

# AN ALTERNATIVE JACKKNIFE PROCEDURE FOR NAEP VARIANCE ESTIMATION

Stanley Weng, Sameena Salvucci, Synectics for Management Decisions Inc.,

Michael P. Cohen, National Center for Education Statistics

Sameena Salvucci, Synectics for Mgmt. Decisions, Inc., 3030 Clarendon Blvd #305, Arlington VA 22201

**Key Words:** Replicate Weights, Complex Sample Survey, Robust Computational Procedures

## 1. Study purpose

This empirical study explores an alternative method for performing jackknife variance estimation which makes better use of the sampling variation than the procedure currently used for the National Assessment of Educational Progress (NAEP), a periodic survey conducted by the National Center for Education Statistics (NCES). Better use of the sampling variation should improve the accuracy of the NAEP variance estimates. The alternative method should also make it possible to implement systematic computational procedures to conduct NAEP jackknife variance estimation.

## 2. NAEP Sample Design

The basic primary sampling unit (PSU) sample design for the main NAEP assessment is a stratified probability sample with one PSU selected per stratum with probability proportional to the population. The sampling unit within the PSU is the individual school. Schools are selected systematically with probability proportionate to the assigned measure of size. The sample of students within sampled schools is systematically drawn from school-prepared lists of eligible students.

## 3. Assignment of Sessions to Schools

All sampled students within a school are assigned to assessment sessions based on the following three age/grade eligibility classes:

Age Class 1: Age 9/Grade 4

Age Class 2: Age 13/Grade 8

Age Class 3: Age 17/Grade 12

Print administered reading, writing, and mathematics sessions and tape administered mathematics sessions were conducted at all age classes. The method of determining the number and type of sessions to be administered in a given school varied by age class.

Our study was limited to examining standard errors for grade 8 reading proficiency estimates in the 1992 NAEP main assessment.

## 4. NAEP Jackknife Variance Estimation

The NAEP variance estimation procedure, as used for the 1992 and 1994 NAEP, uses a jackknife variance estimator. This method will be referred to as the original "paired" jackknife procedure.

For the purposes of variance estimation, pairs of first-stage sampling units (FSSUs) or of appropriate aggregates of them are defined in a manner that models the design as one in which two first-stage units are drawn with replacement per stratum. The definition and pairing of the FSSUs are different for the certainty and noncertainty PSUs. Each noncertainty PSU constitutes a single FSSU while each certainty PSU contains two or more sampled FSSUs, each consisting of one or more schools. The  $2N$  noncertainty PSUs are formed into  $N$  pairs of FSSUs, where the pairs are composed of PSUs from adjacent strata and are thus relatively similar on the sample stratification characteristics. Whereas, as described in section 2 above, the actual sample design was to select one FSSU with probability proportional to size from each of  $2N$  strata, for variance estimation purposes the design is regarded as calling for the selection of two FSSUs with probability proportional to size with replacement from each of  $N$  strata. This alteration probably produces a positive bias to estimates of sampling error.

Although the two-PSU-per-stratum jackknife is a simple procedure, it may not perform satisfactorily. The formation of the jackknife replicates greatly changed the original sampling design, and it ignored much of the sampling variation contained in the sample, with a considerable reduction of the degrees of freedom for the estimation space.

## 5. NAEP Student Jackknife Replicates

The NAEP variances are based on a set of student jackknife replicates (replicate weights) contained in each sample. Each main NAEP sample

dataset contains a set of 56 jackknife replicates: 30 replicates reflect the amount of sampling variance contributed by the noncertainty strata of PSUs, and 26 reflect the variance contribution of the certainty PSU samples. The replicates were formed in the following way. The 60 noncertainty PSUs, drawn from 60 strata, were formed into 30 pairs, each pair composed of PSUs from adjacent strata within each subuniverse of sampling (thus the strata were relatively similar on the characteristics of stratification). The 26 replicates from the 34 certainty PSUs were created in a more complex way: the seven largest PSUs were assigned to ten replicates, the next five largest PSUs were assigned to one replicate each, and the remaining 22 were paired and assigned to 11 replicates.

## 6. Alternative jackknife variance estimation

We propose an alternative jackknife procedure to better incorporate the data sampling structure into jackknifing and hence to catch more of the sample variation, and to be able to implement systematic computational procedures. NAEP's sample design has one PSU selected per stratum; therefore, there is no direct way to estimate sampling variance at the PSU level without collapsing strata. The alternative jackknife procedure performs jackknifing at the next sampling level, the school level; that is, the alternative procedure is a general stratified jackknife performed to schools within PSU. Since the sampling fraction of schools within PSU is small we assume they are independent. We expected the alternative to provide improved accuracy for the variance estimates.

In proposing the alternative jackknife procedure, we reviewed the jackknife variance estimation methodology (Shao and Tu, 1995, Shao and Wu, 1989).

## 7. Analysis and results

### Data

The 1992 NAEP Main Assessment Reading Test Age 13/Grade 8 data were used to conduct the alternative jackknife variance estimation. A SAS data set was created from the raw data in the 1992 NAEP National Assessment CD-ROM. The five composite variables for reading proficiency ("Plausible NAEP reading value") were used as response variables to estimate average reading proficiency for the nation and for the domains defined by Region (Northeast,

Southeast, Central, West) and Type of School (Public, Private, Catholic), respectively. Missing cases for the response variables were deleted.

### Estimation

We performed jackknife variance estimation using (1) our alternative jackknife procedure and (2) the original "paired" jackknife procedure. Since the our alternative jackknife variance estimation does not include nonresponse, trimming, and poststratification adjustments, we calculated comparable "unadjusted" variances using the original "paired" jackknife procedure. Therefore, in implementing the original "paired" jackknife procedure we used WesVar PC to develop a set of jackknife replicate weights based on the NAEP final student weight instead of using the student jackknife replicate weights available on the NAEP file because these weights already included nonresponse, trimming, and poststratification adjustments. We used the VPLX software (Fay, 1995) for implementing our alternative procedure and as stated above WesVar PC for the original procedure. VPLX has been shown to produce reliable jackknife estimates in a previous study (Weng et al., 1995).

The grade 8 national and domain average reading proficiency estimates and their associated standard errors from the two jackknife procedures in comparison are presented in tables A-1, A-2, and A-3, respectively.

For reference, table A-4 lists the grade 8 average reading proficiency and associated standard errors provided by Mullis et al. (1993). However, note that these standard errors were based on the NAEP student replicate weights which were created to include nonresponse, trimming, and poststratification adjustments. Thus, these standard errors are not directly comparable to the standard errors that we calculated in our analyses.

### Discussion

It can be seen from tables A-1 and A-3 that the standard error for average reading proficiency using our alternative jackknife procedure is just a little greater than that from the original jackknife procedure (except in Catholic schools). In addition, in table A-2, the variance for the Central region using our alternative method is almost one third higher than when using the original method. This result conforms with our belief that the alternative jackknife would

catch sampling variation ignored by the original jackknife. In comparing variances across the other domains, it can be seen that the variances are very similar. Also, since the alternative method has more degrees of freedom than the original method, the variance estimate precision is improved. Also, Shao and Tu (1995) discuss that the jackknife has some robustness properties against the violation of the school independence assumption.

Note, however, that the alternative jackknife can not estimate the sampling variation at the NAEP PSU level within strata: the variance estimates provided by this procedure would generally be underestimated.

The two-PSU-per-stratum “paired” version of the jackknife procedure, as implemented in the WesVar software (Westat, 1995) now available on the Internet, has almost been adopted as a standard version of jackknife. It is in wide use for NCES survey variance estimation. This study provides useful information on the performance of such a jackknife procedure. The results of this analysis may be interesting as NCES considers how to improve jackknife variance estimation practice.

## 8. Further steps

The alternative jackknife procedure for NAEP variance estimation seems promising. This study is only the first step in exploring how to improve jackknife variance estimation for NAEP. Further steps may be taken according to the following methodological consideration: Shao and Wu (1989) and Wu (1990) discussed the more general delete- $d$  version of jackknife procedure, which, with appropriately chosen  $d$ , can be used to improve the performance of the variance estimation and make the jackknife variance estimator more robust.

## 9. References

Fay, R.E. (1995). *VPLX*. Washington, DC: U.S. Bureau of the Census.

Mullis, I.V.S., Campbell, J.R. and Farstrup, A.E. (1993). *NAEP 1992 Reading Report Card for the Nation and the States*. Washington, DC: National Center for Education Statistics.

Rust, K., Burke, J., Fahimi, M. And Wallace, L. (1992). *1990 NAEP Assessment of Educational Progress, Sampling and Weighting Procedures*,

Part 2 - National Assessment. Rockville, MD: Westat, Inc.

Shao, J. and Tu, D. (1995). *The Jackknife and the Bootstrap*. New York: Springer-Verlag.

Shao, J. and Wu, C.F.J. (1989). A general theory for jackknife variance estimation. *The Annals of Statistics*, 17, 1176-1197.

Wallace, L. and Rust, K. (1994). *1992 National Assessment of Educational Progress, Sampling and Weighting Procedures, Final Report*. Rockville, MD: Westat, Inc.

Weng, S.S., Zhang, F. and Cohen, M.P. (1995). Variance estimates comparison by statistical software, *ASA 1995 Proceedings of the Section on Survey Research Methods*, vol. 1, 333-338. Alexandria, VA: American Statistical Association.

Westat, Inc. (1996). *A User's Guide to WesVarPC, Version 2.0*. Rockville, MD: the author

**Table A-1. National Grade 8 Average Reading Proficiency and Jackknife Variance Estimates**

Variable	Average proficiency	Standard Error Calculated by		
		Alternative Method	Original Method	Alternative s.e./ Original s.e.
Reading proficiency 1	254.465	0.952	0.853	1.116
Reading proficiency 2	253.995	0.976	0.912	1.070
Reading proficiency 3	254.975	0.948	0.916	1.035
Reading proficiency 4	254.383	0.938	0.902	1.040
Reading proficiency 5	255.011	0.978	0.933	1.048
Average	254.566	0.958	0.903	1.062

**Table A-2. Domain Grade 8 Average Reading Proficiency and Jackknife Variance Estimates: by region**

Domain	Average proficiency	Standard Error Calculated by		
		Alternative Method	Original Method	Alternative s.e./ Original s.e.
<b>Northeast</b>				
Reading proficiency 1	257.226	2.341	2.013	1.163
Reading proficiency 2	256.939	2.176	2.050	1.061
Reading proficiency 3	257.660	2.142	1.985	1.079
Reading proficiency 4	257.285	2.246	1.930	1.164
Reading proficiency 5	258.033	2.273	2.108	1.078
Average	257.429	2.236	2.017	1.109
<b>Southeast</b>				
Reading proficiency 1	247.418	2.111	2.265	0.932
Reading proficiency 2	246.601	2.109	2.421	0.871
Reading proficiency 3	247.707	2.059	2.458	0.838
Reading proficiency 4	247.526	2.012	2.434	0.827
Reading proficiency 5	247.524	2.178	2.331	0.934
Average	247.355	2.094	2.382	0.880
<b>Central</b>				
Reading proficiency 1	259.105	1.605	1.195	1.343
Reading proficiency 2	259.283	1.728	1.369	1.262
Reading proficiency 3	260.425	1.543	1.261	1.224
Reading proficiency 4	259.249	1.611	1.329	1.212
Reading proficiency 5	260.392	1.651	1.459	1.132
Average	259.691	1.628	1.323	1.235
<b>West</b>				
Reading proficiency 1	254.250	1.511	1.629	0.928
Reading proficiency 2	253.350	1.681	1.715	0.980
Reading proficiency 3	254.263	1.683	1.742	0.966
Reading proficiency 4	253.691	1.575	1.754	0.898
Reading proficiency 5	254.302	1.637	1.809	0.905
Average	253.971	1.617	1.730	0.935

**Table A-3. Domain 8th Grade Average Reading Proficiency and Jackknife Variance Estimates:  
by type of school**

Domain	Average proficiency	Standard Error Calculated by		
		Alternative Method	Original Method	Alternative s.e./ Original s.e.
<b>Public</b>				
Reading proficiency 1	252.219	1.042	0.937	1.112
Reading proficiency 2	251.813	1.074	0.981	1.095
Reading proficiency 3	252.783	1.037	0.986	1.052
Reading proficiency 4	252.185	1.034	0.972	1.064
Reading proficiency 5	252.800	1.075	1.036	1.038
Average	252.360	1.052	0.982	1.072
<b>Private</b>				
Reading proficiency 1	280.323	2.853	2.817	1.013
Reading proficiency 2	279.919	2.627	2.421	1.085
Reading proficiency 3	280.862	2.812	2.538	1.108
Reading proficiency 4	279.618	2.457	2.497	0.984
Reading proficiency 5	281.336	3.037	2.800	1.085
Average	280.412	2.757	2.615	1.055
<b>Catholic</b>				
Reading proficiency 1	272.527	1.683	1.723	0.977
Reading proficiency 2	271.064	1.683	1.869	0.900
Reading proficiency 3	272.209	1.742	1.846	0.944
Reading proficiency 4	272.098	1.631	1.773	0.920
Reading proficiency 5	272.262	1.635	1.633	1.001
Average	272.032	1.675	1.769	0.948

**Table A-4. Grade 8 Average Reading Proficiency and Standard Error**

Domain	Average Proficiency	Standard error
Nation <sup>1</sup>	260	0.9
Region <sup>2</sup>		
Northeast	263	1.8
Southeast	254	1.7
Central	264	2.2
West	260	1.2
Type of School <sup>3</sup>		
Public	258	1
Private	283	3
Catholic	275	1.9

Source: Mullis et al. (1993), <sup>1</sup>table 1, <sup>2</sup>table 3, <sup>3</sup>table 2.