

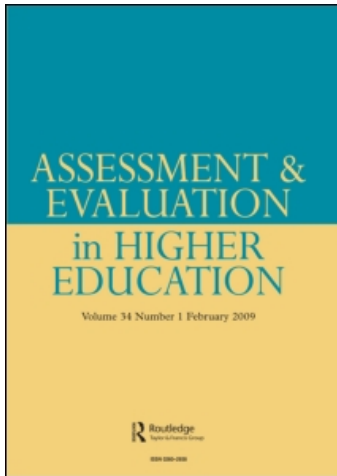
This article was downloaded by: [University of Florida]

On: 15 July 2009

Access details: Access Details: [subscription number 906459425]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Assessment & Evaluation in Higher Education

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713402663>

Using NCES national datasets for evaluation of postsecondary issues

Debbie L. Hahs-Vaughn^a

^a University of Central Florida, USA

Online Publication Date: 01 June 2007

To cite this Article Hahs-Vaughn, Debbie L.(2007)'Using NCES national datasets for evaluation of postsecondary issues',Assessment & Evaluation in Higher Education,32:3,239 — 254

To link to this Article: DOI: 10.1080/02602930600896373

URL: <http://dx.doi.org/10.1080/02602930600896373>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Using NCES national datasets for evaluation of postsecondary issues

Debbie L. Hahs-Vaughn*

University of Central Florida, USA

The purpose of this study is to review 10 NCES databases that can be used for researching postsecondary issues and provide lesser-known facts to using the datasets that are important but may not be widely understood. Issues addressed include: (1) access; (2) statistical issues; (3) database nuances; and (4) database training opportunities. A concise review of each database is also provided which includes: (1) a general overview of the survey; (2) formats in which the dataset is available; and (3) research areas (which include key variables that can be used as a basis for research themes along with examples of how the dataset has been used to answer research questions). The databases provide rich sources of information for national as well as international comparative analysis studies.

Introduction

Many US federal agencies have collected data from a wide array of individuals and organizations on a multitude of topics. The National Center for Education Statistics (NCES) has seven survey and program areas for which they have collected data or have reports on topics from elementary education to libraries. Within the research area of postsecondary studies alone, NCES has 10 survey databases. While the data has been collected on entities within the US, the datasets hold value for international researchers as well (although this has been seemingly an untapped resource internationally). Recognizing this, a crosswalk has been developed between the International Standard Classification of Education (ISCED) and the Integrated Postsecondary Education Data System (IPEDS; reviewed herein) and the Common Core of Data (CCD) to increase accessibility and utility of US data to the worldwide community (Hunt, 1993). Internationally, ISCED is the recognized standard for reporting and interpreting education program data. In addition to providing rich data for conducting comparative studies, international scholars may be interested in using one or more datasets to analyze internationalization in higher education in the US (e.g., who are

*University of Central Florida, PO Box 161250, Orlando, FL 32816-1250, USA. Email: dhahs@mail.ucf.edu

the students who enroll in US higher education and where do they call home; what are the trends of enrollment by international students; what are the characteristics of institutions that enroll large numbers of international students).

Milam (1999) conducted an extensive review of the national datasets available for faculty studies. In that same vein, the purpose of this study is to review ten NCES databases that can be used for researching and adding to our knowledge of postsecondary issues as well as provide modifying details and lesser-known facts to using the datasets that are important but may not be widely known or published. Details that will be expounded include: (1) access; (2) statistical issues; (3) database nuances; and (4) database training opportunities. A concise review of each database is also provided which includes: (1) a general overview of the survey; (2) formats in which the dataset is available; and (3) key variables (which can be used as a basis for research themes).

Modifying details and little known facts

Access to national datasets has increased dramatically through Internet availability of studies (Milam, 1999). With the multitude of datasets available, however, it may be difficult to discern which dataset is most appropriate for a particular research question. Beyond that, understanding the intimate characteristics of the datasets is also important (Milam, 1999). It is anticipated that most researchers who have used a national dataset will confirm that there are lots of lesser-known facts in using the data. Introducing these issues is not meant to discourage but to enlighten prospective users. Modifying details and lesser-known facts of using national datasets presented herein include: (1) access; (2) statistical issues; (3) dataset nuances; and (4) training opportunities.

Access

Most NCES postsecondary databases are available through public access, the Data Analysis System, and restricted use. The Data Analysis System (DAS) is a web format or Windows software application that provides public access to NCES survey data (NCES, n.d. a). There is a separate DAS for each survey available through the system. The postsecondary surveys accessible through the DAS web site include: Baccalaureate and Beyond (B&B), Beginning Postsecondary Students Longitudinal Study (BPS), High School and Beyond (HS&B), National Household Education Survey (NHES), National Longitudinal Study of the High School Class of 1972 (NLS), National Postsecondary Student Aid Study (NPSAS), and the National Study of Postsecondary Faculty (NSOPF).

DAS users can generate tables of percentages, means, or correlation coefficients by choosing the DAS variables (based on survey questionnaire items) that they would like to appear in a table. The file produced provides the table numbers and the corresponding standard errors that have been calculated taking into account the complex sampling procedures used in the surveys. If you prefer not to download your public

access data from the DAS web site, most surveys still offer the data on free CD-ROMs that can be ordered through the NCES web site.

If the DAS system does not provide the detail of data needed for your analysis, some datasets (e.g., NSOPF:93) have public microdata CD-ROMs available that provide access to a limited number of variables. To protect confidentiality, public microdata excludes a large number of variables so that the identification of a particular individual is impossible. When available, however, public microdata is a great tool for graduate students to use in class projects, theses, and dissertations. These files also provide researchers an opportunity to get acclimated to the dataset, conduct preliminary analyses, and determine if the licensed data is required.

Due to NCES confidentiality legislation, access to complete raw data files is available only through a restricted data license, for which researchers must submit an application. Information on applying for a restricted license is detailed on the NCES web site, along with a listing of the datasets that can be obtained via license (NCES, n.d. b). Very user-friendly, the site also provides sample license agreements. Included in the application process for a restricted license must be the name of the database, purpose of the loans of the data, length of time for use, a security plan, list of authorized users, and an affidavit from each person using the data for confidentiality. The NCES postsecondary databases for which restricted licenses are possible include: B&B, BPS, HS&B, NHES, NPSAS, NSOPF, Postsecondary Education Quick Information System (PEQIS) and Recent College Graduates Study (RCG). Should you apply for a restricted license, anticipate the process taking a few months.

Statistical issues

There are a few fundamental statistical issues to consider before plunging into using national datasets. For example and related to access to the datasets, while the DAS makes access to the data very easy, will your research questions be answerable through percentages, means, or correlation coefficients? If not, you will need to apply (if possible) for a license to obtain restricted use data that will provide the raw data from the surveys or obtain (if available) public microdata.

By nature, national datasets are complex in design because they employ sampling designs that are usually multistage, cluster, or stratified—or a combination of these sampling techniques. In addition, a specific subset of the population is often over-sampled to ensure good representation and thus more accurate parameter estimates (Kaplan & Ferguson, 1999; Thomas & Heck, 2001; Stapleton, 2002). For example, the NSOPF:93 oversampled faculty that were full time women, Black non-Hispanics and Hispanics, Asian/Pacific Islanders, and faculty in four National Endowment for the Humanities disciplines (Selfa *et al.*, 1997). If you are using the public microdata or licensed data, it is important to understand weights and design effects to ensure the data are used correctly—i.e., that the analyses has been accommodated for oversampling and homogenous clusters that may exist due to multi-stage sampling. Ignoring disproportionate sampling may result in biased parameter estimates and poor performance of test statistics and confidence intervals (Pfefferman, 1993) as the weights are

required to produce estimates which are representative of the intended population (US Department of Education, 2002). Because of the similarities within clusters, ignoring the complex sampling design results in the assumption of independence being negated (Kish & Frankel, 1974) and the true population variance being underestimated (Selfa *et al.*, 1997; Hox, 1998).

Because most of the datasets are not simple random samples, weights are used to ensure the data is representative of the population. Using weights to adjust for over-sampling is many times the easiest way to incorporate unequal probability of selection (Stapleton, 2002). Longitudinal datasets, such as the BPS, have numerous weights from which to select depending on what data you have used (e.g., cross sectional or longitudinal). While intended to be nationally representative of some specific population, failing to employ proper techniques to compensate for the complex sampling design creates analyses that reflect only the sample and not the intended population (Thomas & Heck, 2001; Hahs, 2003).

Given that most NCES sample designs have stratification and/or clustering, it is important to ensure that appropriate techniques for the estimation of variance in sample surveys are identified, implemented, and documented. By nature, a model-based approach (e.g., multilevel structural equation modeling or hierarchical linear modeling) directly incorporates clustering in the analyses and negates the need to deal with clustered subsets (Thomas & Heck, 2001). However, many researchers do not utilize multilevel models. Employing a design effect, therefore, is also required when single level analyses (e.g., ANOVA, regression, single-level structural equation modeling) is performed. Calculating the average design effect is often left to the researcher since design effects for all variables are not included in the technical reports and no single design effect is universally applicable to any given survey or analyses (NCES, n.d. c). Failure to use weights and design effects will confound the results of your analyses (Thomas & Heck, 2001; Hahs, 2003).¹

Dataset nuances

There are many nuances of datasets as you work from one to others. While all are comprehensive, some datasets (e.g., datasets that have wider use) tend to have more reader-friendly codebooks and technical manuals. For example, the National Education Longitudinal Study of 1988 (NELS:88), which is not a postsecondary dataset, has very detailed technical reports and abundant information for logistics on how to use the database (e.g., how to estimate design effects) (Huang *et al.*, 1996) and is written in a more user-friendly manner than some reports. In addition, databases that are more frequently used may provide an advantage in that problems or issues that you run into may have already been addressed and/or resolved.

Some longitudinal datasets have the initial year of data drawn from a different database. For example, both the B&B and the BPS have the NPSAS as the database from which the first wave of data was drawn. This may affect the variables of interest (for example, attitudinal items may be available only in waves beyond the base year).

As most researchers will anticipate, variables usually require extensive recoding as alphanumeric and inappropriately scaled Likert-type items abound (e.g., a value of one is ‘strongly agree’ and a value of five is ‘strongly disagree’). Understanding missing values may also require critical inquiry skills. Researchers may find that there are multiple missing data—missing, missing skip, missing don’t know, etc. Even from the actual survey and codebook, it is sometimes challenging to completely discern what ‘missing’ means (e.g., if a value is missing completely at random or if the respondent didn’t know or skipped a question).

Training in the use of NCES datasets

In cooperation with NCES, national professional organizations are promoting the use of NCES datasets by offering training on the use of NCES datasets. The most comprehensive is offered by the Association for Institutional Research. The AIR/NCES/NSF Summer Data Policy Institute is offered each year. Prospective fellows submit a short proposal, and if accepted, spend two weeks in intensive study in the Washington DC area learning about the NSF and NCES datasets, how to access the resources and how to use them, and hearing from leading national policy experts in special seminars. Fellows receive transportation, hotel accommodations, and per diem for meals and incidental expenses as well as a one-year membership in AIR (AIR, n.d.). NCES also sponsors a number of database-specific workshops throughout the year (NCES, n.d. d).

The American Educational Research Association also sponsors professional development in national datasets in conjunction with its annual meeting. For example, for the April 2004 annual meeting, participants were offered the opportunity to apply to attend a three-day meeting on the Early Childhood Longitudinal Studies and the National Household Education Survey. The sessions include overviews of the design of the studies, computer software demonstrations, and discussions of technical issues (AERA, n.d.).

Review of datasets

The NCES postsecondary datasets that are reviewed include: (1) Baccalaureate and Beyond (B&B); (2) Beginning Postsecondary Students Longitudinal Study (BPS); (3) High School and Beyond (HS&B); (4) Integrated Postsecondary Education Data System (IPEDS); (5) National Household Education Survey (NHES); (6) National Longitudinal Study of the High School Class of 1972 (NLS); (7) National Postsecondary Student Aid Study (NPSAS); (8) National Study of Postsecondary Faculty (NSOPF); (9) Postsecondary Education Quick Information System (PEQUIS); and (10) Recent College Graduates Study (RCG). While an extensive review of each database is detailed in Russell and Winter (2002), a more concise summary is provided which includes: (1) a general overview of the survey; (2) formats the datasets is available; and (3) research areas including key variables (which can be used as a basis for research themes) and examples of research questions that have been

answered with the datasets. While many of the datasets have been analyzed extensively within US contexts, they hold rich promise for international comparisons and have seemingly gone untapped in this arena. One goal of this study is to expose the international audience to and assist in broadening the utility of this data to international researchers. While not detailed herein, some examples of the use of the data with various statistical procedures include regression (see Beasley, 2002), logistic regression (see Cofer & Somers, 2000; Diamond *et al.*, 2000), multinomial logit (see Perna, 2003a), single-level structural equation modeling (see Hahs-Vaughn, 2004), and multilevel structural equation modeling (see List & Wolfle, 2000).

To assist in international interpretation, a crosswalk (similar to that of Hunt, 1993) of the datasets is included in Table 1. Readers are referred to Hunt (1993) for extensive detail on IPEDS.

Baccalaureate and Beyond (B&B)

Overview. The B&B is a cross sectional and longitudinal survey that uses the NPSAS:93 as its base year with follow-ups in 1994, 1997 and 2003. Nationally representative of baccalaureate degree completers, the B&B provides information on education and work experiences after completing the bachelor's degree and on entry into, persistence and progression through, and completion of graduate level education following up on the cohort over a 12-year period (NCES, n.d. e).

Format. The B&B is available in both public access format (DAS) and restricted license.

Research areas. Key variables relate to major field of study in college, financial aid, degree and employment expectations, job search strategies, employment after degree, family responsibilities, community service and loan repayment. A special emphasis of the B&B is on career paths and patterns of elementary and secondary teachers with variables that look at expectations for entering the teaching profession, career path as a teacher, and teacher certification. Variables within B&B can be applied in a variety of studies. For example, Millett (2003) examined how decisions of recent bachelor's degree recipients to enroll in post-baccalaureate study (graduate or professional school) were related to student background characteristics, collegiate performance and financial debt. Joy (2003) analyzed salary gaps as a function of gender, college major, grades, college credits and postsecondary institution.

Beginning Postsecondary Students Longitudinal Study (BPS)

Overview. The BPS is a cross sectional and longitudinal study that uses the NPSAS as its base year with follow-ups in subsequent years. There are two BPS studies: BPS:90/92/94 and BPS:96/98/01. The BPS is nationally representative of all individuals who entered postsecondary education and collects data on persistence of

Table 1. Crosswalk for NCES postsecondary datasets

Dataset	Type of program ¹	Education level(s) of respondents ²	Access
B&B	General	First university degree	DAS and restricted license
BPS	General	Award not equivalent to a first university degree through postgraduate university degree or equivalent	DAS and restricted license
HS&B (sophomore cohort)	General	Grade 10 through postgraduate university degree or equivalent	DAS and restricted license
HS&B (senior cohort)	General	Grade 12 through postgraduate university degree or equivalent	DAS and restricted license
IPEDS	General	Institution level data collected on institutions that award degrees ranging from award not equivalent to a first university degree through postgraduate university degree or equivalent	Online (guest level access and peer analysis system)
NHES	General	Pre-kindergarten through postgraduate university degree or equivalent	DAS and restricted license
NLS	General	Grade 12 through postgraduate university degree or equivalent	DAS (through B&B and BPS) and CD-ROM
NPSAS	General	Award not equivalent to a first university degree through postgraduate university degree or equivalent	DAS and restricted license
NSOPF	General	First university degree or equivalent through postgraduate university degree or equivalent*	DAS and restricted license
PEQUIS	General	Institution level data collected on institutions that award degrees ranging from first university degree or equivalent through postgraduate university degree or equivalent	Restricted license and limited online access
RCG	Teacher training	First university degree or equivalent through post-graduate university degree or equivalent	Restricted license

Notes: (1) *General* refers to data collected on activities not defined by a specific program or subject area such as data on enrollments as a certain grade or degree level. *Teacher training* refers to data collected on college graduates qualified to teach. These programs correspond to ISCED broad field codes of 01 and 14 respectively. (2) ISCED classifications for education level include the following: (a) 1st level, level 0: pre-kindergarten and kindergarten; (b) 1st level, level 1: Grades 1–6; (c) 2nd level, 1st stage, level 2: Grades 7–9; (d) 2nd level, 2nd stage, level 3: Grades 10–12; (e) 3rd level, 1st stage, level 5 (award not equivalent to a first university degree); postsecondary award, certificate, or diploma, associate's degree; (f) 3rd level, 1st stage, level 6 (first university degree or equivalent): bachelor's degree or post-baccalaureate certificate; (g) 3rd level, 2nd stage, level 7 (postgraduate university degree or equivalent): master's degree, post-master's certificate, doctor's degree, first-professional degree, first-professional certificate (post degree). Because NCES datasets often have collected data that spans multiple levels, the table includes the range of education levels available within the dataset.

postsecondary education, relationships between work and education, and other outcomes of postsecondary education (e.g., enrollment in graduate or professional school, degree aspirations four years after entering college) (NCES, n.d. f).

Format. The BPS is available in both public access format (DAS) and restricted license.

Key variables. Key variables relate to persistence and degree attainment, financial aid and financing college, and aspirations for education. The BPS dataset has been used to evaluate various postsecondary student issues such as differences between first-generation and non first-generation students on characteristics they have upon entering college, experiences during college, and outcomes four years after entering college (see Hahs-Vaughn, 2004) and the frequency of part time enrollment and the persistence of students who enroll part time (O'Toole *et al.*, 2003).

High School and Beyond (HS&B)

Overview. The HS&B is one of three studies of the National Education Longitudinal Studies (NELS) program and followed two cohorts of students: the 1980 high school senior class and the 1980 sophomore class. Both cohorts were surveyed every two years through 1986. In addition, the sophomore class was surveyed again in 1992 (NCES, n.d. g).

Format. The HS&B is available in both public access format (DAS) and restricted license.

Research areas. Key variables relate to educational attainment, employment, and family structure in addition to various life activities (such as voting patterns). Betts and Grogger (2003) used the HS&B to evaluate the effects of grading standards on student academic achievement, educational attainment, and earnings. Perna (2003b) used the HS&B to identify the earning premium attributed to a bachelor's degree and then analyzed how the earning premium varied based on characteristics such as ability, experience and occupation as well as gender, race and socioeconomic status. The HS&B has also been used to examine the labor market value of a GED for women as compared to high school dropouts with no GED (Tyler *et al.*, 2003).

Integrated Postsecondary Education Data System (IPEDS)

Overview. IPEDS was established as the foundation for the postsecondary databases of NCES and collects data from postsecondary education providers (academic, vocational and continuing education). This dataset collects information from both public and private institutions (NHES, n.d. h).

Format. IPEDS is available in zipped files online in a downloadable system. A peer analysis system is also available online that allows comparison of an institution to comparable peer institutions. Institutions who have submitted data have the highest level of authorization for extracting data from the peer analysis system. A guest level access is available that permits the public to view adjudicated data.

Research areas. Key variables relate to student enrollment and completion, institutional characteristics (e.g., housing), faculty, staff and financial data. The IPEDS data has been used, for example, to build a classification system for two-year institutions based on variables related to enrollment, student demographics and institutional characteristics (McCormich & Cox, 2003) and to examine funding of public institutions (Schuh & Shelley, 2001). A more detailed accounting of how IPEDS can be used to compare postsecondary institutions is available from Schuh (2002).

National Household Education Surveys Program (NHES)

Overview. The NHES dataset collects information on educational activities and experiences of persons in the US and covers the whole range of ages from early childhood learning to adult learning. Educational trends are monitored by collecting measures of the same phenomena in various years and one-time surveys are also conducted (NCES, n.d. i).

Format. NHES is available in both public access format (DAS) and restricted license.

Research areas. NHES surveys have focused on adult education, before- and after-school programs, civic engagement, participation in early childhood programs, household library use, parent and family involvement, school readiness, and school safety. The NHES was used to determine factors (e.g., participation in cultural activities, attendance at center-based programs such as Head Start, involvement by parents with their child, and sociodemographic risk variables such as parent's education levels) that influenced cognitive readiness in preschoolers (Beasley, 2002). Other researchers have used the NHES to examine parents' conceptions of their child's readiness for kindergarten and parents' decision to delay their child's entry into kindergarten as predicted by family background (e.g., parents' education and race) and home activities (e.g., watching educational programming on television and learning at home) (Diamond *et al.*, 2000).

National Longitudinal Study of the High School Class of 1972 (NLS)

Overview. NLS is the 'grandmother' of NCES longitudinal studies and is considered 'probably the richest archive ever assembled on a single generation of Americans'

(NCES, n.d. e). Survey participants generally completed the initial survey as seniors in high school in the spring of 1972. Follow-up surveys were conducted in 1973, 1974, 1976, 1979 and 1986 and NLS also includes high school records and postsecondary transcripts that were collected in 1984. This dataset has been used extensively for cohort comparisons of later generations (e.g., HS&B) (NCES, n.d. j).

Format. NLS is available in public access format (DAS) through the B&B and BPS surveys. For the complete NLS dataset, a CD-ROM can be purchased from NCES.

Research areas. Key variables relate to attendance and achievement. This study has been used to evaluate the impact of father's presence in a household, student's ability, high school curriculum and high school grades on postsecondary educational attainment and how this differs based on race (List & Wolfe, 2000). Using the NLS, Trent (1997) examined how school desegregation impacted later earnings and occupations of Black students and Riordan (1994) analyzed how female students who attended a women's college differed on education, occupation and income as compared with female students attending coeducational colleges.

National Postsecondary Student Aid Study (NPSAS)

Overview. The NPSAS collects data on how postsecondary education is paid by students and families, gathering data from students who receive financial aid as well as those who do not. The nationally representative sample includes all levels of students (undergraduates through professional) and students attending all types of institutions (public or private, less than two-year to major universities) (NCES, n.d. k).

Format. NPSAS is available in both public access format (DAS) and restricted license.

Research areas. Key variables relate to the cost of education, distribution of financial aid and traits of postsecondary students and families. Student and parent interviews were also conducted in some years. Heller (2001) analyzed data from the NPSAS to determine how institutional awards of need- vs. non-need-based grants to undergraduate students had changed, to determine the socioeconomic characteristics of students receiving the grants, and to determine how institutional and student characteristics can predict receipt of a grant. The NPSAS has also been used to evaluate persistence in postsecondary programs. Persistence at two-year colleges was analyzed as a determinant of student background (e.g., age and dependent status), aspirations (e.g., aspiration for a college degree and advanced degree), high school degree attainment, college experiences (e.g., residency of student, degree- or non-degree-seeking status, and GPA), cost of college (e.g., tuition and fees, grants and scholarships) and college debtload (high or low debt threshold) (Cofer & Somers, 2000).

National Study of Postsecondary Faculty (NSOPF)

Overview. The NSOPF collects data on full time and part time faculty and instructors at postsecondary institutions and was created in response to a need to better understand the individuals who were directly impacting postsecondary education. This study is ‘the most comprehensive study of faculty in postsecondary educational institutions ever undertaken’ (NCES, n.d. l).

Format. The NSOPF is available in both public access format (DAS) and restricted license.

Research areas. Key variables relate to background of faculty and instructors, workloads, teaching methods, salaries and benefits, attitudes, and future plans. Data was also collected from institutional and department levels (1988 was the only year department level data was collected) on faculty composition, turnover, recruiting efforts, retention and policies toward tenure. Various questions were addressed by Toutkoushian and Bellas (2003) with the NSOPF including how gender and factors such as marital status, parental status, and non-employment income affect preference for part time faculty employment and how gender, part time status, and other factors influence faculty satisfaction with benefits, salary and job. Perna (2003a) used the NSOPF data to examine gender and racial relationships to employment outcomes (e.g., type of institution employed, employment status, salary, academic rank, tenure status) controlling for human capital, structural and market explanations for the relationships.

Postsecondary Education Quick Information System (PEQIS)

Overview. PEQIS was developed because of the need for a survey venue that would allow for quick data collection, and it is also used to test the feasibility of large-scale data collection on specific topics or as supplements to other postsecondary surveys. The PEQIS allows for quick and timely data to be collected with minimal burden on the respondents by using a standing sample of approximately 1600 two-year and four-year postsecondary institutions that is nationally representative of public and private bachelors, masters and doctoral degree granting institutions. PEQIS can also conduct surveys at the state level for higher education (NCES, n.d. m).

Format. PEQIS is available in restricted license format, and some select PEQIS data is downloadable from the web site including Occupational Programs and the Use of Skill Competencies at the Secondary and Postsecondary Levels, 1999, and Distance Education at Postsecondary Education Institutions, 1997–98.

Research areas. Key variables relate to distance education, remedial education, occupational programs, campus crime and security, and students with disabilities. The

PEQUIS has been used by researchers outside of NCES to a much lesser extent than the other databases. However, published reports by NCES that capture the vitality of this database are available. For example, Lewis and Farris (1995) presents results from postsecondary institutions on how changes in the 1992 Higher Education Act reauthorization on federal student financial aid impacted institutional verification policies and applicant verification. National estimates on distance education (e.g., number of distance education courses, distance education course enrollments, accommodations for students with disabilities in distance education) at two- and four-year institutions are provided in Waits (2003).

Recent College Graduates Study (RCG)

Overview. The RCG focuses on college graduates (bachelors and masters level degree recipients from US institutions) who are qualified to teach at the elementary and secondary levels. The RCG was conducted sporadically from 1976 to 1991, and in 1993, the B&B replaced the RCG (NCES, n.d. n).

Format. RCG is available in restricted license only.

Research areas. Key variables relate to post-degree employment immediately after postsecondary degree attainment and characteristics of that employment (including income), field of study and financial aid. Henke and Zahn (2001) used data from the RCG to determine if teachers were more or less likely than workers in other occupations to leave the workforce or work in a different occupation.

Conclusion

While a comprehensive review of the NCES datasets available for postsecondary research is available (see Russell & Winter, 2002), a more concise summary of the postsecondary datasets along with analytical issues and lesser-known facts for using the datasets and examples of research questions that have been analyzed with the extant data is not available in a single source. With easy access to NCES national datasets (including through the DAS as well as public microdata), understanding the options available for postsecondary research, key research areas that each focus on, and the formats available will be helpful to national and international researchers as well as policy analysts. Possibly more important however is having an upfront understanding of some of the issues, limitations and lesser-known facts of the datasets. The importance of attending to the details (e.g., addressing the issue of weighting and design effects) cannot be understated however it is often overlooked. A review of research articles that utilized data from national datasets ($N = 34$) published between 1999–2003 in *Research in Higher Education*, *Journal of Higher Education* and *Review of Higher Education* found that 32% ($n = 11$) applied weights to the analysis and provided sufficient detail for the reader to discern appropriate application and only 12% ($n = 4$)

noted whether and how design effects were used (Hahs-Vaughn, in press). Attendance at a respective training institute for the database helps acquaint the researcher with these statistical aspects and also provides the opportunity to work with the data so that nuances (e.g., recoding) become evident. As described herein, a number and variety of database training sessions are offered each year by NCES and other research organizations (e.g., AERA and AIR) in order to build a league of scholars equipped with the skills to address policy and applied research using the datasets.

The databases previewed provide a wealth of data that will address a near infinite number of research questions relating to postsecondary issues appropriate for and of interest to both national and international scholars. While many of the datasets have been analyzed within US contexts, the data provide a rich resource for comparative studies for scholars interested in international issues and are essentially untapped in that respect. This study will assist in fostering the enlightened use of NCES postsecondary databases nationwide and worldwide.

Notes

1. While beyond the scope of this manuscript, review of Thomas and Heck (2001) and Hahs-Vaughn (2005) in conjunction with the dataset technical manuals and statistical software manuals make understanding and using weights and design effects manageable. In addition, there are several excellent studies that have reviewed weights and design effects within the methodology for specific statistical procedures (e.g., regression, structural equation modeling) using simulation (see Korn & Graubard, 1995; Stapleton, 2002) and using extant data (see Hahs, 2003) in relation to national studies. To produce valid and reliable results, therefore, it is critical to understand how the data is collected in order that the correct statistical accommodations be made (i.e., weights and design effects).

Notes on contributor

Debbie Hahs-Vaughn joined the Department of Educational Research, Technology and Leadership within the College of Education at the University of Central Florida in August 2003. She completed her Ph.D. in educational research at the University of Alabama. Her current teaching load includes courses in quantitative statistics and research methods. Her research interests related to complex surveys include methodological issues associated with applying quantitative statistical methods to the data and the application of complex data to studying research questions within education. Other research interests include using evaluation methods to understand and strengthen programs, particularly at the post-secondary level, and understanding practitioner use of research and quantitative statistics to inform their practice.

References

- American Educational Research Association (n.d.) *Annual meeting 2004*. Available online at: www.aera.net/meeting/profdev/index.html (accessed 24 December 2003).

- Association for Institutional Research (n.d.) *Summer data policy institute*. Available online at: www.airweb.org/page.asp?page=42 (accessed 24 December 2003).
- Beasley, M. T. (2002) Influence of culture-related experiences and sociodemographic risk factors on cognitive readiness among preschoolers, *Journal of Education for Students Placed at Risk*, 7(1), 3–24.
- Betts, J. R. & Grogger, J. (2003) The impact of grading standards on student achievement, educational attainment, and entry-level earnings, *Economics of Education Review*, 22(4), 343–352.
- Cofer, J. & Somers, P. (2000) Within-year persistence of students at two-year colleges, *Community College Journal of Research & Practice*, 24(10), 785–807.
- Diamond, K. E., Reagan, A. J. & Bandyk, J. E. (2000) Parents' conceptions of kindergarten readiness: Relationships with race, ethnicity, and development, *Journal of Educational Research*, 94(2), 93–100.
- Hahs, D. L. (2003) *The utilization of sample weights in structural equation modeling: An application using the Beginning Postsecondary Students Longitudinal Study 1990/92/94*. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.
- Hahs-Vaughn, D. L. (2004) The impact of parents' education level on college students: An analysis using the Beginning Postsecondary Students Longitudinal Study 1990–92/94, *Journal of College Student Development*, 45(5), 483–500.
- Hahs-Vaughn, D. L. (2005) A primer for using and understanding weights with national datasets, *Journal of Experimental Education*, 73(3), 221–248.
- Hahs-Vaughn, D. L. (in press) Weighting omissions and best practices when using national datasets in educational research, *Association for Institutional Research Professional File*.
- Heller, D. E. (2001) Race, gender, and institutional financial aid awards, *Journal of Student Financial Aid*, 31(7), 7–24.
- Henke, R. R. & Zahn, L. (2001) Attrition of new teachers among recent college graduates: Comparing occupational stability among 1992–93 graduates who taught and those who worked in other occupations, *Education Statistics Quarterly*, 3(2), 69–76.
- Hox, J. J. (1998) Multilevel modeling: when and why, in: I. Balderjahn, R. Mathar & M. Schader (Eds) *Classification, data analysis, and data highways* (New York, Springer Verlag), 147–154.
- Huang, G., Salvucci, S., Peng, S. & Owings, J. (1996) *National Educational Longitudinal Study of 1988 (NELS:88) research framework and issues*. Working paper No. 96–03 (Arlington, VA, Synetics for Management Decisions).
- Hunt, E. S. (1993) *A guide to the international interpretation of US education program data: CIP, IPEDS, CCD and ISCED*. (NCES Publication No. OR-93-3223) (Washington, DC, US Department of Education, National Center for Education Statistics).
- Joy, L. (2003) Salaries of recent male and female college graduates: educational and labor market effects, *Industrial and Labor Relations Review*, 56(4), 606–621.
- Kaplan, D. & Ferguson, A. J. (1999) On the utilization of sample weights in latent variable models, *Structural Equation Modeling*, 6(4), 305–321.
- Kish, L. & Frankel, M. P. (1974) Inference from complex samples, *Journal of the Royal Statistical Society, Ser. B*, 36, 1–37.
- Korn, E. L. & Graubard, B. I. (1995) Examples of differing weighted and unweighted estimates from a sample survey, *The American Statistician*, 49(3), 291–295.
- Lewis, L. & Farris, E. (1995) *Outcomes of recent changes in federal student financial aid. Statistics in brief* (NCES Publication No. NCES-96-231) (Washington, DC, National Center for Education Statistics).
- List, J. H. & Woffle, L. M. (2000) The effect of a father's presence on postsecondary educational attainment among Whites and Blacks, *Research in Higher Education*, 41(5), 623–636.
- McCormick, A. C. & Cox, R. D. (Eds) (2003) Classification systems for two-year colleges, *New Directions for Community Colleges*, 122, 7–15.
- Milam, J. (1999) Using the national datasets for faculty studies, *Association for Institutional Research Professional File*, 70, 1–19.

- Millett, C. M. (2003) How undergraduate loan debt affects application and enrollment in graduate or first professional school, *Journal of Higher Education*, 74(4), 386–427.
- National Center for Education Statistics (n.d. a) *Data analysis system*. Available online at: <http://nces.ed.gov/dasol/> (accessed 7 May 2004).
- National Center for Education Statistics (n.d. b) *Electronic catalog*. Available online at: <http://nces.ed.gov/pubsearch/licenses.asp> (accessed 24 December 2004).
- National Center for Education Statistics (n.d. c) *Statistical standards*. Available online at: http://nces.ed.gov/statprog/2002/std5_2.asp (accessed 11 July 2003).
- National Center for Education Statistics (n.d. d) *Conference/training connection*. Available online at: www.nces.ed.gov/conferences/ (accessed 7 May 2004).
- National Center for Education Statistics (n.d. e) *Baccalaureate and Beyond*. Available online at: www.nces.ed.gov/surveys/b&b/ (accessed 24 December 2004).
- National Center for Education Statistics (n.d. f) *Beginning Postsecondary Students Longitudinal Study*. Available online at: www.nces.ed.gov/surveys/bps/ (accessed 24 December 2004).
- National Center for Education Statistics (n.d. g) *High School and Beyond*. Available online at: www.nces.ed.gov/surveys/hsb/ (accessed 24 December 2004).
- National Center for Education Statistics (n.d. h) *Integrated Postsecondary Education System*. Available online at: www.nces.ed.gov/ipeds/ (accessed 24 December 2004).
- National Center for Education Statistics (n.d. i) *NHES: National Household Education Surveys Program*. Available online at: www.nces.ed.gov/nhes/ (accessed 24 December 2004).
- National Center for Education Statistics (n.d. j) *National Longitudinal Study of 1972*. Available online at: www.nces.ed.gov/surveys/nls72/ (accessed 24 December 2004).
- National Center for Education Statistics (n.d. k) *National Postsecondary Study Aid Study (NPSAS)*. Available online at: www.nces.ed.gov/surveys/npsas/overview.asp (accessed 24 December 2004).
- National Center for Education Statistics (n.d. l) *National Study of Postsecondary Faculty*. Available online at: www.nces.ed.gov/surveys/nsopf/ (accessed 24 December 2004).
- National Center for Education Statistics (n.d. m) *PEQIS: Postsecondary Education Quick Information System*. Available online at: www.nces.ed.gov/surveys/peqis/
- National Center for Education Statistics (n.d. n) *Recent College Graduates*. Available online at: www.nces.ed.gov/surveys/rcg/ (accessed 24 December 2004).
- O'Toole, D. M., Stratton, L. S. & Wetzels, J. N. (2003) A longitudinal analysis of the frequency of part time enrollment and the persistence of students who enroll part time, *Research in Higher Education*, 44(5), 519–537.
- Perna, L. W. (2003a) The status of women and minorities among community college faculty, *Research in Higher Education*, 44(2), 205–241.
- Perna, L. W. (2003b) The private benefits of higher education: an examination of the earnings premium, *Research in Higher Education*, 44(4), 51–72.
- Pfeffermann, D. (1993) The role of sampling weights when modeling survey data, *International Statistical Review*, 61(2), 317–337.
- Riordan, C. (1994) The value of attending a women's college: education, occupation, and income benefits, *Journal of Higher Education*, 65(4), 486–510.
- Russell, A. B. & Winter, S. B. (2002) *Compendium of national data sources on higher education* (Denver, CO, State Higher Education Executive Offices). Available online at: www.sheeo.org (accessed 9 July 2003).
- Schuh, J. H. (2002) The integrated postsecondary education data system, *New Directions for Higher Education*, 118, 29–38.
- Schuh, J. H. & Shelley, M. C. (2001) A longitudinal analysis of funding for student affairs in public institutions, *Journal of College Student Development*, 42(5), 456–464.
- Selfa, L. A., Suter, N., Myers, S., Koch, S., Johnson, R. A., Zahs, D. A., Kuhr, B. D. & Abraham, S. Y. (1997) *1993 National Study of Postsecondary Faculty (NSOPF:93) methodology report*

- (NCES Publication No. 97-467) (Washington, DC, US Department of Education, National Center for Education Statistics).
- Stapleton, L. M. (2002) The incorporation of sample weights into multilevel structural equation models, *Structural Equation Modeling*, 9(4), 475–502.
- Thomas, S. L. & Heck, R. H. (2001) Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs, *Research in Higher Education*, 42(5), 517–540.
- Toutkoushian, R. K. & Bellas, M. L. (2003) The effects of part time employment and gender on faculty earnings and satisfaction, *The Journal of Higher Education*, 74(2), 172–195.
- Trent, W. L. (1997) Outcomes of school desegregation: findings from longitudinal research, *Journal of Negro Education*, 66(3), 255–257.
- Tyler, J. H., Murnane, R. J. & Willett, J. B. (2003) Who benefits from a GED? Evidence for females from ‘high school and beyond’, *Economics of Education Review*, 22(3), 237–247.
- US Department of Education (2002) *User’s manual for the ECLS-K first grade public-use data files and electronic codebook* (NCES Publication No. 2002-135) (Washington, DC, National Center for Education Statistics).
- Waits, T. (2003) *Distance education at degree-granting postsecondary institutions: 2000–2001* (NCES Report No. NCES-2003-017) (Washington, DC, National Center for Education Statistics).