Effective Use of Secondary Data Analysis in Gifted Education Research: Opportunities and Challenges

Christian E. Mueller  
*University of Memphis*

Caroline O. Hart  
*University of Memphis*

Follow this and additional works at: [http://docs.lib.purdue.edu/giftedchildren](http://docs.lib.purdue.edu/giftedchildren)

**Recommended Citation**  
Available at: [http://docs.lib.purdue.edu/giftedchildren/vol4/iss2/3](http://docs.lib.purdue.edu/giftedchildren/vol4/iss2/3)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

This is an Open Access journal. This means that it uses a funding model that does not charge readers or their institutions for access. Readers may freely read, download, copy, distribute, print, search, or link to the full texts of articles. This journal is covered under the [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/).
Secondary analysis of archived data had an early beginning in the field of gifted education with the influential work of Lewis Terman and his Genetic Studies of Genius. In the present article, we briefly trace this history, focusing particularly on how these data have been analyzed in subsequent research. This is followed by a discussion of the benefits, challenges, and limitations of conducting secondary data analysis in gifted education research. Lastly, we review and describe two archived datasets (the National Longitudinal Study of Adolescent Health and the Educational Longitudinal Study) that offer potential sources of data on targeted samples of gifted and talented students for graduate students and established researchers alike.

Secondary data analysis of existing or archived data has a rich history in the fields of economics, demography, and sociology (Duncan, 1991). Friedman (2007) also notes an increased frequency within the field of psychology and attributes this growth to two main factors: (1) the creation and maintenance of large longitudinal data sets by governmental and other funding agencies, and (2) a growing emphasis on interdisciplinary collaborations between psychologists and other social science researchers. Combined, these factors have made it possible for researchers in multiple disciplines to explore topics and design studies utilizing the best (and worst) practices and methodologies from previously disparate fields. Further, Brooks-Gunn, Phelps, and Elder (1991) encouraged developmental psychologists to embrace secondary data analysis as a cost-effective way to “study lives in context over time” (p. 899). Indeed, archived data sources for secondary analysis provide exciting opportunities and afford access to large population samples not otherwise available for most independent giftedness researchers at the university level. The present article is divided into three parts: first, we will provide a brief historical exploration of secondary data analysis in the field of gifted education; second, we will discuss some of the benefits, challenges, and limitations of conducting secondary data analysis in gifted education research; and third, we will review and describe two archived data sets that offer potential sources of data on targeted samples of gifted and talented adolescents.

Secondary data analysis had an early beginning within the field of gifted education, beginning with the influential work of Lewis Terman and his colleagues at Stanford University. In the 1920’s, Terman and colleagues launched the Genetic Studies of Genius, which resulted in a total of five volumes of work (e.g., Terman & Oden, 1959). Some consider this work to be the first longitudinal study ever conducted in the field of psychology (Cravens, 1992). Using a longitudinal design, Terman and colleagues collected and archived large amounts of data on approximately 1,500 gifted children (IQ>140) over the course of several decades. Many of these data can still be accessed by contemporary researchers and have resulted in several studies over the past decade (e.g., Crosnoe & Elder, 2004; Zuo & Tao, 2001). Seagoe (1975) points out that part of Terman’s enduring legacy was his desire to have his data archived and available for future researchers in order to advance the study of the gifted, as well as the study of measurement of cognitive abilities.

Although the use of Terman’s data has decreased in recent years, current researchers in gifted education and other disciplines continue to examine these data in order to provide new insights into the lives of the gifted (e.g., Crosnoe & Elder, 2002, 2004; Lester, 1991; Zuo & Cramond, 2001). Shanahan, Elder, and Miech (1997) noted, for example, that data from Terman’s gifted cohorts “have been used to study history and a range of life-course processes ... so their use in our study contributes to a larger, emerging picture of these groups” (p. 58). Additionally, Cravens (1992) noted that much of Terman’s work relied on assumptions about the gifted—namely the belief in the fixed nature of IQ and that individuals develop in a larger social context—that continue to remain relevant today. Despite the numerous advantages, gifted education researchers will also find challenges and limitations in conducting secondary analysis with archived data. These opportunities, challenges, and limitations are outlined in greater detail in the next section.

Secondary data analysis is distinct from primary analysis in both form and function. Windle (2010) suggests that primary data analysis is mainly used to collect and analyze first-time data using originally derived research questions and methodology. Secondary data analysis, on the other hand, is useful as a way to explore alternate relationships among variables or from different research perspectives (e.g., sociological versus psychological viewpoints), as well as to conduct research studies using statistical methodology that may not have been available at the time of the original data collection. For example, researchers using the two datasets described later (i.e., Add Health and ELS:2002) have explored both sociological (e.g., Owens, 2010) and psychological (e.g., Shahar & Henrich, 2010) influences on adolescent
development. Further, Shanahan et al. (1997) used structural equation modeling, an advanced statistical procedure, to predict educational attainment in a portion of Terman’s original sample. As both examples illustrate, secondary analysis of archived data can and should be viewed as a complement to, rather than as a replacement for, primary data analysis.

Many authors have outlined the benefits of conducting secondary data analysis in relation to family research (e.g., Hofferth, 2005), deaf education (e.g., Kluwin & Morris, 2006), school counseling and social work (e.g., Bryan, Day-Vines, Holcomb-McCoy, & Moore-Thomas, 2010; Williams, 2008), and developmental psychology (e.g., Brooks-Gunn et al., 1991; Duncan, 1991; Friedman, 2007). Here, we summarize some of those benefits and relate them directly to the field of gifted education. First, archived data, especially from nationally representative data sets, make large data sets readily available to gifted researchers at little or no cost (Williams, 2008). Given the current state of dwindling funding sources for gifted education research, secondary data offer an important alternative to extending the field of gifted education in the 21st century. Second, large government-funded research studies typically employ data collection of multiple individuals over multiple time points and in multiple contexts, allowing for exploration of larger systemic or ecological influences on development (e.g., neighborhood or school). As will be discussed in more detail, many of these data are readily accessible to researchers in multiple domains (Friedman, 2007). Third, use of nationally representative data sets allows for broad selection and sampling and, thus, increases external validity of findings (Duncan, 1991; Mueller, 2009). Last, secondary analysis of large archived data sets allows access to an array of variables relevant to various psychology disciplines such as developmental psychology, cognitive psychology, social psychology, or educational psychology. Secondary data analysis therefore facilitates interdisciplinary research (Williams, 2008), which can have a profound impact at the level of policy and practice in gifted education (Mueller, 2009). It is not surprising that secondary data analysis provides a great opportunity for interdisciplinary research, as interdisciplinary collaboration often resides at the root of secondary data analysis. That is, researchers from various disciplines often collaborate in designing data collection endeavors. In the area of developmental psychology, for example, Brooks-Gunn et al. (1991) note that an increased involvement of developmental psychologists in the design and implementation of large-scale data collection projects resulted in more useful data being collected and archived for use by developmental researchers across the U.S. Even with all of the benefits outlined here, there are several limitations and unique challenges that researchers face while conducting secondary analysis.

Limitations and challenges in the analysis of secondary data are best characterized as those of training, measurement, and cohort or sample composition (Brooks-Gunn et al., 1991). Conducting secondary analysis on existing data sets can sometimes be difficult, especially when one lacks specific training at the graduate level. Many times, students and researchers have to convert data provided on CDs to a format compatible with statistical software packages (e.g., SPSS or SAS), and several attempts might be required before the data can be transformed into a usable format. Perhaps more important, there are numerous issues with measurement when conducting secondary analysis on archived data. For example, data often need to be recoded (e.g., eliminate missing cases or variables) and a priori scales often do not exist and must be developed through statistical techniques (e.g., factor analysis). Additionally, extreme care needs to be given to establishing reliability of newly created scales, especially when being used for the first time with gifted samples. Lastly, cohort or sample composition can be especially challenging when using archived data because most large, longitudinal studies were not designed with identification of gifted students in mind. This may impact the types of research questions and variables that can be explored, particularly given the controversy that exists around identifying and defining giftedness (e.g., Brown et al., 2005; Sternberg & Zhang, 1995). Specific procedures for helping researchers deal with some of these challenges are provided next as we review two national datasets previously used in gifted education research: the National Longitudinal Study of Adolescent Health (Add Health) and the Educational Longitudinal Study of 2002 (ELS:2002). Both data sets are accessible through the Inter-University Consortium for Political and Social Research (ICPSR).

**Inter-University Consortium for Political and Social Research (ICPSR)**

According to the ICPSR website (http://www.icpsr.umich.edu/ICPSRweb/ICPSR/), the ICPSR is a national and international consortium of approximately 700 academic and research organizations. The ICPSR archives more than 500,000 research files in most domains of the social sciences, including psychology and education. Brooks-Gunn et al. (1991) note, “In 1962, the ICPSR was founded by the Survey Research Center at the University of Michigan and 21 other universities in the United States to serve as a central repository for machine-readable social science data” (p. 905). Students and researchers who attend or work at participating institutions may access both public-use and restricted data bases by signing up to be a member on the ICPSR website (visit http://www.icpsr.umich.edu/cgi-bin/newacct).

One example of archived data that has already been discussed is Terman’s Genetic Studies of Genius (i.e., Life-Cycle Study of Children with High Ability, 1922-1991). Even though several studies have subsequently been published since the data were originally collected (e.g., Crosnoe & Elder, 2002, 2004; Zuo & Cramond, 2001), the historical nature of Terman’s data may limit its usefulness for some contemporary researchers (see Cravens, 1992, for a discussion of reasons why these data may still be relevant for the field of giftedness). For the remainder of this section, we review two data sets that may provide useful archived data for secondary analysis for a variety of gifted education researchers: the National Longitudinal Study of Adolescent Health (Add Health) and the Educational Longitudinal Study (ELS:2002). Each section includes a brief description of the study, how gifted researchers can select subsamples of gifted and talented students, and the types of variables that can be explored.

**National Longitudinal Study of Adolescent Health (Add Health)**

Add Health is the most comprehensive and systematic study of adolescent development ever undertaken in the United States. Funded primarily through a grant from the National
Institute of Child Health and Human Development (NICHD), Add Health employed a longitudinal and multi-level research design to collect survey data from 20,745 7th- through 12th-graders at 80 high schools across the U.S. on demographic, physical, and psychosocial well-being variables. Contextual data were also collected on parents and siblings, neighborhoods, schools and teachers, peers, and romantic-partners, allowing researchers a unique opportunity to explore how social environments impact achievement and health-related outcomes from early adolescence to young adulthood. Specific information about the Add Health study is found at http://www.cpc.unc.edu/projects/addhealth. Here, we present a brief overview of Add Health, especially as it relates to the gifted education researcher.

Add Health data were collected over four time points: (1) wave I (1994-1995), which focused on 7th- through 12th-graders’ family, school, and personal demographics; psychosocial factors (e.g., self-esteem), and physical and behavioral characteristics (e.g., delinquency-related behaviors); (2) wave II (1996), which surveyed approximately 15,000 of those students one year later across many of the same areas, but included additional information on physical health such as nutritional habits; (3) wave III (2001-2002), which placed more emphasis on college and work issues, as participants were ages 18-26, and most recently; (4) wave IV (2007-2008), which surveyed the participants as they faced issues of young adulthood, including many health and lifestyle choices (e.g., marital and occupational choices). These data have been used to track developmental trends across adolescence and young adulthood (e.g., Bromann, 2009; Ueno, 2010), as well as to compare patterns across different groups (e.g., ethnicity, Almgren, Magarati, & Mogford, 2009; gifted and non-gifted, Mueller, 2009).

Data are available in both restricted and public-use forms. Public-use data are available through two sources: ICPSR and Sociometrics. Researchers wanting a CD-ROM of the public-use data must pay a small fee and order this through Sociometrics, or data can be downloaded for free directly from ICPSR (http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/21600/sda). At present, waves I-III are available in public-use form. Researchers interested in accessing restricted-use data must have an IRB-approved security plan for storing sensitive data and must also sign a contract ensuring that data will be kept confidential. There is a cost for anyone wanting to use the restricted-use data. Additional details for accessing restricted Add Health data are found at http://www.icpsr.umich.edu/icpsrweb/DSDR/access/add-health.jsp. Although Add Health did not explicitly sample students who were labeled as gifted and talented, the research design did include a proxy measure of verbal intelligence that has been used to identify gifted students in previous studies (e.g., Halpern, Joyner, Udry, & Suchindran, 2000; Mueller, 2009). The Peabody Picture Vocabulary Test-Revised (PPVT-R) was adapted and used as an indicator of verbal intelligence in Add Health. Halpern et al. (2000) note that the PPVT-R is often used in large field studies such as Add Health because of short administration time, little need for specialized training for administration, and because overall performance is not dependent upon reading ability. Further, the PPVT-R test has shown moderate correlations with other IQ measures (e.g., .62, Stanford-Binet; .64, Wechsler Intelligence Scale for Children) and uses a scoring metric similar to other IQ tests (Mean = 100, SD = 15). Despite the ongoing debate about using standardized scores for identifying gifted students (e.g., Borland, 2009), Add Health continues to provide one of the richest archived data sets available for use for gifted education researchers. This is especially true when one considers that these individuals can be identified in adolescence (i.e., 1994-1995) and their developmental trajectories tracked into their 30’s.

As an example of recent gifted research using Add Health data, Mueller (2009) identified verbally gifted adolescents by selecting a subset from wave I of the larger Add Health sample of 20,745 students, using several criteria. First, the standardized scores from the Add Health PVT (variable: AH_PVT) were used as the proxy measure of verbal intelligence. Mueller identified students as verbally gifted if their AH_PVT scores were in the top five percent of overall scores. This yielded a total sample of 762 participants whose scores ranged from 123 to 146, with a mean of 125.88 (SD = 3.08). Ages ranged from 12 to 19 years old, with a mean of 15.70 years (SD = 1.65); there were more males (52.6%) than females (47.4%); and the sample was predominantly White (75.6%). Other researchers have also utilized the standardized AH_PVT score as a proxy measure for intelligence (e.g., Guo & Stearns, 2002; Halpern et al., 2000), although not specifically identifying a subsample of gifted students. To date, the study by Mueller is the only study to have identified a gifted subsample utilizing Add Health data.

Add Health contains variables related to disciplines such as psychology, sociology, and epidemiology that may be of interest to gifted researchers. As discussed previously, Add Health collected data at the individual level (e.g., self-concept, depression), as well as at the larger contextual level (e.g., family, schools, and neighborhood). For example, gifted researchers who have questions about academic achievement are provided with four variables that can be combined to create a proxy for GPA (wave I variables: H1ED11, H1ED12, H1ED13, H1ED14). In order, these variables represent the most recent self-reported grades in English/language arts, math, history/social studies, and science. For those researchers who may be more interested in psychosocial development among gifted students, Add Health contains 19 items designed specifically to measure depression (wave I variables: H1F51-H1F519). Resnick et al. (1997) and Anderman (2002), among others, discuss additional procedures for developing scales for use with Add Health data. There were literally thousands of questions asked of adolescents, their families, and school officials during the four waves of data collection, allowing researchers to explore multiple aspects of health and well-being among a nationally representative sample.

**Educational Longitudinal Study (ELS:2002)**

The Education Longitudinal Study of 2002 (ELS:2002) was conducted by the Research Triangle Institute (RTI) on behalf of the National Center for Education Statistics (NCES) of the United States Department of Education. It was designed to monitor the transition of a cohort of more than 15,000 10th graders through high school and on to postsecondary education and/or the workplace. In addition to being a longitudinal study, the ELS:2002 is a multi-level study. That is, data are collected not only from students but also from...
their parents, teachers, schools, and librarians. By surveying these multiple respondent populations over time, the ELS:2002 offers great opportunities for researchers to investigate the various social factors that could exert an influence on students. The NCES website (http://nces.ed.gov/surveys/els2002/) provides extensive information about the ELS:2002.

The ELS:2002 consists of four waves: (1) base-year of 2002, which focused on 10th graders’ demographic information, school experiences, attitudes, and beliefs; (2) first follow-up (2004), which focused on students’ school and work experiences, achievement gain in mathematics, and plans for the future; (3) second follow-up (2006), which focused on issues of college access and choice; and (4) third follow-up (scheduled for 2012), which will assess outcomes such as persistence, higher education attainment, and transition into the workplace. Because the ELS:2002 was designed with the goal of maintaining comparability with previous longitudinal studies such as the National Longitudinal Study of 1972 (NLS-72), High School and Beyond (HS&B), and the National Education Longitudinal Study of 1988 NELS:88 (Owings, Wirt, & Brown, 2007), the ELS:2002 trends and outcomes can be compared to those of cohorts from previous studies.

Data are available in both restricted and public-use form. Access to the public-use data files is gained through the online downloading tool, EDAT (http://nces.ed.gov/edat/). In a very few clicks, researchers are able to access all public-use data files from the first three waves of data collection (2002, 2004, and 2006), as well as syntax files for eight programming languages (SAS, Stata, SPSS, Sudaan, R, S-Plus, ASCII, and CSV) and more than 100 composite variables. Researchers who wish to use data from the first follow-up transcript variables and college entrance test scores need to know that these types of data are restricted-use only, as they contain individually identifiable information that are confidential and protected by law. However, it is relatively easy to apply for a restricted-use license. The following link explains how to apply for a license: http://nces.ed.gov/surveys/els2002/obtainingrest.asp. Provided there is no issue with the license paperwork, it usually takes three weeks for researchers to receive the restricted-use data CD-ROM.

Even though the ELS:2002 does not label gifted and talented students as such, there are several options for researchers to investigate such populations using the various questionnaires available. For instance, using the data from the base-year student questionnaire, researchers could identify any students who had ever been in advanced placement programs (variable #1089) or who planned to take advanced placement tests (variable #1222). Students who had received advanced training in English (variable #2550) and math (variable #2642) could be selected from the base-year teacher questionnaires, while the percentage of 10th graders in college prep programs (variables #, 2763, 3641) could be found in the base-year administrative and school administrator questionnaires. Unfortunately, the first follow-up school administrator questionnaires do not distinguish the percentage of 12th graders in college prep programs from the percentage of 12th graders enrolled in other specialized academic program (variables #3108, 3986). However, these same questionnaires do provide the percentage of the student body in advanced placement courses (variables #3133, 4011). Similar to the base-year student questionnaire, the first follow-up student questionnaire identifies students who took or planned to take advanced placement tests (variable #1411).

While there are several ways for researchers to identify gifted and talented students from the various ELS:2002 questionnaires, perhaps the most valuable information resides in resources other than the data collected from questionnaires. The first follow-up transcript variables and college entrance test scores, which consist of data collected from transcripts and report cards, offer extremely valuable information. For example, the transcript variables provide an array of data related to students’ academic achievement, such as SAT information (variables #369 to 374), numbers of advanced placement courses (variables #473 to 490), and GPA data from 9th to 12th grade (variables #509 to 533). The college entrance test scores offer information regarding ACT scores (variables #536 to 551, 554, 555), SAT scores (variables #552, 553, 556 to 562, 632 to 669), and AP exam scores (variables #564 to 631). Using data from both the first follow-up transcript variables and college test scores, researchers could identify gifted and talented students. That is, one could, for example, consider “gifted students” those who scored above a predetermined level on a specific test.

There are abundant research questions that could be answered using the ELS:2002 that encompass social background information (e.g., demographics, family income, family structure, parent education and employment, parental aspirations for child, health history), context information (i.e., home, school, and community environment) and outcome information (e.g., academic achievement scores, engagement in school, socioemotional development, postsecondary attainment, labor market outcomes, family information, and citizenship). Given the amount of ELS:2002 data available for researchers interested in gifted and talented students, it seems surprising that gifted education research using the ELS:2002 is still in its infancy. A review of the ELS:2002 bibliography as well as our own computer searches of online bibliographic databases uncovered only two studies related to the field of giftedness that employed the ELS:2002: Barber and Torney-Purta’s (2008) study, which investigated teachers’ nomination for advanced programs, and Well, Lohman, and Marron’s (2009) study of grade acceleration.

Using ELS:2002 data, Barber and Torney-Purta (2008) drew on Gagne’s Differentiated Model of Giftedness and Talent as a framework to investigate whether high-achieving English and math high school students’ social perceptions, individual motivation, and demographic background influenced their likeliness to be nominated for advanced English and math programs by their teachers. Well et al. (2009) took advantage of the comparability between the ELS:2002 and NELS:88 to identify personal, familial, and scholastic factors that are correlated with student grade acceleration. Both Barber and Torney-Purta (2008) and Well et al.’s (2009) studies illustrate the applicability of secondary analysis to research on gifted and talented students.
Summary and Conclusion

In the present article, we have described how secondary analysis of archived data presents researchers in gifted education a valuable alternative to original data collection and analysis. As stated previously, secondary data analysis should not be viewed as a replacement for primary data collection and analysis, but rather, as an alternative source for gifted education researchers. As many researchers have already found, numerous benefits await those who are patient enough to engage in this growing methodology. Much as Terman did, we understand the importance of collecting and archiving longitudinal data on gifted and talented individuals as absolutely essential for advancing the field of gifted education research well into the 21st century. By highlighting the resources available in secondary data analysis for the field of gifted education, we hope that this article will inspire gifted education researchers to conduct more research using such data.

References


