1. Are the ratings that are being made relatively reliable (i.e., > .70)? Are they absolutely reliable?
   1. Relative reliability (Based on G and Phi coefficients)
      1. **Current relative reliability is .575, and therefore not relatively reliable.**
      2. **Current absolute reliability is .487, and therefore not absolutely reliable.**

**For these items, I utilized the less conservative G coefficient for decision making. This assumes that rank ordering is important, but absolute scores are not.:**

1. If not, how many additional items would be required to obtain generalizable ratings?
   1. **Retaining 4 levels of facet one (items), the D study predicts a minimum of 6 occasions to exceed a .7 G coefficient, achieving a G coefficient of .730.**
2. If not, how many additional rating occasions would be required to obtain generalizable ratings?
   1. **Retaining 3 levels of facet two (occasions), the D study predicts a minimum of 17 items to exceed a .7 G coefficient, achieving a G coefficient of .702.**
3. What is the optimal addition/subtraction of items and occasions to ensure good generalizability?
   1. **To retain only the best quality test items, and to balance test fatigue with cost of administration, I recommend utilizing 5 items and 5 occasions, which produces an estimated a G coefficient measure of .717.**