Online Appendix to accompany Sterba, S.K. (2009).

Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration, *Multivariate Behavioral Research, 44,* 711-740.

Software Implementation of the Hybrid Model/Design-Based Framework.

Commonly used analytic procedures in standard model-based software packages such as SAS 9's Proc REG, Proc MIXED, and Proc CALIS (SAS Institute, Cary, NC); or SPSS 17's REGRESSION, AMOS, and MIXED (SPSS, Inc., Chicago, IL) cannot, at present, implement the hybrid design/model-based framework. Even though these procedures may include a WEIGHT option, the programs think that the weight variable is either a variance weight or a frequency weight, not a sampling weight. Even though these procedures may include an option for robust standard errors (e.g., Proc Mixed's EMPIRICAL), these standard errors are robust only to heteroscedasticity, and do not account for unmodeled stratification and clustering.

Instead, there are three main options for implementing the hybrid design/model-based framework. The first option is a set of newly-released survey modules in standard model-based software packages. In SAS 9, these include Proc surveyLOGISTIC and Proc surveyREG. In SPSS 17 these include GLM and LOGISTIC procedures within the Complex Samples module. In STATA 10 (StataCorp, College Station, TX), these include SVY:REGRESS and SVY:LOGIT. However, these survey modules of standard software packages do not accommodate, for example, structural equation, multilevel, or mixture modeling.

The second option is to use traditional design-based software packages that accommodate some modeling (e.g., SAS-callable SUDAAN from RTI International, Inc.). SUDAAN's procedures MULTILOG (for generalized multinomial logit models), LOGISTIC (for logistic regression), LOGLINK (for log linear models), and REGRESS (for linear regression) allow broader modeling possibilities than the survey modules of standard software packages, but again cannot accommodate many popular models, such as structural equation, multilevel, and mixture models.

The third option is to use specialized psychometric software programs that were once purely model-based, but have recently added the capability for sample-weighted point estimation and design-adjusted (linearized) variance estimation. Table 1 reviews such point and variance estimators provided by LISREL 8.8 (Jöreskog, Sörbom, du Toit, & du Toit, 2001), M*plus* 5 (Muthén & Muthén, 1998-2007), GLLAMM (Rabe-Hesketh, Skrondal, & Pickles, 2004), MLwiN 2.1 (Rasbash, Browne, Goldstein, Yang, Plewis, & Healy et al., 2000), and HLM 6 (Raudenbush, Bryk, & Congdon, 2007). These modeling programs are very flexible and can accommodate, for example, design/model-based analyses of: structural equation models (e.g., M*plus*, LISREL, GLLAMM) generalized linear single-level and multilevel models (all), and mixture models (GLLAMM, M*plus*). The performance of these programs was compared by Asparouhov (2004, 2005), Asparouhov and Muthén (2006), Bell-Ellison and Kromrey (2007), and du Toit, du Toit, Mels and Cheng, (2005).

Table 1. Psychometric software programs that account for complex sampling designs via the hybrid framework.

	Single-level model (Design-based adjustments for clustering)				Multilevel model (Modeling clustering)			
	Continuous Outcomes: Point Estimation	Categorical Outcomes: Point Estimation	Continuous Outcomes: Variance Estimation	Categorical Outcomes: Variance Estimation	Continuous Outcomes: Point Estimation	Categorical Outcomes: Point Estimation	Continuous Outcomes: Variance Estimation	Categorical Outcomes: Variance Estimation
LISREL	PML: Design-based weighting for D	PML: Design-based weighting for D	Linearization: Design-based adjustment for S,C; weighting for D	Linearization: Design-based adjustment for S,C; weighting for D	PWIGLS: S and C modeled (w/ random effect by default); Design-based weighting for D	PWIGLS S and C modeled (w/ random effects by default); Design-based weighting for D	Linearization:** Model accounts for S,C; Design-based weighting for D	Linearization:** Model accounts for S,C; Design-based weighting for D
M*plus*	PML:* Design-based weighting for D	PML:* Design-based weighting for D	Linearization: Design-based adjustment for S,C; weighting for D	Linearization: Design-based adjustment for S,C; weighting for D	MPML:* C modeled (w/ random effect by default); Design-based adjustment for S; weighting for D	MPML:* C modeled (w/ random effect by default); Design-based weighting for D, S	Linearization: Design-based adjustment for S,C; weighting for D	Linearization: Design-based adjustment for S,C; weighting for D
GLLAMM	PML: Design-based weighting for D	PML: Design-based weighting for D	Linearization: Design-based adjustment for S,C; weighting for D	Linearization: Design-based adjustment for S,C; weighting for D	MPML: C modeled (w/ random effect by default); Design-based adjustment for S; weighting for D	MPML: C modeled (w/ random effect by default); Design-based weighting for D, S	Linearization: Design-based adjustment for S,C; weighting for D	Linearization: Design-based adjustment for S,C; weighting for D
HLM					MPML: S and C modeled (w/ random effect by default); Design-based weighting for D	W-PQL: S and C modeled (w/ random effect by default); Design-based weighting for D	Linearization: Model accounts for S,C; Design-based weighting for D	Linearization: Model accounts for S,C; Design-based weighting for D
MLwiN					PWIGLS: S and C modeled (w/ random effect by default); Design-based weighting for D	W-PQL: S and C modeled (w/ random effect by default); Design-based weighting for D	Linearization:** Model accounts for S,C; Design-based weighting for D	Linearization:** Model accounts for S,C; Design-based weighting for D

Notes. S=stratification; C=clustering; D=disproportionate probabilities of selection; MPML=multilevel pseudo-maximum likelihood; PML=pseudo-maximum likelihood; PWIGLS=probability weighted iterative generalized least squares; W-PQL=weighted penalized quasi-likelihood; *Note that M*plus* allows many different probability-weighted estimators, not just PML. For example, for categorical outcomes, weighted least squares with mean and variance adjustment (WLSMV) can be modified to incorporate probability weights. See Muthén and Muthén (1998-2007, p. 457) for details. **Note that some programs (e.g., LISREL, MLwiN) automatically employ linearized standard errors *only* if weights are included. This can be overridden (e.g., in LISREL by specifying WEIGHT1=intcept). In all programs, multilevel weighted analyses allow for weights to be included at each level of the hierarchy (i.e., between-level weights and within-level weights). To counter biases induced when within-level weights are used with small cluster sizes, within-level weights are scaled. The method used to scale within-level weights differs across program (see Chantala, Blanchette, & Suchindran, 2006 for a comparison). One difference among these programs, in the case of multilevel analyses, is that some (LISREL, MLwiN, HLM) require strata to be entered as a level-3 random effect in a multilevel model, with clusters as the level-2 random effect. Other programs (M*plus*) automatically use clusters as level-2 random effects, and automatically use strata only to adjust standard error calculations. Still other programs (e.g., GLLAMM) allow either option. In contrast, Skinner, Holt, and Smith (1989) had suggested including strata as fixed effects. Finally, note that the first four columns are sometimes called an “Aggregated analysis” and the next four columns are sometimes called a “Disaggregated analysis” because of how these methods differentially handle clustering.

Variables used: (labels in caps are actual HSB datafile names)

SCHLID: cluster indictor from 1982 school datafile

Lev2wt: the level-2 weight: $\frac{1}{\pi_j}$, the inverse of the probability that cluster *j* is selected, which is SCHLWT in the 1980 school datafile.

Lev1wt: the level-1 weight: $\frac{1}{\pi_{i|j}}$, which is the inverse of the probability individual *i* selected given cluster *j* selected. In public-use datasets like HSB, this variable is often not provided. Rather, only a level-2 weight, $\frac{1}{\pi_j}$, and a total weight, $\frac{1}{\pi_{i|j}} \times \frac{1}{\pi_j}$, are available. But the total weight can be divided by the level 2 weight to yield the level 1 weight, $\frac{1}{\pi_{i|j}}$. In the HSB dataset $\frac{1}{\pi_j}$ is labeled SCHLWT in the 1980 school datafile and $\frac{1}{\pi_{i|j}} \times \frac{1}{\pi_j}$ is labeled RAWWT in the 1982 student datafile.

cses: school-mean centered BYSES, i.e. base-year student socioeconomic status, from the 1982 student datafile

sector: author-constructed variable denoting public or private school, constructed from the stratification variable SCHSAMP on the 1982 school datafile

meanses: author-constructed school means of BYSES

black: author-constructed variable denoting whether school had $\geq$30% Black enrollment, from school-level dataset variable SB0094S

hispanic: author-constructed variable denoting whether school had $\geq$30% Hispanic enrollment, from school-level dataset variable SB0093S

sectorXblack: author-constructed variable; product of *sector* x *black*

sectorXhisp: author-constructed variable; product of *sector* x *Hispanic*

mathach: student math achievement; BBMATHFS on the 1982 student datafile

M*plus* 5.2 Code for Model 1 from hybrid HSB analysis

```
data: file is hsbdata.dat;
variable: names are schlid lev2wt lev1wt cses mathach
sector meanses Black Hispanic sectorXblack sectorXhisp;
usevariables are mathach cses sector meanses;
missing are .;
within=cses;
between =sector meanses;
cluster = schlid;
analysis: type = meanstructure twolevel random ;
MODEL:
%WITHIN%
s1 | mathach ON  cses ;
%BETWEEN%
mathach on sector meanses;
s1 on sector meanses;
mathach with s1;
```

M*plus* 5.2 code for Model 2 from hybrid HSB analysis (additions to Model 1 shown in **bold**)

```
data: file is hsbdata.dat;
variable: names are schlid lev2wt lev1wt cses mathach
sector meanses Black Hispanic sectorXblack sectorXhisp;
usevariables are mathach cses sector meanses;
missing are .;
within=cses;
between =sector meanses;
cluster = schlid;
weight=lev1wt;
bweight=lev2wt;
analysis: type = meanstructure twolevel random ;
MODEL:
%WITHIN%
s1 | mathach ON  cses ;
%BETWEEN%
mathach on sector meanses;
s1 on sector meanses;
mathach with s1;
```

M*plus* 5.2 code for Model 3 from hybrid HSB analysis (additions to Model 2 shown in **bold**)

```
data: file is hsbdata.dat;
variable: names are schlid lev2wt lev1wt cses mathach
sector meanses Black Hispanic sectorXblack sectorXhisp;
usevariables are mathach cses sector meanses Black
Hispanic sectorXblack sectorXhisp;
```

```
missing are .;
within=cses;
between =sector meanses
black hispanic sectorXblack sectorXhisp;
cluster = schlid;
weight=lev1wt;
bweight=lev2wt;
analysis: type = meanstructure twolevel random ;
MODEL:
%WITHIN%
s1 | mathach ON  cses;
%BETWEEN%
mathach on sector meanses black
hispanic sectorXblack sectorXhisp;
s1 on sector meanses;
mathach with s1;
```

Appendix References

Asparouhov, T. (2004). *Stratification in multivariate modeling*: *Mplus Web Notes No. 9*. Unpublished manuscript. Retrieved from http://www.statmodel.com

Asparouhov, T. (2005). Sampling weights in latent variable modeling, *Structural Equation Modeling*, *12*, 411-434.

Asparouhov, T., & Muthén, B. O. (2006). *Multilevel modeling of complex survey data*. ASA Section on Survey Research Methods. Retrieved from http://www.statmodel.com

Bell-Ellison, B. A., & Kromrey, J. D. (2007). Alternatives for analysis of complex sample surveys: A comparison of SAS, SUDAAN, and AM software. Poster presented at the SAS Global Forum 2007.

Chantala, K., Blanchette, D., & Suchindran, C. M. (2006). Software to compute sampling weights for multilevel analysis. Unpublished manuscript, UNC-Chapel Hill. Retrieved from: http://www.cpc.unc.edu/restools/data_analysis/ml_sampling_weights/Compute%20Weights%20for%20Multilevel%20Analysis.pdf

du Toit, S. H. C., du Toit, M., Mels, G., & Cheng, Y., (2005). *Analysis of complex survey data with LISREL: Chapters 1-5.* Unpublished manual. Retrieved from: www.ssicentral.com

Jöreskog, K. G., Sörbom, D., du Toit, S., & du Toit, M.(2001). *LISREL 8: New statistical features. (Third printing with revisions).* Chicago: Scientific Software International.

Muthén, L. K. & Muthén, B. O. (1998-2007). *Mplus users guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., & Lewis, T. (2000). *A user's guide to MLwiN* (Version 2.1). London: Institute of Education.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2000). *HLM 6: Hierarchical linear and nonlinear* *modeling.* Chicago: Scientific Software International.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). *GLLAMM manual.* Technical Report 160. Division of Biostatistics, University of California, Berkeley. Retrieved from: http://www.bepress.com/ucbbiostat/paper160/

Skinner, C. J., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex surveys.* New York: Wiley.