

English Discipline Interrater Reliability Study Fall Day 2009

Process

The six faculty in the day program came together and read five papers selected by Justin Wyble, the Writing Program Assessment Coordinator. The faculty applied the current EN 102 Final Paper Common Rubric, which is shared by all instructors of EN 102 courses. The faculty applied the rubric independently and then discussed why they gave the ratings they did for each of the papers.

Results

The interrater reliability analysis report is below along with the data tables. First, only three papers had data points for all four standards, which made the sample very small. Two faculty did not complete the fifth paper, the fourth paper was missing one standard and could not be used in the analysis.

The size of the sample and variation produced an average measure of intraclass correlation of .4121, which is low and indicates of a lack of interrater reliability at present.

The overall picture of data supported the notion that while the rubric is understood, the faculty did attach different expectations and therefore weight to each of the standards and the relative levels of achievement in a given paper.

Action Plans

Based on the results of our Fall 2009 inter-rater reliability exercise, the English faculty have begun to discuss possible revisions to our common grading rubric for the EN 102 final research project. We believe that many of the differences in scoring can be attributed either to ambiguous and/or vague descriptions of student expectations or to the lack of particular categories in the current rubric. During our follow-up meeting, we also discussed how our pedagogical theories and practices may influence how much weight we assign certain categories.

Our immediate goal for Spring 2010 is to revise the current common grading rubric based on this semester's inter-rater reliability exercise and our related discussions. Once we agree upon a revised common grading rubric, we will conduct another inter-rater reliability exercise before the close of Spring 2010. The long-term goal of this on-going, collaborative process is to develop a grading rubric that will help our students to better understand our expectations and improve their writing skills, while allowing us to evaluate their writing more effectively.

RELIABILITY ANALYSIS - SCALE (ALPHA)

		Mean	Std Dev	Cases
1.	FACA	6.5000	3.7749	3.0
2.	FACB	10.5000	2.2913	3.0
3.	FACC	6.0000	.0000	3.0
4.	FACD	8.0000	2.6458	3.0
5.	FACE	6.8333	2.0207	3.0
6.	FACF	9.6667	1.4434	3.0

Intraclass Correlation Coefficient

Two-Way Mixed Effect Model (Consistency Definition):

People Effect Random, Measure Effect Fixed

Single Measure Intraclass Correlation = .1046*

95.00% C.I.: Lower = -.1296 Upper = .9167

F = 1.7009 DF = (2, 10.0) Sig. = .2313 (Test Value = .0000)

Average Measure Intraclass Correlation = .4121**

95.00% C.I.: Lower = -2.2080 Upper = .9851

F = 1.7009 DF = (2, 10.0) Sig. = .2313 (Test Value = .0000)

*: Notice that the same estimator is used whether the interaction effect

is present or not.

** : This estimate is computed if the interaction effect is absent, otherwise ICC is not estimable.

Reliability Coefficients

N of Cases = 3.0

N of Items = 6

Alpha = .4121

Data Tables

Faculty member	Paper One	Kosreaan Migrants			Mean	Standard deviation	Overall grade
	Integration of resources	clarity and rhetoric	Organizational development	Grammar and Mechanics			
A	3	3	2	2.5	2.63	0.479	2.5
B	3.5	3.5	3.25	2.25	3.13	0.595	3.38
C	2	1	1	2	1.50	0.577	1
D	3	2	2	2	2.25	0.500	2.38
E	1	2	1.5	2	1.63	0.479	1.5
F	2.5	3	2.5	2.5	2.63	0.250	2.5
mean	2.50	2.42	2.04	2.21	2.29		2.21
standard deviation	0.894	0.917	0.781	0.246	0.631		0.841

	Paper Two		Differences in Societies					
Faculty member	Integration of resources	clarity and rhetoric	Organizational development	Grammar and Mechanics	Mean	Standard deviation	Overall grade	
A	2	1	2	1	1.50	0.577	1.25	
B	3.25	2.25	3.5	2	2.75	0.736	3.25	
C	2	1	2	1	1.50	0.577	1	
D	2	1	1	1	1.25	0.500	1.25	
E	1	1	1	2	1.25	0.500	1	
F	2.25	2	2.25	1.5	2.00	0.354	2.25	
mean	2.08	1.38	1.96	1.42	1.71		1.67	
standard deviation	0.719	0.586	0.928	0.492	0.579		0.904	

	Paper Three Cut Roots						
Faculty member	Integration of resources	clarity and rhetoric	Organizational development	Grammar and Mechanics	Mean	Standard deviation	Overall grade
A	1	0	0	2	0.75	0.957	1
B	1.5	2	3	1.5	2.00	0.707	2
C	2	1	1	2	1.50	0.577	2
D	2	3	3	2	2.50	0.577	2.5
E	2	2	3	2	2.25	0.500	2
F	2.5	2.5	3.25	2.25	2.63	0.433	2.5
mean	1.83	1.75	2.21	1.96	1.94		2.00
standard deviation	0.516	1.084	1.364	0.246	0.706		0.548

	Paper Four Flying with Wings						
Faculty member	Integration of resources	clarity and rhetoric	Organizational development	Grammar and Mechanics	Mean	Standard deviation	Overall grade
A	3	3	3	2	2.75	0.500	2.5
B	1	1	1.5	1	1.13	0.250	1
C	1	2	2	1	1.50	0.577	1
D	1	1	1	1	1.00	0.000	1
E	na	0	0	na	0.00	0.000	0
F	1.25	1.5	1.5	0.5	1.19	0.473	1.25
mean	1.45	1.42	1.50	1.10	1.26		1.13
standard deviation	0.873	1.021	1.000	0.548	0.890		0.802

	Paper Five Discrimination						
Faculty member	Integration of resources	clarity and rhetoric	Organizational development	Grammar and Mechanics	Mean	Standard deviation	Overall grade
A	3	3	3	2	2.75	0.500	3
B	na	na	na	na	na	na	na
C	1	1	1	1	1.00	0.000	1
D	2	3	2	2	2.25	0.500	2
E	na	na	na	na	na	na	na
F	2.25	2.5	3	2	2.44	0.427	2.5
mean	2.06	2.38	2.25	1.75	2.11		2.13
standard deviation	0.826	0.946	0.957	0.500	0.768		0.854

Overall Totals							
Faculty member	Integration of resources	clarity and rhetoric	Organizational development	Grammar and Mechanics	Mean	Standard deviation	Overall grade
A	2.40	2.00	2.00	1.90	2.08	0.222	2.05
B	2.31	2.19	2.81	1.69	2.25	0.462	2.41
C	1.60	1.20	1.40	1.40	1.40	0.163	1.20
D	2.00	2.00	1.80	1.60	1.85	0.191	1.83
E	1.33	1.25	1.38	2.00	1.49	0.344	1.13
F	2.15	2.30	2.50	1.75	2.18	0.318	2.20
mean	1.97	1.82	1.98	1.72	1.87		1.80
standard deviation	0.419	0.477	0.583	0.214	0.359		0.531