are indeed stable. They appear very good for a short assessment. The item probabilities average 0.75, which is very high. These weights are the equivalent to two point items of the MBTI form G and the percentage of items of this magnitude on the Do What You Are assessment outstrips the percentage on the MBTI.

The factor analysis supports the item analysis finding and shows that the item group with their respective scales. Further analysis here is continuing.

The reliability results are very respectable for an instrument with indices this short. While the longer MBTI scales achieve higher levels reliability due to their length, the Do What You Are assessment's reliabilities are quiet respectable. The 0.10 difference in magnitude appears to be attributable to difference in the lengths of the scales. It is harder for a shorter measure to push people away from the midpoint.

The rescoring shows improved reliability coefficients. The assessment gains modest benefit from the additional predictive power of the items.

All the intercorrelations of the scales, except one, are well below the 0.30 threshold demonstrating virtually no overlap with each other. Only the SN/JP correlates somewhat close to the threshold in the low 0.20 range. The Do What You Are assessment benefits from the shorter scales which reduce the chances of getting higher correlations.

The validity test examining fit, while not the exact same test as performed with the MBTI, shows that the Do What You Are assessment has comparable results. The weakness at this point is that this is the only validity data. Broader and deeper studies will need to continue.

One unique approach taken by the Do What You Are assessment is that subjects with very close scores can examine descriptions and chose a type that they believe fits best. The comparison of reported goodness for subjects accepting their type results versus subjects picking another type show that people who agreed with their assessment reported the highest goodness of fit with the description. While these differences are statistically significant, it is modest and likely improved because of the approach although this has not been empirically demonstrated.

The SRTT analysis showing a large and statistically significant difference in the type distribution of this sample with samples drawn to estimate the US population can not be explained by any of the analysis conducted in this study. The most likely explanation is that is due the voluntary nature of the sample. The MBTI literature has documented the leaning of NF types to be drawn to psychological and personal discovery kinds of activities. Choosing to take a personality assessment on the internet clearly fits that description.

While validation is a process and one study alone cannot absolutely prove validity, this study, with this very large sample, takes a very large first step in this process and shows that the Do What You Are assessment has stable and admirable psychometric properties.

Appendix A

Item Analysis Results for the Do What You Are Assessment

Index	Item Number	Females Predictive Probability		Males Predictive Probability		Gender Differences	
		E	I	E	I	E	I
EI	1	0.85	0.90	0.85	0.88	0.00	0.02
	2	0.73	0.85	0.73	0.85	-0.01	0.00
	3	0.82	0.81	0.83	0.81	-0.01	0.00
	4	0.80	0.77	0.80	0.79	0.00	-0.02
	5	0.76	0.70	0.80	0.71	-0.04	-0.01
	6	0.82	0.73	0.83	0.76	-0.01	-0.03
	7	0.72	0.73	0.68	0.73	0.04	0.00
	8	0.71	0.74	0.73	0.71	-0.02	0.02
	9	0.68	0.64*	0.72	0.66	-0.03	-0.01
	10	0.82	0.75	0.87	0.77	-0.05	-0.02
	11	0.66	0.74	0.71	0.76	-0.05	-0.03
SN	12	0.72	0.76	0.78	0.74	-0.07	0.02
	13	0.76	0.82	0.77	0.83	-0.01	-0.01
	14	0.70	0.67	0.67	0.74	0.03	-0.07
	15	0.74	0.68	0.76	0.70	-0.01	-0.02
	16	0.76	0.80	0.79	0.77	-0.03	0.03
	17	0.66	0.67	0.67	0.70	-0.02	-0.03
	18	0.69	0.69	0.66	0.68	0.03	0.01
	19	0.71	0.74	0.70	0.73	0.01	0.00
	20	0.76	0.77	0.78	0.79	-0.01	-0.02
	21	0.78	0.66	0.79	0.65*	-0.01	0.01
	22	0.69	0.69	0.68	0.70	0.01	-0.01

Appendix A (continued)

Item Analysis Results for the Do What You Are Assessment

Index	Item Number	Females Predictive Probability		Males Predictive Probability		Gender Differences	
		T	F	T	F	T	F
TF	23	0.86	0.75	0.84	0.79	0.03	-0.04
	24	0.79	0.80	0.76	0.74	0.04	0.06
	25	0.75	0.83	0.72	0.78	0.02	0.05
	26	0.72	0.62*	0.71	0.64*	0.01	-0.03
	27	0.75	0.77	0.69	0.75	0.06	0.02
	28	0.76	0.75	0.68	0.77	0.08	-0.01
	29	0.78	0.76	0.73	0.75	0.05	0.01
	30	0.82	0.76	0.77	0.82	0.05	-0.06
	31	0.69	0.79	0.67	0.75	0.02	0.04
	32	0.59	0.63**	0.61	0.60**	-0.02	0.03
	33	0.80	0.63*	0.80	0.67	0.00	-0.04
JP		J	P	J	P	J	P
	34	0.79	0.66	0.76	0.64*	0.03	0.01
	35	0.87	0.83	0.89	0.76	-0.02	0.07
	36	0.85	0.84	0.81	0.81	0.04	0.02
	37	0.80	0.73	0.87	0.68	-0.07	0.05
	38	0.74	0.82	0.78	0.78	-0.04	0.04
	39	0.70	0.64*	0.71	0.70	0.00	-0.07
	40	0.65	0.78*	0.66	0.77	-0.02	0.02
	41	0.92	0.74	0.86	0.70	0.06	0.03
	42	0.71	0.80	0.70	0.84	0.01	-0.04
	43	0.75	0.82	0.78	0.79	-0.03	0.03
44	0.68	0.84	0.67	0.82	0.01	0.02	

* weak items

** failed item