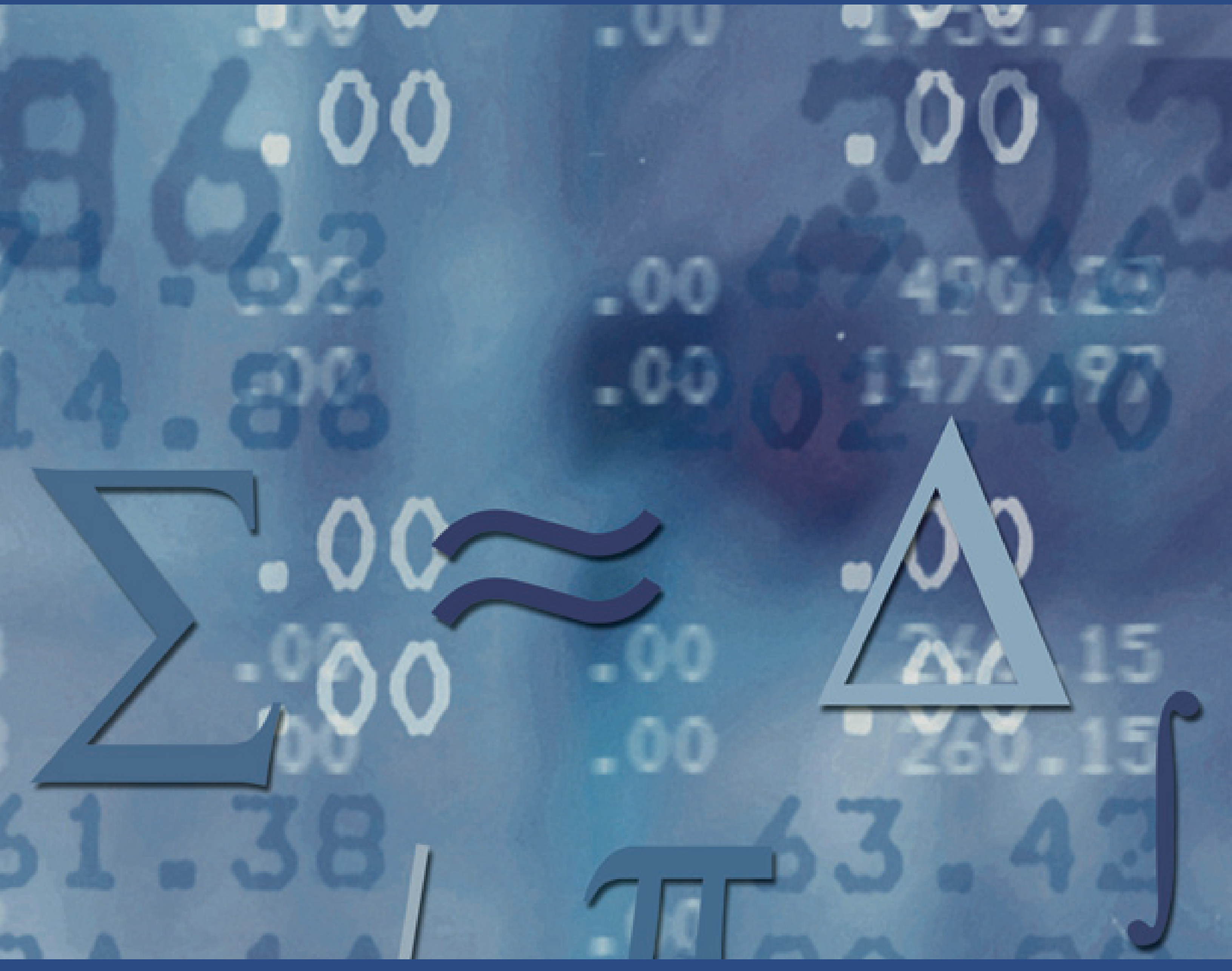
ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

# National Institute of Statistical Sciences Configuration and Data Integration for Longitudinal Studies Technical Panel

Final Report



NCES 2011-607

U.S. DEPARTMENT OF EDUCATION

# National Institute of Statistical Sciences Configuration and Data Integration for Longitudinal Studies Technical Panel

Final Report

**MARCH 2011**

**Alan F. Karr**
National Institute of Statistical Sciences

NCES 2011-607
**U.S. DEPARTMENT OF EDUCATION**

NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

**U.S. Department of Education**
Arne Duncan
*Secretary*

**Institute of Education Sciences**
John Q. Easton
*Director*

**National Center for Education Statistics**
Jack Buckley
*Commissioner*

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

> NCES, IES, U.S. Department of Education
> 1990 K Street NW
> Washington, DC 20006-5651

March 2011

The NCES Home Page address is http://nces.ed.gov.
The NCES Publications and Products address is http://nces.ed.gov/pubsearch.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES World Wide Web Publications and Products address shown above.

This report was prepared for the National Center for Education Statistics under Contract No. ED-05-CO-0044 with the American Institutes for Research. The content of this report does not necessarily reflect the views or policies of the U.S. Department of Education. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

**Suggested Citation**
Karr, A.F. (2011). N*ational Institute of Statistical Sciences Configuration and Data Integration Technical Panel: Final Report* (NCES 2011-607). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from http://nces.ed.gov/pubsearch.

**Content Contact**
Andrew A. White
(202) 502-7472
andrew.white@ed.gov

## Executive Summary

This is the final report of the National Institute of Statistical Sciences (NISS) Technical Panel on Configuration and Data Integration for Longitudinal Studies (hereafter, CDI).

The principal recommendations regarding configuration are as follows:

1. The National Center for Education Statistics (NCES) should configure grades K-12 studies as a series of three studies: (1) a grades K-5 study, followed immediately by (2) a grades 6-8 study, followed immediately by (3) a grades 9-12 study. One round of such studies, ignoring postsecondary follow-up to the grades 9-12 study, requires 13 years to complete.
2. Budget permitting, NCES should initiate a new round of grades K-12 studies every 10 years. This can be done in a way that minimizes the number of years in which multiple major assessments occur.

The technical panel finds that there is no universal strategy by means of which NCES can institutionalize data integration across studies. One strategy was examined in detail: continuation of students from one study to the next. Based on experiments conducted by NISS, the technical panel finds as follows:

- The case for continuation on the basis that it supports cross-study statistical inference is weak. Use of high-quality retrospective data that are either currently available or are likely to be available in the future can accomplish nearly as much at lower cost.
- Continuation is problematic in at least two other senses. First, principled methods for constructing weights may not exist. Second, no matter how much NCES might advise to the contrary, researchers are likely to attempt what is likely to be invalid or uninformative inference on the basis of continuation cases alone.

The technical panel stops short of a categorical recommendation against continuation. If the continuation group was a representative sample, then it might provide meaningful results, albeit with large variability. The technical panel urges that, as an alternative means of addressing specific issues that cross studies, NCES consider the expense and benefit of small, targeted studies that target specific components of students’ trajectories.

The technical panel was not charged to examine in detail the articulation between grades K-12 and postsecondary studies. It does, however, note that the current once-every-4-years frequency of the National Postsecondary Student Aid Study (NPSAS) is not congruent with the 10-year cycle of recommendation 1. By contrast, a 5-year frequency for NPSAS would allow every other NPSAS to follow immediately after a grades 9-12 study.

This page intentionally left blank.

## Contents

**Page**

## List of Tables

**Table** **Page**

## List of Figures

**Figure** **Page**

This page intentionally left blank.

# 1 Technical Panel Membership and Charge

Members of the Technical Panel on Configuration and Data Integration for Longitudinal Studies (CDI) were Susan Ahmed (Mathematica Policy Research), James Chromy (RTI International), Lyle Jones (University of North Carolina at Chapel Hill), Alan Karr (National Institute of Statistical Sciences [NISS]; chair), Jennifer Madans (National Center for Health Statistics), and Jerome Reiter (Duke University). Andrew White was National Center for Education Statistics (NCES) liaison. NISS postdoctoral fellow Satkartar Kinney conceived, designed, and performed the experiments described in section 4.

The technical panel was asked by NCES to address two related issues:

1. how NCES can configure the timing of its longitudinal studies (e.g., Early Childhood Longitudinal Study [ECLS], Education Longitudinal Study [ELS], High School Longitudinal Study of 2009 [HSLS:09]) in a maximally efficient and informative manner; and, since the main but not sole focus was at the primary and secondary levels,
2. what NCES can do to support data integration for statistical and policy analyses that cross breakpoints between longitudinal studies.

The issues intersect in the obvious sense that some configurations are more supportive of data integration than others. Subtle problems lie at the edges: for instance, "continuation" years of studies needed to accommodate students moving at slower paces may aid in data integration. NCES was particularly interested in the roles of synthetic cohorts and record linkage in data integration.

The technical panel met in person in Washington on September 27-28, 2007, and conducted the remainder of its activities by e-mail. The agenda for that meeting appears in appendix B.

## 1.1 Background

A summary of longitudinal studies considered by the technical panel is contained in appendix A. A longitudinal study is defined by

- a calendar year of initiation, N;
- a grade cohort, G: the study cohort is, with weights, a nationally representative sample of students who enter grade G in year N; and
- a design duration, D, the nominal number of years for which the cohort is followed.

The K-12 studies and the Beginning Postsecondary Students Longitudinal Study (BPS) fit this model.

## 1.2 The visual metaphor

The technical panel considers one of its contributions to be a robust way of thinking about configuration and data integration problems over the long term, as well as presenting visual metaphors for representing them. One such metaphor is shown in figure 1, where the horizontal coordinate represents calendar time and the vertical coordinate represents data subjects—students.

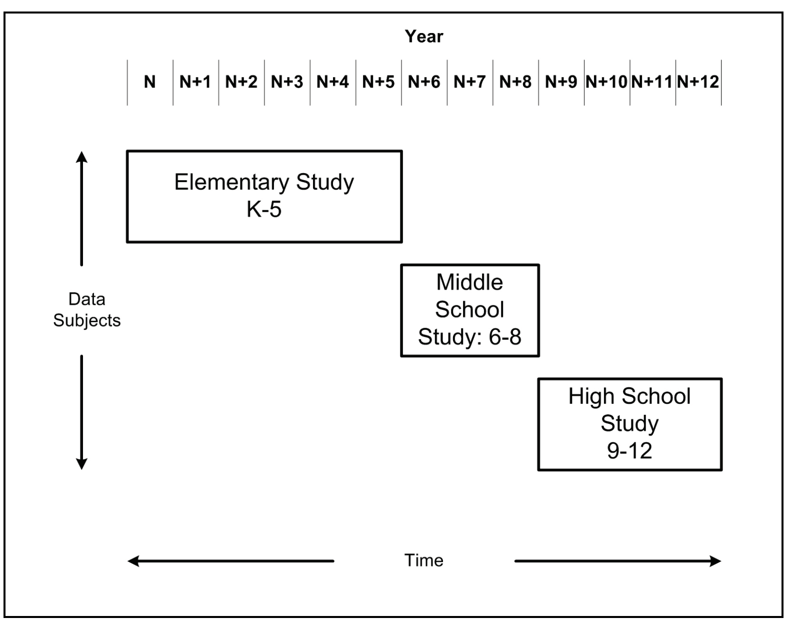
A single longitudinal study is then a rectangle, which represents the data for a particular set of subjects followed for a certain span of time, starting in a certain year, when they enter a certain grade. Figure 1 shows three studies:

- an elementary study, spanning years N through N+5, of students entering kindergarten in year N;
- a middle school study, in years N+6 through N+8, of students entering grade 6 in year N+6; and
- a high school study, in years N+9 through N+12, of students entering grade 9 in year N+9.

It is important to note that figure 1 does not depict temporal intrastudy changes to the set of data subjects. In reality, of course, there are multiple changes, such as controlled attrition, uncontrolled attrition, and controlled addition, or "freshening." Moreover, figure 1 does not depict continuation of a study to accommodate students who proceed at a slower than nominal pace. Nor does figure 1 represent weights.

Figure 1 also depicts two characteristics that are not logical necessities. First, it shows studies covering contiguous grade blocks as abutting in time. Therefore, except for students' differing paces of progress through grades, all three studies treat the same grade cohort—students born in year N-5. The importance of this is discussed in sections 2.2 and 4. Second, by design no student participates in more than one of the studies.

**Figure 1. CDI technical panel visual metaphor for longitudinal studies. The horizontal coordinate represents time, and the vertical coordinate represents data subjects.**



# 2 Configuration

In the context of the National Institute of Statistical Sciences (NISS) Configuration and Data Integration for Longitudinal Studies (CDI), the “c” component of CDI addresses the question of how multiple longitudinal studies spanning a set of grades should be configured.

## 2.1 What is a configuration?

The technical panel identified three dimensions of configuration:

- **Partition:** How many studies are there, and which grades are the breakpoints between them? In figure 1, there are three studies spanning K-12, and the partition is [K-5|6-8 |9-12], corresponding to elementary, middle school, and high school.
- **Alignment:** Given the partition, how do studies align in time? Figure 1 shows a “seamless” alignment with neither gaps nor overlaps between studies.
- **Frequency:** How frequently does a new set of studies need to be initiated? No frequency is shown in figure 1.

In a literal sense, therefore, configuration addresses only the horizontal component in figure 1.

## 2.2 Configuration recommendations for K-12 studies

These recommendations deal with the 13-year grade span from kindergarten through 12th grade.

- **Alignment:** The technical panel recommends that regardless of partition and frequency, K-12 studies *be aligned temporally as in figure 1, without overlaps or gaps*.

  *Rationale*: Alignment creates significant potential benefits in terms of data integration. In particular, successive studies that abut directly pertain to the same target population, except for temporal changes in that population[1]. This at least makes data integration conceptually consistent. A practical consequence is discussed in section 4.3.

- **Frequency:** The technical panel finds that frequency is independent of partition and alignment, and anticipates that NCES decisions regarding frequency will be based on multiple considerations, including budget and policy needs. Consistent with plans presented to it by NCES, the panel recommends a *10-year frequency*.

  *Rationale:* The technical panel believes that a frequency of less than “one cycle each 10 years” raises the risk of NCES losing touch with important developments in, and important components of, the state of the K-12 educational system.

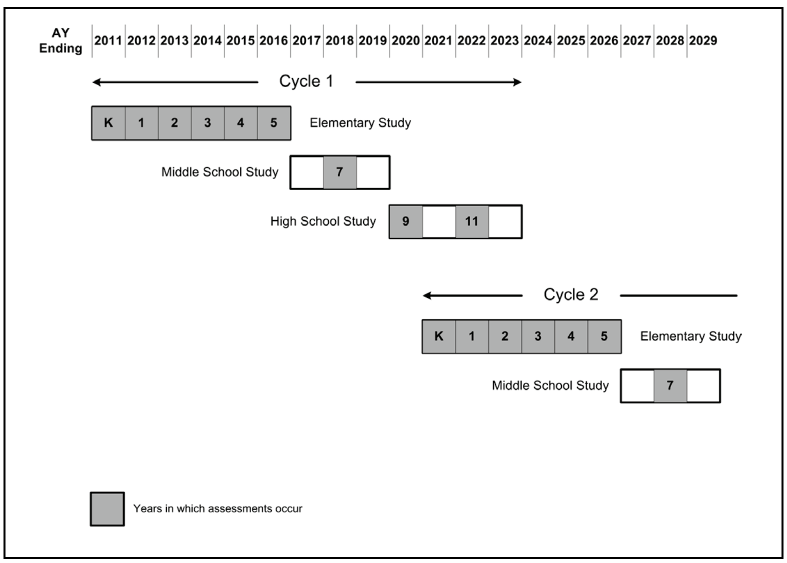
[1]These changes may result, for instance, from immigration or emigration and from grade retentions and advancements.

- **Partition:** The technical panel recommends that NCES should employ the partition shown in figure 2, with *three studies* treating K-5 (elementary), 6-8 (middle school), and 9-12 (high school).

  *Rationale:* In the past, NCES has, in effect, employed the partition shown in figure 3—two studies treating K-8 and 9-12, respectively. The recommended partition opts for higher within-study quality and a clearly representative middle school sample, at the expense of losing full K-8 longitudinal data. As discussed further in section 2.3, it is likely to be less costly. The recommended partition is consistent with some plans shared with the technical panel by NCES.

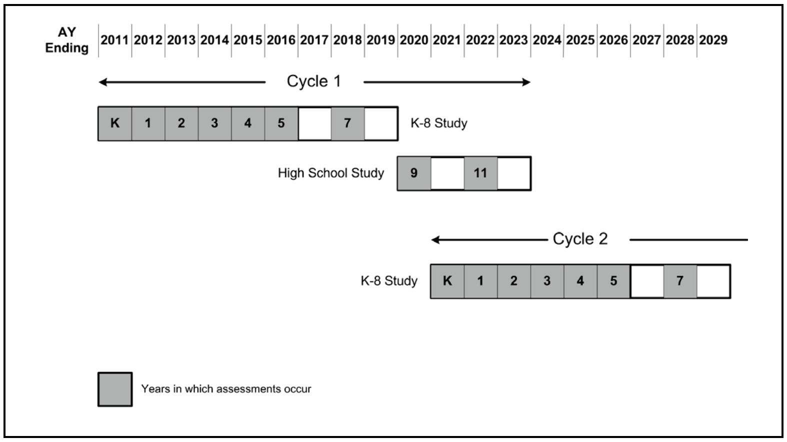
In figures 2 and 3, calendar years in which major assessments would occur are shaded[2] because of the budgetary implications in years with multiple assessments. For both partitions, only once in every 13 years do two assessments occur.

**Figure 2. Technical panel-recommended partition of [K-5|6-8|9-12], shown with a 10-year frequency starting in AY 2011. Years in which major assessments occur are shaded.**




[2]These are derived from information provided to the technical panel by NCES personnel.

**Figure 3. Partition of [K-8|9-12] with 10-year frequency, starting in AY 2011. Years in which major assessments occur are shaded.**




## 2.3 Elaboration

The technical panel believes that NCES should view selection of a configuration as a decision problem in which the *fundamental tradeoff is between cost and data quality*. In the visual metaphor of figure 1, the cost of a study is related to the area of its associated rectangle. Although more formal modeling of costs is possible, as described in section 3.2, these statements apply:

- It is reasonable that the cost of a study is a linear function of (i.e., is directly proportional to) the number of data subjects—the height of the rectangle.
- Cost is likely to be a discontinuous, convex function[3], increasing more rapidly than linearly, of the width of the rectangle—the grade span of the study. Reasons include freshening, "structural" changes of school (e.g., from primary to middle) that affect all students, and move-based changes of school that affect only some students[4].

Data quality is more nebulous, in part because inference-based measures are lacking. It does seem credible that quality is an increasing function, probably concave[5], of both the number of data subjects[6] and the grade span of the study.

Data quality is also strongly user dependent. To some users, having full K-8 or even K-12 student trajectories, no matter how few or unrepresentative, may be extremely important, while to users uninterested in student-level longitudinal effects, quality may decrease as a function of grade span, if only because higher quality cross-sectional data could have been collected instead.

[3]In appendix C we illustrate linear, superlinear (convex), and sublinear (concave) functions graphically.
[4]Such phenomena "break" NCES's cost-effective students-within-schools model of data collection.
[5]A concave function increases more slowly than linearly.
[6]As one illustration, quality measured by precision of estimates is proportional to the square root of the number of subjects. Many other inference-based measures of quality appear in the statistical disclosure limitation literature.

The technical panel finds both [K-5|6-8|9-12] and [K-8|9-12] partitions acceptable in the event that there is no student continuation across studies. The [K-5|6-8|9-12] partition in figure 2 is recommended for these reasons:

- It is consistent with the [elementary|middle|secondary] structure of many school systems.
- It produces a nationally representative sample of sixth-graders, which is nearly impossible in a [K-8|9-12] partition.
- Because cost is a convex function of duration, the figure 2 partition is expected to be cheaper than the figure 3 partition.

The main argument in favor of the partition in figure 3 is that it produces intact K-8 longitudinal data records. Such records would, for instance, capture the elementary-to-middle-school transition for all data subjects.

### 2.4 Early initiation and late termination

It is not necessary that grades be strictly partitioned among studies. For instance, the configuration in figure 2 could be replaced by that in figure 4, in which the 6-8 middle school study is replaced by a 5-8 middle school study, but temporal relationships are preserved. The same could be done at the middle school–high school interface.

One justification for this configuration is that it captures the elementary-to-middle-school transition for all students. The technical panel cannot state definitively that the configuration in figure 4 is better than that in figure 2; data quality and cost both increase, perhaps significantly. It is easier to argue that it is better than the configuration in figure 3, especially if freshening is performed to ensure a nationally representative sample in grade 6 of the 5-8 middle school study[7]. However, freshening in the final year of a multiyear study is expensive and yields only incomplete data on the new subjects.

In the sense that figure 4 represents early initiation of studies, figure 5 corresponds to late termination: the K-5 study becomes K-6. In some ways, late termination may be preferable, because studies already continue in order to collect data about students who progress more slowly than the nominal rate. This configuration also avoids the need for freshening.

The technical panel finds that early initiation and late termination are sound, albeit potentially costly, strategies as means by which NCES could ensure that transition data are not lost. An alternative approach is noted in section 5.

[7]This is the same way that freshening currently takes place in grade 1 of K-X studies.

**Figure 4.** **Configuration with partition of [K-5|5-8|9-12], corresponding to early initiation of the middle school study.**

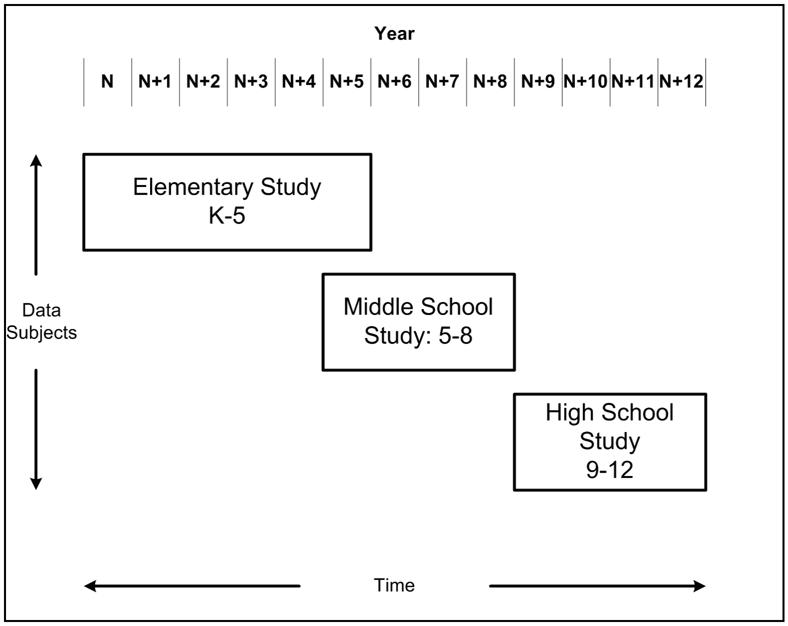
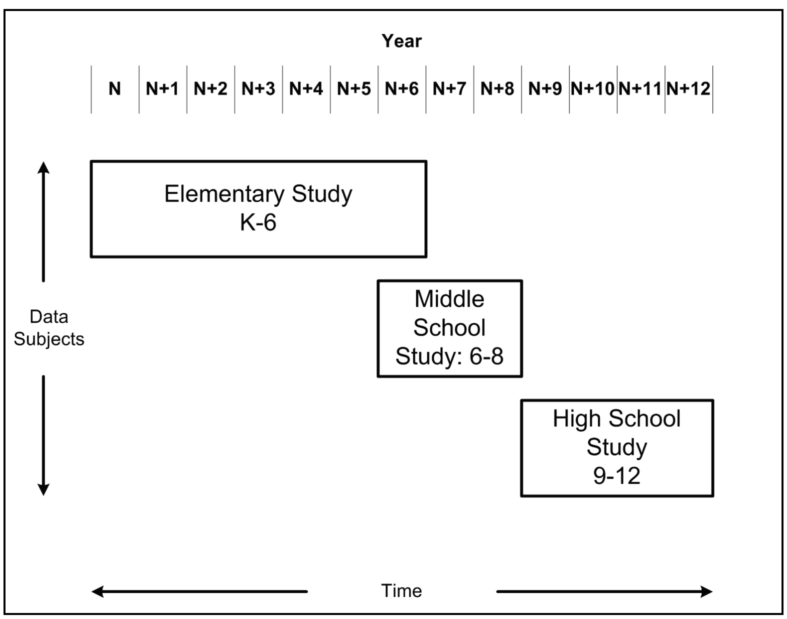



**Figure 5.** **Configuration with [K-6|6-8|9-12] partition, corresponding to late termination of the elementary school study.**




## 2.5 Articulation of K-12 and postsecondary studies

As noted in section 1, the technical panel concentrated on K-12 studies. Here, without specific recommendations, are some observations regarding the secondary–postsecondary interface.

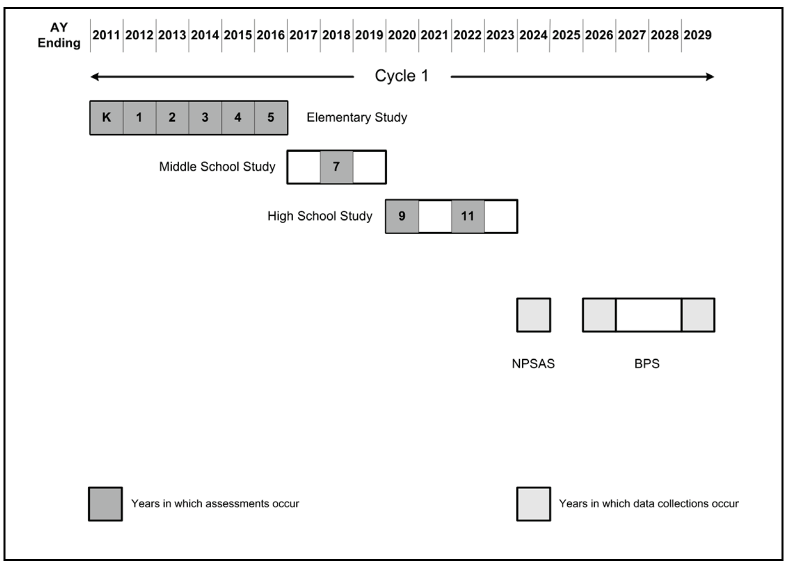
Currently, the National Postsecondary Student Aid Study (NPSAS) occurs every 4 years, and is followed alternately by, and feeds, the following studies:

- Beginning Postsecondary Students Longitudinal Study (BPS), using NPSAS subjects entering postsecondary study in the NPSAS year; and
- Baccalaureate and Beyond Longitudinal Study (B&B), using NPSAS subjects completing bachelor's degrees in the NPSAS year.

The 4-year frequency for NPSAS is incongruent with the 10-year K-12 cycle recommended in section 2.2. The technical panel acknowledges that changing the frequency of NPSAS may not be feasible; however, it notes that from an articulation point of view, a 5-year frequency for NPSAS would interface cleanly with the 10-year frequency for K-12 studies recommended in section 2.2. This is especially attractive if each NPSAS associated with BPS were to be initiated the year following completion of a 9-12 study (in figure 2). This reconfiguration is shown in figure 6. The "other" NPSASs, which would occur in years ending in 9 and would be associated with B&Bs, are not shown.

The technical panel is also aware that secondary studies such as the High School Longitudinal Study of 2009 (HSLS:09) plan significant follow-up of students after high school graduation. It is not clear how these follow-ups relate to other postsecondary data collections such as NPSAS.

**Figure 6. NPSAS/BPS articulated with the [K-5|6-8|9-12] partition in figure 2.**




# 3 Data Integration Across Studies

Warning: none of the figures in this section attempt to depict attrition during the course of studies.

## 3.1 General considerations

The configurations shown in figures 2 and 3 describe studies that are disjoint with respect to both grade span and data subjects. In the early initiation and late termination configurations in figures 4 and 5, studies overlap in time in order to capture specific phenomena, but remain disjoint with respect to data subjects.

The technical panel realizes that some members of the research and policy communities would like to have full K-12 student trajectories, and that there exist important scientific and policy issues that span multiple studies. For instance, to what extent do student performance or school characteristics in elementary school affect performance in middle school and high school?

Such questions cannot be addressed directly by configurations such as that in figure 2, and therefore it is essential to ask how NCES can support data integration for analyses that involve more than one study.

## 3.2 Data integration by means of continuation

The technical panel recommends that NCES conceptualize the problem of integrating data across two studies—for concreteness, a K-5 elementary school study and a 6-8 middle school study—as shown schematically in figure 7[8]. The principal difference between this and the configurations in section 2 is the presence of *continuation cases*—subjects from the K-5 study who are continued into the 6-8 study—that are represented by the horizontal band running across the full time span. In effect, then, there are three classes of data subjects, for whom different sets of data are available[9]:

- **K-5 only**, for whom data collected by NCES during the study are available for grades K-5, and "prospective data" relating to a student following completion of his or her study participation exist for grades 6-8, but are not available. The available data correspond to block A in figure 7;
- **both**, for whom data are collected for grades 1-8, corresponding to block B in figure 7; and
- **6-8 only**, for whom data are collected for grades 6-8, corresponding to block C in figure 7 as well as some *retrospective* data pertaining to a student's history for grades K-5, prior to participating in the 6-8 study. These correspond to the "recoverable retrospective data" in figure 7, which are explained below.

The difference between retrospective data and prospective data is that some of the former can be collected within a study, but none of the latter can be collected without specific actions by NCES. Sources of retrospective and prospective data include study participants, families, schools, and state-level administrative databases[10].

---

[8]The assumption of disjointness in time is made in order to avoid unnecessary complexity.
[9]"Ordinary" missing data are ignored in figure 7.
[10]Developments subsequent to the principal activities of the technical panel have increased the likelihood that state-level data systems will be constructed sooner rather than later.

There is an important distinction between recoverable and unrecoverable retrospective data. The former consist of elements such as family status and student grades that can be obtained from administrative data or survey instruments when the 6-8 study is initiated. The unrecoverable data consist of elements such as performance on assessments that were not administered. The experiments in section 4 illustrate this distinction concretely.

The purpose of continuation cases is to improve inference regarding issues that involve both studies. Therefore, the problem formulation is decision-theoretic. The decision variables are the number of continuation cases—the vertical coordinate in figure 7—and the means by which they are selected, which is not shown in figure 7. The decision criterion is a tradeoff between data quality—the capability of continuation to improve inference, and cost—continuation cases are costly, especially because they will rarely follow NCES's "sample students within sampled schools" design model.

To illustrate, the cost associated with a specific situation might be calculated as follows:

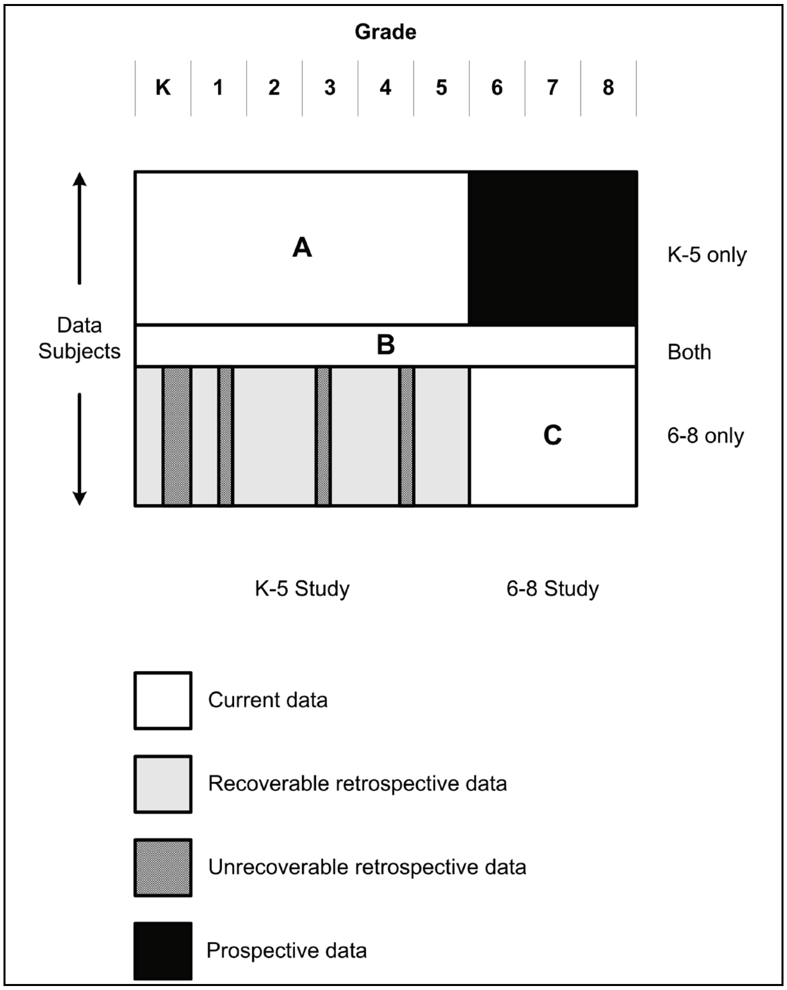
$$
\begin{aligned}
& c_1 * (\text{height of rectangle A}) * (\text{width of rectangle A})^2 \\
+ & c_2 * (\text{height of rectangle B}) * (\text{width of rectangle B})^2 \\
+ & c_3 * (\text{height of rectangle C}) * (\text{width of rectangle C})^2 \\
+ & c_4 * (\text{height of rectangle C}) + c_5
\end{aligned}
$$

In this expression, the $c_i$ are per-student costs defined in detail below. For simplicity, each cost component is assumed to be linear in the number of subjects. The exponent 2 in this expression is a simple form of the "convexity in the length of study" discussed in section 2. The components of the total cost are

- the cost of the K-5 only cases in the K-5 study, where $c_1$ is the cost per student in the initial year of the K-5 study;
- the cost of the K-8 cases, where $c_2$ is the cost per student in the initial year of the K-8 study;
- the cost of the 6-8 only cases in the 6-8 study, where $c_3$ is the cost per student in the initial year of the 6-8 study;
- the cost of recoverable retrospective data for those in the 6-8 study only, where $c_4$ is the cost per student to recover the recoverable retrospective data; and
- all other costs, which are equal to $c_5$.

The extent to which, and in what manner, continuation cases increase data quality by improving inference is more elusive. Any such improvement is both contextual and empirical. The experiments discussed in section 4 provide some insight.

**Figure 7. Data matrix when there is continuation from a K-5 study to a 6-8 study.**




## 3.3 Configurations with continuation

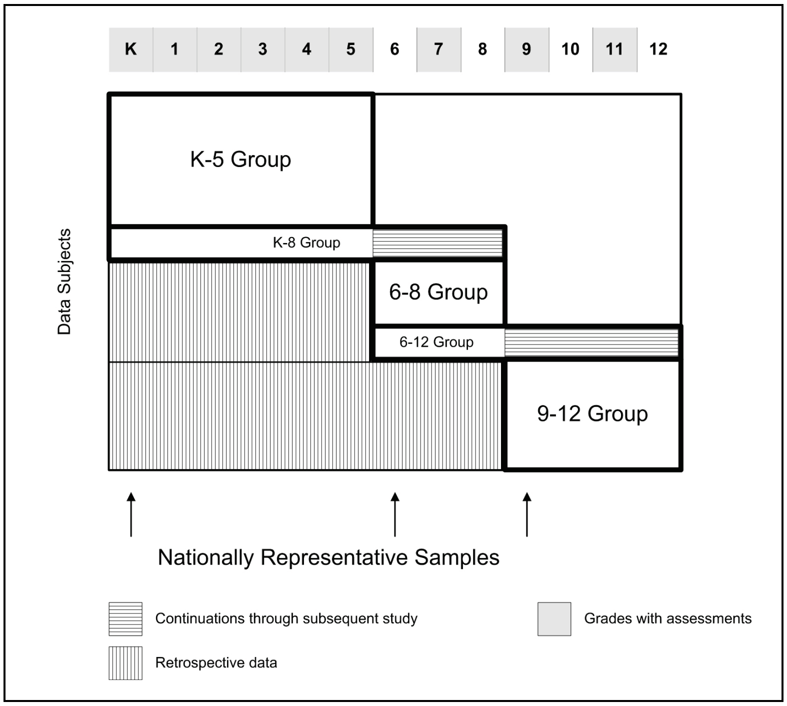
For reasons discussed in section 3.2, the technical panel does not provide categorical recommendations regarding continuation cases. It does, however, propose principles on which NCES can base its decisions:

- By controlling the size, and potentially the method of selection, of the continuation set, NCES can actively discourage use of this set alone for statistical purposes[11].
- NCES should place little reliance on *prospective* data, which are likely to be costly, inconsistent, and of questionable quality.
- Sources of retrospective data other than records maintained by a student's current school and state-level databases are problematic[12].

[11]Underlying this is the presumption, as figure 7 suggests, that the continuation set is "small" compared to the size of the studies. The technical panel anticipates that budget constraints would force this to be the case.

Figure 8 illustrates a configuration with continuation that seems particularly attractive.

**Figure 8. Configuration with continuation into (one) subsequent study.**




Underlying this figure is figure 2: K-5, 6-8, and 9-12 studies are conducted. There are five distinct groups of students, in chronological order as follows:

- **K-5:** students in only the K-5 study;
- **K-8:** students in both the K-5 study and the K-8 study, for whom the primary-to-middle-school transition is observed;
- **6-8:** students in only the 6-8 study;
- **6-12:** students in both the 6-8 study and the 9-12 study, for whom the middle-to-high-school transition is observed; and
- **9-12:** students in only the 9-12 study.

We assume that no student participates in all three studies. Therefore, no intact K-12 records are produced.

It is important, as indicated in figure 8, that the sample be nationally representative at each of grades K, 6, and 9, in the sense that

[12]The technical panel accepts the view of some NCES personnel that even if state-level administrative databases are not currently promising sources of retrospective data, they will be such in the future.

- at grade K, the K-5 and K-8 groups collectively are a nationally representative sample of grade K;
- conditional on the K-8 group, it, combined with 6-8 and 6-12, forms a nationally representative sample of grade 6 at that time. This implies that the study whose data subjects are the union of K-8, 6-8, and 6-12 and whose grades are 6-8 stands on its own; and
- conditional on the 6-12 group, 6-12 and 9-12 combined form a nationally representative sample of grade 9 at that time.

Consistent with the above principles, figure 8 depicts continuation as the exception, not the rule. Of course, there are technical issues underlying figure 8 that require resolution:

- Can selection of the K-8 and 6-12 continuation groups and the new samples in grades 6 and 9 be done in a way that "national representation" is ensured?
- Students in the continuation groups are likely to be weighted differently within each of the two overlapping studies[13]. Whether this actually constitutes a problem, and, if so, what are its implications and resolution, is not clear.
- The nature and use of retrospective data are not inherently well defined.

In fact, figure 8 does not depict a "configuration" but rather a three-parameter family of configurations depending on the numerical size of the K-8 and 6-12 continuation groups and the length of the two continuations. Figure 8 shows continuation through the entire next study, but there are other possibilities. For instance, continuation might be through the first assessment in the succeeding study.

### 3.4 Continuation from K-12 to postsecondary studies

The technical panel finds the case for continuation from 9-12 studies (e.g., the High School Longitudinal Study of 2009 and of 2019 [HSLS:09, HSLS:19]) to the Beginning Postsecondary Students Longitudinal Study (BPS) less compelling than continuation within K-12 but, at the same time, less problematic. There is little cost to doing so, but possibly little gain, because HSLS:09 (and presumably, successors) plans follow-up 2 and 8 years following grade 12. This follow-up will capture those both in and not in postsecondary education. BPS does not capture individuals who never enter postsecondary education; it does capture and follow those individuals who start postsecondary education but do not complete their education as well as those who do complete their postsecondary education.

### 3.5 Data integration in the absence of continuation

Should, as evidence in section 4 suggests, the data quality benefits of continuation not justify the cost, NCES might implement other strategies to facilitate data integration. Here, the technical panel comments on several such strategies.

[13]This presents perhaps further reason that K-8 and 6-12 cannot serve as standalone databases.

### *3.5.1 Imputation*

As figure 7 makes explicit, the problem of cross-study data integration is, at the core, one of missing data, but one in which the missingness is on a massive scale and for structural reasons. In particular, standard “missing (completely or not) at random” assumptions do not apply. The data quality experiments discussed in section 4 address the use of imputation as a tool for data integration, with and without continuation.

### *3.5.2 Record linkage*

The technical panel was requested to consider whether NCES should itself create synthetic records that represent data records spanning multiple studies, albeit not for real data subjects. Examples of methodologies that might be employed to accomplish this are probabilistic record linkage, creating what some would call synthetic cohorts, although a more precise term is synthetic records, and (possibly multiple-) imputation-based methods.

The technical panel recommends that NCES *not create or release synthetic records*. The justification is both methodological and practical. First, there is not a sound statistical basis for doing so. Probabilistic and other record linkage methods are designed for situations where the data sets being linked can be linked conceptually, in the sense that they are known to represent the same subjects[14]. However, if NCES was to attempt to link across longitudinal studies with differing data subjects, the information allowing definitive linkage (e.g., a foreign database key) would be absent. A central practical, and also methodological, impediment is sample weights, which differ across studies in ways that seem to make sensible linkage difficult to impossible[15].

## 4 Data Quality Experiments

If NCES were to implement the cost–quality tradeoff approach to selection of a configuration of the form shown in figure 7, then it would need to quantify both costs and data quality gains associated with different choices of the size and duration of the continuation group. The technical panel is not able to provide informed estimates of costs to NCES, but under its direction the National Institute of Statistical Sciences (NISS) has conducted experiments designed to yield insight regarding data quality.

The principal conclusion from those experiments is that *from a statistical perspective, continuation is of limited value*. Together, use of recoverable retrospective data for students in the second study and imputation of unrecoverable retrospective data for those students are as effective as continuation levels as high as 20 percent. Moreover, the cost of the imputation strategy is virtually certain to be less than that of continuation.

We stress that the experiments reported here do not employ weights, which are, however, discussed in section 4.3.

---

[14]Or, they represent at least partially overlapping sets of subjects.

[15]It is possible to argue that such linked data could be used (only) for unweighted analyses, but even then issues would remain. For instance, what population would such a data set purport to describe?

## 4.1 Conceptual structure of the experiments

The experiments share a common conceptual structure. We take an existing NCES longitudinal database—specifically, the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)—and split it into two simulated, successive, shorter-term longitudinal studies by suppressing some data. As in figure 7, some students are continued from the first study to the second.

This process is illustrated for ECLS-K in figure 9, with the split into K-5 and 6-8 studies[16].

Consider a research investigation that involves attributes from both the shorter studies, for instance, the relationship between student performance in grade 1 and student performance in grade 8 as a function of gender and race. This relationship can be estimated from both

- the full data, using appropriate statistical methods; and
- the two-study data, using statistical approaches that cope with the missing data[17].

The results can be compared in order to understand how replacing the single study by two studies degrades data quality, in the sense of attenuating the relationship, distorting it, or increasing uncertainty about it.

There is another path to estimation, which is to use the continuation group alone. As discussed in section 3, the technical panel feels that when there is continuation, NCES should discourage inference based on continuation cases alone. The experiments show that the statistical effectiveness of this approach cannot be dismissed out of hand. Nevertheless, seemingly insuperable difficulties remain; see the discussion in section 4.3.

The experimental structure shown in figure 9 has two parameters whose effect can be assessed. The first of these is the size of the continuation group. The second is more subtle, but also may be more important: the way in which the continuation group is selected. The possibilities range from a simple random sample to a weight-based random sample to using concepts from experimental design—specifically, space-filling designs meant to “cover” high-dimensional spaces with few design points. The latter two were not explored within the NISS experiments.

In reality, the simulated data sets in figure 9 are too “all-or-nothing.” For students in the 6-8 study, some K-5 attributes are readily coverable from either survey instruments or state-level pupil tracking systems. Using such attributes produces a simulated two-study data set of the form shown in figure 10, where all available data are indicated by white. This approach introduces a third parameter, the *selection of the retrospective attributes,* which may reflect both quality and cost considerations. Of course, some retrospective attributes may be problematic in terms of quality, cost, or both. The utility of retrospective data is discussed in section 4.2.4.

---

[16]This is consistent with the recommendations in section 2.

[17]In the experiments described in section 4.2, the missing data are imputed, and then exactly the same methodology is used.

**Figure 9. Schematic representation of the experiments. Top: full data from K-8 study. Bottom: data from simulated K-5 and 6-8 studies, with suppressed data shown in gray.**

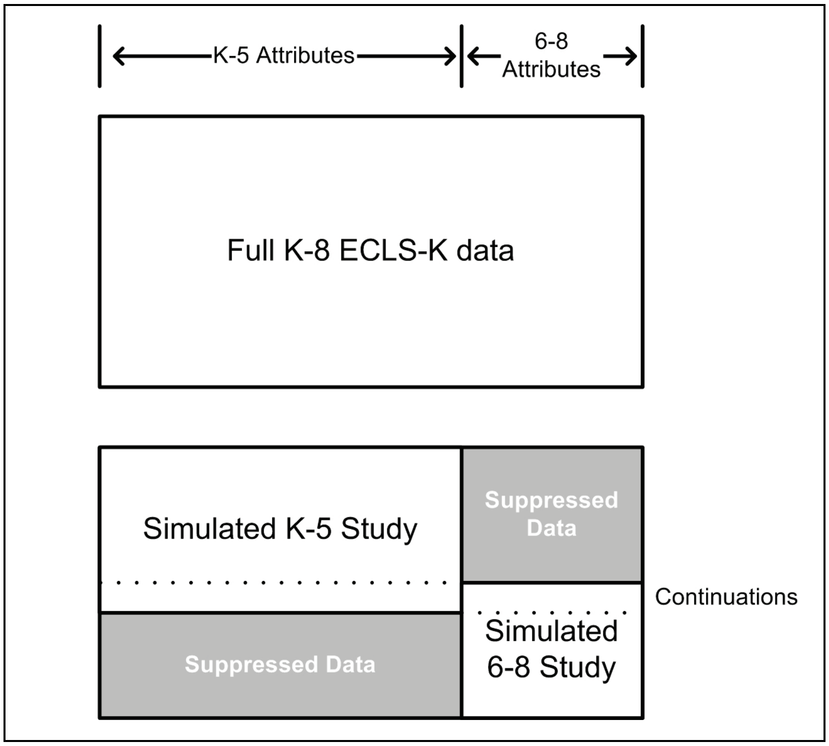
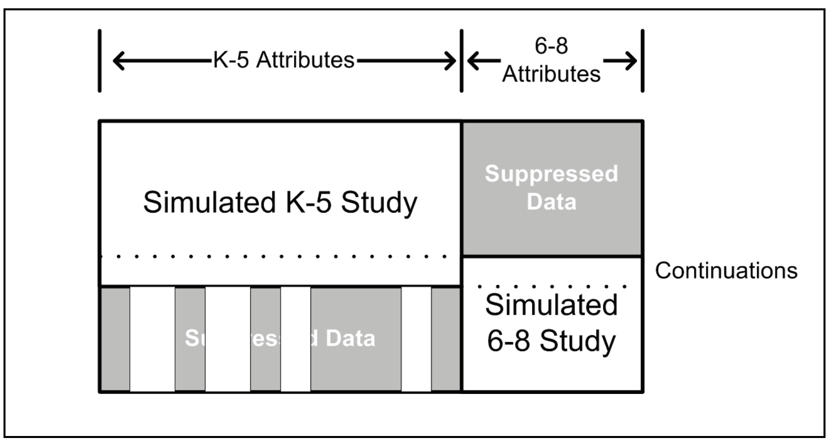



**Figure 10. Final form of simulated two-study data.**




## 4.2 An illustrative experiment

This experiment is of the form described in section 4.1, but simplified in order to allow investigation of the underlying research question, the size of the continuation group, and the choice of retrospective attributes. There is also one important difference: because of constraints on availability of data, in this section we consider the effect of splitting a K-5 study into two studies that comprise the following elements:

- **K-2 Study:** Covering years K-2 ,and based on kindergarten and first grade panels; and
- **3-5 Study:** Covering grades K-3, and based on third- and fifth-grade panels.

For completeness, this is illustrated in figure 11: following selection of the continuation group, the remaining cases were split randomly and equally into a K-2 group whose 3-5 data were suppressed and a 3-5 group whose K-2 data were suppressed.

Then, all suppressed data—everything in gray in figure 11—were imputed, following which exactly the same analysis was performed on both the original K-5 data and the two-study + imputed data shown in figure 12.

- *Group 1:* Fully observed in **K-2**, completely unobserved in **3-5**.
- *Group 2:* Fully observed in both **K-2** and **3-5**.
- *Group 3:* Completely observed in **3-5** and, *possibly*, partially observed in **K-2**, in the form of retrospective data.

**Figure 11. Creation of data sets for the experiments. Data in light gray are suppressed.**

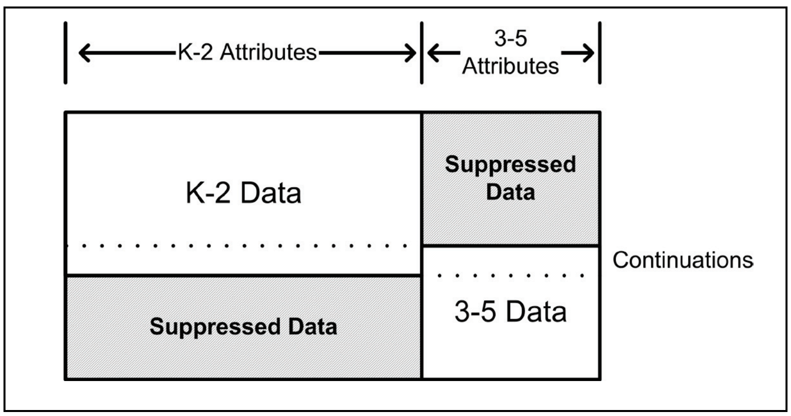



**Figure 12. Two-study data following imputation of suppressed data in figure 11. Data in light gray are imputed.**

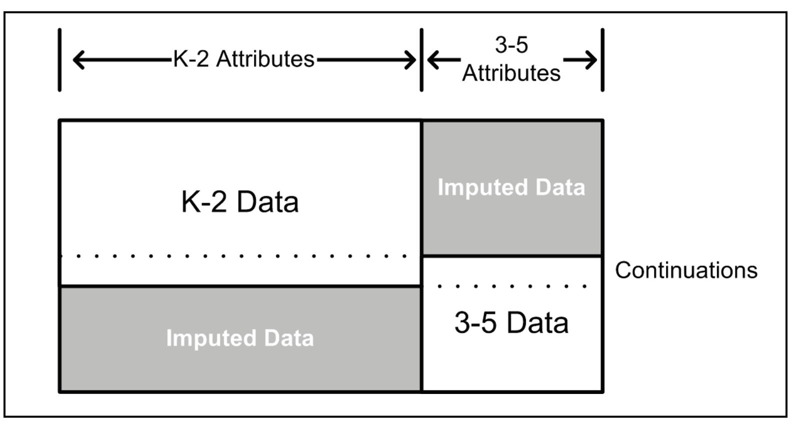
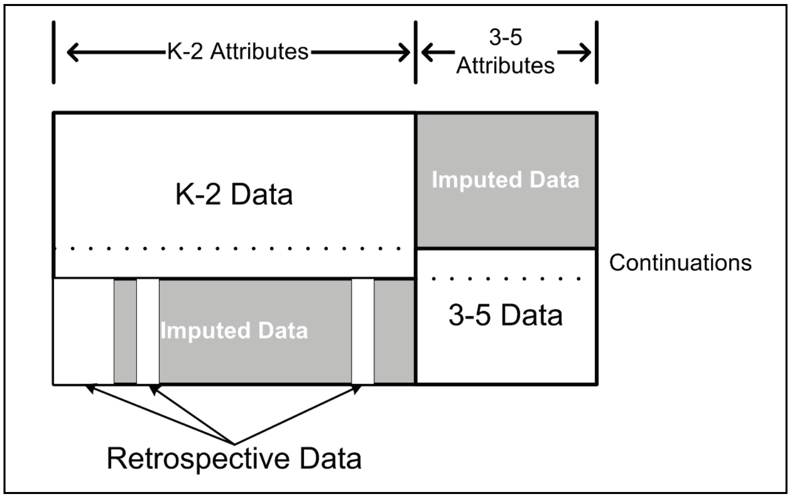


**Figure 13. Two-study data when retrospective data are present, following imputation of suppressed data. Data in light gray are imputed.**




We note that the conceptual scheme associated with figures 11-13 is discussed in a different order in section 4.2. Section 4.2.3 deals with data of the form shown in figure 13, with emphasis on the size of the continuation group. Section 4.2.4, which addresses the utility of retrospective data, deals with data of the form shown in figure 12, which corresponds to Case 3 there.

### *4.2.1 Data set construction*

The base data set, representing a full K-5 study, was created from ECLS-K. A merged file was constructed containing only records for students with test scores recorded in the kindergarten, first-grade, third-grade, and fifth-grade cohorts, with a total of 9,940 students. Negative and extraneous codes for the variables were changed to missing values. All missing values were completed by means of a single imputation with IVEware[18]. The result is a completed data set that can be used for the experiments.

### *4.2.2 Scenarios*

Each scenario below is defined by two characteristics: the proportion of records allocated to each group and the completeness of the retrospective data in Group 3. These are varied to yield insight concerning the questions of interest.

For each scenario, there are suppressed data corresponding to the gray regions in figure 11. These were multiply imputed using IVEware. The number of predictors employed in imputation was limited to 20, in order to improve computational efficiency.

[18]Visit http://www.isr.umich.edu/src/smp/ive for more information. Percentages were logit-transformed, imputed, and transformed back to percentages. To speed up computation, the number of predictors was limited to five. This one-time imputation of data missing in ECLS-K should not be confused with other imputations that “replace” suppressed data in the context of figures 11-13.

#### *4.2.3 Effects of the size of the continuation group*

In this scenario, the best case for Group 3 is assumed: all retrospective variables except assessments are observed. This case corresponds to figure 13. The proportion of students in the continuation group was varied from 0 percent to 20 percent, and Groups 1 and 2 were of equal size. Assignment to groups was *random*. To reduce the effect of group assignment on the experiment results, the groups are nested within each other. For example, the 5 percent continuation group was selected by randomly drawing one-quarter of the 20 percent continuation group, with the remaining three-quarters randomly assigned to Group 1 or Group 3.

Two research questions were considered; one addressed univariate distributions of attributes and the other addressed correlations between attributes from different studies. To summarize, the panel arrived at these results, *which pertain to the entire imputed dataset*:

- Means and standard deviations were preserved well in all cases. Minima and maxima were not preserved as well, but this could be improved by using imputation models that can handle skewed data more effectively.
- Correlations between K-2 and 3-5 attributes degrade as the size of the continuation group decreases.
- At the extreme cross-study overlap of 20 percent, direct estimation of correlations yields, by comparison with imputation, estimates with less bias but higher variability.

The variable names in the following tables are chosen for clarity. Appendix D provides a mapping of these onto ECLS-K variable names.

Table 1 presents results for means and their multiple imputation standard errors (Rubin's rules) and standard deviations (square root of average variance across imputations) for selected attributes, using five imputations. A size of 0 percent for the continuation group means that no full-length cases are available for imputation purposes.

In this experiment, the principal effect on single attributes of splitting the study is a reduction in sample size. In reality, both of the separated studies would be larger. Even so, it seems clear that for single attributes there is effectively no benefit from any level of continuation. With or without continuation, multiple imputation appears to be a promising strategy for obtaining weighted estimates of means.

Parallel results for correlations, both within and across the two studies, are presented in table 2. The correlations shown for the imputed cases are averages across five imputations. As expected, the correlations within one study (first and last rows) are well preserved and do not depend on the continuation percentage.

For the between-study correlations (rows 2-5 in table 2, which are italicized), the results are again as expected. Correlations are attenuated as compared to the full data, in some cases dramatically. For instance, consider the second row, which represents the correlation between kindergarten and third-grade mathematics assessments. The true correlation of .72 is substantially underestimated even at a continuation group size of 20 percent. The correlations do improve, but only modestly, with increasing continuation group size.

Table 3 shows a similar lack of benefit from having a continuation group in a regression analysis. With increasing continuation group size we see similar results without much trend toward the complete data results. The standard errors (MI StdErr) shown for the imputed cases are computed using Rubin's rules for multiply-imputed data.

**Table 1. Summary of statistics as a function of the size of the continuation group.**

| | | Size of Continuation Group | | | | |
|---|---|---|---|---|---|---|
| | | **Full Data** | **0%** | **1%** | **5%** | **20%** |
| K-Math | Mean | 28.88 | 29.09 | 29.04 | 29.15 | 29.06 |
| | MI StdErr | | 0.11 | 0.12 | 0.14 | 0.12 |
| | StdDev | 8.58 | 8.97 | 8.92 | 9.01 | 8.96 |
| G1-Reading | Mean | 57.28 | 57.49 | 57.36 | 57.50 | 57.48 |
| | MI StdErr | | 0.19 | 0.23 | 0.22 | 0.19 |
| | StdDev | 13.02 | 13.90 | 13.73 | 13.91 | 13.82 |
| G3-Math | Mean | 86.48 | 85.95 | 86.49 | 86.15 | 86.66 |
| | MI StdErr | | 0.66 | 0.59 | 0.51 | 0.63 |
| | StdDev | 17.09 | 17.67 | 17.78 | 17.77 | 17.71 |
| G5-Reading | Mean | 141.72 | 141.31 | 141.34 | 140.62 | 141.67 |
| | MI StdErr | | 1.08 | 0.94 | 0.93 | 0.28 |
| | StdDev | 21.89 | 22.83 | 22.65 | 22.70 | 22.65 |

**Table 2. Summary of correlations as a function of the size of the continuation group. Pairs of variables that cross studies are italicized.**

| Correlation Between Variables | | Size of Continuation Group | | | | |
|---|---|---|---|---|---|---|
| | | **Full Data** | **0%** | **1%** | **5%** | **20%** |
| K-Math | G1-Reading | .65 | .65 | .64 | .65 | .65 |
| *K-Math* | *G3-Math* | *.72* | *.38* | *.38* | *.40* | *.47* |
| *K-Math* | *G5-Reading* | *.61* | *.40* | *.40* | *.41* | *.46* |
| *G1-Reading* | *G3-Math* | *.62* | *.33* | *.33* | *.34* | *.41* |
| *G1-Reading* | *G5-Reading* | *.70* | *.37* | *.37* | *.38* | *.46* |
| G3-Math | G5-Reading | .70 | .68 | .69 | .69 | .69 |

As noted at the beginning of this section, the results here are based on continuation by inclusion of all retrospective data other than unrecoverable assessments. We next look more closely at the utility of retrospective data.

### *4.2.4 The utility of retrospective data*

To some extent, given the conclusions in section 4.2.3, assessment of the value of the retrospective data is not essential, but we include the experimental results for completeness. The size of the continuation is taken to be 5 percent.

**Table 3. Summary of regression coefficients for a multiple regression predicting G5-Reading from other variables, as a function of the size of the continuation group. The predictor variables were transformed as indicated.**

| Predictor | | Size of Continuation Group: Full Data | 0% | 1% | 5% | 20% |
|---|---|---|---|---|---|---|
| log(G5-Math) | Estimate | 0.438 | 0.536 | 0.528 | 0.532 | 0.516 |
| | StdErr | 0.007 | 0.010 | 0.008 | 0.012 | 0.008 |
| log(K-Math) | Estimate | 0.025 | 0.048 | 0.044 | 0.045 | 0.035 |
| | StdErr | 0.006 | 0.009 | 0.007 | 0.011 | 0.010 |
| log(K-Reading) | Estimate | 0.148 | 0.048 | 0.049 | 0.046 | 0.080 |
| | StdErr | 0.005 | 0.008 | 0.009 | 0.009 | 0.010 |
| logit(G5-MinorityPercent) | Estimate | -0.002 | -0.002 | -0.002 | -0.003 | -0.002 |
| | StdErr | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| logit(G5-FreeLunch) | Estimate | -0.005 | -0.005 | -0.005 | -0.005 | -0.004 |
| | StdErr | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |

The rationale for this part of the experiment is based on the assumptions that some retrospective data (see figure 10) may not be readily available at low cost, even when statewide pupil tracking systems become ubiquitous, and that the quality of some retrospective data may be low. In the setting of these experiments, retrospective data were divided, somewhat realistically, into these categories:

- **Family Data,** representing roughly what is on the ECLS-K parent questionnaire;
- **Student Data,** such as date of birth, age, gender, and race; and
- **School Data,** about the school(s) in which the student was enrolled during the "retrospective" time period.

Oversimplifying, these are increasingly easy to obtain and of increasing quality.

Three cases were considered:

- **Case 1:** Retrospective data contain all information other than results of assessments, resulting in a scenario that is identical to a 5 percent continuation group in section 4.2.3.
- **Case 2:** Retrospective data contain only school data and student data.
- **Case 3:** No retrospective data are used, which corresponds pictorially to figure 12.

The results in tables 4, 5, and 6, which are analogs of tables 1, 2 and 3, respectively, show that retrospective data seem to be of some, but not immense, value for improving the imputations. Retrospective data on schools and students are relatively easy to collect and have other potential uses, and so appear worth the effort to collect. School data can be obtained from the data frame if past schools can be identified. Student variables other than assessments can be reconstructed, and additional variables may become available from statewide tracking databases.

Comparing cases 1 and 2, it does not appear that efforts to reconstruct family variables such as socioeconomic status during the retrospective period are likely to be cost effective.

**Table 4. Summary statistics as a function of completeness of the retrospective data.**

| | | Retrospective Data | | | |
|---|---|---|---|---|---|
| | | **Full Data** | **Case 1** | **Case 2** | **Case 3** |
| K-Math | Mean | 28.88 | 29.15 | 29.06 | 29.15 |
| | MI Std Err | | 0.14 | 0.12 | 0.20 |
| | StdDev | 8.58 | 9.01 | 8.90 | 8.92 |
| G1-Reading | Mean | 57.28 | 57.50 | 57.42 | 57.55 |
| | MI Std Err | | 0.22 | 0.23 | 0.19 |
| | StdDev | 13.02 | 13.91 | 13.72 | 13.80 |
| G3-Math | Mean | 86.48 | 86.15 | 86.28 | 86.81 |
| | MI Std Err | | 0.51 | 1.63 | 0.63 |
| | StdDev | 17.09 | 17.77 | 17.97 | 17.97 |
| G5-Reading | Mean | 141.72 | 140.62 | 140.54 | 141.61 |
| | MI Std Err | | 0.93 | 1.01 | 0.58 |
| | StdDev | 21.89 | 22.70 | 22.63 | 22.76 |

**Table 5. Summary of correlations as a function of the completeness of the retrospective data. Pairs of variables that cross studies are italicized.**

| Correlation Between Variables | | Retrospective Data | | | |
|---|---|---|---|---|---|
| | | **Full Data** | **Case 1** | **Case 2** | **Case 3** |
| K-Math | G1-Reading | .70 | .64 | .64 | .64 |
| *K-Math* | *G3-Math* | *.68* | *.40* | *.38* | *.38* |
| *K-Math* | *G5-Reading* | *.61* | *.41* | *.39* | *.41* |
| *G1-Reading* | *G3-Math* | *.52* | *.35* | *.33* | *.33* |
| *G1-Reading* | *G5-Reading* | *.58* | *.40* | *.36* | *.38* |
| G3-Math | G5-Reading | .73 | .69 | .68 | .69 |

**Table 6. Summary of regression coefficients for a multiple regression predicting G5-Reading as a function of the completeness of the retrospective data. The predictor variables were transformed as indicated.**

| | | Retrospective Data | | | |
|---|---|---|---|---|---|
| **Predictor Variable** | | **Full Data** | **Case 1** | **Case 2** | **Case 3** |
| log(G5-Math) | Estimate | 0.438 | 0.532 | 0.529 | 0.531 |
| | StdErr | 0.007 | 0.012 | 0.010 | 0.008 |
| log(K-Math) | Est | 0.025 | 0.045 | 0.043 | 0.048 |
| | StdErr | 0.006 | 0.011 | 0.007 | 0.008 |
| log(K-Reading) | Est | 0.148 | 0.046 | 0.046 | 0.049 |
| | StdErr | 0.005 | 0.009 | 0.007 | 0.007 |
| logit(G5-MinorityPercent) | Est | -0.002 | -0.003 | -0.002 | -0.002 |
| | StdErr | 0.000 | 0.000 | 0.000 | 0.000 |
| logit(G5-FreeLunch) | Est | -0.005 | -0.005 | -0.005 | -0.004 |
| | StdErr | 0.001 | 0.001 | 0.001 | 0.001 |

### *4.2.5 Direct estimation from continuation cases*

As has been noted previously, the technical panel recommends that, even should NCES decide to conduct studies with continuations, it should strongly discourage direct inference from continuation cases alone. For purposes of scientific understanding, the program of experiments did include computation of such "direct" estimators.

With the full study viewed as the underlying population and with continuation cases constituting a simple random sample, it is clear that estimates constructed from the continuation cases alone are unbiased. However, because of the small size of the continuation set, estimates based only on the continuation cases have high standard errors as estimators of corresponding characteristics of the full data. These latter characteristics are, when weights are accounted for properly—see also section 4.3—can be used as estimators of characteristics of the target population. Table 7 illustrates the first but not—except implicitly via the "sample size"—the second characteristic of the direct estimators.

It is tempting, but ultimately misleading, to construe the difference between the imputation approach employed above and direct estimation to be a tradeoff between bias and variability. This is true only empirically, because there are several problematic aspects of direct inference, including issues associated with weights discussed in section 4.3. A more basic, conceptual question is exactly what target population the set of continuation cases purports to describe, to which there appears to be no credible answer. Finally, there seems to be virtually no possibility that continuation cases alone could support (even unweighted) high-resolution analyses addressing subgroups or geographical effects.

**Table 7. Direct estimates of selected correlations, using only the continuation cases. Correlations between variables that cross studies are italicized.**

| Correlation Between Variables | | Full Data | Continuation Level | | |
|---|---|---|---|---|---|
| | | | 1% | 5% | 20% |
| K-Math | G1-Reading | .70 | .75 | .68 | .65 |
| *K-Math* | *G3-Math* | *.68* | *.79* | *.73* | *.73* |
| *K-Math* | *G5-Reading* | *.61* | *.68* | *.64* | *.62* |
| *G1-Reading* | *G3-Math* | *.52* | *.76* | *.64* | *.63* |
| *G1-Reading* | *G5-Reading* | *.58* | *.81* | *.72* | *.71* |
| G3-Math | G5-Reading | .73 | .74 | .74 | .72 |

### *4.2.6 Implications for NCES*

Of course, there is no certainty that these results hold for all data sets and all statistical questions, but the experiments suggest five broad conclusions:

1. NCES should expect that splitting studies will lead to potentially significant attenuation of relationships that cross studies, even in the presence of continuation.
2. Modest levels of continuation (1 percent to 5 percent) decrease attenuation of relationships as compared to no continuation, but not dramatically.
3. Larger levels of continuation (20 percent) offer little improvement over modest levels, and clearly would be more costly.

4. Retrospective data do improve imputation.
5. There is no need to collect retrospective data other than the readily available information about students and schools they attended previously.

There are other possibilities. One could, for example, attempt to *model* the relationship between correlation when there is 1 percent overlap and correlation in the full data, and use the model-derived relationship in a predictive sense. For instance, a model might show[19] that correlation in the full data is 2.5 times that in a 1 percent continuation. Such an approach may be difficult for NCES to implement, justify, or explain.

It is possible that applying experimental design principles to selection of the continuation group or paying more careful attention to the imputation models would yield improved results. However, given that this experiment addressed the simplest possible relationship and yielded generally negative results, this strategy did not seem sufficiently promising to merit detailed investigation.

## 4.3 Weights

The experiments described in section 4.2 all comprise statistical summaries and analyses that do not account for case weights, nor do the imputation processes take weights into account.

Consider first the case where there is no continuation but imputation is performed, as in section 4.2.[20] Assume that, consistent with section 2, a K-5 study is followed immediately by a 6-8 study, as shown in figure 11, which for clarity we use for illustration in this section. Importantly, each study addresses the same target population—students who began kindergarten in the first year of the K-5 study[21]. The two studies are simply non-overlapping samples from that population. Each of these studies has associated base weights that account for the complex sample design as well as school-level and student-level nonresponse[22]. Conceptually and in practice, it seems to be justifiable to construct a set of base weights for the "combined study" by linearly re-scaling the study-specific weights so that the sum of the "combined study" weights is "correct," in the sense of matching the size of the target population or satisfying some other calibration criterion.

The imputation necessary to fill in the light gray blocks in figure 11 can be performed, as we did in section 4.2, with the weights ignored. There are also, however, imputation methods that take weights into account. Although it is plausible that there are only modest differences, this cannot be known without further experimentation.

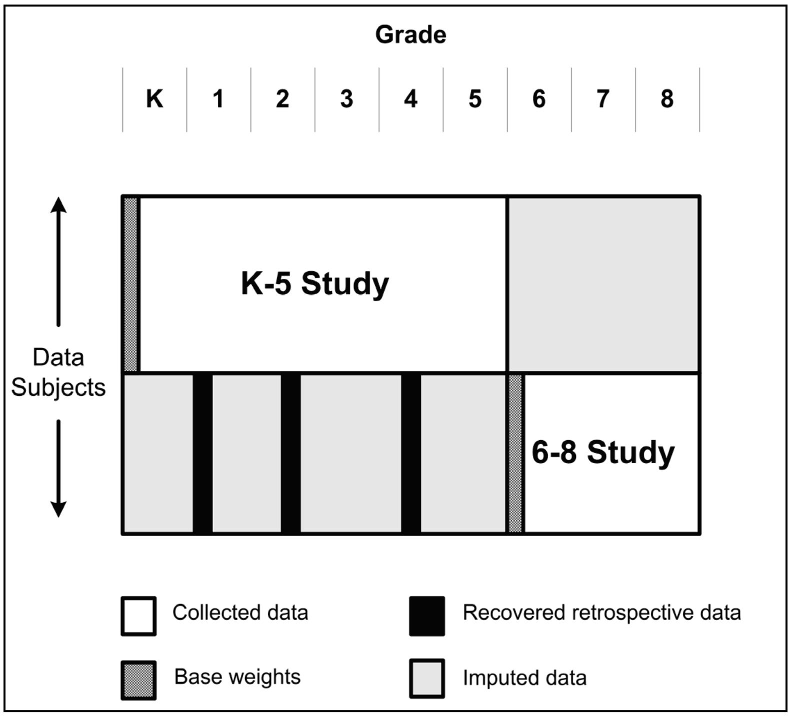
Once the imputation is done—ignoring weights or not—weighted analyses on the data represented in figure 11 can be performed in the usual manner.

[19]This would be the case if one were treating the second row in table 2 as the results of such a model.
[20]In the context of that section, this corresponds to a continuation percentage of zero.
[21]In the strictest sense this is not true, because the target population changes as the result of deaths, immigration and emigration, and slower (or faster) than nominal student progress. However, it is approximately true.
[22]A longitudinal study such as ECLS-K has several other weights that account for changes over time in the set of participants. For simplicity, these are omitted from the current discussion.

The situation changes significantly when there are continuation cases. Each such case has one set of weights associated with each study. These would need to be reconciled, but it is not obvious that methods exist for doing so[23]. As a result, there is no clear path to conducting weighted analyses when there is continuation. The argument that if the level of continuation is small, then any sensible procedure[24] will be acceptable lacks justification.

The difficulty just described is most extreme for inference based on only continuation cases, as discussed in section 4.2.6. Indeed, this may be the strongest argument against using only continuation cases.

**Figure 14. Schematic representation of back-to-back K-5 and 6-8 studies with no continuation, but with weights.**




## 5 Concluding Remarks

To summarize, the technical panel recommends the following practices:

- A 13-year cycle of K-12 studies, partitioned as K-5, 6-8, and 9-12, with no interstudy gaps, and initiated every 10 years; and

[23]NCES and other organizations have devised what are in some cases rather complex methods for adjusting weights within studies, which can accommodate phenomena such as freshening. Rotating panel data collections face similar issues. Nevertheless, it is at least highly uncertain that the issues can be resolved.
[24]An example would be averaging the two (base) weights and then re-scaling everything.

- In the absence of justifications of which the panel is not aware, use of high-quality retrospective data and imputation as the principal means of facilitating data integration across studies.

In particular, the continuation strategies discussed in section 3 present conceptual and practical issues associated with target populations and weights, do not seem to yield high statistical value, and appear likely to be costly.

The technical panel acknowledges that, especially with a split of K-8 into K-5 and 6-8, data regarding what may be central components of students' trajectories will not be collected. To the extent that there is consensus about the scientific or policy importance of understanding such phenomena, NCES might consider conducting a set of small—in both sample size and duration—studies, each targeted at one such phenomenon. For instance, there could be an "Elementary–Middle School Transition Study" that would interface with and leverage both the K-5 study and the 6-8 study, but need not be aligned to either. While such a study would need a sample that is rich enough to support sound inference, the principal goal might be scientific insight rather than preparation of national estimates, reducing sample size and cost.

# Appendix A. Summary of NCES Longitudinal Surveys

## A.1 Early childhood

***Early Childhood Longitudinal Study (ECLS)***
The Early Childhood Longitudinal Study (ECLS) program has been designed to include two overlapping cohorts: a Birth Cohort and a Kindergarten Cohort. The birth cohort follows a sample of children from birth through kindergarten entry. The kindergarten cohort follows a sample of children from kindergarten through the eighth grade.

The ECLS program provides national data on children's status at birth and at various points thereafter; children's transitions to nonparental care, early education programs, and school; and children's experiences and growth through the eighth grade. The ECLS program also provides data to analyze the relationships among a wide range of family, school, community, and individual variables with children's development, early learning, and performance in school.

*Birth Cohort*
The ECLS is designed to provide decision makers, researchers, child care providers, teachers, and parents with detailed information about children's early life experiences. The birth cohort of the Early Childhood Longitudinal Study (ECLS-B) looks at children's health, development, care, and education during the formative years from birth through kindergarten entry.

*Kindergarten Cohort*
The Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) is an ongoing study that focuses on children's early school experiences beginning with kindergarten and following children through middle school. The ECLS-K provides descriptive information on children's status at entry to school, their transition into school, and their progression through eighth grade. The longitudinal nature of the ECLS-K data enables researchers to study how a wide range of family, school, community, and individual factors are associated with school performance. The ECLS-K is a nationally representative sample of kindergartners, their teachers, and schools. Information is collected from children, their families, their teachers, and their schools all across the United States.

## A.2 High school

***National Longitudinal Study***
The NLS-72 describes the transition of young adults from high school through postsecondary education and the workplace. The data span the years 1972 through 1986 and include postsecondary transcripts.

***High School and Beyond (HS&B)***
The HS&B describes the activities of seniors and sophomores as they progressed through high school, postsecondary education, and into the workplace. The data span the years 1980 through 1992 and include data on parents and teachers, high school transcripts, student financial aid records, and postsecondary transcripts, in addition to student questionnaires and interviews.

***National Education Longitudinal Study (NELS)***
The NELS:88, which began with an eighth-grade cohort in 1988, provides trend data about critical transitions experienced by young people as they develop, attend school, and embark on their careers. Data were collected from students and their parents, teachers, and high school principals and from existing school records such as high school transcripts. Cognitive tests (math, science, reading, and history) were administered during the base year (1988), first follow-up (1990), and second follow-up (1992). Third follow-up data were collected in 1994. All dropouts who could be located were retained in the study. A fourth follow-up was completed in 2000.

***Education Longitudinal Study (ELS)***
The Education Longitudinal Study of 2002 (ELS:2002) is a longitudinal survey that monitors the transitions of a national sample of young people as they progress from 10th grade to, eventually, the world of work. ELS:2002 obtains information from students and their school records, and from students' parents, their teachers, their librarians, and the administrators of their schools.

***High School Longitudinal Study (HSLS)***
The HSLS began in 2009 with a cohort of ninth graders and will focus on the decisions students and their parents make as they progress through high school into postsecondary education or work. There is also special interest in students' decisions as they relate to education in science, technology, and mathematics.

## A.3 Postsecondary

***National Postsecondary Student Aid Study (NPSAS)***
The NPSAS is a comprehensive study that examines how students and their families pay for postsecondary education. It includes nationally representative samples of undergraduate, graduate, and first-professional students. It includes students attending public and private less-than-2-year institutions, community colleges, 4-year colleges, and major universities. Students who receive financial aid as well as those who do not receive financial aid are included in NPSAS. Comprehensive student interviews and administrative records, with details concerning student financial aid, are available for academic years 1986-87, 1989-90, 1992-93, 1995-96, 1999-2000, 2003-04, and 2007-08.

***Beginning Postsecondary Students (BPS)***
BPS studies follow students who first begin their postsecondary education in a particular year. Initially, students in the NPSAS surveys are identified as being first-time beginners of undergraduate studies. These students are asked questions about their experiences during, and transitions through, postsecondary education and into the labor force, as well as family formation. Transfers, persisters, stopouts/dropouts, and completers are among those included in the studies. For NPSAS:90, the first cohort of first-time, beginning students was identified in the 1989-90 academic year. These students were followed in 1992 (BPS:90/92) and in 1994 (BPS:90/94). A second cohort of first-time, beginning students was identified in NPSAS:96, with follow-ups performed in 1998 (BPS:96/98) and in 2001 (BPS:96/2001). The third cohort was identified in NPSAS:04 and was followed in 2006 (BPS:04/06) and in 2009 (BPS:04/09). For the third cohort, researchers also collected transcripts from postsecondary schools attended.

### *Baccalaureate and Beyond (B&B)*

B&B studies follow students who complete their baccalaureate degrees. Initially, students in the NPSAS surveys are identified as being in their final year of undergraduate studies. Students are asked questions about their future employment and education expectations, as well as about their undergraduate education. In follow-ups, students are asked questions about their job search activities and their education and employment experiences after graduation. Individuals who showed an interest in becoming teachers are asked additional questions about their pursuit of teaching and, if teaching, about their current teaching position. As part of NPSAS:93, the first cohort of students who completed their bachelor's degrees in the 1992-93 school year was identified. These students were followed up in 1994 (B&B:93/94), 1997 (B&B:93/97), and 2003 (B&B:93/2003). A new B&B cohort began with NPSAS:2000 and involved only a 1-year follow-up in 2001 (B&B:2000/01). The third cohort was identified in NPSAS:08, was followed in 2009 (B&B:08/09), and will be interviewed again in 2012. For the first and third cohorts, researchers also collected transcripts from the postsecondary institutions that awarded the students' bachelor's degrees. Future B&B cohorts will alternate with BPS in using NPSAS surveys as their base.

This page intentionally left blank.

## Appendix B. Agenda for Technical Panel Meeting on September 27-28, 2007

NISS

**National Institute of Statistical Sciences**
PO Box 14006, Research Triangle Park, NC 27709-4006
Tel: 919.685.9300 FAX: 919-685-9310
www.niss.org

Technical Panel on Longitudinal Studies:
Configuration and Data Integration

Technical Panel Meeting
September 27-28, 2007

Agenda

Thursday, September 27

| | |
|---|---|
| 9:00 AM | Welcome and introductions<br>Alan Karr, NISS<br>Andrew White, NCES |
| 9:15 | Review of charge |
| 9:30 | Discussion of configuration of K-12 studies |
| 10:30 | Break |
| 11:00 | Formulate initial recommendations on configuration of K-12 studies |
| 12:00 N | Lunch |
| 1:00 PM | Discussion of configuration of postsecondary studies |
| 2:00 | Formulate initial recommendations on configuration of postsecondary studies |
| 2:30 | Break |
| 3:00 | Discussion of data integration |

What are the needs?
What are the issues?
What are possible solutions?
What are the implications re configuration?

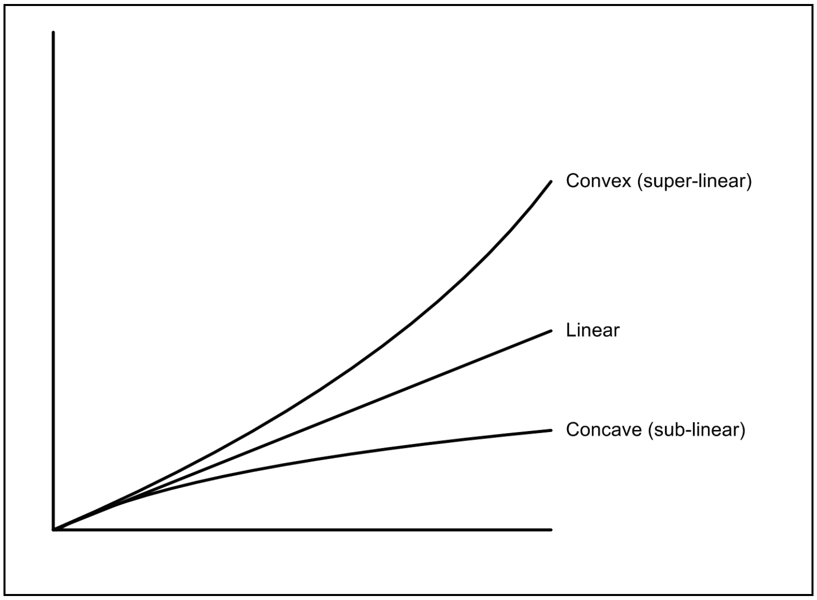
| | |
|---|---|
| 4:45 | Discuss information/presentation needs from NCES for 9/28 |
| 5:00 | Adjourn for the day |

Friday, September 28

| | |
|---|---|
| 9:00 AM | Meet with Commissioner Schneider<br>Summarize initial recommendations<br>Discuss additional action items |
| 10:00 | Formulate initial recommendations on data integration<br>Potential methods<br>Roles played by NCES<br>NISS experiments |
| 10:45 | Break |
| 11:00 | Revise/refine recommendations on configuration |
| 12:00 | Define action items for technical panel and NISS |
| 1:00 | Adjourn |

# Appendix C. Graphical Illustration of Linear, Convex, and Concave Functions

**Figure 15.** **Graphical illustration of linear, convex (superlinear), and concave (sublinear) functions. The function $f(x) = ax^{\alpha}$ is linear when $\alpha = 1$, convex when $\alpha > 1$ and concave when $0 < \alpha < 1$.**



This page intentionally left blank.

## Appendix D. Mapping of Variables in Section 4.2 to ECLS-K

**Table 8. Mapping of variables in section 4.2 to ECLS-K**

| Name in Section 4.3 | ECLS-K Name | Definition |
|---|---|---|
| K-Math | c2mscale | Spring kindergarten math assessment |
| K-Reading | c2rscale | Spring kindergarten reading assessment |
| G1-Reading | c4rrscal | Spring first-grade reading assessment |
| G3-Math | c5r2mscl | Spring third-grade math assessment |
| G5-Math | c6r3mscl | Spring fifth-grade math assessment |
| G5-Reading | c6r3rscl | Spring fifth-grade reading assessment |
| G5-MinorityPercent | g6pmin | Percentage of minority students in fifth-grade class |
| G5-FreeLunch | s6flch_i | Imputed percentage of free-lunch eligible students in fifth-grade class |