Construction of Replicate Weights for Project TALENT

Zhulin He[*†] Alan F. Karr[*] Michael P. Cohen[†] Danielle Battle[†]

Deanna Lyter Achorn[†] Alexander D. McKay[‡]

Abstract

Project TALENT is a large, nationally representative longitudinal study developed by American Institutes for Research and conducted from 1960 to 1974. The goal was to assess the interests, abilities, and demographics of 9^{th}–12^{th} graders and their trajectories into adulthood. More than 1,200 junior and senior high schools participated. Replicate weights were not constructed at the time, preventing the estimation of standard errors. In this paper, the retrospective construction of 118 sets of student-level replicate weights is described. The process entailed adjustment of the original base year (1960) student weights and school weights to better estimate the educational and life experiences that are most important to individuals' life trajectories. CHAID analysis was performed to generate variance strata and variance primary sampling units. Finally, the student-level replicate weights were constructed using a jackknife procedure. The use of replicate weights is illustrated by estimating standard errors for quantiles of composite cognitive scores constructed from student questionnaires.

Key Words: Project TALENT, Weight adjustment, CHAID, Jackknife replicate weights, Survey analysis.

1. Introduction

In this section we provide background on Project TALENT, a groundbreaking longitudinal study of high school students. Project TALENT challenged the limits of technical feasibility for its day because of its large sample size and the quantity of rich data collected on each student sampled. Today Project Talent is being revived to study the physical, cognitive, and social processes of aging. Producing student-level replicate weights, the focus of this paper, is an important step in having the base-year data meet modern standards.

1.1 Project TALENT Overview

Project TALENT (PT) is a large, nationally representative, longitudinal study of about 377,000 American students who were in the 9^{th}–12^{th} grades in 1960. Conducted by American Institutes for Research (AIR), PT was designed to assess how early life experiences, interests, aptitudes, and cognitive abilities affect future educational attainment, occupation, and family formation. Additional data were collected from principals and guidance counselors on their school's programs, policies, and community demographic characteristics to identify the impact that school and guidance programs had on student development. The overarching purpose was to understand and develop the talents of young people so the nation could maximize its human potential (Flanagan et al., 1962).

An extensive amount of information was collected on the students over 2 days of testing. Nearly 400 items were collected on family characteristics, health, activities, and future plans, and personality was assessed with 150 items. A 395-item information test assessed

[*]National Institute of Statistical Sciences, 19 T.W. Alexander Drive, Research Triangle Park, NC 27709
[†]American Institutes for Research, 1025 Thomas Jefferson Street NW, Washington, DC 20007
[‡]BlackRock, 601 Union Street 56th Floor, Seattle, WA 98101

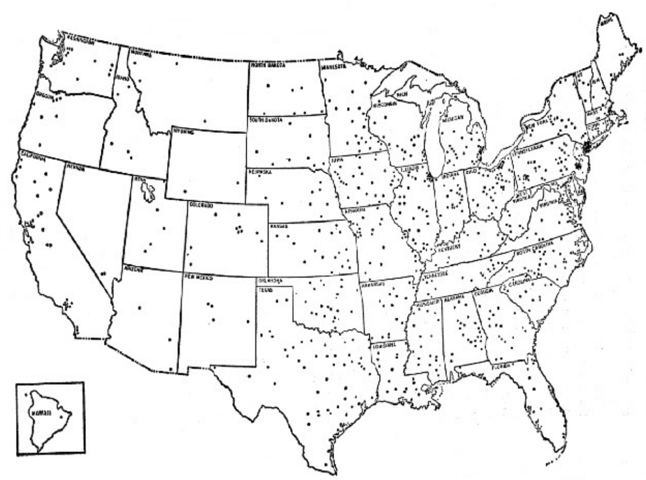


Figure 1: Locations of schools participating in PT (Source: The Project TALENT Data Bank Handbook. (Wise et al., 1979))

knowledge, and interests were assessed with another 205 items. Additional information is available on the PT website (American Institutes for Research, 2013).

In total, 377,015 students were surveyed in the base year, constituting approximately 5% of U.S. high school students in 1960. Participants were surveyed in follow-up collections 1, 5, and 11 years after their expected high school graduation. The primary design and modal ages of the students participating in PT are described in Table 1.

Schools participating in PT were selected primarily from a list provided by the U.S. Office of Education; the Internal Revenue Service provided a supplementary list of schools. The National Catholic Education Association provided a list of parochial schools and the National Center for Education Statistics (NCES) gave AIR a list of other nonpublic schools. AIR sampled schools from these lists (Shaycoft, 1977). The base year survey was conducted in 1960 with students (grades 9–12) from 985 senior high schools and 241 junior high schools. Figure 1 shows the location of these 1,226 schools.

Beyond the stratification variables and sampling rates in Tables 2 and 3, details of how schools were sampled are obscure. In New York City, all public senior and junior high schools were selected. In Chicago, "20 of the 38 academic high schools" were selected (Wise et al., 1979, p. 13).

1.2 Primary Goal

The primary goal of the research reported in this paper is to create student-level replicate weights that are usable for estimating standard errors of national estimates calculated from PT data. Preliminary steps included adjusting base-year weights and generating school-level variance strata and variance primary sampling units. The structure of this paper is as follows. In Section 2, how the base-year weights were computed and adjusted is described. In Section 3, the CHAID analysis is introduced and applied to the PT data. In Section 4, replicate weights are generated using a jackknife methodology. Section 5 shows the results of an illustrative application: the estimation of standard errors for quantiles of several composite scores.

Table 1: Modal age of PT respondents by survey component

	Calendar Year	9^{th} Grade	10^{th} Grade	11^{th} Grade	12^{th} Grade
Base Year	1960	Age 15	Age 16	Age 17	Age 18
1-Year Follow Up	1961				Age 19
	1962			Age 19	
	1963		Age 19		
	1964	Age 19			
5-Year Follow Up	1965				Age 23
	1966			Age 23	
	1967		Age 23		
	1968	Age 23			
11-Year Follow Up	1971				Age 29
	1972			Age 29	
	1973		Age 29		
	1974	Age 29			

2. Base-Year Weights

To compute the sampling weights (school weights and student weights) of the participating schools and students, original PT researchers considered four stratification factors: (1) school type (public, parochial, and private); (2) school geographic location; (3) size of senior class (for public schools only); and (4) retention ratio (for public schools only), which is the number of graduates in the preceding year divided by the number of 10^{th} graders.

The computations of schools weights and student weights are different for parent senior high schools and feeder junior high schools. For parent senior high schools, school weights (W_{sch}) are computed as the inverse of sampling probability in the sampling frame stratum, adjusted for schools that were selected but did not participate. Figures 2 and 3, which are essentially reproduced from original PT documentation (Wise et al., 1979), show details of the stratification and school weight computation. For junior high schools, the situation is more complicated. If a junior high school was clearly and exclusively associated with one senior high school (it is a "feeder" of that senior high school, which is termed its "parent") that participated in PT, then it was selected and all of the 9^{th} graders in it were selected. Otherwise—for example, for junior high schools sending students to both a senior high school selected for PT and one not selected—various judgments were applied, the details of which have not been recovered.

All students in grades 9–12 in sampled schools (other than a small number of absentees and—possibly—refusals), except in Chicago and New York City, were surveyed in the base year. In Chicago, one-tenth of the students in academic and technical public high schools were tested, and in New York City, one-twelve of the students in every public high school were tested. These students are believed to have been selected by systematic sampling. Corresponding student weights (W_{std}) in the senior high schools are then defined as the

				REGIONS																				
				New England			Mideast			Great Lakes			Plains			Southeast			Southwest			Rockies		
		Sampling Ratio	Retention Ratio	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight
Public High Schools	1-24 seniors	1/50	Low	2	2	50.0	2	2	50.0	8	8	50.0	22	23	52.3	17	19	57.1	12	12	52.2	5	5	50.0
			High	2	2		1	1		7	7	50.0	12	13	54.2	11	12		11	11		3	3	
			Unk.													0	1		0	1				
			Spec. Cases													5	5	20.0	1	1	20.0			
	25-99 seniors	1/20	Low	10	10	20.0	21	21	20.0	45	45	20.0	34	35	20.6	77	82	21.3	19	21	22.1	6	8	26.7
			High	6	6	20.0	21	22	21.0	40	40	20.0	27	28	20.7	64	66	20.6	16	18	22.5	5	6	24.0
			Unk.										1	1	20.0	0	1							
			Spec. Cases							1	1	16.2	1	1	16.2	1	1	32.5						
	100 -- 399 seniors	1/20	Low	8	8	20.0	20	22	22.0	16	17	21.3	8	9	22.5	19	20	21.1	6	6	20.0	2	4	40.0
			High	4	4	20.0	13	15	23.1	14	15	21.4	4	5	25.0	17	18	21.2	4	5	25.0	1	1	20.0
			Unk.																					
			Spec. Cases													1	1	17.0						
	400+ seniors	1/13	Low	0	0	13.0	2	3	26.0	2	2	18.6	2	2	13.0	2	2	13.0	1	1	13.0	1	1	13.0
			High	1	1		1	3		5	7		1	1		1	1		2	2		1	1	
			Unk.							0	1					1	1							
			Spec. Cases							1	1	20.0							1	1	3.0			
Parochial High Schools		1/20	N/A	13	13	20.0	23	25	21.7	21	21	20.0	15	16	21.3	13	15	23.1	7	7	20.0	2	2	20.0
Private High Schools		1/20	N/A	**8**	**9**	**22.5**	8	10	25.0	6	7	23.3	4	4	20.0	7	11	31.4	4	4	20.0	2	2	20.0
													1	1	50.0	1	1	50.0						
Totals				54	55		112	124		166	172		132	139		237	257		84	90		28	33	

Figure 2: Proposed base year weights

same as the school weights except for students in public schools in Chicago and New York City. For public schools in Chicago and New York City, student weights are computed as $W_{std} = 10W_{sch}$ and $W_{std} = 12W_{sch}$ respectively.

Figures 2 and 3 were originally printed in 1977 (Wise et al., 1979). In the "original" PT dataset, weights were rounded to three digits ($xx.y$) because of electronic tape limitations at that time. In addition, two schools had no associated student records but were assigned school weights. Our adjustments consisted of recalculating all school weights to the full precision of SAS® (rather than to three decimal places) and dropping those two schools, whose weights were re-assigned to other schools in the same stratum.

Figures 2 and 3 also show that some *ad hoc* adjustments were made for private and parochial schools in some locations. These adjustments are retained in the school weights, together with calculation to full precision. Because the rationale for some of the adjustments could not be identified, they were of necessity dropped in the construction of replicate weights.

3. CHAID Analysis

The key preliminary step in constructing replicate weights (for either schools or students) is to generate school-level variance strata (VSTRATA) and variance primary sampling units (VPSUs). It is common practice to have VSTRATA and VPSUs that differ from the sampling strata and PSUs; in this case, to reduce the number of replicate weights to a more manageable level, fewer VSTRATA were preferred. The approach here was to generate VSTRATA of schools and then randomly split each VSTRATUM into two VPSUs. This approach mirrored the design of PT: other than in Chicago and New York, sampling was at the school level.

Essential characteristics of VSTRATA are internal homogeneity—schools in each VSTRATUM are similar—and size: each VSTRATUM should have neither too few nor too

			CITIES (not included in regional groupings)																			
Far West			Los Angeles			Detroit			Chicago			New York			Philadelphia			Noncontiguous			Totals	
Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited	Weight	Participating	Invited
2	3	75.0									1.9*			1.0*								
1	1	50.0																				
												0	0									
									*Weight for all public high schools			*Weight for all public high schools									124	132
8	9	22.5																				
6	7	23.3																				
									1	1		3	3									
																					413	433
6	10	33.3																				
6	7	23.3																				
			1	1	20.0				9	9		32	32		1	1	20.0	2	2	20.0		
																					194	212
2	3	21.7*	1	2																		
1	4	65.0*	0	1																		
*includes the weight for L.A. 400+ seniors						1	1	13.0	10	10		50	50		1	1	13.0					
																					91	103
					35**				**Weight for all parochial high schools in city groupings													
10	**11**	**22.0**	1	2		3	3		1	3		3	4		0	2		0	0		112	124
					26.7**				** Weight for all private high schools in city groupings													
5	5	20.0	0	0		0	0		1	1		2	2		0	1		1	1	20.0	50	59
47	60		3	6		4	4		22	24		90	91		2	5		3	3		984	1063

Figure 3: Proposed base year weights (continued)

many schools. In this section, Chi-square Automatic Interaction Detection (CHAID) analysis (Kass, 1980), also known as partition modeling, is applied to construct VSTRATA. As noted previously, VPSUs are then constructed by randomly splitting each VSTRATUM into two VPSUs. CHAID was chosen based on its nonparametric nature, ability to handle missing data, and ease of implementing stopping rules, which in this case represent either lack of statistical significance or VSTRATA that are too small. The version of CHAID used was SAS JMP (Jones and Huddlestone, 2009; SAS Institute, 2012). Summarizing briefly, similarity of schools is defined in terms of a single response variable, and binary splits are made on the basis of the predictor variables, one variable at a time. Figures 5 and 6, described in Section 3.1, illustrate the process. The terminal nodes of the partition are taken to be the VSTRATA.

Because of the differences between parent senior high schools and feeder junior high schools, we performed two separate CHAID analyses. Details are explained in Sections 3.1 and 3.2.

3.1 CHAID for Senior High Schools

For the CHAID with senior high schools, the percentage of students in a college preparatory program (CPP) is used as the response variable. This variable was collected in extensive school-level questionnaires that complemented the student-level data collection. Roughly speaking, CPP measures the extent to which a high school is academically oriented, as opposed to preparing students for vocational training or direct employment. As a point of reference, in 1960 approximately 45% of U.S. high school graduates were enrolled in college in the fall following graduation; accounting for dropouts, the enrollment rate was much lower.

The PT CPP variable was categorized into 13 groups: none, 0–9% (0 not included), 10%–19%, 20%–29%, 30%–39%, 40%–49%, 50%–59%, 60%–69%, 70%–79%, 80%–89%, 90%–99%, 100%, and missing information group. The unweighted histogram for

CPP is shown in Figure 4. Five predictors were employed:

Geographic Area: The nine (present-day) U.S. Census Bureau divisions, plus New York City. The nine divisions are New England, Middle Atlantic, East North Central, West North Central, Mountain, Pacific, West South Central, East South Central and South Atlantic.

School type/control: Public, private, or parochial.

Size of the senior class: Total number of 12^{th} graders in each public school.

Minority enrollment[1]: Based on the school questionnaire, the distribution of students by race was categorized as predominantly Black, predominantly White, or predominantly other where predominantly refers to 50% or more enrollment.

Coed: Single-sex (at least 90% male or female enrolled) or coeducational.

The "retention rate" stratification variable was not used because it is available only for some of the public schools, and because the values present in the data could not be explained.

The CHAID/partitioning analysis was straightforward. Splits were made on the five predictor variables until they were not statistically significant or node size was less than 12. Terminal nodes with more than 24 schools were split randomly into sets of size 12–24. Figure 5 shows the first two splits: first on control (public vs. private or parochial); and then for public schools, on minority (predominantly White vs. predominantly Black or another minority). Figure 6 shows the final partition, with 21 nodes, some of which required further splitting to satisfy the "$12 \leq$ size ≤ 24" criterion.

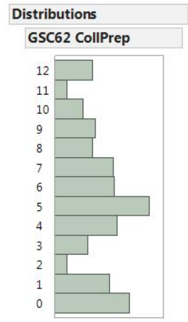
3.2 CHAID for Junior High Schools

Junior high schools not in New York City were simply assigned to the same VSTRATUM as their parent senior high school; this was feasible because each of these schools had only one parent. Junior high schools in New York City do not have parent senior high schools in the dataset; a separate CHAID analysis was conducted for them. As before, the response variable was CPP. Because all of the New York City junior high schools were public schools, we considered three predictors: minority enrollment (Black, White, and other minority); coed (coed and single-sex), and number of 9^{th} grade students.

3.3 VSTRATA and VPSUs

Combining the results of the two CHAID analyses created 52 VSTRATA containing the 985 senior high schools and the 116 associated feeder junior high schools, and 7 VSTRATA containing the 125 public junior high schools in New York City. Each VSTRATUM was split randomly into two VPSUs.

[1]When all items identifying the schools student racial distribution were missing, the categorization was based on the racial distribution of students who responded to the 1-, 5-, or 11-year follow-up. If 60% or more of the students reported Black or White as their race, then the school was categorized as being predominantly Black or predominantly White, respectively. If 60% of the respondents reported a race other than White or Black, then the school was categorized as predominantly other. This higher percentage was based on an analysis that compared the accuracy of using student-reported data (given patterns in nonresponse and changes in reporting options) to school-reported data where school- and student-level information was present.



Frequencies

Level	Count	Prob
0	149	0.12284
1	110	0.09068
2	24	0.01979
3	66	0.05441
4	124	0.10223
5	189	0.15581
6	120	0.09893
7	117	0.09646
8	75	0.06183
9	82	0.06760
10	56	0.04617
11	25	0.02061
12	76	0.06265
Total	1213	1.00000

N Missing 13
13 Levels

Figure 4: Unweighted histogram of values of CPP for PT schools

4. Replicate Weights

Keeping in mind that the VSTRATA and VPSUs (Section 3.3) are at the school level, but that the goal is to construct replicate weights at the student level, a jackknife procedure was used to construct 118 sets of student-level replicate weights. Specifically, for each set of replicate weights, students in schools in one VPSU were omitted and their weights reassigned to students in its "twin," that is, the other VPSU in the same VSTRATUM.

The method to estimate the standard error for a target statistic $\hat{\theta}$ is well-established (Stapleton, 2008; Wolter, 2007). Let $\hat{\theta}_{(ik)}$ be the statistic computed using the replicate weights associated with omitting VPSU k from VSTRATUM i, and let $\hat{\theta}_0$ be the estimate calculated using the full set of base weights. Then, the estimated standard error is

$$\widehat{\mathrm{SE}}(\hat{\theta}) = \sqrt{\sum_{i=1}^{59} \sum_{k=1}^{2} \frac{1}{2} (\hat{\theta}_{(ik)} - \hat{\theta}_0)^2}.$$

5. Application and Results

We applied CHAID analysis and jackknife replicate weights to PT to estimate standard errors of eight composite scores for student-level measures. The results are shown in Table 2. All eight composite scores are generated as weighted combinations of students' item scores, or of other composites and, as indicated by the names in Table 2, they represent ability, knowledge, and achievement measures. The estimates and standard errors of three quantiles of the eight variables were calculated. Coefficients of variation were computed at the same time.

The eight composites are extremely highly correlated, as shown by the principal components analysis in Figure 7. It is expected that, accounting for the different scales, quantiles should be estimated with approximately equal accuracy, which is confirmed by the uniformity of the coefficients of variation in Table 2. The estimated standard errors in Table 2 are consistent with those obtained from replicate weights constructed using successive

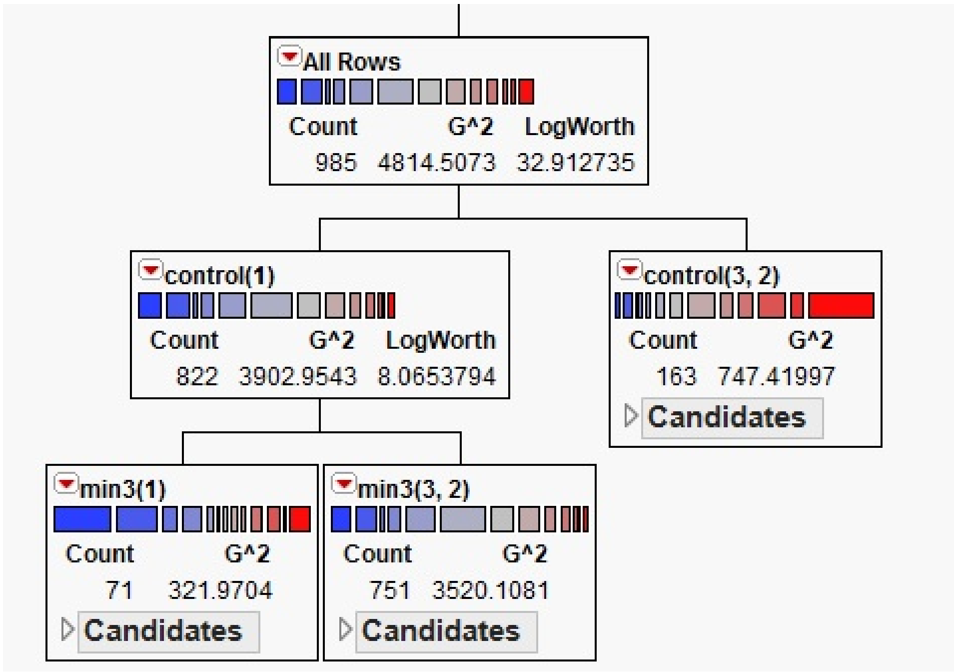


Figure 5: Initial splits in the CHAID analysis for senior high schools

differences (Ash, 2011), details of which are not reported here. The successive difference approach has the advantages of not creating weights equal to zero and of accommodating the school-level sampling in most locations as well as the student-level sampling in Chicago and New York City. Preliminary analyses indicate that the design effects for PT are very large; details will be reported in future papers.

Table 2: Estimated standard errors of eight PT composite scores

Variable	Percentile	Estimate	Std Error	Coeff of Var
IQ composite score	Q1	120.26	1.31	0.0109
	Median	164.61	1.24	0.0075
	Q3	204.91	1.00	0.0049
General academic aptitude composite score	Q1	398.68	2.80	0.0070
	Median	486.35	2.71	0.0056
	Q3	575.37	2.54	0.0044
Verbal composite score	Q1	94.86	0.50	0.0053
	Median	110.54	0.45	0.0041
	Q3	125.16	0.40	0.0032
Quantitative aptitude score	Q1	61.54	0.55	0.0089
	Median	83.30	0.79	0.0095
	Q3	115.69	1.07	0.0092
Mathematics composite score	Q1	47.62	0.42	0.0088
	Median	64.83	0.63	0.0097
	Q3	90.35	0.84	0.0093
Technical aptitude composite score	Q1	30.72	0.27	0.0088
	Median	41.69	0.30	0.0072
	Q3	55.39	0.32	0.0058
Technical information composite score	Q1	18.73	0.18	0.0096
	Median	26.10	0.21	0.0080
	Q3	35.66	0.23	0.0064
Scientific composite score	Q1	362.98	3.41	0.0094
	Median	479.22	3.42	0.0071
	Q3	606.84	3.27	0.0054

6. Discussion

Through a combination of established methodology and statistical detective work, a set of 118 student-level replicate weights were constructed for the PT base year data. Among the complicating factors were the absence of frame information and the inability to reproduce one design-level stratification variable—the student retention ratio. Lack of information about some aspects of the school-level and student-level sampling led to plausible but unverifiable assumptions. These issues notwithstanding, the replicate weights produce sensible and consistent estimates for standard errors and are usable for more detailed analyses of PT data than are reported here.

REFERENCES

American Institutes for Research (last visit: Sep. 9, 2013). Project TALENT Website. Retrieved from http://www.projecttalent.org/

Ash, S. (2011). "Using Successive Difference Replication for Estimating Variances," *JSM 2011 Proceedings, Section on Survey Research Methods*, 3534–3548. Alexandria, VA: American Statistical Association.

Flanagan, J. C., Dailey, J. T., Shaycoft, M. F., Gorham, W. A., Orr, D. B., and Goldberg, I. (1962), *Design for a Study of American Youth*. Boston, MA: Houghton Mifflin Company.

Jones, A. and Huddleston, E. (2009), *SAS/STAT 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

Kass, G. V. (1980), "An Exploratory Technique for Investigating Large Quantities of Categorical Data," *Applied Statistics*, 119–127.

SAS Institute Inc. (2012), *Using JMP 10*. Cary, NC: SAS Institute Inc.

Shaycoft, M. F. (1977), *Project TALENT: A Short History of a L-O-N-G Project*. Palo Alto, CA: American Institutes for Research.

Stapleton, L. M. (2008), "Variance Estimation Using Replication Methods in Structural Equation Modeling with Complex Sample Data," *Structural Equation Modeling*, 15(2), 183–210.

Wise, L. L., McLaughlin, D. H., and Steel, L. (1979), *The Project TALENT Data Bank Handbook*. Palo Alto, California: American Institutes for Research.

Wolter, K. M. (2007), *Introduction to Variance Estimation*. New York, NY: Springer.

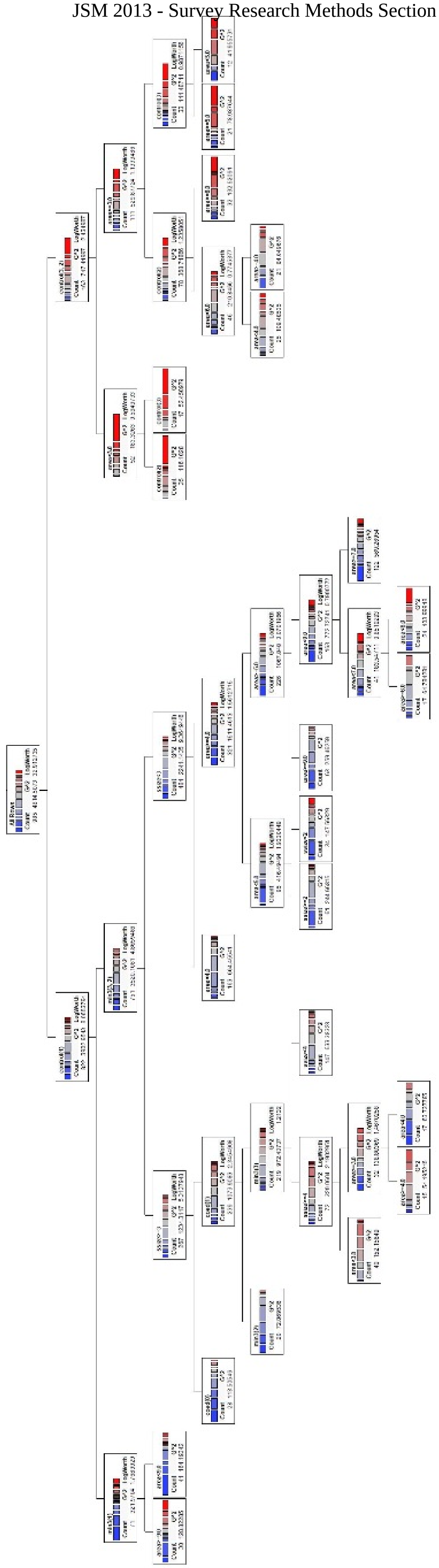


Figure 6: Final partition in the CHAID analysis for senior high schools

Principal Components: on Correlations

Correlations

	BY_C001 IQComp	BY_C002 GAComp	BY_C003 VbComp	BY_C004a QAComp	BY_C004b MaComp	BY_C005a TAComp	BY_C005b TIComp	BY_C006 ScComp
BY_C001 IQComp	1.0000	0.9384	0.8386	0.7790	0.7837	0.6402	0.5874	0.9028
BY_C002 GAComp	0.9384	1.0000	0.9404	0.8691	0.8634	0.6373	0.5939	0.9090
BY_C003 VbComp	0.8386	0.9404	1.0000	0.7258	0.7238	0.5575	0.5182	0.7922
BY_C004a QAComp	0.7790	0.8691	0.7258	1.0000	0.9950	0.6626	0.6318	0.8653
BY_C004b MaComp	0.7837	0.8634	0.7238	0.9950	1.0000	0.6722	0.6423	0.8697
BY_C005a TAComp	0.6402	0.6373	0.5575	0.6626	0.6722	1.0000	0.9758	0.8457
BY_C005b TIComp	0.5874	0.5939	0.5182	0.6318	0.6423	0.9758	1.0000	0.8029
BY_C006 ScComp	0.9028	0.9090	0.7922	0.8653	0.8697	0.8457	0.8029	1.0000

There are 23160 missing values. The correlations are estimated by Pairwise method.

Summary Plots

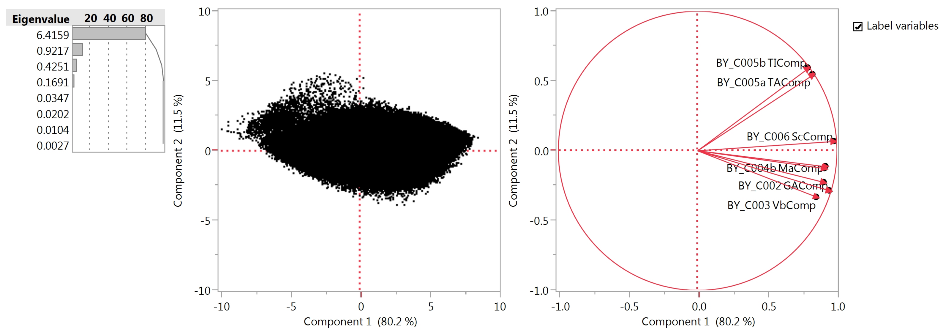


Figure 7: Principal components analysis and correlation matrix for the eight composite scores[2]

[2]'BY_C001IQComp' to 'BY_c006ScComp' represent IQ composite score, general academic aptitude composite score, verbal composite score, quantitative aptitude score, mathematics composite score, technical aptitude composite score, technical information composite score, and scientific composite score respectively.