this is for James, but I'm including the rest in case this is useful  
for you:  James, I'd like to correct the way we combine variables into  
indices in the GD survey to the "correct" way I mentioned the other  
day. As per my previous pre-analysis plan, you're currently using egen  
std to compute z-scored variables and then average them together and  
z-score again. But in averaging them together, there is a trick that  
takes into account the covariance structure between the variables.  
Simply, if two variables that make up an index together with others  
are highly correlated with each other, each of them receives lower  
weight.  
  
I implemented this in the attached ado-file, which is an extension to  
egen. You have to save the file in your personal ado folder (find it  
by typing "sysdir"), and then call it like so:  
  
egen indexvariable = weightave(varlist) [, normby(var)  
  
The output is a new variable (indexvariable) which contains a weighted  
average of  
varlist. The weighted average is computed following Anderson (2011) by  
subtracting the mean,  
normalizing by the standard deviation (usually of the control group;  
see below), computing the covariance matrix, inverting it, summing the  
rows of the covariance matrix, and then weighting each variable with  
its corresponding entry in the summed inverted covariance matrix.  
After that divide by the sum of the weights just to normalize.  
The "normby(var)" option, where var is a dummy variable, specifies  
that the standard deviation  
through which the mean-subtracted variables are divided is computed  
only from those observations  
where var = 1. E.g., we often want to subtract the mean of the whole  
sample, but divide by  
the standard deviation of the control group only; this does that. This  
can be achieved by calling:  
  
egen indexvariable = weightave(varlist), normby(controldummy)  
  
A special case occurs when varlist has only one element. In this case  
this variable receives  
weight 1, and the ouput is simply a z-scored (normalized) version of itself.  
  
Note that step 1 in Anderson's procedure - inverting variables so that  
"positive" outcomes  
go with higher values for all variables - needs to be done by hand  
before calling egen weightave.  
  
Further note that the program does its own z-scoring (as explained  
above), so no need to standardize  
variables before calling it.