Implications: Sparse Domain Coverage

- View: PEs represent “sample” of the full K-12 FW for large scale assessment, specified by NGSS standards writers etc.
- Should intersections (the sample frame) be stable over time?
- If yes, then current choice IS sparse but still larger than can be sampled in one year even with partial matrix, may require 3 years (recall anchor). So any different sample likely as sparse.
- If no and the sampling frame EITHER shifts or is the saturated matrix (all intersections), then reporting likely insufficiently comparable year-to-year for individual reports (OK group).
- Is the problem really the particular PE sample identified rather than the size (doesn’t support naturalistic tasks unless short; not “even”, problematic intersections Gr 5), OR are we willing to grant FW too big and needs instructional priorities?

Implications: Less Concerning

- Nested data structures, assuming clusters not more than half and branching confined to partial credit scoring
- Multidimensionality, unlikely to be unacceptably present
- Subscores: Assuming they are sufficiently reliable and meet measurement standards in this way, is there enough separation? Total score and subscores often share a high degree of correlation. But consider Math and Science example, correlations often as high as .88 but still reported separately because they *signify* important differences of utility to the field (not only predictive). Argue signification *can be* important design value (unless correlation extremely high).
- Transfer concern realistically not an immediate need for large scale because student performance likely so weak just on PEs.

Implications: More Concerning

- Equating especially for sparse designs some states considering
- Clusters only (if more than half of assessment in testlets)
- PLDs very important so take time/effort described by Leslie
- Need MUCH better solutions for small repeated measures throughout year, especially in different sequences and windows
- Perhaps handscoring is BIG cost issue for NGSS, “white elephant in room,” but vendors claim more AI success than we can verify
- Caution on making formal inferences for individuals in setting collaborative: most of variance at group level in work w/Wilson
- “Are psychometricians prepared for oceans of data from Ed Tech?” – heavily attended NCME panel I served on in 2017