A Comparison of Methods for a Survey of High School Students in Iowa

Lu Lu, Michael D. Larsen
Department of Statistics, Iowa State University, Ames, IA, 50011
email: icyemma@iastate.edu, larsen@iastate.edu

Abstract

Iowa's State Board of Education conducted a stratified multi-stage sample survey to study the availability of employment preparation courses and the degree to which students in Iowa's public high schools enroll in those courses. The design and estimation options for the survey motivated a series of research questions, which the authors have explored through simulation.

Keywords: Generalized variance functions; One PSU per stratum sampling; Probability proportional to size sampling; Ratio estimation; Survey costs; Systematic sampling.

1. Introduction

In 2004, representatives of Iowa's State Board of Education approached the Center for Survey Statistics and Methodology (CSSM) at Iowa State University (ISU) for help in planning a series of surveys. The purpose of one of the surveys is to study the availability of employment preparation (EP) courses and the degree to which students in Iowa's public high schools enroll in those courses. EP courses belong to a diverse set of courses including those in information technology, accounting and business, trades and professions, and agricultural management (Bradby et al. 1995). A primary concern of the survey is to assess the degree to which students in Iowa's public school districts, which vary greatly in size, community characteristics, and ruralness, have equal opportunities to prepare in school for employment, college, and life in general.

The sample survey was designed to produce estimates of average numbers of EP courses for the State of Iowa and populations of small (less than 250 students), medium (250 to less than 2500 students), and large (2500 or more students) school districts. Districts in Iowa are organized into twelve area education agencies (AEAs) for the purposes of administration and support. District size and AEA were used as stratifying variables. Districts were sampled with probability proportional to total enrollment size. For political reasons, all schools in selected districts were included in data collection. Only large and one medium school district in Iowa have more than one high school. Due to their extreme size, all large districts were included with certainty. A simple random sample of students was selected in each selected school. The samples were split between grade nine and grade twelve students, so that one could compare the first year high school students in 2005 to the ninth grade records of the seniors in 2005. A stratified three-stage design was proposed and implemented. Population quantities of interest, estimators, and estimators of variance are described in Section 2.

The design and estimation options for the survey given the budget restrictions motivated a series of research questions, which the authors have explored through simulation (Hewitt and Larsen 2005, 2006; Lu 2005). A population database of twelfth grade students was created through simulation. The numbers of EP courses taken by students in a school were generated as independent Poisson random variables with a rate for the school. The Poisson rates were generated independently from a random effects model with main effects due to school size and AEA, which are the factors used for the actual stratification. Based on examining preliminary data, the simulations did a reasonable job of creating a population database not unlike, in terms of number of courses taken by students, that being gathered in the survey. **The results presented in this paper are not actual results from the survey and should not be interpreted as characterizing schools in the State of Iowa**.

The first research question concerns sampling and the fact that schools have very different levels of enrollment. Alternatives for implementing probability proportional to size sampling are investigated under two assumptions concerning the number of primary sampling units. Methods of Murthy, Brewer, and Durbin are reviewed and results using Horvitz-Thompson and ratio estimators are presented in Section 3.

Second, data from only a single school were collected in some strata. Variance estimation using collapsed strata variance estimators followed by synthetic variance redistribution and generalized variance functions for designs with one primary sampling unit per stratum are proposed. Results in Section 4 suggest that it might be possible to produce reasonable estimates of stratum variance in one-per-stratum designs in some circumstances.

Third, resources for conducting the survey could be redistributed from small to big schools. This means that the design is not invariant in the sense of Särndal, Swensson, and Wretman (1992, page 134). Comparisons of variance estimation for Horvitz-Thompson and ratio estimators in an invariant (multi-stage) and a non-invariant (multi-phase) designs which correspond to circumstances with and without resource redistribution are reported in Section 5.

Fourth, future surveys could potentially include more schools. The trade-off between adding more schools and reducing the number of students per school in terms of precision of estimation are presented under different assumptions concerning the cost of adding schools to the survey in Section 6.

Section 7 is a summary, gives recommendations for future surveys, and suggests possible future research work.

2. Iowa's State Board of Education Employment Preparation Survey

The design of Iowa's State Board of Education Employment Preparation (EP) survey was described in general in Section 1. First, estimators for a stratified three-stage design with some strata having only one primary sampling unit (PSU) will be discussed. Second, the process of redistributing resources will be explicated.

2.1 Stratified Three-Stage Design Estimators

Two kinds of estimators are proposed to estimate the total number of employment preparation courses taken by high school students in a stratum. The schools in districts of a particular size within an AEA are the members of a stratum. The first estimator used in this paper is the π-expansion estimator of Horvitz-Thompson (HT; 1952). In a multi-stage sampling design, the inclusion probability is a product of probabilities of selection at all stages. In the EP survey, within strata defined by district size and AEA, the probability of selecting a student is the product of the probabilities of selecting districts, schools within districts, and the student within a school. In a multi-phase sampling, which is relevant for the non-invariant designs discussed in the next subsection and Section 5, an adjusted HT estimator, called the $\pi^\star$ estimator, in the terminology of Särndal, Swensson, and Wretman (1992, page 347), is suggested.

The second estimator used in this paper is the ratio estimator (Cochran 1977, chapter 6). Ratio estimation works well when a convenient and inexpensive auxiliary variable that is correlated with the response variable is available for all units in the population. Enrollment is known for all schools and is positively correlated with the number of enrollments in EP courses in the school.

The estimates of totals in the whole state, size levels, and AEAs are the sum of estimates of totals in all strata contained in those aggregations. The estimates of means are the estimates of totals divided by the number of students in the relevant aggregation.

2.2 Survey Resources Redistribution Process

In the planning phase of the actual survey, ISBE informed CSSM that it would be possible to collect data at sixty (60) schools. In the design with 60 schools, the 22 schools in eight large districts in seven AEAs were taken with certainty. The remaining 38 schools were split evenly between the medium and small school districts; 19 schools were selected from twelve strata in each size level. That means in each size level seven strata were assigned two PSUs and the remaining five strata had only one PSU sampled. The sample in each school included students from four population subgroups defined by two factors: ninth or twelfth grade and general or special (having Individual Education Plans, or IEPs) education groups. To sample a total of 12,000 students in 60 schools, 50 students on average were selected from each group. If the number of students in a group was substantially larger than 50, then initially 50 students were selected from the group. Most schools, however, had fewer than 50 IEP students total in grades 9 and 12. This situation allowed redistribution of resources. Specifically, additional student transcripts were examined in some groups in some schools up to 200 students total in every selected school. Further, resources from schools with fewer than 200 students were assigned to other larger schools so that excess sample was collected in some bigger schools. This was possible, because it usually was feasible to do more data collection in large schools. Students within groups in a sampled school were selected by simple random sampling (SRS) without replacement. This process of redistributing survey resources can be applied to other designs with a different number of schools and students.

3. Unequal Probability Sampling Without Replacement

Cochran (1977, chapter 9A) discusses methods for probability proportional to size (PPS) sampling without replacement (WOR). Murthy's (1957), Brewer's (1963), and Durbin's (1967) schemes are described below. Simulations reported in this section examine their performances on the simulated school population with the two estimators of Section 2. Further discussion of methods for PPS sampling can be found in Särndal, Swensson, and Wretman (1992, section 3.6) and Brewer and Hanif (1983). When the sample size equals one, all methods select a unit with probability proportional to size.

3.1 Murthy's PPS WOR Scheme

In Murthy's (1957) PPS WOR procedure, successive units are drawn with probabilities proportional to size conditioned on the remaining units. It is easy to implement. Although the first order inclusion probability of individual units are not exactly proportional to size when $n > 1$, it is still a good approximation to an exact PPS WOR design when the population is large and the units are not too different in size. Let p_i be the probability of selecting unit i on the first draw. The probability of selecting unit j second, unconditional on the selection of the first unit, is $\sum_{i \neq j} p_i \frac{p_j}{1-p_i}$. As a function of p_j, the derivative of this formula with respect to p_j is $Q + 1 - \frac{1}{(1-p_j)^2}$, where $Q \equiv \sum_i \frac{p_i}{1-p_i}$ is a quantity that is a characteristic of the population and probabilities of selection. Since the derivative decreases as p_j increases, Murthy's method tends to oversample smaller PSUs in the second draw compared with methods that have unconditional probabilities of selection for the second unit exactly proportional to size. It could be expected that Murthy's method tends slightly to lose some information about the population units of larger size and would underestimate means and totals in circumstances of sampling two PSUs per stratum.

To estimate the total and the variance of the estimated total, one simple approach is to approximate the sampling procedure by a PPS *with* replacement design. The degree of bias of the approximation depends on the number of sample units and the variation in unit size. Murthy (1957) suggested an unbiased estimator and its variance and an unbiased variance estimator for his specific design (Cochran 1977, section 9A.9).He also proved his estimator has smaller variance than the ordered estimator suggested by Des Raj (1956).

3.2 Brewer's PPS WOR Scheme

Assume $p_i < 1/2$ for all $i = 1, \cdots, N$. When $n = 2$, Brewer (1963) draws the first unit i with revised probabilities $\frac{p_i(1-p_i)}{D(1-2p_i)}$, where $D = \sum_{i=1}^{N} \frac{p_i(1-p_i)}{1-2p_i}$, and the second unit j with probabilities $\frac{p_j}{1-p_i}$. The first order inclusion probability of unit i in the sample is $2p_i$. Variance and variance estimator formulas follow the standard formulas for the Horvitz-Thompson estimator.

In Brewer's PPS method, the probability of selection for the first sample unit is proportional to $\frac{p_i(1-p_i)}{1-2p_i}$ instead of proportional to size. The slope of the probability of selection at p_i as the derivative with respect to p_i: $\frac{(1-p_i)^2+p_i^2}{(1-2p_i)^2}$ is strictly increasing in p_i for $p_i < \frac{1}{2}$. This indicates that Brewer's PPS method tends to oversample larger PSUs in the first selection when $n = 2$.

3.3 Durbin's PPS WOR Scheme

Durbin's (1967) PPS approach draws the first unit i with probability p_i. Given that unit i is selected first, the probability that unit j is drawn second is $\frac{p_j}{2D}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right]$, where D is the same as in Brewer's method and $p_i < \frac{1}{2}$ is assumed for all i in the population. Even though different sampling procedures are used, both the first and second order inclusion probabilities of Durbin's PPS scheme are exactly the same as Brewer's, and so are the variances and variance estimators.

The unconditional probability of selection in Durbin's (1967) PPS method is p_i for the first draw and, for second draw, $\sum_{i \neq j} p_i \frac{p_j}{2D}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right] = \frac{(1+T)}{2D} p_j$, which is proportional to size because $T \equiv \sum_i \frac{p_i}{1-2p_i}$ is determined only by the population and does not depend on the sample units. Therefore, Durbin's (1967) PPS scheme has its own desirable property that the unconditional probability of drawing a certain unit at either the first or the second draw is exactly proportional to size.

3.4 Comparison of Methods

In the EP survey, the above three PPS methods were used to generate 1,000 independent samples from the simulated school population. Table 1 displays the standard deviation of the total estimators using Murthy's and Durbin's PPS methods for samples of 70 schools. The results using Brewer's method are the same as those using Durbin's, so they are not included in the table. Using Murthy's PPS method, the HT estimator has larger variance than Murthy's estimator. The estimator of variance assuming with replacement sampling (the PWR estimator) also has a significantly larger value than Murthy's estimator in medium districts due to large variation in enrollment size among these districts. The ratio estimator behaves consistently well in the whole state and all size levels. The sampling implementation (Murthy or Durbin) also affects the variation of the total estimators. Both Horvitz-Thompson and ratio estimators have smaller standard deviations using Durbin's PPS scheme compared to Murthy's, but the degree of reduction of variance is different for the two estimators. The standard deviation using the ratio estimator is a little smaller for Durbin's method in medium districts but essentially stays the same at the small level. For the HT estimator, the standard deviation is about 7% smaller using Durbin's method at the medium size level. Based on these simulations, Durbin's method and ratio estimation are recommended for situations such as Iowa's EP high school survey.

4. One PSU per Stratum Variance Estimation

4.1 One PSU per Stratum

For strata containing only large districts, variance estimation uses formulas for a stratified simple random sampling (SRS) without replacement design. Among strata involving medium or small size districts, some contain many districts and two districts are selected from each of them. In such a case, variance estimation will follow standard multi-stage sampling formulas for an invariant sample design and multi-phase sampling formulas for a non-invariant design. Among strata involving medium or small size districts with few districts, due to budget restrictions, only one district can be sampled. For these strata, there is not enough degrees of freedom to make direct variance estimation, and thus standard multi-stage or multi-phase sampling formulas can not be applied. This is a challenging problem. Standard approaches, such as the collapsed strata variance estimation (Cochran 1977, section 5A.12), consider estimation of variances for aggregations of strata, but not for individual strata. Variance estimators for individual strata and models of variances are proposed in the next two sections.

4.2 Collapsing Strata Synthetic Variance Estimation for Strata Variances

If no real average cluster difference exists within a stratum, then approximating the variance of the total estimate within a stratum by a formula for a simple random sample (SRS) from the population in the stratum would be fine. Similarly, using a difference estimator as described in Wolter (1984) for variance estimation in systematic sampling should work well if population units are randomly associated into clusters. However, if there is strong homogeneity within strata, then the SRS estimator will (significantly) underestimate the true variance. In most cases, variance estimators taking the clustering effect into account are needed.

The collapsed strata estimator (Cochran, 1977, section 5A.12) is a well-known estimator of variance estimation in one-per-stratum problem. The procedure collapses strata with one unit per stratum into groups and treats the strata in a group as independent samples from the combined stratum. In the EP survey, collapsing can be accomplished separately among the strata containing small and medium sized districts with one district in the sample. First arrange the strata in a non-increasing sequence based on total enrollment size. Then collapse strata into pairs or groups sequentially. The variance estimator of a group is given by (5A.56) in Cochran's (1977).

After getting the collapsed strata variance estimate for a group of strata, to produce variance estimates of individual strata within the group, the authors propose *proportional redistribution* of variance based on squared total enrollment size. The reason for this redistribution is that estimates of strata within the group are independent, and thus the variance of the estimate of the group equals the sum of variances of estimates of strata. If one assumes that strata in the same group are homogeneous in terms of within strata variation, then the ratio of variances of two strata within the group is proportional to the ratio of squared enrollment sizes, and thus the variance of each strata is a portion of the variance of the group with a weight proportional to squared enrollment size. In the EP survey, total enrollment size instead of enrollment size of twelfth grade students is suggested for redistribution because the former is longitudinally more stable. This redistribution, although not a standard practice, is important in this application for producing estimates of variance for AEAs and individual strata. The method can be referred to as *collapsing strata synthetic variance (CSSV) estimation.*

4.3 Modeling and Generalized Variance Functions

The standard design-based variance estimator is usually relatively unstable with small sample size. In such a circumstance, one may consider using alternative variance estimators based on generalized variance functions. From Valliant (1987), the derivation of generalized variance functions is motivated by its simplicity of computation, approximate unbiasedness, reasonable coverage in confidence intervals, and greater stability than individual direct estimates of variance.

In the EP Survey, even though the collapsing strata synthetic estimator works well for the whole state and size levels, it produces large estimates of variance for small domains. In order to produce better estimates of variance, one can consider using model-based estimators that could involve covariate variables to improve the efficiency and stability of the design-based variance estimators. Generalized variance functions traditionally model the relationship between relative variances and expectations of the total estimators for individual strata and predict the strata variances from the estimated totals through the estimated functions.

Considering that the simulated population of study was generated from a product of Poisson distributions, it seems appropriate to use the traditional generalized variance functions suggested in Valliant (1987). The superpopulation model is

$$V^2 = \alpha + \frac{\beta}{T} \tag{1}$$

where V^2 is the relative variance or rel-variance of the total estimator. The corresponding sample model is then given by $\hat{V}^2 = m\left(\hat{t}\right) + \epsilon$, where $m\left(\hat{t}\right) = \alpha + \frac{\beta}{\hat{t}}$ is the model mean function and ϵ is the error with mean 0 and conditional variance proportional to the model mean. That is $E[\epsilon|\hat{t}] = 0$ and $Var[\epsilon|\hat{t}] = m\left(\hat{t}\right)\sigma^2$, where σ^2 is an unknown constant. One disadvantage of this model is that it could produce negative predictions of variance. Wolter (1985, chapter 5) suggested adding restrictions to the generalized variance function to assure nonnegative predictions generated from it.

Assuming that the districts within the same stratum have the same average number of employment preparation courses taken by an individual student, it should be appropriate to assume that $\hat{t}_{st} = \sum_{i \in st} \frac{\hat{t}_i}{\hat{N}_i} N_i$ exists, where the subscript st means stratum. Let $N_{st} = \sum_{i \in st} N_i$. By imposing the restriction that $V^2_{N_{st}} = 0$, which is equivalent to $\alpha = -\frac{\beta}{N_{st}}$, a restricted generalized variance function can be given by

$$\hat{V}_t^2 = \beta \left(\frac{1}{\hat{t}} - \frac{1}{N} \right) \tag{2}$$

To estimate the unknown parameter, one can use iteratively reweighted least squares estimation or maximum likelihood estimation algorithms such as the Newton-Raphson or Nelder-Mead algorithms.

4.4 Comparisons of Methods

The first comparison of the collapsing strata synthetic variance (CSSV) estimation and generalized variance function (GVF) methods is based on the coefficients of variation (cve's) of the variance estimates using these two methods for the case of 60 sample schools. Results are shown in Table 2. Under all three PPS sampling procedures, the cve's using the GVF method at medium and small size levels are consistently smaller than the cve's using the CSSV technique, which indicates the GVF method works better for variance estimation in these groups. For variance estimation in individual strata and AEAs, the GVF method still performs well. The five strata with one PSU per stratum in samples from the simulated population occur in AEAs 4, 12, 14, 15, and 16 (AEAs are not numbered consecutively). As mentioned before, results are based on simulated data and should not be interpreted as representing Iowa's schools. In simulated AEAs 4 and 12, the GVF method produces standard errors between the true standard deviations and the standard errors produced by CSSV estimation. In AEA 15, the GVF method estimates the standard error almost the same as the real standard deviation, which is much better than the standard error obtained by CSSV estimation. In AEAs 14 and 16, even though both methods overestimate the standard deviations and the GVF even overestimates a little bit more, basically they produce very close variance estimates in these two AEAs.

The second comparison concerns coverage of confidence intervals produced using these two variance estimation methods. Table 3 shows the coverage rate of the confidence intervals of HT and ratio estimators computed by both methods. In all five strata with one PSU per stratum, the confidence intervals computed by the GVF method have significantly higher coverage rates than those using CSSV estimation for both estimators. The improvement by using the GVF method is more prominent for the ratio estimator than for the HT estimator.

Table 4 displays 2.5%, 25%, 50%, 75%, and 97.5% empirical percentiles of the width of confidence intervals using ratio estimators over 1,000 simulations. In AEAs 4 and 9, the GVF method produces consistently narrower confidence intervals than CSSV estimation. In AEAs 14 and 16, even though the medians of the width of confidence intervals by the GVF method are bigger, the third quantiles are smaller. The empirical inter-quantile ranges are all smaller using the GVF method in those five strata. This comparison indicates that the variance estimates produced by the GVF method are more stable than the variance estimates given by CSSV estimation. All the results for strata containing small districts are substantially the same.

In conclusion, numerical results show that the GVF method produces better variance estimates in terms of consistently smaller coefficients of variation of variance estimates, a higher coverage rate for confidence intervals, and more stable performance for a group of estimators. It is clear, however, that results vary more by AEA than by method of estimation. In Section 7, the use of diagnostic plots and comparison to direct estimators of variance to check on the performance of GVF is discussed.

5. An Invariant and a Non-Invariant Design

The standard stratified multi-stage design is invariant that the same subsample design for a PSU is used every time the PSU is included in a first stage selection. If up to 50 students from the four groups in selected schools are sampled independently by simple random sampling with no redistribution in the case of small schools (see Section 2.2), then the design is a stratified multi-stage design and the standard formulas for estimators of means and totals and estimators of variances are applicable. The non-invariant design is easy to operate. The realized overall number of students in the sample, however, is not fixed and would be less than the specified maximum. This will increase the variance contribution in the terminal stage of sampling especially for large schools.

On the other hand, if excess sample is redistributed across districts and schools, the inclusion of certain districts or schools in the sample affects student selection probabilities in other clusters. In the terminology of Särndal, Swensson, and Wretman (1992, page 134), the design is not invariant. The non-invariant multi-stage design can be thought of as a multi-phase sample design. In multi-phase sample designs, the subsample design depends on the entire selected first phase sample. In the EP survey, if further resource redistribution is planned, then the design is a stratified multi-phase design and the standard formulas (see Särndal, Swensson, and Wretman, 1992, chapter 9) for estimating totals and variances of estimators for a multi-phase design should be employed. The estimators of both totals and variances are unbiased. Although the design is more complicated to implement and estimation formulas are more involved, the non-invariant multi-phase design makes use of all available resources and tends to decrease the terminal phase variation due to a larger sample in that phase.

Table 5 shows that the non-invariant design results in smaller cves using either the HT or the ratio estimator and estimating variance by either the GVF or the CSSV estimation method (Section 3). As in Table 3, higher coverage rates were produced by the GVF method than by the CSSV estimation method in the non-invariant design for all five strata with one PSU per stratum. Thus, if it possible to implement, the non-invariant design has some definite advantages.

6. A Cost Evaluation

Iowa's State Board of Education decided on 60 schools instead of 70 as was recommended, because it was operationally feasible in terms of budget, staff, and coordination with schools. Given that the 70 school design has lower variances at most levels of aggregation beyond the schools and districts (assuming fewer students per sampled school), one can investigate the implied costs of adding schools to the sample and hope to quantify the implied trade-off between cost and variance.

Costs in this school board survey generally come from general administration, data processing, sampling districts, and sampling students. Since additional schools will be from the small and medium districts, each additional district adds one school.

If order to examine this issue, it is assumed that all sampled students cost the same in terms of data collection and processing. Call the student cost one cost unit. Also assume that all sampled schools cost the same. Suppose that an additional school "costs" a cost units each. That is, in order to spend resources to code a new course catalog and to interact with school administrators, a fewer students across the whole study have transcripts reviewed. Considering that the between schools variation plays a significant role in the variance of total estimates, it is of interest to study with a fixed overall budget how much one could benefit from increasing the number of sample schools while correspondingly reducing the number of sample students overall and per school.

Based on the numerical results in Table 6, if one samples 65 schools, the variance estimates are reduced about ten percent on average. Larger costs, at least in the range of costs and numbers of districts considered, do not seem to make much of a difference; the number of sample schools is much more influential. After 65 schools, variance estimates decrease about five percent for every five schools added. Besides the decrease in variance, the interquartile range of variance reductions over 1,000 simulations also decreases as the number of schools increases to 75 or 80 schools. This means that there is less variance in variance reductions, which means that variances are estimated with more stability. Therefore, it seems that even accounting for higher costs, adding more schools to the survey would produce better estimates of total and mean and better estimates of variance.

7. Summary And Discussion

A survey for which records on transcripts of Iowa public high school students served as the source of data was used to motivate five examinations of research questions. First, using a ratio estimator improves the precision by using auxiliary variables a lot over Horvitz-Thompson (HT) estimation. The outcome in the survey for Iowa's State Board of Education is the number of employment preparation (EP) courses and the auxiliary variable is the number of students in a grade level. In the EP survey application, the ratio estimator works better than the HT estimator under all three versions of PPS sampling in terms of smaller variance and MSE.

Second, compared to Murthy's (1957) PPS sampling method that tends to oversample small PSUs in the second draw, Durbin's (1967) method has the desirable property that the unconditional probability of drawing a certain unit at either the first draw or the second draw is exactly proportional to size, and hence produces consistently accurate estimates in situations such as the EP survey. Simulations confirmed these advantages.

Third, in the situation that a variance estimate is needed for stratum in which there really are not enough degrees of freedom to make a direct estimate, using a generalized variance function and choosing a reasonable estimate based on some model diagnostics might be possible. The traditional collapsing strata estimator is widely applied for estimating the variance of a total for a group of strata. In some other cases, there are two or more PSUs per strata and an unbiased estimate of variance is possible. A scatter plot of direct estimates and GVF estimates can sometimes provide some indication of the quality of the variance modeling. Taking advantage of auxiliary information to improve the efficiency of the estimator is encouraged. The numerical results show that the GVF method produces better variance estimates in terms of consistently smaller variation, a higher coverage rate of confidence intervals, and a more stable performance for a group of estimators than a new collapsed strata synthetic variance estimation method.

Other methods for variance estimation in one-per-stratum designs will be considered in future work and Iowa education surveys. Shapiro, Singh, and Bateman (1980) use a without replacement estimator under Durbin's (1967) PPS scheme. Hansen, Hurwitz and Madow (1953) have another form of the collapsed strata estimator. Hartley, Rao, and Kiefer (1969) use a regression estimator with auxiliary variables. Bayesian approaches, such as Singh and Sedransk's (1988), are worth considering as well.

Fourth, in the situation that extra resources can be redistributed across sample clusters, a non-invariant stratified multi-phase design can be applied. Compared with an invariant stratified multi-stage design, the non-invariant design makes full use of available resources and reduces the bias and variance of the estimator at a cost of more complicated implementation. Variance estimation using GVFs might achieve even better estimation in a non-invariant design due to the improved direct estimators.

Fifth, the numerical results indicate that the between cluster variation might be more influential than the within cluster variation. To reduce the overall variation, it would be a good idea to increase the number of clusters rather than sampling a greater number of units within clusters. With fixed overall cost, the impact on variance estimates of increasing the number of sample schools while assuming the cost of sampling a district to be in a reasonable interval was studied. The number of sample schools shows a definite effect on the variance estimate. The cost of sampling a district, in the range of alternatives considered, only affects the variance estimate on a small scale. If practically possible, it would be strongly advisable to increase the number of schools in the sample.

Future study related to the EP surveys and other surveys for Iowa's Board of Education will include small area estimation (Rao 2003), which reduces the variance of traditional design-based estimators by "borrowing strength" cross subsets of the districts, and survey regression calibration methods (Deville and Särndal 1992). Calibration equations will be used to produce a model-assisted estimator and corresponding calibration variance estimator by making use of information on auxiliary variables. Better administrative data for covariates might be available in the future as well. Additionally, methods for improving the design to produce good estimators of characteristics in multiple subpopulations (size levels, AEAs, etc.) can be studied. The actual data analysis and cost evaluations will be made when the actual data are completely reviewed and made available for the purpose.

Acknowledgments

The authors wish to thank Iowa's State Board of Education for partial support of this work and bringing this survey to Iowa State University's Center for Survey Statistics and

Methodology. In particular, the authors wish to acknowledgment the involvement of Barb Guy, Pat Sitlington, and Alan Frank from the offices of the State Board and Dianne Anderson and Jan Larson from Survey Research Services in CSSM at ISU. **The results presented in this paper are not actual results from the survey and should not be interpreted as characterizing schools in the State of Iowa. Any opinions, findings, and conclusions, or recommendations expressed on this material are those of the authors and do not necessarily reflect the views of Iowa's State Board of Education.**

References

Bradby, D., Levesque, K.A., Henke, R.R., and Malitz, G. (1995). *A pilot standard national course classification system for secondary education.* U.S. Department of Education, Office of Educational Research and Improvement, NCES 95-480. National Center for Educational Statistics: Washington, D.C.

Brewer, K.W.R. (1963). A model of systematic sampling with unequal probabilities. *Australian Journal of Statistics*, Vol. 5, 5-13.

Brewer, K.R.W., and Hanif, M. (1983). *Sampling With Unequal Probabilities.* New York: Springer-Verlag.

Cochran, W.G. (1977). *Sampling Techniques*, third edition. New York: J. Wiley.

Des Raj. (1956). Some estimators in sampling with varying probabilities without replacement. *Journal of the American Statistical Association, 51, 269-284.*

Deville, J.C., and Särndal, C.E. (1992). Calibration estimators in survey sampling, *Journal of the American Statistical Association*, 87, 376-382.

Durbin, J. (1967). Design of multi-stage surveys for the estimation of sampling errors, *Applied Statistics*, 16, 152-164.

Hansen, M.H., Hurvitz W.N., and Madow W.G. (1953). *Sampling survey methods and theory*, Vol I, 399-401, and Vol II, 218-222. New York: Wiley.

Hartley, H.O., Rao, J.N.K., and Kiefer, G. (1969). Variance estimation with one unit per stratum, *Journal of the American Statistical Association*, 64, 841-851.

Hewitt, K., and Larsen, M.D (2005). Designing a sampling method for a survey of Iowa high school seniors. *2005 Proceedings of the America Statistical Association*, Survey Research Methods Section [CD-ROM]. Alexandria, VA: American Statistical Association.

Hewitt, K.A., and Larsen, M.D. (2006). Design of a Survey of high School Students in Iowa: Studying Design Choices Using Simulation. *Stats* magazine, accepted.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, 47, 663-685.

Lu, L. (2005). Variance Estimation in a Iowa's State Board of Education Survey. *MS creative component paper*, Iowa State University, Department of Statistics, Unpublished.

Murthy, M.N. (1957). Ordered and unordered estimators in sampling without replacement. Sankhyā, 18, 379-390.

Rao, J.N.K. (2003). *Small Area Estimation.* New Jersey: Wiley.

Särndal, C.E., Swensson, B.S., and Wretman, J. (1992). *Model Assisted Survey Sampling.* New York: Springer.

Shapiro, G.M., Singh, R.P., and Bateman, D. (1980). Empirical research involving an alternative variance estimator to the collapsed stratum variance estimator, *Proceedings of the American Statistical Association, Survey Research Methods Section*, 793-798.

Singh, B., and Sedransk, J. (1988). Variance estimation in stratified sampling when strata sample sizes are small, Sankhyā, Series B. 50, 382-393.

Valliant, R. (1987). Generalized variance functions in stratified two-stage sampling, *Journal of the American Statistical Association*, 82, 499-508.

Wolter, K.M. (1984). An investigation of some estimators of variance for systematic sampling, *Journal of the American Statistical Association*, 79, 781-790.

Wolter, K.M. (1985). *Introduction to Variance Estimation.* Springer Series in Statistics.

Table 1: The Standard Deviation of Total Estimators Using Murthy's and Durbin's Sampling Methods for Samples of 70 Schools for Three Aggregations.

	Sampling Method and Estimator					
Aggre-	Murthy				Durbin	
gation	Murthy	PWR	HT	Ratio	HT	Ratio
State	5157	5390	5461	4805	5112	4772
Medium	4998	5230	5310	4678	4952	4644
Small	1155	1193	1160	961	1154	961

Table 2: The cve of Variance Estimators Using Collapsing Strata or GVF Methods for the Ratio Estimator Averaged over 1,000 Simulations for Three Aggregations.

	Collapsing Strata Synthetic Variance Estimation			Restricted Generalized Variance Function		
PPS Scheme:	M	B	D	M	B	D
State	3.35	3.36	3.34	3.25	3.27	3.25
Medium	6.05	6.06	6.05	5.89	5.91	5.90
Small	5.20	5.22	5.18	4.91	4.96	4.92

M = Murthy; B = Brewer; D = Durbin.

Table 3: Number of Confidence Intervals Obtained by Using Collapsing Strata Synthetic Variance (CSSV) Estimation and Restricted GVF Estimation out of 1,000 Covering Totals for Strata with Medium Size Districts. Two Sampling Designs and Two Estimators are Used. Data are Simulated.

Design	Esti-mator		Area Education Agencies 4	12	14	15	16
Non-Invariant	Ratio	CSSV	983	883	968	751	838
		GVF	996	939	1000	834	991
	HT	CSSV	983	892	941	691	916
		GVF	985	942	1000	876	1000
Invariant	Ratio	CSSV	969	882	962	735	823
		GVF	982	927	1000	829	975
	HT	CSSV	972	895	942	744	902
		GVF	981	947	999	868	997

Table 4: Empirical Percentiles of the Width of Confidence Intervals Obtained by Collapsing Strata Synthetic Variance (CSSV) and Restricted GVF Estimation over 1,000 Simulations for Strata with Medium Size Districts Using the Non-Invariant Design with the Ratio Estimator.

	Area Education Agency 4	12	14	15	16
CSSV					
2.5%	423	561	333	113	120
Q1	1485	1969	1169	1166	1243
Median	1989	2638	1565	2343	2498
Q3	3072	4073	2417	4035	4302
97.5%	4165	5523	3277	5953	6347
GVF					
2.5%	812	1191	1048	1446	1856
Q1	1240	1949	1643	2605	2804
Median	1565	2487	1986	3608	3475
Q3	2018	3072	2334	4693	4126
97.5%	2792	4473	3164	6510	5450

Table 5: CVEs of Horvitz-Thompson and Ratio Estimators Under Invariant and Non-Invariant Designs Using Collapsing Strata Synthetic Variance (CSSV) and Restricted GVF Estimation for 60 Sample Schools Using Durbin's PPS Method for Three Aggregations.

Esti-mator	Invariant Design CSSV	GVF	Non-Invariant Design CSSV	GVF
HT				
State	3.60	3.66	3.55	3.48
Medium	6.46	6.54	6.38	6.26
Small	6.29	6.67	6.21	6.01
Ratio				
State	3.42	3.43	3.34	3.25
Medium	6.19	6.19	6.05	5.90
Small	5.27	5.39	5.18	4.92

Table 6: Empirical Percentiles over 1,000 Simulations of the Relative Decrease in Variance Estimates Due to Adding More Schools to the Sample with Fixed Total Cost Per School. Decrease is Relative to a Sample with 60 Schools. Cost is the Reduction in the Total Number of Students in the Sample Per Additional School. IQR is the Inter-Quantile Range of Variance Reductions.

	Number of Schools	Cost Factor Per School 200	150	100	50	0
Students	65	11000	11250	11500	11750	12000
Median		0.118	0.108	0.112	0.105	0.101
IQR		0.315	0.326	0.336	0.344	0.342
Students	70	10000	10500	11000	11500	12000
Median		0.169	0.164	0.160	0.158	0.141
IQR		0.338	0.323	0.308	0.331	0.317
Students	75	9000	9750	10500	11250	12000
Median		0.202	0.212	0.193	0.201	0.196
IQR		0.230	0.213	0.220	0.221	0.236
Students	80	8000	9000	10000	11000	12000
Median		0.238	0.242	0.253	0.252	0.255
IQR		0.192	0.196	0.188	0.189	0.176