



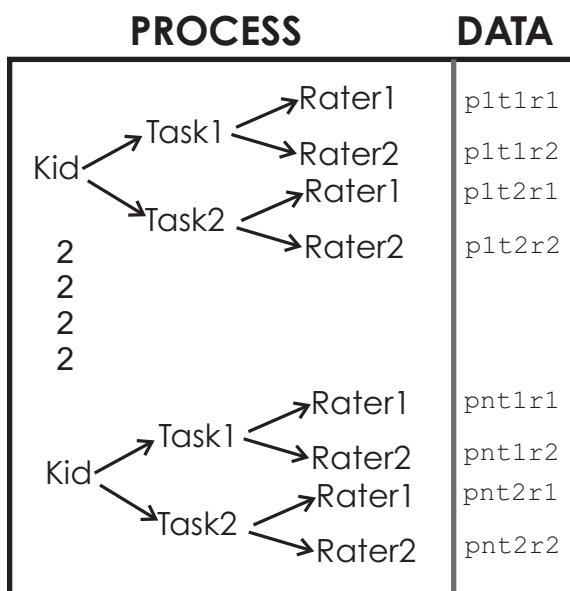
SIZE does matter (some) ... *but* **ORDER** really matters



Crossed-Design
(p x t x r)
n 2 2

DISCRETE FORMULATION

CONTINUOUS FORMULATION



Sidenote: conditional independence of ratings given task, etc

1. Student. For population of students, Population distribution indicated by probabilities of category membership—e.g., a 4-category population distribution "flat" is (1/4, 1/4, 1/4, 1/4).

2. Task. Task Misclassification Matrix. Rows are true category membership, columns the category a perfectly scored task response would receive. Example from artificial formulation

task misclass matrix: t=.3

	1	2	3	4	
1	.70	.30	.00	.00	Allows different matrices for each task; more complex forms.
2	.15	.70	.15	.00	
3	.00	.15	.70	.15	
4	.00	.00	.30	.70	

3. Rater. Rater Misclassification matrix Rows representing the perfectly scored paper and columns the result of the raters (fallible) scoring. Empirical rater misclassification Mathematics Rater Misclassification

	1	2	3	4	
1	.913	.080	.006	.002	Allows different matrices for each rater; more complex forms.
2	.049	.848	.097	.006	
3	.004	.085	.802	.109	
4	.001	.004	.088	.908	

1. Student. For each individual, start with a true ability which for a population (or a school) can have various distributions:

- a. Continuous Uniform on .5, 6.5 (mean 3.5 ; var 3.0)
- b. Triangular on .5, 6.5 (mean 3.5 ; var 1.5)
- c. Skew distribution on [.5, 6.5] (used in examples)

2. Task. Add task wobble. Task wobble most simply is represented by a Gaussian random variable with mean 0 (choice of task introduces no bias in the score) and variance T-- N(0, T).

3. Rater. Add rater smear. Again most basic (unrealistic) rater smear is a mean 0 (no bias) variance R Gaussian r. V. (raters interchangeable to start) . Thus add the N(0, R) r. v. to the score from a perfectly rated task.

Scoring 1-6 scale with category boundaries
1.5 2.5 3.5 4.5 5.5

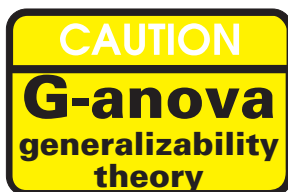
Task-Wobble, Rater-Smear

Base example (support [.5, 6.5])

person triang[.5, 4]
task N(0, .84) [No real t x p interaction]
rater N(0, .77)

CLBH recreation

Person triang[.65, 3.7]
task N(k_t, .69) ; k_t = [.37, 0, -.37]
rater N(k_r*N_{pr}(1,1) + k_{rt}, .59) ;
k_r = [.20, -.20] ;
k_{rt} = [.17, -.17, 0, 0, -.17, .17]



What happens when G-anova is applied?

