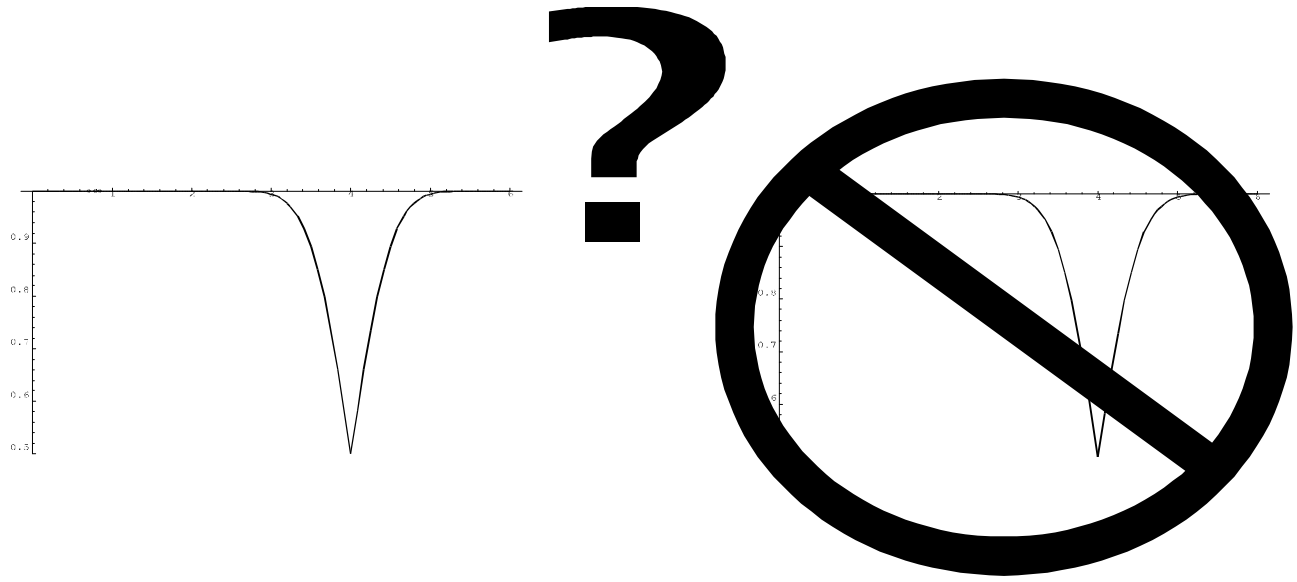


Examples of the Performance of G-theory Extensions for Estimating Error

David Rogosa
Haggai Kupermintz
June 24, 1996

CCSSO National Conference on Large-Scale Assessment

Research supported by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Questions about 'Error' in Assessment Data

Error in what

Individual Scores = > Assessment of Misclassification

Group Summaries = > Bias and precision of
Proportion at or above cut-off (PAC)

Facets of Error **(for pXtXr structures)**

TASKS **RATERS**

Discrete Formulation for pXtXr design using Task and Rater Misclassification

Task Misclassification

Abstract Representation.
Rows are true category membership, columns the category a perfectly scored task response would receive. Diagonal entries are 1 - t and entries above and below the diagonal are t/2 or t--e.g. t = .3 for a six category system , gives the matrix at the right.

task misclass. matrix: t = .3

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	.70	.30	.00	.00	.00	.00
[2,]	.15	.70	.15	.00	.00	.00
[3,]	.00	.15	.70	.15	.00	.00
[4,]	.00	.00	.15	.70	.15	.00
[5,]	.00	.00	.00	.15	.70	.15
[6,]	.00	.00	.00	.00	.30	.70

Rater misclassifications

Rows representing the perfectly scored paper and columns the result of the raters (fallible) scoring. Empirical rater misclassification matrices ("experts" vs regular scorings) from 1994 CLAS Writing (6x6) and Mathematics (4x4)

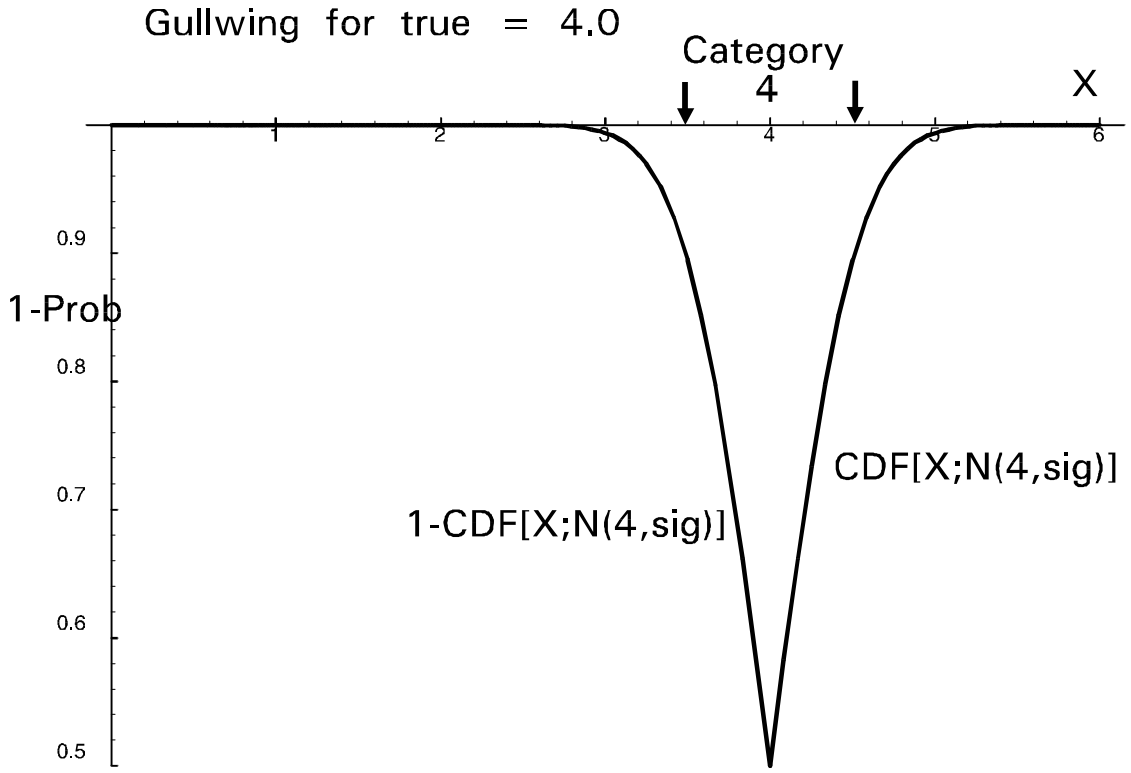
Mathematics Rater Misclassification (9,000 scorings)

	1	2	3	4
1	.913	.080	.006	.002
2	.049	.848	.097	.006
3	.004	.085	.802	.109
4	.001	.004	.088	.908

Writing Rater Misclassification (7,000 scorings)

	1	2	3	4	5	6
1	.191	.515	.243	.043	.009	.000
2	.167	.494	.283	.054	.003	.000
3	.007	.250	.580	.158	.005	.000
4	.003	.071	.300	.461	.154	.011
5	.001	.001	.026	.315	.481	.177
6	.000	.000	.002	.057	.374	.567

Misclassifications from G-theory



The "standard error" obtained from the G-theory variance components is denoted by "sig" in the above.
To compute category misclassifications (e.g. off-by-one-or-more, 1 - hit rate; off-by-two-or-more): Ave over category range of
[Pr{score < category boundary} + Pr{score > category boundary}]
for true score in category range.

Examples Based on the Math ratings

BASE rater

	1	2	3	4
1	.913	.080	.006	.002
2	.049	.848	.097	.006
3	.004	.085	.802	.109
4	.001	.004	.088	.908

base - .10

	1	2	3	4
1	.813	.171	.0128	.0043
2	.081	.748	.161	.0099
3	.006	.128	.702	.164
4	.002	.008	.183	.808

RESULTS pXr anova

var. comp.: p r pr,e

1.096 0 .149

	Hit rate	**off>=2**
emp|2	.846	.0059
emp|3	.801	.0044
gull(.386)	.683	.0014

var. comp.: p r pr,e

.951 0 .251

	Hit rate	**off>=2**
emp|2	.743	.009
emp|3	.70	.0067
gull(.501)	.602	.0094

RESULTS pXtXR anova t = .1

variance components:

p	t	r	pt	pr	tr	ptr,e
.96	0	0	.08	0	0	.151

	Hit rate	**off>=2**
emp|2	.771	.0118
emp|3	.733	.0048
gull(.485)	.613	.0076

variance components:

p	t	r	pt	pr	tr	ptr,e
.85	0	0	.07	0	0	.248

	Hit rate	**off>=2**
emp|2	.690	.0166
emp|3	.651	.0097
gull(.564)	.562	.0183

RESULTS pXtXR anova t = .3

variance components:

p	t	r	pt	pr	tr	ptr,e
.73	0	0	.22	0	0	.154

	Hit rate	**off>=2**
emp|2	.618	.0206
emp|3	.600	.0108
gull(.614)	.532	.0279

variance components:

p	t	r	pt	pr	tr	ptr,e
.65	0	0	.184	0	0	.26

	Hit rate	**off>=2**
emp|2	.567	.0352
emp|3	.556	.0162
gull(.666)	.504	.040

Examples Based on the Math ratings

base - .20

	1	2	3	4
1	.713	.262	.0196	.0065
2	.113	.648	.225	.0139
3	.008	.171	.602	.219
4	.003	.013	.277	.708

base - .30

	1	2	3	4
1	.613	.353	.026	.0088
2	.146	.548	.288	.0178
3	.010	.214	.502	.274
4	.004	.017	.372	.608

RESULTS pXr anova

var. comp.: p r pr,e

.805 0 .336

	Hit rate	**off>=2**
emp|2	.646	.0143
emp|3	.609	.0067
gull(.579)	.552	.0210

var. comp.: p r pr,e

.674 0 .417

	Hit rate	**off>=2**
emp|2	.554	.0182
emp|3	.499	.0097
gull(.645)	.515	.0349

RESULTS pXtXR anova t = .1

variance components:

p	t	r	pt	pr	tr	ptr,e
.71	0	0	.06	0	0	.339

	Hit rate	**off>=2**
emp|2	.616	.0256
emp|3	.561	.0129
gull(.636)	.520	.0327

variance components:

p	t	r	pt	pr	tr	ptr,e
.60	0	0	.05	0	0	.418

	Hit rate	**off>=2**
emp|2	.516	.0301
emp|3	.477	.0153
gull(.686)	.493	.0450

RESULTS pXtXR anova t = .3

variance components:

p	t	r	pt	pr	tr	ptr,e
.56	0	0	.15	0	0	.355

	Hit rate	**off>=2**
emp|2	.517	.0439
emp|3	.498	.0231
gull(.712)	.480	.0524

variance components:

p	t	r	pt	pr	tr	ptr,e
.48	0	0	.13	.01	0	.416

	Hit rate	**off>=2**
emp|2	.465	.0564
emp|3	.451	.029
gull(.747)	.463	.0627

Examples Based on the Writing ratings

BASE Rater misclass

	1	2	3	4	5	6
1	.191	.515	.243	.043	.009	.000
2	.167	.494	.283	.054	.003	.000
3	.007	.250	.580	.158	.005	.000
4	.003	.071	.300	.461	.154	.011
5	.001	.001	.026	.315	.481	.177
6	.000	.000	.002	.057	.374	.567

base + .10

	1	2	3	4	5	6
1	.291	.451	.213	.0377	.008	.000
2	.134	.594	.227	.0433	.002	.000
3	.0053	.190	.680	.120	.004	.000
4	.0024	.0578	.244	.561	.125	.009
5	.0008	.0008	.021	.254	.581	.143
6	.000	.000	.0015	.044	.288	.667

RESULTS pXr anova

var. comp.: p r pr,e

1.592 0 .575

	Hit rate	**off>=2**
emp|3	.585	.0108
gull(.759)	.458	.0662

var. comp.: p r pr,e

1.76 0 .509

	Hit rate	**off>=2**
emp|3	.680	.008
gull(.713)	.480	.053

RESULTS pXtXR anova t = .1

variance components:

p	t	r	pt	pr	tr	ptr,e
1.5	0	0	.05	0	0	.576

	Hit rate	**off>=2**
emp|3	.55	.027
gull(.791)	.444	.077

variance components:

p	t	r	pt	pr	tr	ptr,e
1.7	0	0	.06	0	0	.498

	Hit rate	**off>=2**
emp|3	.63	.023
gull(.747)	.463	.0627

RESULTS pXtXR anova t = .3

variance components:

p	t	r	pt	pr	tr	ptr,e
1.4	0	0	.15	0	0	.573

	Hit rate	**off>=2**
emp|3	.491	.0613
gull(.852)	.419	.097

variance components:

p	t	r	pt	pr	tr	ptr,e
1.5	0	0	.16	0	0	.50

	Hit rate	**off>=2**
emp|3	.541	.0495
gull(.812)	.435	.0835

Examples Based on the Writing ratings

base + .20

	1	2	3	4	5	6
1	.391	.388	.183	.0324	.007	.000
2	.101	.694	.171	.0327	.002	.000
3	.0037	.131	.780	.0828	.003	.000
4	.0019	.0447	.189	.661	.097	.007
5	.0006	.0006	.016	.194	.681	.109
6	.000	.000	.0011	.0307	.201	.767

base + .30

	1	2	3	4	5	6
1	.491	.324	.153	.027	.006	.000
2	.068	.794	.116	.022	.001	.000
3	.002	.071	.880	.045	.001	.000
4	.0013	.032	.133	.761	.068	.005
5	.0004	.0004	.011	.133	.781	.075
6	.000	.000	.0006	.018	.115	.867

RESULTS pXr anova

var. comp.: p r pr,e

1.941 0 .425

	Hit rate	**off>=2**
emp|3	.778	.0066
gull(.652)	.511	.0365

var. comp.: p r pr,e

2.18 0 .317

	Hit rate	**off>=2**
emp|3	.875	.0031
gull(.563)	.562	.0182

RESULTS pXtXR anova t = .1

variance components:

p	t	r	pt	pr	tr	ptr,e
1.9	0	0	.06	0	0	.417

	Hit rate	**off>=2**
emp|3	.723	.0154
gull(.692)	.490	.0468

variance components:

p	t	r	pt	pr	tr	ptr,e
2.1	0	0	.06	0	0	.313

	Hit rate	**off>=2**
emp|3	.808	.0099
gull(.615)	.531	.0281

RESULTS pXtXR anova t = .3

variance components:

p	t	r	pt	pr	tr	ptr,e
1.7	0	0	.18	0	0	.409

	Hit rate	**off>=2**
emp|3	.601	.0353
gull(.767)	.454	.0689

variance components:

p	t	r	pt	pr	tr	ptr,e
1.9	0	0	.19	0	0	.302

	Hit rate	**off>=2**
emp|3	.663	.0218
gull(.703)	.485	.0499

Proportion At or Above Cut-off (PAC)

PAC simulations (group size = 100; reps = 200)

Math (cut = 3)

True Cat dist	rater	task	se(PAC)*n^.5	(pq)^.5	G-anova
Flat	base	perfect	.20080	.499	.217
{10,40,40,10}	base	perfect	.25631	.499	
Flat	base-.30	perfect	.35049	.499	.335
{10,40,40,10}	base-.30	perfect	.42709	.499	
Flat	base	t=.3	.33148	.499	
Flat	base-.30	t=.3	.37085	.499	

Writing (cut = 4)

True Cat dist	rater	task	se(PAC)*n^.5	(pq)^.5	G-anova
Flat	base	perfect	.29869	.499	.290
{5,15,30,30,15,5}	base	perfect	.35670	.497	
Flat	base+.20	perfect	.22445	.499	.2427
{5,15,30,30,15,5}	base+.20	perfect	.27800	.498	
Flat	base	t=.3	.34124	.499	
Flat	base+.20	t=.3	.29265	.499	

References

Cronbach, L.J., Bradburn, N.M & Horvitz, D.G. Sampling and Statistical Procedures used in the California Learning Assessment System. Report of the Superintendent's Select Committee July 1994. Sacramento, CA: California State Department of Education.

Cronbach, L.J., Linn, R.L., Brennan, R.L., & Haertel, E. (1995). Generalizability analysis for educational assessment. Evaluation Comment, Summer 1995. Los Angeles: CRESST

Rogosa, D. R. Summarizing group performance using cutting scores. California Department of Education (Research, Evaluation, Technology Division), May 1993.

Rogosa, D. R. Misclassification in student performance levels. In *Technical Report: California Learning Assessment System 1993.* CTB/McGraw-Hill, May 1994.

http://www-leland.stanford.edu/~rag
http://www-leland.stanford.edu/~haggaik