



Weighting Omissions and Best Practices When Using Large-Scale Data in Educational Research

Debbie L. Hahs-Vaughn
Assistant Professor, Educational Research
University of Central Florida

Abstract

Federal agencies conduct large, national studies that provide abundant data easily accessible to researchers. While these datasets provide access to rich data for analyzing a multitude of topics, certain skills are necessary for appropriate use. Specifically, over sampling and multistage cluster sampling must be accommodated in the analyses to ensure the estimates are accurate. The purpose of this study is to examine recent research that uses national datasets to determine the extent to which researchers use the weight and design effect compensation and to highlight best practices in reporting analyses using large-scale data. Research articles using large-scale datasets from five years of *Research in Higher Education*, *Journal of Higher Education*, and *Review of Higher Education* were reviewed to determine the data's appropriate use by reviewing the use of weights and design effects. These journals were selected because these are official journals of relevant professional organizations and/or are considered to be leading scholarly journals on topics related to higher education. Best practices in using national datasets for authors and journal editors are discussed.

The National Science Foundation (NSF) and the National Center for Education Statistics (NCES), among other federal agencies, conduct large, national studies that provide abundant data easily accessible to researchers (e.g., NCES, n.d.; NSF, n.d.). While the survey data's free accessibility is a benefit in itself, the data are becoming more highly sought after because of the comprehensiveness (e.g., Roberts, 2000), incorporating a large and representative population

sample unfeasible to gather by an independent researcher or even a small team of researchers. In addition, support to learn how to use the datasets is available through a national summer database training institute (AIR, n.d.a). Professional organizations, including the Association for Institutional Research (AIR) and the American Educational Research Association (AERA), offer grants and fellowships for projects that employ large-scale data (AERA, n.d.a; AIR, n.d.b).

Superficially, a researcher might examine the datasets and find that there is nationally representative data on an almost unlimited variety of individuals that are free and accessible and all one must do is generate a research topic and plug the data into a statistical software package. This superficial perception of the data is a reason that database training institutes are offered each summer. While the national datasets have an abundance of data for analyzing a multitude of topics, certain statistical skills are necessary to appropriately analyze that data. As stated by the AIR (n.d.a), "the uses of these resources have, for a variety of reasons, lagged behind national-level data collection and organization."

Some of the statistical methodological characteristics associated with using the datasets may be the reason use has lagged. This study examines recent research that utilizes one or more national datasets to determine the extent to which the datasets are analyzed appropriately and to highlight best practices in reporting analyses using large-scale data. By nature, these datasets are complex in design and methodological measures must be taken when conducting statistical analyses to ensure that two

features are accommodated: non-simple random sampling or cluster sampling and over-sampling of groups.

Non-Simple Random Sampling (Clustering)

Most complex surveys employ sampling designs that are multistage, cluster, or stratified—or a combination of these sampling techniques—because a simple random sample is not feasible. For example, many national datasets involve multistage probability sampling where geographic regions are first selected, then institutions, and finally students (Pratt, et al., 1996). An example of this multistage probability sample design was used to collect data for the Early Childhood Longitudinal Study Kindergarten Class of 1998-99 (ECLS-K). In the ECLS-K, geographic areas (specifically counties) were the primary sampling unit (PSU), followed by schools within counties as the second-stage sampling unit, and then students within schools as the last sampling unit (U.S. Department of Education, 2004).

Understanding the challenges associated with clustering is critical to correctly analyzing the data (Hahs-Vaughn, 2005; Stapleton, 2002; Thomas and Heck, 2001) as statistical procedures have assumptions underlying the methods that must be addressed to ensure accurate parameter estimates (e.g. independence) (Hahs-Vaughn.; Kaplan and Ferguson; Stapleton; Thomas and Heck). Conventional formulas for calculating standard errors assume independence (Skinner, 1989). When a simple random sample is not used, homogeneities exist between and amongst the sampling units that negate the assumption of independence (Kish and Frankel, 1974) and that can lead to underestimated standard errors (Muthen and Satorra, 1995; Skinner). Using the ECLS-K data as an example, it is likely that students selected within the same school share similarities (i.e., are more homogenous) than had a simple random sample been drawn of all kindergarten students. Thus, standard errors will be smaller when the sampling design is not accommodated in statistical calculations. Using data from the ECLS-K and the National Study of Postsecondary Faculty (NSOPF:93), showed this to be the case (Hahs-Vaughn).

Over-Sampling

Many times, some specific subset of the population is over-sampled to ensure good representation of the group(s) and thus more accurate parameter estimates (Kaplan and Ferguson; Stapleton; Thomas and Heck). Using the ECLS-K as an example, Asian and Pacific Islanders were over sampled in the base year (U.S. Department of Education). This creates unequal selection probability because of sampling the population at different rates (Stapleton). If left unattended, the parameter estimates and standard errors may be overly influenced by the over-sampled units (Kaplan and Ferguson). Survey

weights can be applied to data when generating statistical analysis to deal with over-sampled groups and are a necessary condition to prevent bias created by the sample design (Potthoff, Woodbury, and Manton, 1992). A weight is simply the inverse of the probability of selection (Kish, 1965). This weight is often also adjusted for nonresponse and poststratification (Kalton, 1989; Korn and Graubard, 1995). Methodology reports that accompany national datasets include information on calculation of weights, available weights within the dataset, and information to discern which weight is appropriate given the type of analyses (e.g., cross sectional or longitudinal). For example, the ECLS-K methodology report indicates that weight C5CPTW0 should be used when conducting analysis that involves data from the child, parent and teacher. Specifically, children who have spring third grade assessment data, whose parent completed the spring third grade FSQ part of the interview, and whose teacher completed part B of the teacher questionnaire (U.S. Department of Education).

Failure to Accommodate Complex Sample Design in Statistical Analysis

If cluster sampling and over-sampling are ignored in the statistical analysis, the results can be biased. This has been shown using various statistical methods including regression techniques (e.g., DuMouchel and Duncan, 1983; Hahs-Vaughn; Korn and Graubard; Lee, Forthofer, and Lorimor, 1989; Skinner, Holt, and Smith, 1989) and structural equation modeling (Hahs, 2003; Kaplan and Ferguson; Stapleton). The bias can include biased parameter point estimates including smaller standard errors and larger test statistics (Hahs-Vaughn; Stapleton) and poor performance of test statistics and confidence intervals (Pfeffermann, 1993) including narrower confidence intervals (Hahs-Vaughn). As an example of the influence of over-sampled groups, Hahs-Vaughn (2005) analyzed data from the NSOPF: 93. Proportions of faculty by sex were reviewed by computing the mean of sex (where sex was a dummy coded variable with male = 0 and female = 1). When reviewing weighted to unweighted proportions, she found that the proportion of women faculty was overestimated when weights were not applied (i.e., the mean for sex was smaller reflecting fewer males). This was because of the oversampling of women faculty which decreased the proportion of women in the weighted distribution ($M = .61$) as compared to the unweighted distribution ($M = .57$).

Ignoring the implications of the complex sampling design (including over-sampling and cluster sampling) leads to questionable generalizations (Hahs, 2003; Thomas and Heck). While intended to be nationally representative of some specific population, failing to employ proper techniques to compensate for the complex sampling design creates analyses that reflect only the

sample with results not able to be inferred to the intended population (Hahs; Thomas and Heck). Various strategies are proposed to deal with complex samples (e.g., Peng, 2000; Thomas and Heck) and an overview is presented here.

Weights

Previous research identified procedures for accommodating the unequal probability of selection when using national datasets. One of the basic elements needed for a complete analysis of survey data is the use of weighting (Lee, Forthofer, and Lorimer, 1989). In the most basic case, a sample weight is the inverse of the probability that the observation will be included in the sample (Kaplan and Ferguson). Estimates based on the unweighted sample will be biased in favor of persons/groups that were over sampled (Thomas and Heck). Using weights to adjust for over sampling is the easiest way to incorporate unequal probability of selection (Stapleton), although the actual role of sampling weights is a subject of controversy (Pfeffermann, 1993). With inferential statistics, there are a variety of opinions on what role, if any, sampling weights should play—"from modelers who view the weights as largely irrelevant to survey statisticians who incorporate the weights into every analysis" (p. 317). More recent recommendations encourage the use of weights with any test of inference (e.g., Hahs-Vaughn; Thomas and Heck).

Sampling weights play a vital role in modeling data by testing and protecting against sampling designs that could cause selection bias (in other words, allowing the researcher to accommodate unequal selection probabilities in the statistical analysis) and by protecting against misspecification of the model (Pfeffermann, 1993). The frequency of using weights to adjust the sampling design in research has yet to be determined (Hahs, 2003). It was recommended that journals "require authors to address how weights [have been] employed in the study to ensure that the parameter estimates, and thus the results of the study, are accurate given the population to which the researcher wants to generalize" (Hahs, 2003, p. 266). Similar to reporting effect size (e.g., McLean and Kaufman, 1998), failure to report the use or omission of weights when using a national dataset prevents the reader from critically evaluating the research.

Design Effects

Another issue to consider is the clustered sampling design. If adjustments for clustering are not made, standard errors may be seriously underestimated (Gregorich, 1997). Multilevel approaches, sometimes called model-based approaches (Kalton, 1983) or disaggregated approaches (Muthen and Satorra), negate the need to act on the clustered design because these designs take clustering into account when producing

estimates (Thomas and Heck). Examples of multilevel approaches include hierarchical linear modeling and multilevel structural equation modeling (Hahs-Vaughn). Use of multilevel models to account for multistage sampling has received much attention (e.g. Hox and Kreft, 1994; Kaplan and Elliott, 1997; Muthen, 1994). Multilevel models allow the researcher to define, for example, student-level equations that will predict student-level outcomes, while at the same time providing student-level indicators as a function of between-class or between-school indicators (Kaplan and Elliott). Mathematically, the variance of the dependent variable is partitioned into within- and between-group components. The variances at each level are then explained by including predictor variables hierarchically; (Heck and Mahoe, 2004) however, over-sampling must still be dealt with (Hahs-Vaughn).

Although it has been argued that the appropriate statistical method to analyze multistage samples is a multilevel model, multilevel modeling is not always appropriate for analyzing complex surveys (Kaplan and Elliott, 1997). For example, not all secondary datasets have the appropriate second-level variables required to conduct a multilevel model (i.e., that allow the researcher to connect the level one variables, such as student level variables, with level two variables, such as institution level variables) and not all researchers may ask questions that require the use of a multilevel model (Kaplan and Elliott). When a multilevel model is not appropriate, a single-level model can be employed. Single-level analyses must accommodate multistage cluster sampling and the resulting homogeneity of clusters that will lead to biased standard errors (Stapleton) as well as accommodate over-sampling. Design effect (DEFF) is the ratio of the estimated variance using the complex sample design to the exact variance that would have been obtained from a simple random sample drawn of the same size (U. S. Department of Education,) and can be used in single-level analysis to accommodate clustering:

$$DEFF = \frac{\text{Variance}_{\text{Complex Sample}}}{\text{Variance}_{\text{Simple Random Sample}}}$$

How to Accommodate Complex Sample Design in Statistical Analysis

Accommodating Over-Sampled Groups

As stated previously, weights are one element needed to correctly analyze survey data (Lee, Forthofer, and Lorimer, 1989) and available weights for a given survey can be obtained by reviewing the methodological reports. Once the research determines the correct weights, these must be appropriately applied to the data. While a comprehensive discussion of the types of weights can be

found in Hahs-Vaughn (2005), a concise overview is presented here. The raw weight is the weight included in the dataset which sums to the population size (West and Rathburn, 2004). Parameter estimates computed using raw weights reflect the population rather than the sample size and thus are not appropriate to directly apply to analyses. Relative or normalized weights are the raw weights divided by the mean of the raw weights (Peng, 2000; Thomas and Heck) which sum to the sample size (Kaplan and Ferguson; Pfeffermann, Skinner, Holmes, Goldstein, and Rasbash, 1998). Parameter estimates produced using relative weights would be accurate had the sample been drawn randomly (i.e., a simple random sample) (Thomas and Heck). Given the multistage probability sample, this is not the appropriate weight to directly apply to analyses. Design effect adjusted weights accommodate both the cluster sampling and over-sampling and are computed as the normalized weight divided by DEFF (Hahs-Vaughn). This new design effect adjusted weight is then applied to the data.

Accommodating Clustering

Thomas and Heck (2001) have suggested alternatives to correct for clustered samples in single-level analyses including: a) utilizing a special software program to analyze the data, b) adjusting the analyses with a design effect either by adjusting the standard errors upward as a function of the design effect or adjusting the relative weight downward as a function of the design effect, or c) employing a more conservative critical value (i.e., by decreasing the alpha level). While using a specialized software program (e.g., SUDAAN, WesVar) is the most highly recommended strategy, the software is often cost prohibitive and/or difficult to use and thus this strategy is the least often used (Thomas and Heck). The exception to this is a relatively new software program, AM, that can be downloaded free of charge from <http://am.air.org> (see Hahs-Vaughn, for a guide to using AM). However AM is limited in the statistical procedures available (Hahs-Vaughn). Adjusting analyses with DEFFs are the next best option.

Design effect corrections are used to make adjustments for the variances produced by the analysis (Peng, 2000). Adjustments to the standard errors can be made by multiplying the standard errors in the model by the square root of the design effect (DEFT; see formula below) and then using the new standard errors to calculate new *t* test statistics.

$$DEFT = \sqrt{DEFF} = \frac{SE_{Complex\ Sample}}{SE_{Simple\ Random\ Sample}}$$

Design effects are easily obtainable by reviewing methodological reports of the surveys or, in the case of the NCES surveys, using the Data Analysis System (DAS) (Thomas and Heck). While the DAS is a powerful interface that allows researchers to create tables and correlation matrices, not all NCES datasets are accessible and no raw data is available through DAS thus many complex questions cannot be answered.

The information presented provides a brief overview into the complexities associated with analyzing data from complex samples. For a comprehensive guide to applying weights and design effects using SPSS or AM software (including determining how to select an appropriate weight, offering steps for computing and applying weights and design effects, and providing recommendations for strategies to accommodate both over-sampled groups and cluster sampling), readers are encouraged to refer to Hahs-Vaughn (2005). To determine the extent to which the use of weights was discussed and to highlight best practices in reporting the use of weights in journal articles, published research studies that use one or more national datasets will now be examined.

Method

The data for this study were obtained from *Research in Higher Education*, *Journal of Higher Education*, and *Review of Higher Education*. All research articles published between the years 1999-2003 were examined. *Research in Higher Education* was selected as it is the official journal of AIR. As discussed previously, AIR is an advocate of national datasets, offers the National Summer Data Policy Institute, as well as grants and fellowships for the use of national datasets. This journal regularly publishes research articles within the field of higher education that utilize national datasets. *Review of Higher Education* was selected because it is the official journal published through Association for the Study of Higher Education (ASHE), also a leading organization for the study of higher education. *Journal of Higher Education* was selected because it is considered a leading scholarly journal on topics related to higher education (*The Journal of Higher Education*, n.d.)

Excluding memoriams, acknowledgements, editorials, presidential addresses, book reviews, similar non-research publications, and one anniversary issue of *Journal of Higher Education* (September/October 1999) that was devoted to short essays, a total of 378 articles were reviewed. Only articles that used one of the NCES or NSF *student* or *faculty* databases in the methodology were included in the sample (i.e., the Integrated Postsecondary Education Data System was excluded). A breakdown of the datasets utilized in the published studies¹ and the use and/or omission of weights is provided in Table 1.

**Weight Reporting in Research in Higher Education, Journal of Higher Education,
and Review of Higher Education, 1999-2003**

Dataset	Weight Utilization (Frequency)	
	Reported	Unreported
Baccalaureate and Beyond (B&B)	3	1
Beginning Postsecondary Students Longitudinal Study (BPS)	2	1
High School and Beyond (H&B)	2	2
National Education Longitudinal Study (NELS)	1	0
National Longitudinal Study of High School Class (NLS)	0	1
National Postsecondary Student Aid Study (NPSAS)	1	0
National Study of Postsecondary Faculty (NSOPF)	11	3
Survey of Doctorate Recipients (SDR)	4	1
Total	24	9

Results

Weights

Of the 33 articles reviewed, nine (27%) did not discuss weights. While nine is not a relatively large number, it constitutes more than one-fourth of the articles reviewed. While the reader cannot know for certain whether weights were used, erring on the side of caution leads the reader to assume weights were not used. Twenty-four articles recognized that the national datasets have weights and either discussed how or why weights were used or not used in the analyses. Approximately 30% ($n = 10$) of the articles reviewed used weights and provided sufficient detail for the reader to discern appropriate application. Two articles (6%) did not use weights but noted such and provided a justification for not weighting. Twelve of the 33 (36%) articles discussed the use of weights at a minimal level. Operational definitions of "sufficient detail" and "minimal" are provided in the discussion section.

Design Effects

In addition to weights, design effects must also be considered in studies that do not employ a multilevel approach if a specialized software program is not used to accommodate the sampling design. Only four studies

noted whether and how design effects were employed. One of the four studies used a blanket design effect that was selected because a previous researcher had used it in various models and the author saw the number as conservative.

Discussion

Regarding minimal discussion of weighting, many of the studies reported that the results were weighted using normalized weights so that the sample size was maintained. A minimal discussion of the use of weights does not provide sufficient information for the reader to evaluate how effective the weight was used and to discern which weight was used. Most of the datasets offer multiple weights, and while the methodological reports provide guidelines on which weight to use given the type of analysis (e.g., cross sectional vs. longitudinal), it is up to the researcher to select the appropriate weight. A more informative approach in presenting research that uses a national dataset is presented by Cabrera and La Nasa (2001) who devoted an entire paragraph under the heading, *weights employed*, to a discussion on how they utilized survey weights. The authors specify which weight was used (e.g., "panel weight (F2PNLWT)," p. 126), how

it was adjusted to minimize the effect on the standard error because of the large sample size, and a comparison of how the sample size changed based on the adjustment (Cabrera and La Nasa). Failing to provide this type of detail is analogous to reporting score reliability and validity without reporting the type(s) of reliability and validity evidence (including corresponding coefficients). In essence, the reader is offered the information at face value without being able to put it into context (i.e., “the author says they used weights, so I have to assume they used the correct weights and applied these correctly”). A better practice, therefore, is to report the weight information in detail and allow the reader to critique its use. This also provides the level of detail needed for future replication.

Also interesting to note is the recognition and omission of the use of weights in two studies. In one study (referred to as “study A”), weights were not used because of the use of a subset, rather than the entire sample of the dataset. The authors used a nonrandom subset of the sample and justified omitting weights because applying weights would have resulted in a non-representative sample. They go on to report that this is consistent with another study (referred to as “study B”) which also used the same dataset and a subset thereof. However, it is important to understand that when weights are not used when conducting tests of inference, the results can be interpreted only for the sample of individuals who completed the survey (Hahs, 2003) because more weight is provided in the analyses to the over sampled groups thus distorting the true population (Thomas and Heck). While the authors of study B chose not to weight their data, they recognized that their decision gave disproportionate weight to the over sampled groups in the dataset. The authors of study B go on to state that while their results may be valid for the subset they selected, they cannot generalize to the population intended because the results are skewed, reflecting disproportionately some groups of students.

The extent to which the average reader understands the treatment of weights by studies A or B is not known. While a subhead paragraph explaining the estimation methods is provided in study B, only one paragraph is provided as a footer in the notes section of study A. The authors of study A do not provide additional details to help the reader understand that the omission of weights will disproportionately weight some groups of students and will substantially limit to whom the results can be generalized. On a more positive note, study B did test the sensitivity of the use or omission of weights, and this was recommended by others (DuMouchel and Duncan). How selection of a subset of the sampled population interfaces with weights and design effects was not documented. Therefore, if a researcher elects not to use weights on the basis that a subset of the dataset was used, a best practice is to make that fact explicitly known

to the reader as well as to test and report how the use or omission of weights impacts the final results. For example, in a different study that omitted weights, the authors recognized that weights should be used when conducting inferential statistics, noted that weights were not necessary given the descriptive nature of their study, and stated that failing to weight the data could result in problems with generalization.

Design Effects

A best practice for using design effects may be to follow the recommendations provided in the dataset technical manuals or articles that provide detailed guidelines on adjusting models with design effects (e.g., Hahs-Vaughn; Thomas and Heck). Design effects are reported in the technical manuals of national datasets (see for example, Huang, Salvucci, Peng, and Owings, 1996; U.S. Department of Education) and may include both the DEFF and DEFT (e.g., U.S. Department of Education). When the design effect for a dependent variable used in a study is not reported in the technical reports, the design effect for a similar variable, the average design effect averaged over a set of variables, or the average design effect of the dependent variable averaged over subgroups of the independent variable are appropriate to use (Huang, Salvucci, Peng, and Owings). Another best practice is to note how the technical manuals have recommended applying design effects and then report how design effects were applied in the study so that the reader has the information she or he needs to critically evaluate design effect application. For multilevel studies where design effects are not required, authors should explicitly state that design effects were not needed because of the study’s design. Likewise, researchers who utilize specialized software that directly accommodates the sampling design should report the software used and how the design issues were addressed (e.g., reporting strata and cluster variables applied).

Conclusion

Recent published research that utilized one or more national datasets was examined to determine the extent to which the use of weights was reported in the research and to highlight best practices in reporting the use of weights. To this end, five years of articles from *Research in Higher Education*, *Journal of Higher Education*, and *Review of Higher Education* were reviewed. Findings revealed that the reporting of weights is not consistent in use or in detail. Of the studies, 27% do not discuss weights, and 36% report only minimal information on the use of weights. This suggests that proper accommodation of the non-simple random sampling design and unequal selection probability when using national datasets, and the informative practice of reporting how these issues were addressed through the analysis may not be

understood by authors, journal editors, and peer reviewers. The result may be miscommunication to the readers, which can lead to multiple problems. 1) Readers who may have limited understanding of methodological considerations in using the datasets are not informed about the technical matters related to using these resources—given that most of the data is freely accessible upon request and/or online, this can lead to misuse of the data by failing to incorporate these “extra steps.” 2) Readers who *do* understand the technical logistics are not provided enough detail to either make an informative decision regarding whether the non-simple random sampling design and unequal selection probability were addressed properly or replicate the study. 3) Serious limitations to inferential tests may exist that prohibit correct parameter estimates, and thus inaccuracies in the analyses are reported.

Suggested best practices for authors include:

1) specifying which weight is used (including the type of weight, such as panel or cross-sectional weight, and the variable name within the dataset); 2) if weights are not used, testing and reporting how the omission of weights (i.e., failing to accommodate unequal selection probability) impacts the final results including providing details that will help the reader understand a) why weights are not used, b) that the omission of weights will disproportionately weight some groups of respondents, c) to whom the results may be generalized; and 3) specifying in detail, if single-level models are used, the adjustment to correct for homogeneity within clusters (i.e., the use of specialized software or design effects to accommodate the sampling design).

Suggested best practices for journals include:

1) requiring authors who used a national dataset to report explicitly and in relative detail how the non-simple random sampling design and unequal selection probability were accommodated in the analysis; 2) specifying criteria for authors, members of editorial review boards, and peer reviewers on what should be reported in relation to accommodating non-simple random sampling design and unequal selection probability and how to effectively critique these reports.

Suggested additional best practices include: 1) wider dissemination of the importance of accommodating non-simple random sampling design and unequal selection probability via database institutes and national conferences; and 2) wider dissemination of how to accommodate non-simple random sampling design and unequal selection probability when using complex surveys through advanced statistical classes.

This study involved the examination of research studies published in three leading higher education journals during a five-year period. As such, it is not known whether the findings generalize to previous years and to other journals that publish studies that make use of national

postsecondary datasets. If three of the preeminent scholarly higher education journals (one of which the official journal of the organization closely tied to NCES and NSF for offering training and grants for using national datasets) do not require authors to consistently report in detail accommodation of non-simple random sampling design and unequal selection probability, it is likely that other journals also do not have this requirement. Replicating this study, therefore, with a more extensive sampling of scholarly journals devoted to policy issues that encourage national dataset use is needed. Regardless, it appears that the omission of a detailed discussion accommodating the features of complex samples when analyzing data from large-scale datasets prevails in scholarly journals. To rectify these shortcomings, several best practices for authors and journals are provided.

Endnote

¹ A list of the articles reviewed can be obtained by contacting the author.

References

- American Educational Research Association. (n.d.). *AERA grants program*. Retrieved December 2, 2003, from <http://www.aera.net/grantsprogram/>
- Association for Institutional Research. (n.d.a). *Summer data policy institute*. Retrieved December 2, 2003, from <http://www.airweb.org/page.asp?page=42>
- Association for Institutional Research. (n.d.b). *Grants*. Retrieved December 2, 2003, from <http://www.airweb.org/page.asp?page=3>
- Cabrera, A. F. & La Nasa, S. M. (2001). On the path to college: Three critical tasks facing America's disadvantaged. *Research in Higher Education*, 42 (2), 119-149.
- DuMouchel, W.H. & Duncan, G.J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association*, 78 (383), 535-543.
- Gregorich, S. (personal communication, 1997, December 26).
- Hahs, D.L. (2003). *The utilization of sample weights in structural equation modeling: An application using the Beginning Postsecondary Students Longitudinal Study 1990/92/94*. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.
- Hahs-Vaughn, D.L. (2005). A primer for using and understanding weights with national datasets. *The Journal of Experimental Education*, 73 (3), 221-248.
- Heck, R. H. & Mahoe, R. (2004). *An example of the impact of sample weights and centering on multilevel SEM models*. Paper presented at the American Educational Research Association annual meeting, San Diego, CA.
- Hox, J.J. & Kreft, I. G. G. (1994). Multilevel analysis methods. *Sociological Methods & Research*, 22 (3), 283-299.
- Huang, G., Salvucci, S., Peng, S., & Owings, J. (1996). *National Educational Longitudinal Study of 1988 (NELS:88) research framework and issues*. (Working Paper No. 96-03). Arlington, VA: Synetics for Management Decisions.
- Journal of Higher Education*. (n.d.). Retrieved July 21, 2005, from <http://www.ohiostatepress.org/>
- Kalton, G. (1983). Models in the practice of survey sampling. *International Statistical Review*, 51, 175-188.
- Kalton, G. (1989). Modeling considerations: Discussion from a survey sampling perspective. In D. Kasprzyk, G. Duncan, G. Kalton, & M. Singh. (Eds.). *Panel surveys* (pp. 575-585). New York: Wiley.
- Kaplan, D. & Elliott, P.R. (1997). A model-based approach to validating education indicators using multilevel structural equation modeling. *Journal of Educational and Behavioral Statistics*, 22 (3), 323-347.
- Kaplan, D. & Ferguson, A. J. (1999). On the utilization of sample weights in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (4), 305-321.
- Kish, L. (1965). *Survey sampling*. New York: Wiley.
- Kish, L. & Frankel, M.P. (1974). Inference from complex samples. *Journal of the Royal Statistical Society, Ser. B*, 36, 1-37.
- Korn, E. L. & Graubard, B. I. (1995). Examples of differing weighted and unweighted estimates from a sample survey. *The American Statistician*, 49 (3), 291-295.
- Lee, E. S., Forthofer, R. N., & Lorimor, R. J. (1989). *Analyzing complex survey data*. Newbury Park, CA: Sage Publications, Inc.
- McLean, J.E. & Kaufman, A.S. (Editors). (1998). Statistical significance testing [Special issue]. *Research in the Schools*, 5(2).
- Muthen, B. O. (1994) Multilevel covariance structure analysis. *Sociological Methods & Research*, 22 (3), 376-398.
- Muthen, B. O. & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25, 267-316.
- National Center for Education Statistics. (n.d.). *Survey and program areas*. Retrieved December 2, 2003, from <http://www.nces.ed.gov/surveys/>
- National Science Foundation. (n.d.). *Databases*. Retrieved December 2, 2003, from <http://www.nsf.gov/sbe/srs/srsdata.htm>
- Peng, S.S. (2000, June). *Technical issues in using NCES data*. AIR/NCES National Data Institute on the Use of Postsecondary Databases, Gaithersburg, MD.
- Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61 (2), 317-337.
- Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society*, 60 (Series B), 23-40.
- Potthoff, R. F., Woodbury, M. A., & Manton, K. G. (1992). "Equivalent sample size" and "equivalent degrees of freedom" refinements for inference using survey weights under superpopulation models. *Journal of the American Statistical Association*, 87 (418), 383-396.
- Pratt, D. J., Whitmore, R. W., Wine, J. S., Blackwell, K. M., Forsyth, B. H., Smith, T. K., et al. (1996) *Beginning postsecondary students longitudinal study second follow-up (BPS:90/94) Final Technical Report*. (NCES Publication No. 96-153). Washington, DC: U.S. Government Printing Office.
- Roberts, J.K. (2000, November). *The pitfalls of ignoring multilevel design in national datasets*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Bowling Green, KY.
- Skinner, C. J. (1989). Introduction to Part A. In C. J. Skinner, D. Holt, & T. M. F. Smith (Eds.), *Analysis of complex surveys* (pp. 23-57). New York: Wiley.
- Skinner, C. J., Holt, D. & Smith, T. M. F. (1989).

Conclusions. In C. J. Skinner, D. Holt, & T. M. F. Smith (Eds.), *Analysis of complex surveys* (pp. 285-290). New York: Wiley.

Stapleton, L. M. (2002). The incorporation of sample weights into multilevel structural equation models. *Structural Equation Modeling*, 9(4), 475-502.

Thomas, S. L. & Heck, R. H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education*, 42 (5), 517-540.

U.S. Department of Education. (2004). *User's manual for the ECLS-K third grade public-use data files and electronic code book*. (NCES Publication No. 2004-001). Washington, DC: National Center for Education Statistics.

West, J. & Rathburn, A. (2004, April). *ECLS-K technical issues*. Presentation at the AERA Institute on Statistical Analysis for Education Policy, San Diego, CA.

THE AIR PROFESSIONAL FILE—1978-2006

A list of titles for the issues printed to date follows. Most issues are "out of print," but microfiche or photocopies are available through ERIC. Photocopies are also available from the AIR Executive Office, 222 Stone Building, Florida State University, Tallahassee, FL 32306-4462, \$3.00 each, prepaid, which covers the costs of postage and handling. Please do not contact the editor for reprints of previously published Professional File issues.

- Organizing for Institutional Research* (J.W. Ridge; 6 pp; No. 1)
Dealing with Information Systems: The Institutional Researcher's Problems and Prospects (L.E. Saunders; 4 pp; No. 2)
Formula Budgeting and the Financing of Public Higher Education: Panacea or Nemesis for the 1980s? (F.M. Gross; 6 pp; No. 3)
Methodology and Limitations of Ohio Enrollment Projections (G.A. Kraetsch; 8 pp; No. 4)
Conducting Data Exchange Programs (A.M. Bloom & J.A. Montgomery; 4 pp; No. 5)
Choosing a Computer Language for Institutional Research (D. Strenglein; 4 pp; No. 6)
Cost Studies in Higher Education (S.R. Hample; 4 pp; No. 7)
Institutional Research and External Agency Reporting Responsibility (G. Davis; 4 pp; No. 8)
Coping with Curricular Change in Academe (G.S. Melchiori; 4 pp; No. 9)
Computing and Office Automation—Changing Variables (E.M. Staman; 6 pp; No. 10)
Resource Allocation in U.K. Universities (B.J.R. Taylor; 8 pp; No. 11)
Career Development in Institutional Research (M.D. Johnson; 5 pp; No. 12)
The Institutional Research Director: Professional Development and Career Path (W.P. Fenstermacher; 6pp; No. 13)
A Methodological Approach to Selective Cutbacks (C.A. Belanger & L. Tremblay; 7 pp; No. 14)
Effective Use of Models in the Decision Process: Theory Grounded in Three Case Studies (M. Mayo & R.E. Kallio; 8 pp; No. 15)
Triage and the Art of Institutional Research (D.M. Norris; 6 pp; No. 16)
The Use of Computational Diagrams and Nomograms in Higher Education (R.K. Brandenburg & W.A. Simpson; 8 pp; No. 17)
Decision Support Systems for Academic Administration (L.J. Moore & A.G. Greenwood; 9 pp; No. 18)
The Cost Basis for Resource Allocation for Sandwich Courses (B.J.R. Taylor; 7 pp; No. 19)
Assessing Faculty Salary Equity (C.A. Allard; 7 pp; No. 20)
Effective Writing: Go Tell It on the Mountain (C.W. Ruggiero, C.F. Elton, C.J. Mullins & J.G. Smoot; 7 pp; No. 21)
Preparing for Self-Study (F.C. Johnson & M.E. Christal; 7 pp; No. 22)
Concepts of Cost and Cost Analysis for Higher Education (P. T. Brinkman & R.H. Allen; 8 pp; No. 23)
The Calculation and Presentation of Management Information from Comparative Budget Analysis (B.J.R. Taylor; 10 pp; No. 24)
The Anatomy of an Academic Program Review (R.L. Harpel; 6 pp; No. 25)
The Role of Program Review in Strategic Planning (R.J. Barak; 7 pp; No. 26)
The Adult Learner: Four Aspects (Ed. J.A. Lucas; 7 pp; No. 27)
Building a Student Flow Model (W.A. Simpson; 7 pp; No. 28)
Evaluating Remedial Education Programs (T.H. Bers; 8 pp; No. 29)
Developing a Faculty Information System at Carnegie Mellon University (D.L. Gibson & C. Golden; 7 pp; No. 30)
Designing an Information Center: An Analysis of Markets and Delivery Systems (R. Matross; 7 pp; No. 31)
Linking Learning Style Theory with Retention Research: The TRAILS Project (D.H. Kalsbeek; 7 pp; No. 32)
Data Integrity: Why Aren't the Data Accurate? (F.J. Gose; 7 pp; No. 33)
Electronic Mail and Networks: New Tools for Institutional Research and University Planning (D.A. Updegrave, J.A. Muffo & J.A. Dunn, Jr.; 7pp; No. 34)
Case Studies as a Supplement to Quantitative Research: Evaluation of an Intervention Program for High Risk Students (M. Peglow-Hoch & R.D. Walleri; 8 pp; No. 35)
Interpreting and Presenting Data to Management (C.A. Clagett; 5 pp; No. 36)
The Role of Institutional Research in Implementing Institutional Effectiveness or Outcomes Assessment (J.O. Nichols; 6 pp; No. 37)
Phenomenological Interviewing in the Conduct of Institutional Research: An Argument and an Illustration (L.C. Attinasi, Jr.; 8p p; No. 38)
Beginning to Understand Why Older Students Drop Out of College (C. Farabaugh-Dorkins; 12 pp; No. 39)
A Responsive High School Feedback System (P.B. DUBY; 8 pp; No. 40)
Listening to Your Alumni: One Way to Assess Academic Outcomes (J. Pettit; 12 pp; No. 41)
Accountability in Continuing Education Measuring Noncredit Student Outcomes (C.A. Clagett & D.D. McConochie; 6pp; No. 42)
Focus Group Interviews: Applications for Institutional Research (D.L. Brodigan; 6 pp; No. 43)
An Interactive Model for Studying Student Retention (R.H. Glover & J. Wilcox; 12 pp; No. 44)
Increasing Admitted Student Yield Using a Political Targeting Model and Discriminant Analysis: An Institutional Research Admissions Partnership (R.F. Urban; 6 pp; No. 45)
Using Total Quality to Better Manage an Institutional Research Office (M.A. Heverly; 6 pp; No. 46)
Critique of a Method For Surveying Employers (T. Banta, R.H. Phillippi & W. Lyons; 8 pp; No. 47)
Plan-Do-Check-Act and the Management of Institutional Research (G.W. McLaughlin & J.K. Snyder; 10 pp; No. 48)
Strategic Planning and Organizational Change: Implications for Institutional Researchers (K.A. Corak & D.P. Wharton; 10 pp; No. 49)
Academic and Librarian Faculty: Birds of a Different Feather in Compensation Policy? (M.E. Zeglen & E.J. Schmidt; 10 pp; No. 50)
Setting Up a Key Success Index Report: A How-To Manual (M.M. Sapp; 8 pp; No. 51)
Involving Faculty in the Assessment of General Education: A Case Study (D.G. Underwood & R.H. Nowaczyk; 6 pp; No. 52)

THE AIR PROFESSIONAL FILE—1978-2006

- Using a Total Quality Management Team to Improve Student Information Publications (J.L. Frost & G.L. Beach; 8 pp; No. 53)
- Evaluating the College Mission through Assessing Institutional Outcomes (C.J. Myers & P.J. Silvers; 9 pp; No. 54)
- Community College Students' Persistence and Goal Attainment: A Five-year Longitudinal Study (K.A. Conklin; 9 pp; No. 55)
- What Does an Academic Department Chairperson Need to Know Anyway? (M.K. Kinnick; 11 pp; No. 56)
- Cost of Living and Taxation Adjustments in Salary Comparisons (M.E. Zeglen & G. Tesfagiorgis; 14 pp; No. 57)
- The Virtual Office: An Organizational Paradigm for Institutional Research in the 90's (R. Matross; 8 pp; No. 58)
- Student Satisfaction Surveys: Measurement and Utilization Issues (L. Sanders & S. Chan; 9 pp; No. 59)
- The Error Of Our Ways; Using TQM Tactics to Combat Institutional Issues Research Bloopers (M.E. Zeglin; 18 pp; No. 60)
- How Enrollment Ends; Analyzing the Correlates of Student Graduation, Transfer, and Dropout with a Competing Risks Model (S.L. Ronco; 14 pp; No. 61)
- Setting a Census Date to Optimize Enrollment, Retention, and Tuition Revenue Projects (V. Borden, K. Burton, S. Keucher, F. Vossburg-Conaway; 12 pp; No. 62)
- Alternative Methods For Validating Admissions and Course Placement Criteria (J. Noble & R. Sawyer; 12 pp; No. 63)
- Admissions Standards for Undergraduate Transfer Students: A Policy Analysis (J. Saupe & S. Long; 12 pp; No. 64)
- IR for IR—Indispensable Resources for Institutional Researchers: An Analysis of AIR Publications Topics Since 1974 (J. Volkwein & V. Volkwein; 12 pp; No. 65)
- Progress Made on a Plan to Integrate Planning, Budgeting, Assessment and Quality Principles to Achieve Institutional Improvement (S. Griffith, S. Day, J. Scott, R. Smallwood; 12 pp; No. 66)
- The Local Economic Impact of Higher Education: An Overview of Methods and Practice (K. Stokes & P. Coomes; 16 pp; No. 67)
- Developmental Education Outcomes at Minnesota Community Colleges (C. Schoenecker, J. Evens & L. Bollman; 16 pp; No. 68)
- Studying Faculty Flows Using an Interactive Spreadsheet Model (W. Kelly; 16 pp; No. 69)
- Using the National Datasets for Faculty Studies (J. Milam; 20 pp; No. 70)
- Tracking Institutional leavers: An Application (S. DesJardins, H. Pontiff; 14 pp; No. 71)
- Predicting Freshman Success Based on High School Record and Other Measures (D. Eno, G. W. McLaughlin, P. Sheldon & P. Brozovsky; 12 pp; No. 72)
- A New Focus for Institutional Researchers: Developing and Using a Student Decision Support System (J. Frost, M. Wang & M. Dalrymple; 12 pp; No. 73)
- The Role of Academic Process in Student Achievement: An Application of Structural Equations Modeling and Cluster Analysis to Community College Longitudinal Data¹ (K. Boughan; 21 pp; No. 74)
- A Collaborative Role for Industry Assessing Student Learning (F. McMartin; 12 pp; No. 75)
- Efficiency and Effectiveness in Graduate Education: A Case Analysis (M. Kehrhahn, N.L. Travers & B.G. Sheckley; No. 76)
- ABCs of Higher Education-Getting Back to the Basics: An Activity-Based Costing Approach to Planning and Financial Decision Making (K. S. Cox, L. G. Smith & R.G. Downey; 12 pp; No. 77)
- Using Predictive Modeling to Target Student Recruitment: Theory and Practice (E. Thomas, G. Reznik & W. Dawes; 12 pp; No. 78)
- Assessing the Impact of Curricular and Instructional Reform - A Model for Examining Gateway Courses¹ (S.J. Andrade; 16 pp; No. 79)
- Surviving and Benefitting from an Institutional Research Program Review (W.E. Knight; 7 pp; No. 80)
- A Comment on Interpreting Odds-Ratios when Logistic Regression Coefficients are Negative (S.L. DesJardins; 7 pp; No. 81)
- Including Transfer-Out Behavior in Retention Models: Using NSC Enrollment Search Data (S.R. Porter; 16 pp; No. 82)
- Assessing the Performance of Public Research Universities Using NSF/NCES Data and Data Envelopment Analysis Technique (H. Zheng & A. Stewart; 24 pp; No. 83)
- Finding the 'Start Line' with an Institutional Effectiveness Inventory¹ (S. Ronco & S. Brown; 12 pp; No. 84)
- Toward a Comprehensive Model of Influences Upon Time to Bachelor's Degree Attainment (W. Knight; 18 pp; No. 85)
- Using Logistic Regression to Guide Enrollment Management at a Public Regional University (D. Berge & D. Hendel; 14 pp; No. 86)
- A Micro Economic Model to Assess the Economic Impact of Universities: A Case Example (R. Parsons & A. Griffiths; 24 pp; No. 87)
- Methodology for Developing an Institutional Data Warehouse (D. Wierschem, R. McBroom & J. McMillen; 12 pp; No. 88)
- The Role of Institutional Research in Space Planning (C.E. Watt, B.A. Johnston, R.E. Chrestman & T.B. Higerd; 10 pp; No. 89)
- What Works Best? Collecting Alumni Data with Multiple Technologies (S. R. Porter & P.D. Umback; 10 pp; No. 90)
- Caveat Emptor: Is There a Relationship between Part-Time Faculty Utilization and Student Learning Outcomes and Retention? (T. Schibik & C. Harrington; 10 pp; No. 91)
- Ridge Regression as an Alternative to Ordinary Least Squares: Improving Prediction Accuracy and the Interpretation of Beta Weights (D. A. Walker; 12 pp; No. 92)
- Cross-Validation of Persistence Models for Incoming Freshmen (M. T. Harmston; 14 pp; No. 93)
- Tracking Community College Transfers Using National Student Clearinghouse Data (R.M. Romano and M. Wisniewski; 14 pp; No. 94)
- Assessing Students' Perceptions of Campus Community: A Focus Group Approach (D.X. Cheng; 11 pp; No. 95)
- Expanding Students' Voice in Assessment through Senior Survey Research (A.M. Delaney; 20 pp; No. 96)
- Making Measurement Meaningful (Carpenter-Hubin, J. & Hornsby, E.E., 14 pp; No. 97)
- Strategies and Tools Used to Collect and Report Strategic Plan Data (Blankert, J., Lucas, C. & Frost, J.; 14 pp; No. 98)
- Factors Related to Persistence of Freshmen, Freshman Transfers, and Nonfreshman Transfer Students (Perkhounkova, Y, Noble, J & McLaughlin, G.; 12 pp; No. 99)
- Does it Matter Who's in the Classroom? Effect of Instructor Type on Student Retention, Achievement and Satisfaction (Ronco, S. & Cahill, J.; 16 pp; No. 100)

The *AIR Professional File* is intended as a presentation of papers which synthesize and interpret issues, operations, and research of interest in the field of institutional research. Authors are responsible for material presented. The *File* is published by the Association for Institutional Research.

Editor:
Dr. Gerald W. McLaughlin
Director of Planning and
Institutional Research
DePaul University
1 East Jackson, Suite 1501
Chicago, IL 60604-2216
Phone: 312/362-8403
Fax: 312/362-5918
gmclaugh@depaul.edu

Associate Editor:
Ms. Debbie Dailey
Associate Director of Planning and
Institutional Research
Georgetown University
303 Maguire Hall, 37th & O St NW
Washington, DC 20057
Phone: 202/687-7717
Fax: 202/687-3935
daileyd@georgetown.edu

Managing Editor:
Dr. Terrence R. Russell
Executive Director
Association for Institutional Research
1435 E. Piedmont Drive
Suite 211
Tallahassee, FL 32308
Phone: 850/385-4155
Fax: 850/385-5180
air@mailier.fsu.edu

AIR Professional File Editorial Board

Dr. Trudy H. Bers
Senior Director of
Research, Curriculum
and Planning
Oakton Community College
Des Plaines, IL

Ms. Rebecca H. Brodigan
Director of
Institutional Research and Analysis
Middlebury College
Middlebury, VT

Dr. Harriott D. Calhoun
Director of
Institutional Research
Jefferson State Community College
Birmingham, AL

Dr. Stephen L. Chambers
Director of Institutional Research
and Assessment
Coconino Community College
Flagstaff, AZ

Dr. Anne Marie Delaney
Director of
Institutional Research
Babson College
Babson Park, MA

Dr. Gerald H. Gaither
Director of
Institutional Research
Prairie View A&M University
Prairie View, TX

Dr. Philip Garcia
Director of
Analytical Studies
California State University-Long Beach
Long Beach, CA

Dr. David Jamieson-Drake
Director of
Institutional Research
Duke University
Durham, NC

Dr. Anne Machung
Principal Policy Analyst
University of California
Oakland, CA

Dr. Marie Richman
Assistant Director of
Analytical Studies
University of California-Irvine
Irvine, CA

Dr. Jeffrey A. Seybert
Director of
Institutional Research
Johnson County Community College
Overland Park, KS

Dr. Bruce Szelest
Associate Director of
Institutional Research
SUNY-Albany
Albany, NY

Authors interested in having their manuscripts considered for the *Professional File* are encouraged to send four copies of each manuscript to the editor, Dr. Gerald McLaughlin. Manuscripts are accepted any time of the year as long as they are not under consideration at another journal or similar publication. The suggested maximum length of a manuscript is 5,000 words (approximately 20 double-spaced pages), including tables, charts and references. Please follow the style guidelines of the *Publications Manual of the American Psychological Association, 4th Edition*.
