



Attribute Agreement analysis (*Appraisal Data*)

Main Resources:

- Create Attribute Agreement Analysis Worksheet
- Attribute Agreement Analysis
- Attribute Gage Study (Analytic Method)

DOES BOTTLED WATER ACTUALLY TASTE BETTER?



Fiji[®]



Zephyrhills[®]



7-11[®] Generic



Filtered Tap Water

In order to answer this question, we setup an experiment to see if people could tell the difference between four different types of water

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

Each person was given a sample of the four waters at the beginning of the test, and told which one was which, so they knew how each water tasted. At any time during the test, they were allowed to go back to the samples and re-taste them.

After tasting each sample, they were given 12 unmarked cups of water, and asked to select the correct water based upon its taste and smell. Each of the four water brands were provided three times in the study

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

Cup #	John	Peter	Mary	Actual	% Correct
1	Generic	Tap	Fiji	Generic	33%
2	Zephyrhill	Generic	Tap	Tap	33%
3	Fiji	Fiji	Generic	Fiji	67%
4	Fiji	Fiji	Generic	Zephyrhill	0%
5	Tap	Tap	Zephyrhill	Fiji	0%
6	Zephyrhill	Zephyrhill	Tap	Tap	33%
7	Fiji	Fiji	Zephyrhill	Generic	0%
8	Tap	Generic	Fiji	Zephyrhill	0%
9	Tap	Tap	Zephyrhill	Tap	67%
10	Generic	Generic	Generic	Generic	100%
11	Generic	Zephyrhill	Zephyrhill	Fiji	0%
12	Fiji	Fiji	Zephyrhill	Zephyrhill	33%
Overall	42% (4)	33% (3)	42% (4)		8% (1)

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

Cup #	John	Peter	Mary	Actual	% Correct
1	Generic	Tap	Fiji	Generic	33%
2	Zephyrhill	Generic	Tap	Tap	33%
3	Fiji	Fiji	Generic	Fiji	67%
4	Fiji	Fiji	Generic	Zephyrhill	0%
5	T				
6	Z				
7	F				
8	T				
9	T				
10	e				
11	e				
12	F				
Overall	4				

Having each brand show up more than once allows us to test how repeatable each tester is. In other words, if one tester correctly chooses the Fiji water the first time, but chooses it incorrectly the other two times, then it shows that the first selection may have been more of a lucky guess, rather than strong evidence that the tester could differentiate between the water.

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

Create Attribute Agreement Analysis Worksheet

Sample standard/attribute in text Options...

Number of samples: 12 Number of appraisers: 3

Sam	Sample N	Text Stan
1	1	Generic
2	2	Tap
3	3	Fiji
4	4	Zephyrhills
5	5	Fiji
6	6	Tap
7	7	Generic
8	8	Zephyrhills
9	9	Tap

Appra	Appraiser Na
1	John
2	Peter
3	Mary

Create Attribute Agreement Analysis Worksheet: Options

Do not randomize

Randomize all runs

Randomize runs within appraisers

Store standard run order in worksheet

Number of replicates: 1

Help

In order to apply statistical analysis to this experiment, we used Minitab's **Attribute Agreement Analysis** test. For those of you not familiar with this technique, it is a method for determining how well different people can select the correct answer from a list of choices.

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

water.mtw

C1	C2-T	C3-T	C4-T	C5-T
RunOrder	Samples	Appraisers	Assessments	Standards
1	1	John	Generic	Generic
2	1	Peter	Tap	Generic
3	1	Mary	Fiji	Generic
4	2	John	Zephyrhills	Tap
5	2	Peter	Generic	Tap
6	2	Mary	Tap	Tap
7	3	John	Fiji	Fiji
8	3	Peter	Fiji	Fiji
9	3	Mary	Generic	Fiji
10	4	John	Fiji	Zephyrhills
11	4	Peter	Fiji	Zephyrhills
12	4	Mary	Generic	Zephyrhills
13	5	John	Tap	Fiji
14	5	Peter	Tap	Fiji
15	5	Mary	Zephyrhills	Fiji
16	6	John	Zephyrhills	Tap
17	6	Peter	Zephyrhills	Tap
18	6	Mary	Tap	Tap
19	7	John	Fiji	Generic
20	7	Peter	Fiji	Generic
21	7	Mary	Zephyrhills	Generic
22	8	John	Tap	Zephyrhills
23	8	Peter	Tap	Zephyrhills
24	8	Mary	Zephyrhills	Zephyrhills

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

Attribute Agreement Analysis

C1	RunOrder
C2	Samples
C3	Appraisers
C4	Assessments
C5	Standards

Data are arranged as

Attribute column: Assessments

Samples: Samples

Appraisers: Appraisers

Multiple columns:

(Enter trials for each appraiser together)

Number of appraisers:

Number of trials:

Appraiser names (optional):

Known standard/attribute: Standards (Optional)

Categories of the attribute data are ordered

Select

Information...

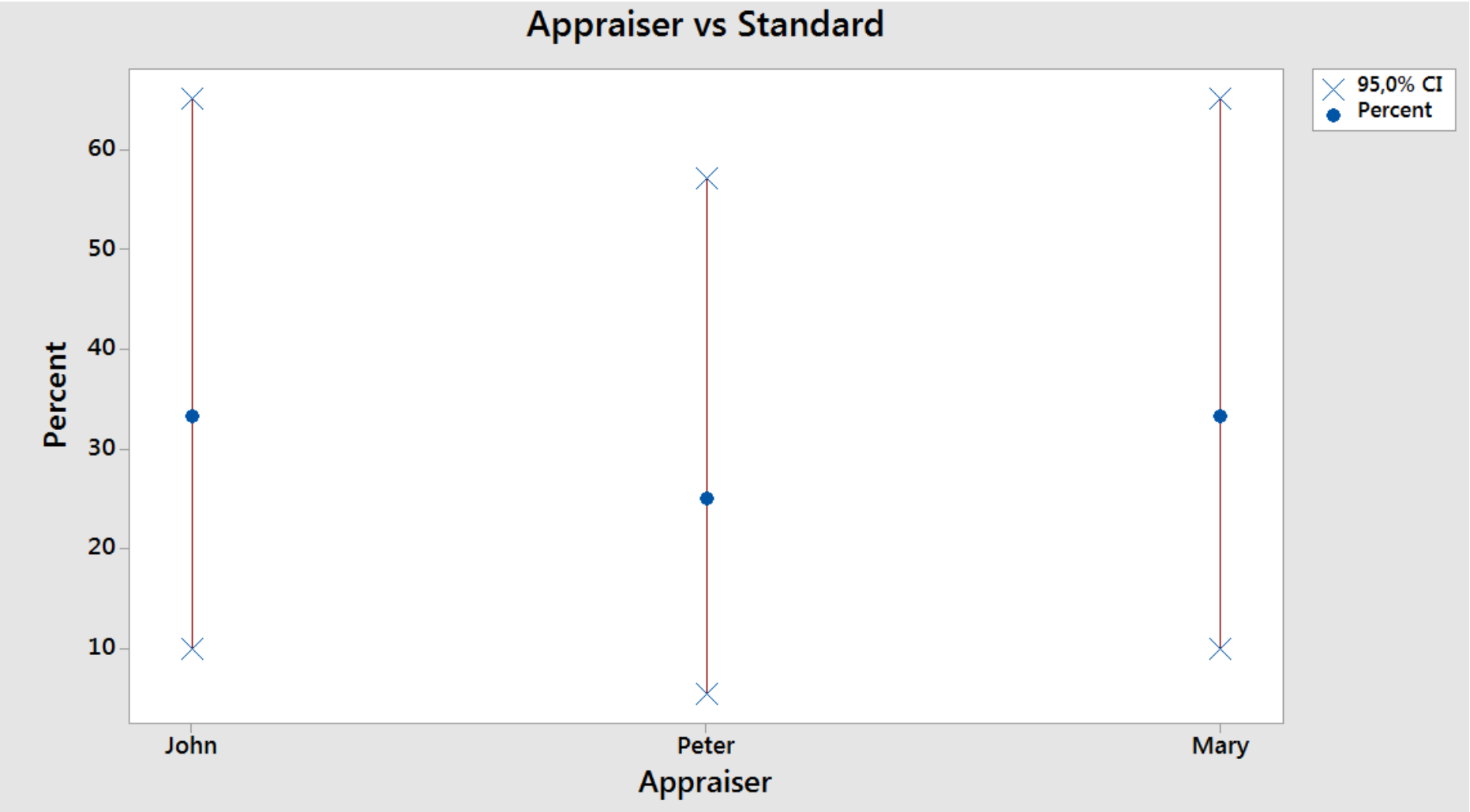
Options...

Graphs...

Results...

OK

DOES BOTTLED WATER ACTUALLY TASTE BETTER?



Attribute Agreement Analysis

1) Agreement within appraisers

Kappa statistics should be greater than 0.7 for consistency.

Use the p-values to choose between two opposing hypotheses, based on your sample data:

- H_0 : The agreement within appraiser is due to chance (appraiser is not consistent with himself)
- H_1 : The agreement within appraiser is not due to chance (appraiser is consistent with himself) → **P-value < 0.05**

2) Agreement between appraiser (check graph and agreement assessment)

3) Correlation between responses

Kendall's coefficient of concordance can range from 0 to 1. **The higher the value of Kendall's, the stronger the association. (ordinal data)**

Use the p-values to choose between two opposing hypotheses, based on your sample data:

- H_0 : There is no association among multiple ratings made by an appraiser (Responses are not correlated)
- H_1 : The ratings are associated with one another (responses are correlated) → **P-value < 0.05**

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

Each Appraiser vs Standard

Assessment Agreement

Appraiser	# Inspected	# Matched	Percent	95% CI
John	12	4	33,33	(9,92; 65,11)
Peter	12	3	25,00	(5,49; 57,19)
Mary	12	4	33,33	(9,92; 65,11)

Matched: Appraiser's assessment across trials agrees with the known standard.

Fleiss' Kappa Statistics

Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
John	Fiji	-0,008403	0,288675	-0,02911	0,5116
	Generic	0,555556	0,288675	1,92450	0,0271
	Tap	0,111111	0,288675	0,38490	0,3502
	Zephyrhills	-0,263158	0,288675	-0,91161	0,8190
	Overall	0,106977	0,167447	0,63887	0,2615
Peter	Fiji	-0,008403	0,288675	-0,02911	0,5116
	Generic	0,111111	0,288675	0,38490	0,3502
	Tap	0,111111	0,288675	0,38490	0,3502
	Zephyrhills	-0,263158	0,288675	-0,91161	0,8190
	Overall	-0,004651	0,167447	-0,02778	0,5111
Mary	Fiji	-0,263158	0,288675	-0,91161	0,8190
	Generic	0,111111	0,288675	0,38490	0,3502
	Tap	0,747368	0,288675	2,58896	0,0048
	Zephyrhills	-0,125000	0,288675	-0,43301	0,6675
	Overall	0,098592	0,168671	0,58452	0,2794

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

Between Appraisers

Assessment Agreement

# Inspected	# Matched	Percent	95% CI
12	1	8,33	(0,21; 38,48)

Matched: All appraisers' assessments agree with each other.

Fleiss' Kappa Statistics

Response	Kappa	SE Kappa	Z	P(vs > 0)
Fiji	0,169231	0,166667	1,01538	0,1550
Generic	0,111111	0,166667	0,66667	0,2525
Tap	0,035714	0,166667	0,21429	0,4152
Zephyrhills	-0,037037	0,166667	-0,22222	0,5879
Overall	0,072165	0,096424	0,74841	0,2271

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

All Appraisers vs Standard

Assessment Agreement

```
# Inspected  # Matched  Percent      95% CI
           12           1      8,33  (0,21; 38,48)
```

```
# Matched: All appraisers' assessments agree with the
```

Fleiss' Kappa Statistics

Response	Kappa	SE Kappa	Z	P(vs > 0)
Fiji	-0,093322	0,166667	-0,55993	0,7122
Generic	0,259259	0,166667	1,55556	0,0599
Tap	0,323197	0,166667	1,93918	0,0262
Zephyrhills	-0,217105	0,166667	-1,30263	0,9036
Overall	0,066972	0,096912	0,69106	0,2448

A kappa value greater than 0.7 is considered acceptable, meaning that our testers are able to adequately select that brand from the rest of them. As you can see, there are no brands with kappa values greater than 0.7, therefore we conclude that with an overall kappa value of 0.067, the testers are not able to determine a difference between the brands of water.

In fact, since some of the values were close to zero, it means that they could have done just as well if they guessed (random chance), than actually tasting the water and making a selection. Some brands were actually below zero, which means that they were **worse** than random chance, so the testers would have done better by simply guessing.

DOES BOTTLED WATER ACTUALLY TASTE BETTER?

Bottom line: Stop buying bottled water, just reuse your water bottles by filling them up with filtered tap water (not recommended for long term use). Not only will it help your own pocketbook, but you'll help the environment, by preventing the creation of new bottles and reduce the transportation costs associated with getting the bottles to your local store.



Guide to a good study of Attribute Agreement Analysis



- Example of items: Good / Bad, -2 / -1 / 0/1/2 (likert), A / B / C / D / E
- Evaluators should evaluate approximately the same number of items;
- Establish the known pattern for each item;
- Select at least three evaluators for the study;
- Evaluators should evaluate each item at least twice;
- Evaluators should evaluate items under normal conditions;
- Evaluators should evaluate items in a random order to minimize bias;
- Give ample time between classifications to minimize memorization.

Create Attribute Agreement Analysis Worksheet

Example: An educational testing company is training 5 new professors to evaluate writing for high school students. Each professor will evaluate 15 essays using a five-point scale (-2, -1, 0, 1, 2). The company wants to assess whether evaluators perform well. The study doesn't have replication.

Sample standard/attribute in numeric values

Options...

Number of samples: 15

Number of appraisers: 5

Sampl	Sample Name	Numeric Standard
1	1	2
2	2	-1
3	3	0
4	4	-2
5	5	1
6	6	2
7	7	1
8	8	0
9	9	1
10	10	2

Appraise	Appraiser Name
1	Pedro
2	Paulo
3	Lucas
4	Maria
5	Nina

Number of replicates: 1

Random

Worksheet

Interpretation:

The summary of the experimental plan is given in the session:

Attribute Agreement Analysis Worksheet

Samples: 15 Appraisers: 5

Replicates: 1 Total runs: 75

Use Options to set the base of the random number generator to generate the same random sample more than once.

↓	C1	C2-T	C3-T	C4	C5
	RunOrder	Samples	Appraisers	Assessments	Standards
1	1	2	Pedro		-1
2	2	4	Pedro		-2
3	3	8	Pedro		0
4	4	14	Pedro		2
5	5	3	Pedro		0
6	6	5	Pedro		1
7	7	13	Pedro		1
8	8	7	Pedro		1

→ Professor's reviews are inserted in this column

Example –Attribute Agreement Analysis

An educational testing company is training 5 new professors to evaluate writing for high school students. Each professor will evaluate 15 essays using a five-point scale (-2, -1, 0, 1, 2). The company wants to assess whether evaluators perform well. The study doesn't have replication.

The screenshot shows the 'Attribute Agreement Analysis' dialog box. The 'Data are arranged as' section has 'Attribute column:' selected with 'Rating' in the text box. Below it are 'Samples:' with 'Sample' and 'Appraisers:' with 'Appraiser'. The 'Multiple columns:' section is unselected. There are buttons for 'Information...', 'Options...', 'Graphs...', and 'Results...'. At the bottom, there is a 'Select' button, a 'Known standard/attribute:' field with 'Attribute' and '(Optional)', a checked checkbox for 'Categories of the attribute data are ordered', and an 'OK' button.

Agreement Analysis
Essay.mtw

Professor x Standard

Assessment Agreement

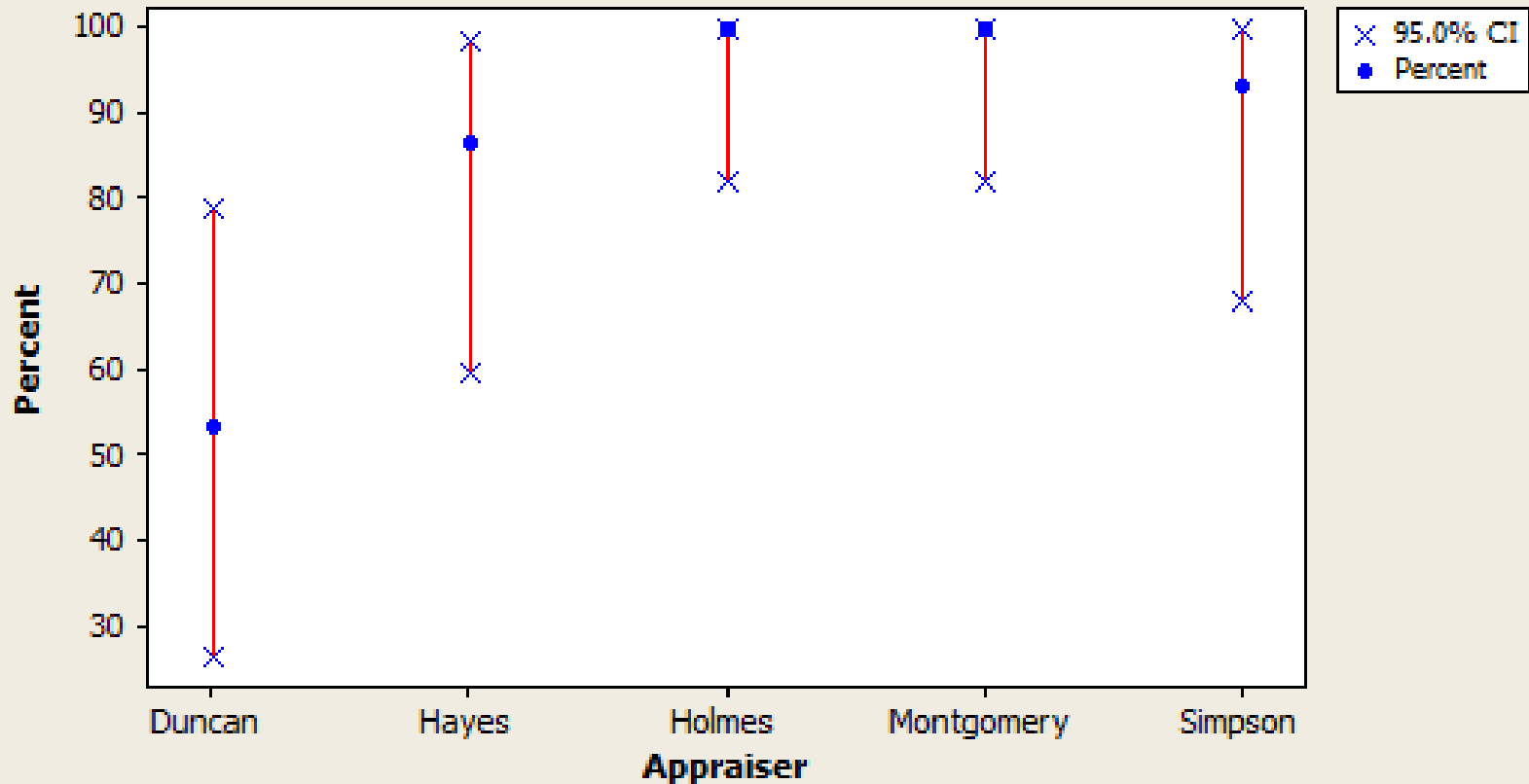
Date of study:

Reported by:

Name of product:

Misc:

Appraiser vs Standard



Professor x Standard

Attribute Agreement Analysis for Rating

Each Appraiser vs Standard

Assessment Agreement

Appraiser	# Inspected	# Matched	Percent	95% CI
Duncan	15	8	53.33	(26.59, 78.73)
Hayes	15	13	86.67	(59.54, 98.34)
Holmes	15	15	100.00	(81.90, 100.00)
Montgomery	15	15	100.00	(81.90, 100.00)
Simpson	15	14	93.33	(68.05, 99.83)

Matched: Appraiser's assessment across trials agrees with the known standard.

Fleiss' Kappa Statistics

Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
Duncan	-2	0.58333	0.258199	2.25924	0.0119
	-1	0.16667	0.258199	0.64550	0.2593
	0	0.44099	0.258199	1.70796	0.0438
	1	0.44099	0.258199	1.70796	0.0438
	2	0.42308	0.258199	1.63857	0.0507
	Overall	0.41176	0.130924	3.14508	0.0008
Hayes	-2	0.62963	0.258199	2.43855	0.0074
	-1	0.81366	0.258199	3.15131	0.0008
	0	1.00000	0.258199	3.87298	0.0001
	1	0.76000	0.258199	2.94347	0.0016
	2	0.81366	0.258199	3.15131	0.0008
	Overall	0.82955	0.134164	6.18307	0.0000
Holmes	-2	1.00000	0.258199	3.87298	0.0001
	-1	1.00000	0.258199	3.87298	0.0001
	0	1.00000	0.258199	3.87298	0.0001
	1	1.00000	0.258199	3.87298	0.0001
	2	1.00000	0.258199	3.87298	0.0001
	Overall	1.00000	0.131305	7.61584	0.0000
Montgomery	-2	1.00000	0.258199	3.87298	0.0001
	-1	1.00000	0.258199	3.87298	0.0001
	0	1.00000	0.258199	3.87298	0.0001
	1	1.00000	0.258199	3.87298	0.0001
	2	1.00000	0.258199	3.87298	0.0001
	Overall	1.00000	0.131305	7.61584	0.0000
Simpson	-2	1.00000	0.258199	3.87298	0.0001
	-1	1.00000	0.258199	3.87298	0.0001
	0	0.81366	0.258199	3.15131	0.0008
	1	0.81366	0.258199	3.15131	0.0008
	2	1.00000	0.258199	3.87298	0.0001
	Overall	0.91597	0.130924	6.99619	0.0000

Kendall's Correlation Coefficient

Kendall's Correlation Coefficient

Appraiser	Coef	SE Coef	Z	P
Duncan	0.87506	0.192450	4.49744	0.0000
Hayes	0.94871	0.192450	4.88016	0.0000
Holmes	1.00000	0.192450	5.14667	0.0000
Montgomery	1.00000	0.192450	5.14667	0.0000
Simpson	0.96629	0.192450	4.97151	0.0000

The high value of the Kendall coefficient means that the evaluators are essentially applying the same standard when evaluating the essays.

Between Appraisers

Assessment Agreement

# Inspected	# Matched	Percent	95% CI
15	6	40.00	(16.34, 67.71)

Matched: All appraisers' assessments agree with each other.

Fleiss' Kappa Statistics

Response	Kappa	SE Kappa	Z	P (vs > 0)
-2	0.680398	0.0816497	8.3331	0.0000
-1	0.602754	0.0816497	7.3822	0.0000
0	0.707602	0.0816497	8.6663	0.0000
1	0.642479	0.0816497	7.8687	0.0000
2	0.736534	0.0816497	9.0207	0.0000
Overall	0.672965	0.0412331	16.3210	0.0000

Kendall's Coefficient of Concordance

Coef	Chi - Sq	DF	P
0.966317	67.6422	14	0.0000

All Appraisers vs Standard

Assessment Agreement

# Inspected	# Matched	Percent	95% CI
15	6	40.00	(16.34, 67.71)

Matched: All appraisers' assessments agree with the known standard.

Fleiss' Kappa Statistics

Response	Kappa	SE Kappa	Z	P (vs > 0)
-2	0.842593	0.115470	7.2971	0.0000
-1	0.796066	0.115470	6.8941	0.0000
0	0.850932	0.115470	7.3693	0.0000
1	0.802932	0.115470	6.9536	0.0000
2	0.847348	0.115470	7.3383	0.0000
Overall	0.831455	0.058911	14.1136	0.0000

Kendall's Correlation Coefficient

Coef	SE Coef	Z	P
0.958012	0.0860663	11.1090	0.0000

* NOTE * Single trial within each appraiser. No percentage of assessment agreement within appraiser is plotted.

Agreement Analysis

Interpretation:

In general, these statistics suggest good agreement.

Kendall values range from 0 to 1. The higher the value of Kendall, the stronger the association. Generally Kendall coefficients of 0.9 or above are considered very good. The high value of the Kendall coefficient means that the evaluators are essentially applying the same standard when evaluating the essays.

A closer look at Each Appraiser vs. Standard indicates that the Duncan and Hayes ratings are not consistent with the standard. Holmes and Montgomery, however has a perfect match with standards.

Duncan, Hayes and Simpson could have additional training