Article

A Web Based High School Transcript and Course Catalog Keying and Coding System

by Donghui Wang, Tiffany Mattox, Brian Evans, Daniel Pratt and Chinh Nguyen

2009

Statistics Canada Statistique Canada

Canada

A Web Based High School Transcript and Course Catalog Keying and Coding System

Donghui Wang, Tiffany Mattox, Brian Evans, Daniel Pratt and Chinh Nguyen[1]

Abstract

RTI International is currently conducting a longitudinal education study. One component of the study involved collecting transcripts and course catalogs from high schools that the sample members attended. Information from the transcripts and course catalogs also needed to be keyed and coded. This presented a challenge because the transcripts and course catalogs were collected from different types of schools, including public, private, and religious schools, from across the nation and they varied widely in both content and format. The challenge called for a sophisticated system that could be used by multiple users simultaneously. RTI developed such a system possessing all the characteristics of a high-end, high-tech, multi-user, multitask, user-friendly and low maintenance cost high school transcript and course catalog keying and coding system. The system is web based and has three major functions: transcript and catalog keying and coding, transcript and catalog keying quality control (keyer-coder end), and transcript and catalog coding QC (management end). Given the complex nature of transcript and catalog keying and coding, the system was designed to be flexible and to have the ability to transport keyed and coded data throughout the system to reduce the keying time, the ability to logically guide users through all the pages that a type of activity required, the ability to display appropriate information to help keying performance, and the ability to track all the keying, coding, and QC activities. Hundreds of catalogs and thousands of transcripts were successfully keyed, coded, and verified using the system. This paper will report on the system needs and design, implementation tips, problems faced and their solutions, and lessons learned

1. Introduction

RTI International was the prime contractor on the Education Longitudinal Study (ELS), a study designed to monitor the transition of a national sample of young people as they progress from tenth grade through high school and on to postsecondary education and/or the world of work. An essential part of the study was collecting transcripts and course catalogs from high schools that the sample members attended, then keying and coding information from them into an electronic data collection system. This task was challenging because the transcripts and course catalogs came from a wide variety of institutions, including public, private, and religious schools from across the nation. They also varied widely in both content and format.

To meet this challenge, RTI developed the ELS Keying and Coding System (ELS-KCS) a sophisticated system that could handle multiple users performing keying and coding activities as well as management activities. In addition to multitasking, the ELS-KCS is user-friendly, low cost, and requires low maintenance. The system is web based and has three major components: keying and coding, keying quality control (keyer-coder end), and coding QC (management end). Because transcript and catalog keying and coding is a complex task, we designed the system to be user friendly and cost efficient. It can transport keyed and coded data throughout the system to minimize rekeying time, logically guide users through all the pages that a type of keying or coding activity required, display appropriate information to help keying performance, and track all the keying, coding, and QC activities. Keyers entered hundreds of catalogs and thousands of transcripts successfully into the system. This paper will report on the system needs and design, give implementation tips, and report on problems, solutions, and lessons learned.

---

[1] Donghui Wang, Tiffany Mattox, Brian Evans, Daniel Pratt and Chinh Nguyen. RTI International, RTP, NC, USA

2. Study background and keying preparation

2.1 Transcript and course catalog data collection

As an indication of the amount of data our system had to handle, from December 2004 through June 2005, we sent transcript request materials to 2,032 schools attended by sample members, which included 759 schools that had participated either in the base-year or first follow-up survey. School staff were asked to provide basic enrollment, testing, and course-taking information for each student as well as the school's grading and graduation policies and requirements. Examples of the types of data we requested include type of diploma and date awarded, date and reason student left school (e.g., graduated or transferred), course title and number of each course taken, and credits earned and grade awarded for each course taken.

Course catalogs were collected during the base year (2001-02) when sampled students were enrolled in the 10th grade, and during the first follow-up year (2003-04). In all, we collected course catalogs for four school years covering 2000-2004. During the transcript request activities, schools were prompted for catalogs that we had not received. Overall, the large number of catalogs and transcripts and associated information was a driving factor in how we designed the ELS-KCS system.

2.2 Data collection results

The ELS study was very successful: the unweighted school participation rate was 79.3 percent for transcript collection, and the base year school weighted response rate was 94.5 percent. The final student coverage rate was ninety-one percent (90.7 percent, weighted) of the entire student sample with at least one transcript with one course on it. For course catalog collection, the response rate for base year schools was 88.0 percent. After the data were received, reviewed, and logged into the control system, approximately 15,700 transcripts and 700 course catalogs were routed to keying and coding, an enormous amount of data.

Table 2.2-1
Data collection response rate

	Unweighted* school	Base year school (weighted*)	Student transcript (weighted*)	Base year school catalog
Response rate (%)	79.3	94.5	90.7	88.0

Unweighted* --- Unweighted rates are based on frequency counts.
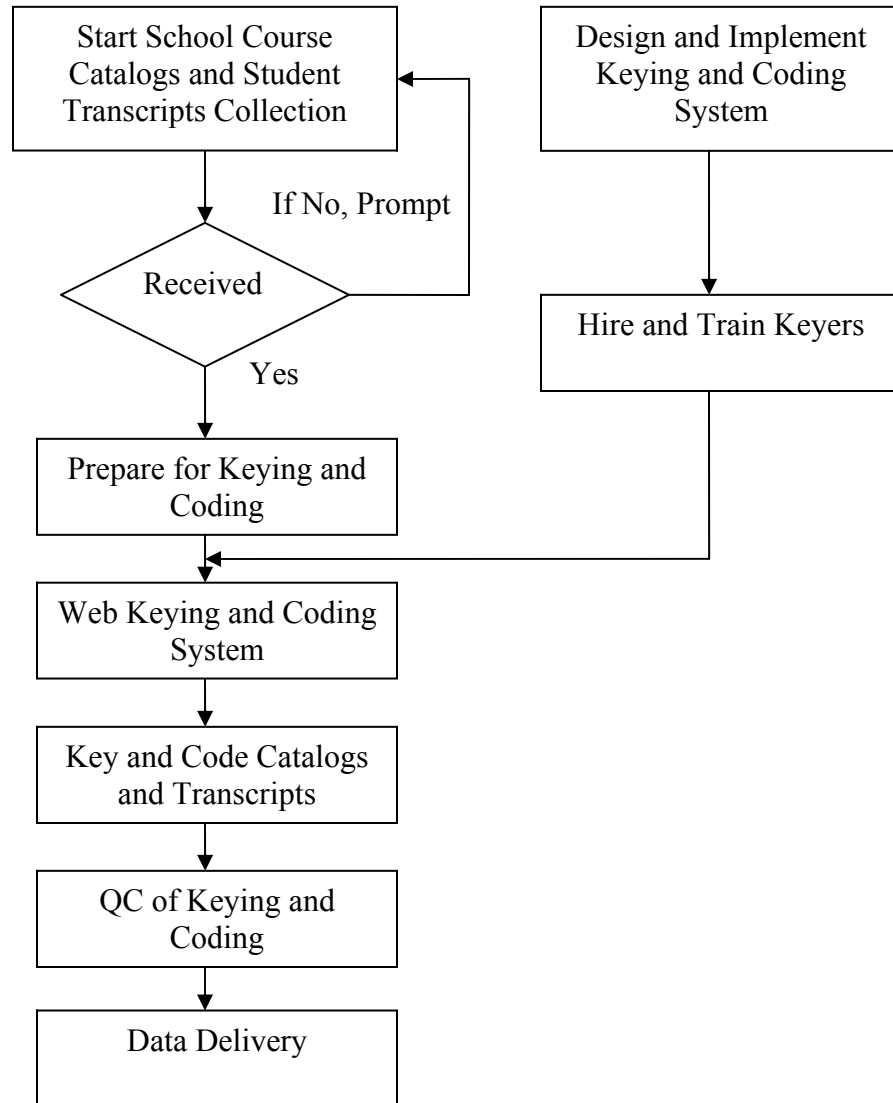Weighted*   --- Weighted rates are based on weighted estimates.

2.3 Preparation for keying and coding data

As the project staff prepared for keying and coding transcript and course catalog data, they examined the activities and information necessary to design the system. In addition, before any data was entered into the system, each catalog and transcript course had to be coded using a scheme that classified and grouped subject areas and courses. Next, specially trained keyers were needed to abstract data from the course catalogs and transcripts. Keyers had to become familiar with hundreds of data elements, including school-, student-, and course-level data. Next, quality control activities on the keying and the coding were essential. Finally, project staff had to generate many complex reports, including assignment, production, and quality reports.

Table 2.3-1
Data preparation volume

	Total	Pages per unit (Average)	Total pages
Student transcripts	15,700	40	628,000
School catalogs	700	150	105,000

2.4 System diagram

```
┌─────────────────────┐          ┌─────────────────────┐
│  Start School Course│          │ Design and Implement│
│  Catalogs and Student│         │   Keying and Coding │
│  Transcripts Collection│◄─┐     │       System        │
└─────────────────────┘   │      └─────────────────────┘
           │              │                 │
           ▼         If No, Prompt          │
        ╱─────────╲        │                ▼
       ╱  Received  ╲──────┘      ┌─────────────────────┐
        ╲─────────╱               │ Hire and Train Keyers│
           │                      └─────────────────────┘
          Yes                               │
           ▼                                │
┌─────────────────────┐                     │
│  Prepare for Keying │                     │
│     and Coding      │                     │
└─────────────────────┘                     │
           │◄───────────────────────────────┘
           ▼
┌─────────────────────┐
│  Web Keying and Coding│
│        System       │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│  Key and Code Catalogs│
│     and Transcripts │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│   QC of Keying and  │
│        Coding       │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    Data Delivery    │
└─────────────────────┘
```

3. The web-based Keying and Coding System

3.1 The Keying and Coding System

The ELS-KCS web system is the transcript study's centerpiece. Highly trained keying staff entered standardized information through the web system into the database. Project management team members also used the system daily to monitor progress and solve any problems and issues discovered by web system reports.

3.1.1 The characteristics of the system

Based on the complexity of the study, we designed the web based keying-coding system to keep cost under control by using common web programming language and SQL server. Our design and implementation also called for making keying pages user friendly by making them clean, straightforward, and easy to understand. Another aspect of user friendliness was making the system's basic knowledge base available by equipping coding pages with look-up documents, helpful drop-down lists and text messages.

To make the system as efficient as possible, we put three major components of the web system together under one login, including keying/coding, user QC, and management QC. The system routed each user to his/her task page based on the role each login had.

We also made the system efficient by having keyed and preloaded data immediately available for future keying/coding activities and putting a full range of checks in place to reduce keying errors. To manage the work flow, we made a full set of stages and statuses available as well as full set of hard-copy and ad hoc reports to help both management team and keyers. In addition, we deployed a full line of database triggers to manage all database activities and we normalized the database to reduce data redundancy. To improve the speed of the system as keyers and management personnel were added, queries were optimized and tables indexed to improve data retrieval and data update. The team also devised and incorporated many pop up screens to make the keying process fast and efficient. The system enforced database validations to avoid non-valid data being keyed in and used different statuses for transcripts and catalogs to make the keying process flow through different roles. Finally, the system applied integrated keying QC within the keying process and used formulated algorithms to select courses for coding QC.

3.1.2 The effectiveness of the system

During the study's peak, there were as many as 80 keyers using the system simultaneously. Management team members were also online to monitoring the system. By using the tools the system provided, keyers were able to maintain high productivity. The project management team was also able to effectively monitor the entire system by using the system reports and other tools. For example, the management team could check the hours spent by a keyer on a certain day against the work he/she had done on the same day to discover whether or not the keyer had issues. If an issue was found, the keyer would receive help right away.

In addition, quality control was important from the very beginning of the keying process because of the volume of data entered. ELS-KCS is well equipped to handle quality control issues, from double-keying of content to automatically selecting verification courses. Ultimately, both keyers and management personnel felt very comfortable that data entered was the highest quality.

3.1.3 The report system

One vital component of the web system was the Process Report System, which was programmed to monitor the entire keying/coding process and to ensure the quality of the work. The reports were designed to run nightly (static) and interactively (dynamic) to help the management team and the keyers verify the quality and quantity of the data entered. Managers used the reports to track each keyer's work and hours spent. For example, we developed the Keyers Performance Report to help managers maintain their budget. We also developed a nested detail summary report that was generated nightly and included over 200MB of data, including catalog and transcript keying and coding. Project personnel used key words to drill down to the content they wanted. For quality control purposes, a QC report was developed that cross-checked over 1/2 million data fields in the database each day to catch all data discrepancies.

4. Lessons learned

Because we had developed an excellent understanding of the study itself and the system requirements, we designed a robust system that we were able to implement successfully. As with any system, we had our share of minor bugs and setbacks. Our largest problem was database performance. The ELS-KCS system slowed down as more keyers were added to the task and more records were added to the tables. We resolved this problem by optimizing queries and indexing all the tables. Using 20/20 hindsight, we understood that we should have done this earlier to prevent the problem in the first place.

5. Summary

The Education Longitudinal Study (ELS) was a multi-stage and multi-task study and the transcript and catalog collection effort was an essential part of it. The transcript study required collecting, keying and coding a large number of student transcripts and school catalogs from variety of schools. A carefully planned pre-keying and pre-coding preparation had to be performed to ensure the success of the keying and coding process. Based on the complexity of the study, the web ELS-KCS system was designed to be user-friendly and cost efficient. It used clean logic and was straight-forward and easy for keyers and management personnel to use, and allowed multi-tasking. The ELS-KCS system ensured that the keying and coding progress, budget, and quality of data entered were all under control. The ELS-KCS system helped the ELS study produce high quality data study.