Methodological Issues in the Sampling Design of the
Third Wave of TEPS*

Tony Tam
Academia Sinica

(First Draft) October 23, 2004

Approx. Word Count: 6,300

* This research was supported by grants from Academia Sinica through the Learning 2000 project, National Academy of Educational Research and National Science Council through the Taiwan Educational Panel Survey (TEPS) project. I thank Yung-Tai Hung, Jing-Shiang Hwang, Lung-An Li, and Sam Peng for detailed comments and other members of the Sampling Design Committee of TEPS for helpful discussion. Any opinions expressed in the technical report series are those of the authors, not those of Academia Sinica and other funding agencies. Direct correspondence to Tony Tam (譚康榮), Academia Sinica, Institute of European and American Studies, Nankang, Taipei, 11529, TAIWAN (email: tam@sinica.edu.tw; tel: (02)3789-7249).

Copyright ©2004 by Tony Tam  All rights reserved.
Methodological Issues in the Sampling Design of the
Third Wave of TEPS

ABSTRACT

This paper identifies the basic methodological issues involved in the sampling design of the third wave of Taiwan Educational Panel Survey (TEPS). Central to the design are three design questions (1) sampling of schools, (2) the sampling of panel members from the second wave of junior high students for two more waves of follow-up, and (3) sampling of classes within school. A systematic discussion of successive hypothetical scenarios contributes to resolving potential confusions and concerns with the sampling design. The motivation for a partially fixed school sample design is discussed and justified. The proposed solutions to the three design issues are tightly integrated and can simultaneously address the need for intercohort tests of institutional and policy impacts, efficient estimation of school effects, avoidance of following up an overly dispersed panel, and maximizing the potential for obtaining class contexts not just for the fresh K11 sample but also for the continuing panel.
Methodological Issues in the Sampling Design of the

Third Wave of TEPS

Taiwan Educational Panel Survey (TEPS) is a multistage stratified sampling survey of Taiwanese high school students (Chang 2003). We take for granted that an appropriate sampling design should be one that can efficiently achieve the main analytic objectives for which the sample is designed. All sampling designs entail costs. The design with the lowest operational cost does not usually have the best efficiency in achieving the analytic objectives. An optimal design is one that strikes a good balance between analytic imperatives and cost concerns. It is assuring to know that the sampling design for the first two waves of TEPS was not only to economize on cost but also to be optimized for uniquely significant analytic objectives such as the multilevel study of the relative roles of school, class, teacher, peer, and family influences on students.

From inception, the sampling committee has made the decision to give up the common practice of maintaining approximately equal probability sampling or probability proportional to size sampling. The payoff is the degree of freedom for achieving multiple practical and analytic goals without complicating the usual procedure and requirement of applying sampling weights. The application of sampling weights in descriptive statistical analysis of TEPS should be routinely done. Weighting is not mandatory for statistical modeling but applying sampling weight is the most conservative practice (DuMouchel and Duncan 1983; Winship and Radbill 1994) and it can be routinely carried out for virtually all common modeling procedures, for instance, in the widely used software Stata without requiring extra programming or adjustment of standard errors (StataCorp 2001, p. 263-265). This paper addresses basic issues of sampling design and weight construction for the third wave of TEPS — after the junior high cohort (grade seven in the fall of 2001 or the 2001-K7 cohort) promotes to year two (grade eleven in the fall of 2005, i.e., becoming the 2005-K11 cohort) in senior high, vocational high, or five-year junior college programs. The central sampling design question is how to draw a new sample of students (the 2005-K11 sample) to be compared with previously sampled students in the same grade level in the 2001 (the 2001-K11 sample).

Before a systematic discussion of the design questions, it is important to discuss and clarify a number of methodological issues that may not be obvious to all readers.

KEY METHODOLOGICAL ISSUES

Irrelevance of Sampling by Nature

Throughout the following discussion of sampling design, we will repeatedly come across issues about the construction of sampling weights. It is useful to state at the outset two
guiding principles in constructing sampling weights. First, the sampling design must consist of a well-defined and replicable procedure of randomly drawing sample from a target population. A well-defined sampling procedure is one that permits the assignment of unambiguous sampling weights appropriate for each case in the sample. Second, properly normalized sampling weights reflect how many cases in the population each case in a sample should represent. The weighting question is determined by the sampling procedure.

The implication of the second guiding principle requires some explanation. To appreciate the implication, it is useful to distinguish between sampling steps introduced by the researcher and sampling processes introduced by nature. (By nature we mean every causal process not subject to the researcher's intervention.) The former is what sampling weights are all about. The latter is what statistical modeling should be concerned with but totally irrelevant to the construction of sampling weights. Improper attention to sampling by nature would only misguide and often disable the construction of sampling weights. To underscore this statement, it is useful to call the Principle of Irrelevance of Sampling by Nature.

We have talked about explicit and implicit sampling steps introduced by the researcher. It is worth giving two specific examples of sampling by nature. One is the distribution of students across geographic areas, the other is the allocation of students across schools. Students have different probabilities of residing in northern or southern Taiwan, depending on where their grandparents grew up, what their parents do, and even on how well they did on senior high school entrance exam. The probability depends on the area and the student. Similarly, students have different probabilities of attending any given school, depending on such factors as the public exam scores and locational preference. The probability also depends on the student and the school. Even with a perfect list of universe of students with their residence area and school at hand, the researcher cannot possibly estimate the probability of any student living in his/her area of residence or attending his/her school.

Sampling by nature is ubiquitous. In fact, the probability of any sampling unit falling into any stratum in a typical stratified sampling design is determined by nature and varies from unit to unit. Fortunately, the ignorance of the probabilities is perfectly fine. It does not handicap the construction of sampling weights at all because these probabilities are determined by nature, not by anything the researcher does. Sampling weights are designed to aid statistical inference based on a statistic from a sample — constructed by the researcher via a reproducible procedure of sampling from a target population — to the corresponding parameter in the target population. Whatever nature does to the distribution of sampling units (such as students) across aggregate units (such as geographic areas, schools), it does not tell us the sampling procedure executed by the researcher. All we have to know for proper inference is how each case is drawn from the target population. To the extent that what nature does would not affect how a case is drawn, it is absolutely irrelevant to the construction of sampling weights. This is the Principle of Irrelevance of Sampling by Nature.
Fixed School Sample Design: Setup A

To facilitate further exposition, it is useful to separately consider three hypothetical and idealized setups (A, B, and C) of the sampling of schools before considering the actual design to be applied to the third wave of TEPS. Each hypothetical setup allows us to discuss and resolve some basic issues without unnecessary complications. Successive scenarios introduce complications and new issues. By the time we get to the TEPS design, the reader will appreciate the nature and solution to various basic issues that would contribute to a proper understanding of the TEPS design.

The first scenario is setup A and it is the simplest:

1. The population of schools in W3 is identical to the 800 schools in W1.
2. The mix of programs offered by, and other stratifying attributes of, each school are absolutely stable.
3. In W1, 300 schools (call it W1-300 henceforth) were sampled using nonproportional stratified random sampling according to (a) program tracks, (b) over two dozen major administrative districts defined by the Executive Yuan as of January 1, 2000, (c) a maximum of four levels of urbanization, and (d) private and public sectors.
4. In W3, 300 schools (call it W3-300) are to be sampled.

If the ideal is to maintain (a) as much an overlap with and (b) as close a replicate of sampling design between W1-300 and W3-300 as possible, setting W3-300 to be the W1-300 set can exactly achieve the ideal. This W3-300 set is by definition 100 percent overlap with the W1 set and its sampling design is identical to the one in W1. Note that the identical sample implies that the sampling weight of each school is identical to the weight for W1 inference to the population. We shall call the use of this ideal set the fixed school sample design (FXD).

A Cautionary Tale on Implicit Sampling

In the computation of sampling weights, it is important to take account of all sampling steps explicitly or implicitly introduced by the researcher. An oversight of any step would lead to erroneous weights and even self-contradictory anomalies.

Consider the following three-part reasoning for constructing sampling weights for W3-300. (P1) First, the ideal set of W3-300 can be regarded as an outcome of the following stratified sampling procedure: (a) Step 1.—Stratify the population of schools for W3 into two strata: stratum S1 composed of the 300 schools from W1, stratum S2 composed of the remaining 500 schools in the population. (b) Step 2.—Sample all schools in S1, sample none from S2. W3-300 is the result. (P2) Second, the sampling weight for S1 is identically one for everyone of the 300 schools, meaning that each of these 300 represents only one school in the population or, more precisely, only representative of stratum S1. (P3) Third, since none is
sampled from stratum S2, this two-stage sampling procedure shows that the sample fails to represent the population as it leaves stratum S2, in fact, 5/8 of the population, totally unrepresented.

How sound is this argument? It implies that the ideal above is anything but ideal — using the identical set of schools from W1 sample results in a grossly unrepresentative sample for the W3 population. However, the argument has self-contradictory implications. On the one hand, it acknowledges that W1-300 is representative of the W1 population, hence also acknowledging that W3-300 (being identical to W1-300) is representative of the W1 population. On the other hand, it asserts that W3-300 is grossly unrepresentative of the W3 population, even though the W1 and W3 populations are simply identical. In other words, the argument implies the following:

(A1) W1 population $\equiv$ W3 population
(A2) W1-300 $\equiv$ W3-300
(A3) Sampling weights for W3-300 $\neq$ those for W1-300
(A4) W1-300 is representative of W1 population
(A5) W3-300 is grossly unrepresentative of W3 population

A1-A5 cannot be all true, thus the argument is self-contradictory.

The error in the argument is due to the oversight of the second part (P2) of the argument. It ignores the fact that in order to sort the population into two strata in step 2, the researcher must implicitly undertake a random sampling procedure that should have been taken into account in defining sampling weight. Imputing an equal weight of one to every one of the W3-300 schools erroneously ignores the implicit sampling step before one can sort the W3 population into strata S1 and S2. Explicit account of the implicit sampling step would require using the W1 sampling weight for each of the 300 schools, implying that each school in fact represents more than one school in the population and vindicate that W3-300 is representative of the entire W3 population. Thus the representativeness of W3-300 for the W3 population is identical to that of W1-300 for the W1 population, and hence there is no self contradiction. Schematically, the argument A1-A5 should be replaced by B1-B5:

(B1) W1 population $\equiv$ W3 population.
(B2) W1-300 $\equiv$ W3-300
(B3) Sampling weights for W3-300 $=$ those for W1-300
(B4) W1-300 is representative of W1 population
(B5) W3-300 is representative of W3 population.

By now it should be clear that there is nothing wrong with the two-step stratified sampling conception of the W3-300 sample. What is wrong is the imputation of sampling weights in P2 of the argument that ignores the implicit sampling step, leading to self contradictory implications. To be exact, the two-step stratified sampling conception should have been formulated as a three-step conception, the first step of which is exactly the
stratified sampling procedure for W1-300. Viewed in this way, the three-step conception adds nothing but two redundant and distracting steps.

**Partially Fixed School Sample Design: Setup B**

The FXD in setup A does have common parallels in the sampling design of many other large-scale surveys. Consider, for instance, the survey design of many micro surveys conducted by the government in Taiwan. Local geographical sampling units are usually fixed for an extended period of time and across different surveys. Only sampling within geographical units is conducted afresh for different surveys. These geographical units are analogous to the schools in the FXD.

The FXD can be easily modified, as is done in two of the cross-sectional labor surveys most well-known to Taiwanese economists — the Manpower Utilization Surveys (MUS) here in Taiwan (Yu 2002; Lin and Chang 2003) and the Current Population Surveys (CPS) in the U.S. (Angrist and Krueger 1999) The MUS and CPS retain a fixed proportion of their sampled households from wave to wave. The households there are analogous to the schools here, only that a fraction of the households will be replaced in consecutive waves.

Thus the second setup introduces a partially fixed school sample design, similar to what happens in the MUS and CPS. Instead of adopting the entire W1-300 sample, it adopts only a subset (N1) of the W1-300 schools is retained for W3. A new set of schools (N2) outside the W1-300 sample is added. For specificity, consider the following scenario:

1. The population of schools in W3 is identical to the 800 schools in W1.
2. The mix of programs offered by, and other stratifying attributes of, each school are absolutely stable.
3. In W1, 300 schools (call it W1-300 henceforth) were sampled using nonproportional stratified random sampling according to (a) program track, (b) over two dozen major administrative districts defined by the Executive Yuan as of January 1, 2000, (c) a maximum of four levels of urbanization, (d) private-public sector.
4. In W3, 250 of the 300 schools (call it W1-250) are to be sampled and 50 new cases are drawn from the 500 schools outside W1-300 via the same stratified sampling procedure in the following way: (a) randomly select 25 strata out of a total of 40 school strata; (b) within each stratum for the W1-300 sample, each unit has been assigned a random number from a generator for the uniform distribution; (c) all units are sorted within stratum according to the random numbers, which are used to determine their priority order of being sampled; (d) for example, if stratum S4 of W1-300 the sampling procedure stopped at the 5th unit, now draw the 6th and 7th units and drop the 4th and 5th units, hence replacing two of the cases in W1-300 without altering the sampling design; (e) repeat the procedure for all 25 strata.
If all strata have at least two schools sampled and a minimum of two schools not in W1-300, the sampling weights for W1-250 are simply their weights in W1-300. The sampling weights for the new 50 years should be the same as those they replace within their strata. In general, however, this may not work out so neatly and the number of schools from a stratum may be different for W1 and W3. In any event, the adjustment of weights involves no more than going back to the W1 stratification scheme and recalculate the sampling proportions wherever appropriate.

In the specific setup above, N1+N2=300. In fact, the partially FXD can be modified so that the number of new schools is larger, or smaller, than the number dropped from the W1-300 sample. This minor extension does not introduce any new issue. It does necessitate the recalculation of all sampling weights.

A Shifting Target Population: Setup C

We now introduce yet another complication to the previous setup: instead of an absolutely stable population, the W3 population is allowed to be different than the W1 population. Table 1 provides a schematic summary of setups A, B, and C. Again, to fix ideas, consider the following scenario for the third setup:

1. The population of schools in W3 is (a) the 800 schools in W1 plus (b) 100 new schools. We call these strata S800 and S100, respectively.
2. The mix of programs offered by, and other stratifying attributes of, each of the 800 W1 schools are absolutely stable.
3. In W1, 300 schools (call it W1-300 henceforth) were sampled using nonproportional stratified random sampling according to (a) program track, (b) over two dozen major administrative districts defined by the Executive Yuan as of January 1, 2000, (c) a maximum of four levels of urbanization, (d) private-public sector.
4. In W3, 250 of the 300 schools (call it W1-250) are to be sampled and 50 new cases are drawn from the 500 schools outside W1-300 via the same stratified sampling procedure in the following way: (a) randomly select 25 strata out of a total of 40 school strata; (b) within each stratum for the W1-300 sample, each unit has been assigned a random number from a generator for the uniform distribution; (c) all units are sorted within stratum according to the random numbers, which are used to determine their priority order of being sampled; (d) for example, if stratum S4 of W1-300 the sampling procedure stopped at the 5th unit, now draw the 6th and 7th units and drop the 4th and 5th units, hence replacing two of the cases in W1-300 without altering the sampling design; (e) repeat the procedure for all 25 strata.
5. In addition, 40 schools are drawn via a separate stratified random sampling procedure from stratum S100.
Although the change in the population (parts 1 and 5) may appear to be a major complication, in fact it is not. Part 4 of this scenario is clearly parallel to setup B and the implication for the construction of sampling weights has been resolved. What is new here is part 5. Yet the implication for sampling weight is also standard, as in any stratified sampling design.

An important extension of the scenario is to allow S800 to shrink, that is, some of S800 disappear in W3. For specificity, assume that 50 of S800 have disappeared from the population, leaving only 750 of the original schools in the W3 population and, say, 20 sample schools are in the dropout list. What does this complication do to the computation of sampling weights? As it turns out, the complication is already anticipated in setup B. There is no new issue raised here.

Moreover, the complication is exactly analogous to the common situation in which a survey has to go through post-sampling adjustment of sampling weights after discovering from field work that the sampling frame has changed just before interviews are actually conducted or after interviews have been completed and a substantial refusal rate was recorded. In the present context, we have to go back to the stratified sampling scheme to adjust the sampling proportion for each stratum in light of the cases lost from S800.

We now turn to the case of TEPS and the central questions of sampling design for third wave of TEPS.

**SAMPLING DESIGN FOR THE THIRD WAVE OF TEPS**

To obtain the new 2005-K11 sample of students for the third wave (W3) of TEPS for (1) intercohort comparative analysis and (2) a follow up of wave one (W1 conducted in 2001) TEPS students, we have to address three major questions. The first one should be considered without regard to following up on any TEPS students from W1, the second and third are in different ways related to W1 panel students. The questions are: (I) sampling of W3 schools, (II) subsampling of W1 and W2 TEPS junior high students for W3 and W4 follow-up survey, and (III) sampling of W3 classes to obtain a fresh new sample of K11 students in 2005. The sampling of schools is crucial first step because it is a necessary step toward obtaining the sampling weights for classes and students within school.

**The Setting**

TEPS is entering the third wave. In the first two waves, two cohorts of students were the target populations. One is the 2001-K11 cohort consisting of students entering year two of senior high, vocational high, or five-year junior college program in the fall of 2001. The other is the 2001-K7 cohort consisting of students entering year one of junior high in the fall of 2001. The 2001-K7 cohort was chosen to be surveyed again in the fall of 2005 to enable rigorous intercohort comparisons. The long-term nature of TEPS entails certain imperatives
on all waves in general and the third wave in particular. The context of data collection is also in flux, entailing some design complications specific to the third wave. This subsection will briefly review the imperatives and the changing context before addressing the three main design issues mentioned at the start.

**Imperatives**

As a long-term survey intended to provide opportunities for evaluating the institutional and policy impacts on high school students, the design of the third wave of TEPS must meet certain constraints on collecting data that would permit (1) intercohort comparison of learning experience under different institutions and policies and (2) long-term panel analysis of learning over varying class and school contexts. Simply put, there are four

1. All of the fundamental design features of the 2001-K11 sample should be maximally replicated in the 2005-K11 sample to avoid unnecessary comparability problems — including the way multistage stratified sampling is done and, most significantly, sampling of classes within school and multiple students within class.

2. School is a critically important source of institutional influences on student learning, and it is a key component in the organization of survey operation. We need an efficient sampling design for the estimation of school effects and lay out the proper computation of the sampling weights for schools. Hence it would be ideal to retain as many W1 schools as possible to provide the maximal potential for measuring fixed-effects of schools on student learning.

3. As many of the 2001-K7 sample should be followed up in the third wave.

4. W3 questionnaire items must be sufficiently comparable with those in W1 to permit intercohort comparative analysis.

**Changing Contexts**

Even if all four imperatives can be met, we cannot control the context in which the sampling is to be conducted. Not only the institutional and policy contexts affecting K11 students are changing, the population of high schools for K11 is shifting as well.

It is worth noting that the sampling frame for the 2001-K11 sample was based on the latest Ministry of Education data on the program offerings by every high school in Taiwan. In order to start data collection in the fall of 2001, TEPS had to complete sampling, conduct pilot testing, and obtain authorization to survey a school well ahead of September 2001. At the time of the first sampling operation, the Ministry of Education could only provide information up to the academic year of 1998-1999. In other words, the target population for the 2001-K11 sample consists of schools on record in the Ministry of Education database of schools and program offerings for the academic year 1998-1999, not 2001-2002 even though the field work was conducted in the fall of 2001. Similarly, at the time of the follow-up field
operation (fall of 2004) in preparation for the third wave, the latest sampling frame of high schools available is for academic year **2003-2004 (Min92)**. This list provides the sampling frame for the third wave schools. It implies that only schools offering K10 classes as of academic year 2003-2004 would be eligible for survey at the end of year 2005. Post-survey sampling weights will take into account the possibility that some of those eligible schools may drop out of the population targeted by the survey.

A careful comparison of the two sampling frames identifies a substantial number of schools offering new programs, especially senior high or comprehensive program, and a substantial number of schools terminating old programs, notably vocational high or five-year junior college program. Some schools were closed since we completed our W2 survey. Some continuing schools simply closed the program for which TEPS sampled the schools. A few schools changed from privately owned to publicly own. This feature of an unstable target population of schools is exactly what distinguishes setup C from the other two setups. Table 2 presents the differences between the two sampling frames, the preliminary results from the 2004 tracking survey of the 2001-K7 sample, and projections for two opposite sampling designs for schools (for simplicity, without breakdown into the four program tracks). It shows that it is highly likely that the FXD would capture at least 7,000 of the nearly 20,000 TEPS junior high sample (2001-K7), a finding of key interest when we address issues II and III of the sampling design.

**Issue I. Sampling of Schools**

Given the imperatives for the third wave of TEPS, the key decision is to decide between two extreme options: either adopt the FXD or draw a fresh new sample of schools following as closely as feasible the sampling procedure for the 2001-K11 sample.

Since the target populations of the W1 and W3 surveys are partly disjoint, the discussion of setup C in the previous section is relevant to the problem here. The discussion shows that, even when the FXD or a partially FXD is desired, the shifting population necessitates the selection of a supplementary sample of schools to make sure that the new W3 school sample is representative of the entire target population. The other necessary adjustment is the sampling weights for all schools from the 2001-K11 sample. But these adjustments are standard, not involving any new methodological issue.

The issue then boils down to the merits and demerits of the FXD relative to the fresh school sample design. Why is so good about this FXD? Surveying the same schools obviously economize on the costs of learning about a new school and building working relationship with its staff. In addition to savings on operational costs, there are two major virtues related to the estimation of school effects.
The first virtue is about the efficient estimation of observed school attributes while the second virtue is about the efficient estimation of all school attributes stable from W1 to W3 and possible W4 using the maximum number of cases from the two cohorts of students. To understand the first virtue of the FXD, it may be instructive to draw a comparison with the classical statistical analysis of linear regression. Assuming everything else is held the same, it is well-known that estimation efficiency is highest if X is fixed throughout the repeat sampling process in which only the unobserved heterogeneity variable (usually called errors) is drawn afresh from a distribution. If X is not fixed but repeatedly sampled together with the errors, the sampling distribution of the estimator of coefficients will have larger variance than if X is fixed. Hence estimation is less efficient in this case than in the case where X is fixed in the repeat sampling process. Similarly, when the schools are fixed from W1 to W3, we maintain maximal efficient.

The second virtue is related to the control for school fixed effects, as in a study of student level processes in which the researcher needs to partial out ALL stable school-level influences — whether or not measured by TEPS in any way. The more overlap between the W1 and W3 schools, the more students would be available to the fixed-effects analysis. All students in W1-only or W3-only schools cannot possible contribute any information for school fixed-effects and are necessarily left out of the fixed-effects analysis. (For simplicity and illustration, consider the numbers specified in setup A, a 100-percent overlap sample of W3 schools means all students and schools contribute to the fixed-effects analysis. By contrast, if W1 and W3 are independent samples of the population, the expected number of schools co-present in the W1 and W3 samples is $800 \times (3/8) \times (3/8) = 112.5$, many fewer than the 300 ideal.) Notice that if school fixed effects exert powerful influences on a student outcome, so that student outcomes within school are highly correlated or homogeneous, there may be little residual variance available for the statistical identification of any effects outside the school fixed effects (see, e.g., Griliches [1986] for a generic econometric exposition and Tam [1997] for the problem in a major substantive literature). The penalty of losing nearly two-third of the schools can be very high. The loss not only threatens the representativeness of the analytic sample, but also may aggravate the unreliability of statistical estimation in a fixed-effects analysis.

In sum, there are large and unique benefits to adopting the FXD of setup A. When the FXD is not feasible due to, e.g., shifting target population, a partially FXD such as the one specified in setup C would still be superior to the fresh school sample design. As a result, TEPS will adopt a partially fixed school sample design similar to the one in setup C.

Issue II. Sampling of Existing TEPS Panel

While the first design issue focuses on the objective of intercohort comparative analysis, another key objective of TEPS is to provide a long-term panel of learning experience in high
school. To provide a long-term panel, it is imperative that TEPS follow up as many of the 2001-K7 sample as is feasible. With an initial sample size of nearly 20,000 students in the 2001-K7 sample, the odds is heavily in favor of the bet that the continuing panel of students would be dispersed into practically every school in the target population of schools with K11 programs, and within school these students are likely to be distributed widely over the universe of classes in K11. The practical implication of following up every case would mean surveying virtually every school with a K11 program. The cost would be much higher than the budget limit TEPS has been working with, and it would be much less cost effective than the operation in the first two waves because the number of students available per school will go down, while the number of classes and teachers involved will go much higher. In short, the cost consequence of following the entire panel is prohibitively high; TEPS must subsample the 2001-K7 panel for long-term follow-up.

It is obviously simple to randomly subsample the TEPS panel. Two obviously natural ways to do so would be to subsample a pre-set proportion of the panel with equal probability for everyone. More efficiently and cost effective would be to stratify the panel members according to measured attributes, such as family background and degree of urbanization residence area, and then select pre-set proportions of each stratum. In any case, however, this kind of procedure would not prevent the problem of going all across Taiwan to many K11 schools for follow-up survey of the panel.

Recall that we proposed a partially fixed school sample design to address the first design issue. The same proposal suggests itself as the natural solution for this design issue as well. Is it not natural to subsample the panel members first at the level of K11 schools? This school-based subsampling automatically resolves the practical problem of an overly dispersed panel sample. There is also the invaluable convenience of piggybacking on the sampling of schools for the 2005-K11 cohort. We have discussed the important virtues of the partially FXD for the 2005-K11 sample. We shall not repeat them here. Indeed we decide to adopt the partially FXD for subsampling the 2001-K7 panel and we set the school sample for this purpose to be identical to the school sample for the fresh 2005-K11 sample.

The construction of sampling weights is standard. Given the chosen design, the sampling weight for each panel member in W3 would be based on the product of two probabilities: the sampling probability \(p_1\) of a member in W1 and the conditional sampling probability \(q_3\) of the member in W3. The reciprocal of the product \(p_3 = p_1 \times q_3\) will be the appropriate the sampling weight for the member.

**Issue III. Sampling of Classes**

After choosing the design for the sampling of W3 schools for the 2005-K11 sample and the subsampling of the 2001-K7 panel, there remains the problem of sampling classes within schools.
The source of the problem is related to the fact that many of the 2001-K7 panel will be in the roughly 300 W3 schools across four program tracks. But they are randomly distributed across classes. As stated earlier, there are at least 7,000 TEPS panel members who happen to be continuing their schooling in the W3 TEPS schools. On average, each W3 school will capture over 23 TEPS students. These students are likely to be distributed over most of the classes rather than, say, concentrated in one-third of the classes in K11. In addition, TEPS will follow up all or a random subset of the 2001-K7 panel members in W3 TEPS schools but, contrary to the design in W1 and W2, will not sample any classmates of the panel to provide class contexts. Sampling classmates are costly, the number of classes needed may be much larger than we can handle, and so greatly limit the size of panel TEPS can feasibly follow.

Fortunately, parallel to the TEPS panel entering K11 is a fresh sample of the 2001-K7 cohort who are now in K11 — the 2005-K11 sample of TEPS. We can have one stone for two birds if we draw a new sample of classes for the 2005-K11 sample by oversampling classes with the 2001-K7 panel members, especially classes with multiple panel members.

The natural question is of course whether this efficient oversampling is statistically legitimate. The oversampling is perfectly legitimate because the procedure is equivalent to the standard use of stratified sampling. As far as sampling is concerned, identifying a student as member of TEPS and classifying the student into a stratum for TEPS panels does not implicitly introduce any sampling. Moreover, whether a student is a member of the TEPS panel is uncorrelated with any characteristic of the classes, therefore the population of classes containing any panel member or not should be indistinguishable from each other. This equivalence renders the stratified sampling procedure as efficient as one without this stratification.

Within class, further sampling may or may not apply stratification of students who are members of the TEPS panel or not. With stratification, however, we can sample all TEPS panel members, i.e., with probability one and less than one for non-panel members. This two-stage stratified sampling within school can maximize the overlap between the two samples, the 2005-K11 sample and the 2001-K7 panel sample, while retaining their respective sampling designs. The panel members in these classes can be used in two ways: (1) as part of the 2005-K11 sample and (2) as part of the 2001-K7 panel. Teacher evaluations for the subsample of TEPS panel can be collected simply as part of the 2005-K11 sample.

**Conclusion**

Taken together, the sampling design for the third wave of TEPS is optimal for simultaneously addressing the need for intercohort tests of institutional and policy impacts, efficient estimation of school effects, subsampling of the TEPS panel from junior high, and maximizing the potential for obtaining class contexts not just for the fresh K11 sample but also for the TEPS panel. The simultaneous achievement of these goals is a unique
accomplishment of the proposed design for the third wave of TEPS. We are not aware of similar accomplishments in a large national panel survey of high school (or primary school) students anywhere in the world. If properly exploited, the distinctive design features discussed in this paper will bring value-added to future research based on TEPS.

REFERENCES
Table 1. Summary of Three Hypothetical Setups

<table>
<thead>
<tr>
<th>Setup</th>
<th>W1 &amp; W3 Populations of Schools</th>
<th>W3 Sample of Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>{W3-800} = {W1-800}</td>
<td>{W3-300} = {W1-300}</td>
</tr>
<tr>
<td>B</td>
<td>{W3-800} = {W1-800}</td>
<td>{W3-300} = {W1-250} + {50}</td>
</tr>
<tr>
<td>C</td>
<td>{W3-900} &gt; {W1-800}</td>
<td>{W3-340} = {W3-300} of setup B + {40 out of 100 new schools}</td>
</tr>
<tr>
<td></td>
<td>i.e. added 100 new schools</td>
<td>{W3-320} = {280 of W3-300 of setup B} + {40 out of 100 new schools}</td>
</tr>
<tr>
<td></td>
<td>{W3-850} \neq {W1-800}</td>
<td>i.e. W1-800 left with 750</td>
</tr>
</tbody>
</table>

Table 2. Comparing the Sampling Frames of Wave One and Wave Three

<table>
<thead>
<tr>
<th></th>
<th>Total # of Schools</th>
<th>W1 only</th>
<th>W1 &amp; W3</th>
<th>W3 only</th>
<th>Total # of Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>\text{W1}</td>
<td>\text{W3}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior-high</td>
<td>275</td>
<td>21</td>
<td>254</td>
<td>74</td>
<td>328</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>61</td>
<td>3</td>
<td>58</td>
<td>105</td>
<td>163</td>
</tr>
<tr>
<td>Vocational</td>
<td>301</td>
<td>47</td>
<td>254</td>
<td>5</td>
<td>259</td>
</tr>
<tr>
<td>Five-year junior college</td>
<td>69</td>
<td>7</td>
<td>62</td>
<td>11</td>
<td>73</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>706</strong></td>
<td><strong>78</strong></td>
<td><strong>628</strong></td>
<td><strong>195</strong></td>
<td><strong>823</strong></td>
</tr>
</tbody>
</table>

Counts as if October 23, 2004