Appendix 1-B
APS Data Access Guidelines

Researchers will be able to access data needed for their research through the APS database. However, it is important that we, as APS researchers–individually and as a community, handle APS data responsibly.

This data access guideline is intended to facilitate adherence to IRB research procedures and obligations. **The goal is to minimize the likelihood for accidental data sharing with those for whom such data access may violate conflicts of interests, IRB approved research protocols, and privacy laws.** Toward this end, the procedures presented here are designed to help researchers identify the data type, recognize data handling risks, and acknowledge responsibilities.

1 Data Types

Some APS data sets are more sensitive than others. To help us differentiate, we can classify our APS data sets as 3 different types: raw, anonymized, or aggregated.

> **Raw** data describes individual student attributes or performance in a way that allows a researcher to identify individual students -- by voice, by appearance (picture), by name, by a standardized identifier used in the study (CAEE_ID), or any other similar means. Thus, data files that include the student's CAEE_ID in its filename must also be considered raw. *Raw data is highly sensitive.* It must be carefully secured at all times, and should not be shared with anyone unnecessarily. Typically, only those researchers involved in the specific data collection process would need access to the respective raw data set.

> **Anonymized** data can be prepared from raw data by either stripping all student identifiers, or replacing identifiers with newly created codes which would be unknown to any other researcher. Using a private set of pseudonyms is one way of anonymizing student data. While real student identities cannot be traced, data of individual students can be collected and tracked. Thus, *anonymized data is moderately sensitive.*

> **Aggregated** data is prepared by grouping, processing, categorizing data of multiple students statistically, numerically, or descriptively. In such case, no individual students are distinguishable or identifiable. With any reasonably large grouping, aggregated data is descriptive of its constituent population. No individual student information can be extracted. Thus, *aggregated data is the least sensitive*.

2 Risks and Responsibilities

Prerequisite: all researchers that have (or are expected to have) contact with the study's subjects and/or their data must first complete human subjects research (IRB) training.

Researchers are individually responsible for continuously reviewing their own risks and concerns, when accessing and handling APS data. To avoid accidentally exposing information to an inappropriate recipient, individual researchers should make their own sets of recognized risks known to others. For example, if researcher_X should not receive information_Y because of a particular conflict of interest risk, researcher_Z who has information_Y should be alerted. In this way, APS researchers share responsibility to alert one another.

3   Research Data Access Requests by Category of Researcher:

A data sharing arrangement for the duration of the active grant was developed near the start of the project and is described in general for the three researcher categories below. Any deviation from these data sharing arrangements could occur with permission from the project leads and subsequent final approval from APS lead, Sheri Sheppard. Data sharing following the end of the grant period is outlined in section 5 Data Academic Pathways Study Data Sharing Agreement below.

Category 1. APS School Researcher: School PI approved
Each school PI would enable access to APS researchers at their own campus to their school's data. It would be up to this PI to decide what types (e.g., method, and raw/anonymized, aggregated) each of his/her researchers should have access to, and to ensure that each researcher has completed human subjects training and has been apprised in the APS risks and responsibility statement. New campus researchers would be added by the PI, as needed.
• LONGTERM APPROACH: PIs would enable access for each researcher via a web form. This form would automatically enable access and would post the status of the researcher's access privileges on the database (so that everyone knows of everyone else's access).
• SHORTTERM APPROACH: until the web form is up and active, PIs should send a list of their campus researchers to the database manager.

Category 2. APS Method Researcher: Method lead compiled, Cognizant APS Lead Reviewed.
Each method lead (ethnography, surveys, structured interviews, performance task, academic transcripts, and Cohort 2) would provide cognizant APS Lead PI with a list of researchers who should have access to data associated with a particular method across ALL schools for the purpose of cross-campus analysis on a single method. In general these will be a list of researchers who have responsibility for single-method analysis at all institutions and who are working closely with the method lead. Cognizant APS Lead PI would then enable access to APS researchers to that data type. It would be up to this PI to decide what types (e.g., method, and raw/anonymized, aggregated) each of his/her researchers should have access to, and to ensure that each researcher has completed human subjects training and has been apprised in the APS risks and responsibility statement. New cross-campus researchers would be added by the cognizant APS Lead PI, as needed.
• LONGTERM APPROACH: Cognizant APS Lead PIs would enable access for each researcher via a web form. This form would automatically enable access and would post the status researcher's access privileges on the database (so that everyone knows of everyone else's access).

• SHORTTERM APPROACH: until the web form is up and active, PIs should send a list of their campus researchers to the database manager.

Category 3. APS Researcher.
APS researcher who wants data across all four campuses for the purposes of multi-method and/or multi-institution APS analysis, submits a web form to the APS Lead team. The form would ask that the researcher briefly describe the research question that they are pursuing, how it relates to core APS research questions, and what type of analysis they are undertaking. This form is reviewed on one of the monthly/bi-monthly APS Lead calls. If urgent, the review could happen via email (championed by the campus APS lead). Each school has the right to veto access to his/her school's data. Approved Category 3 web form is then forwarded to database manager who enables access (if there is way to automate this enabling, we will).

4   Recommended Data Handling Guidelines

1. Whenever possible, data should be transferred via the APS database.
2. Unencrypted data should NOT be transferred via unsecured mail or email.
   Sending data via secure mail or email guarantees that only the intended recipient is able to receive and read the contents of the delivered package. For email, this typically means applying digital key encryption and digital signatures on the data package before sending. For mail, this may mean physically locking the package in such a way so that an unintended recipient would be prevented from opening it.
3. All computers and media containing APS data must be carefully secured and protected from compromise by hackers/virus/spyware/Trojan horse programs/network attacks.
4. Here are a few things to review:
   • is your computer secured from theft?
   • is your computer used or accessible by other people?
   • is your computer data files accessible by other users?
   • is your computer ever connected to a network?
   • is your computer protected by a hardware or software network firewall?
   • does your computer run a virus checker with daily or weekly updates?
   • is your computer running a "malware" checker with daily or weekly updates?
   • are your data files being backed up?
   • are your backups protected from theft?
   • are your backups accessible by other people?
5. In general, researchers should avoid sharing any raw APS data, student contact information and even names of students' participation in the study. For example, in cases with faculty researchers, knowledge of study participant's identities may pose a conflict of interest, or a potential conflict of interest in future years of the study. These represent undesirable outcomes. Therefore, assume APS raw data, student participation or contact information is not sharable. Such information should not be shared publicly, or shared in an insecure forum.

5 Academic Pathways Study Data Sharing Agreement

This agreement establishes the policy to address data sharing among the Academic Pathways Study leads during the concluding phase of the grant (beginning with the initial distribution of CD#1 in early 2009) and continuing after the grant ends (September 30, 2009).

1) The APS Core Data set consists of a 4-box CD set:
        CD#1 (Cohort 1) (late 2008/early 2009)
        CD#2 (Cohort 2) (3rd quarter 2009)
        CD#3 (Cohort 3) (first quarter 2009)
        CD#4 (Cohort 4) (2nd quarter 2009)

Each APS co-PI from the four primary schools receives a 4-box CD set of Core APS data. In addition, the APS co-PI from the remaining school receives CD#2 and CD#4.

2) When a new topic/question is being explored with the Core APS data that involves cross-school data, email notification to the APS co-PI team is required. If the topic/question is to be explored using only data from the cognizant APS Lead's school, notification is recommended.

3) Any publications coming out of said research to have an APS co-PI as a co-author. In addition, any and all references to schools should be via the established pseudonyms.

4) Any research using this core data needs to be under the supervision of one of the APS co-PIs. The supervising APS co-PI is responsible for ensuring that proper IRB approval and training have been obtained for all associated researchers.

5) Secondary data (e.g., Cohort 1', NSSE) to be held by cognizant APS co-PI, and requests to use that data to be sent to that APS co-PI.

6) Any publications coming out of said research should include a statement acknowledging NSF Grant No. ESI-0227558.

Supplement A:
Research Data Access (RDA) Request Form (Category 3)

What:

Data Set ID: (see Appendix B) (data_type + school(s) + collection_event_date) [highlight in table]	[*Survey* | *Structured_Interview* | *Performance_Task* | *Exit_Interview* | *Ethnographic_Interview* | *Academic_Record* | [Cohort 2 Interviews…] [*school identifier*] [*Event_ID*] NOTE: Data Set ID would have format similar to APS data filenaming convention.
Data Classification:	[*Raw* | *Anonymized* | *Aggregated*]

Who:

Researcher Name:	

Researcher Institution:	
Researcher Role(s) (relative to data set):	[*APS Data Owner* | *Campus Data Owner* | *APS_Researcher* | *CAEE_Researcher*]
Affiliated CAEE PI (if different than researcher)	
Has researcher completed IRB training?	(circle one) YES NO

Duration:

Date when data set access will end:	

Benefits:

Briefly describe analysis plan for use of data set:
Briefly explain expected outcome of analysis, why data access is beneficial or required:

Risks:

Identify potential risks associated with researcher's data access and describe how these are (or will be addressed):
Identify persons who should not see data set(s), and why, if any:

Expected Outcomes:

What type of analysis results can or will be made available to others?
    Analysis results data file format (i.e. software applications used)?
    Data type of results: raw, anonymized, or aggregated?