



Reliability of Student Grades in General Education Courses



Stephen C. Light, Ph.D.
SUNY-Plattsburgh

Issues

- Are grades a consistent measure of student learning (reliability)?
- Do grades measure student learning, or something else (validity)?

Reliability and Validity

- Reliability = Consistency
- Reliability = Same score on same measure at different times
- (versus Validity: the item measures what it is intended to measure) – not in this analysis

In a Perfect World...

- Ideally, grades should measure student learning rather than something else, such as:
 - instructor's grading philosophy
 - seniority of the instructor
 - tenured vs. non-tenured instructor
 - full-timer vs. adjunct
 - other factors...

Reliability of Grades

- Multiple sections of same course.
- Same course in different semesters.
- Same course with different instructors.

- Ideally, the average grade in each section should be the same. Is it?...

**Table 1: *Average
Grades in Selected
General Education
Courses: 4-
Semester Analysis
(grades on 4-point
scale)**

Course	Title	Spring 2006	Fall 2006	Spring 2007	Fall 2007	4-Semester	Number of Sections	Number of Students
		Range of Avg. Grade	Range of Avg. Grade	Range of Avg. Grade	Range of Avg. Grade	Range of Avg. Grade		
BIO100	Concepts in Biology	0.09	0.54	0.09	0.24	0.62	8	560
CMM101	Introduction to Public Speaking	0.80	1.39	1.00	1.00	1.46	73	1510
ENG101	College Writing II	1.26	1.12	1.23	1.16	1.41	90	1802
HIS101	U.S. Civilization to 1877	0.71	0.86	0.33	1.36	1.45	27	1011
HIS102	U.S. Civilization Since 1877	1.03	1.01	1.03	0.96	1.05	26	985
MAT161	Introduction to Statistics	1.63	1.43	1.97	1.39	2.20	55	2009
PSY101	General Psychology	0.98	0.79	0.89	0.91	1.17	25	1371
SOC101	Introduction to Sociology	0.86	0.79	0.91	0.78	1.63	25	1273
SPA112	Elementary Spanish II	1.28	1.42	1.76	2.25	2.46	43	994
							372 sections	11515 students

Reliability Results by Course (Example: Spanish 112-from Table 1)

- SPA112 Elementary Spanish II – (Reliability standard met in 0 of 4 semesters analyzed).
- The range of average grades across 43 sections is 2.46, which is the largest range of all of the courses in this analysis. The range in each of the four semesters is higher than 1.0. A clear instructor effect is apparent from examination of the range for all of the sections combined. The sections with the 11 highest average grades were taught by a single full time instructor. The three lowest average grades were awarded in sections taught by a single adjunct instructor.

Summary of Results

- Of the nine courses examined in this analysis, five met the recommended standard in three of the four semesters analyzed. Four of the courses did not meet the standard.
- Is there too much variation in the average grade across sections? YES

What to Do?

- Presentation of grading data at department meetings
- Training of adjunct faculty and new faculty
- Discussion of grading philosophies
 - Bell curve
 - Contract grading
 - Other?
- Too high or too low average grades?
- Who grades high and who grades low?
 - Why?