

# Familiar G-favorites: $p \times t \times r$ designs

Pictures from Brennan (1996), Generalizability of Performance Assessments, in *Technical Issues in Large-Scale Performance Assessment*

Variance components for CAP data (top)

G Study  $\hat{\sigma}^2$

Persons ( $p$ )	0.298
Tasks ( $t$ )	0.092
Raters ( $r$ )	0.003
$pt$	0.493
$pr$	0.000
$tr$	0.002
$ptr,e$	0.148

## Notable & quotable

Shavelson, Baxter, & Gao (1993) "The variance components for rater, school X rater, task X rater, and person:school X rater rounded to zero. The school X rater X task interaction accounted for about 1% of the total variability. These findings demonstrate that sampling variability due to raters is not a problem for school level assessments but task-sampling is." (p.226) and "...measurement error is introduced by task-sampling variability, and not by variability due to other measurement facets" (p. 229).

Gao et al (1994) "The variance component estimates for rater and the interactions between rater and school, rater and task and between rater and person:school were all close to zero....These findings, consistent with past research, suggest that rater-sampling variability might not be a problem for school-level assessments when raters are well-trained (p.333) Brennan (1996) "The fact that the estimates of the rater, person by rater, and task-by-rater variance components are all close to zero suggests that the rater facet does not contribute much to the variability in observed scores."

## Collection of $p \times t \times r$ studies (below)

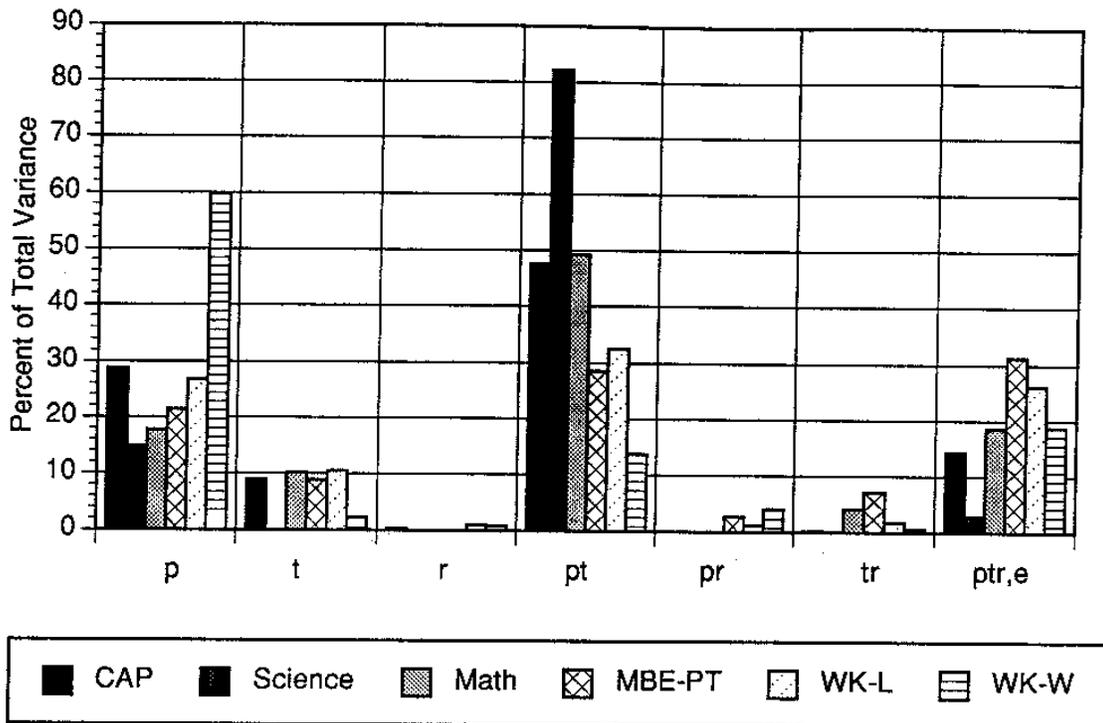


Figure 1. Percents of total variance for variance components for the  $p \times t \times r$  design for six different assessments.

Typical are the results reported for Brennan, Gao, & Colton (1995), where the variance components of this study (as % of total variance) are: p=28; t=10; r=1; pt=32; pr=1; tr=1; ptr,e=27

### Miller (1998), CCSSO

Aside: I tried to create some examples to recreate some examples from David Miller compendium, but found I couldn't understand many of the tables/situations. (e.g. Connecticut, the label "Rater" seems to indicate the high-order interaction confounded with error, also Delaware? North Carolina?) Maybe someone can help me.

Many examples in Miller (1998) he states as being consistent with the examples above [in his words "consistent with Brennan 1996"]: e.g. Alabama, Delaware, North Carolina

## North Carolina Miller, 1998

OBS	TEST	Pupil	Task	Tsk*Pup	Rater
1	MATH03	0.41220	0.07712	0.36213	0.29673
2	MATH04	0.20658	0.15744	0.55551	0.32599
3	MATH05	0.03850	0.06315	0.23790	0.19319
4	MATH06	0.07350	0.48445	0.34844	0.21911
5	MATH07	0.24908	0.32944	0.33966	0.48969
6	MATH08	0.18733	0.24988	0.02150	0.97799

