

DATA PROCESSING METHODS

Data Acquisition and Sources

During the course of its ongoing research activities, the Juvenile Justice Educational Enhancement Program (JJEPP) obtains student-level data from a number of sources each year. These data provide the basis from which to evaluate aggregate student performance in relation to various demographic and program characteristics, and to assist in the specification of facility and student outcomes, such as school success (e.g., credits and diplomas earned, return to school) and continuation of delinquency (e.g., arrest and recommitment rates). Data are provided by means of secure electronic transmission, usually on disk or CD. The student-level data used for the research in this year's annual report were obtained from the following sources:

- Department of Education's (DOE) Survey 5
- Florida Department of Law Enforcement (FDLE)
- Florida Department of Corrections (FDOC)
- Florida Education and Training Placement Information Program (FETPIP)

The content of the submissions from each of these data sources is discussed below.

DOE Survey 5

Survey Five contains a variety of reporting formats, but JJEPP's research initiatives are based on information contained in the following:

- Student Demographics
- Attendance
- Disciplinary Referral
- End-of-Year Status
- Special Education
- Transcript
- Entry/Exit Academic Assessment Testing

FDLE

FDLE was the source of arrest data for the measurement of both the number of prior arrests and whether and when Department of Juvenile Justice (DJJ) youths were arrested subsequent to release from a residential facility. A formal data sharing agreement was first established with FDLE's Statistical Analysis Center (SAC). JJEPP then supplied the SAC with a dataset of the FY2000-01 cohort, which contained offender identifiers, including: last name, first

name, middle initial, social security number, sex, race, and date of birth. Using these identifiers, the SAC matched the cohort to FDLE’s Computerized Criminal History (CCH) database to extract all arrest records for any offender who was in both datasets. Only cases that matched on an appropriate number and type of identifiers, to ensure they were the same person, were retained as legitimate matches. Arrest events with multiple charges were counted as one arrest.

The types of arrest charges reported to FDLE are those submitted by local law enforcement agencies in accordance with section 943.051, Florida Statutes (F.S.).

943.051, F.S. Criminal justice information; collection and storage; fingerprinting.--

3)(a) A minor who is charged with or found to have committed an offense that would be a felony if committed by an adult shall be fingerprinted and the fingerprints shall be submitted to the department in the manner prescribed by rule.

(b) A minor who is charged with or found to have committed the following offenses shall be fingerprinted and the fingerprints shall be submitted to the department:

1. Assault, as defined in s. [784.011](#), F.S.
2. Battery, as defined in s. [784.03](#), F.S.
3. Carrying a concealed weapon, as defined in s. [790.01](#)(1), F.S.
4. Unlawful use of destructive devices or bombs, as defined in s. [790.1615](#)(1), F.S.
5. Negligent treatment of children, as defined in s. [827.05](#), F.S.
6. Assault or battery on a law enforcement officer, a firefighter, or other specified officers, as defined in s. [784.07](#)(2)(a) and (b), F.S.
7. Open carrying of a weapon, as defined in s. [790.053](#), F.S.
8. Exposure of sexual organs, as defined in s. [800.03](#), F.S.
9. Unlawful possession of a firearm, as defined in s. [790.22](#)(5), F.S.
10. Petit theft, as defined in s. [812.014](#)(3), F.S.
11. Cruelty to animals, as defined in s. [828.12](#)(1), F.S.
12. Arson, as defined in s. [806.031](#)(1), F.S.
13. Unlawful possession or discharge of a weapon or firearm at a school-sponsored event or on school property as defined in s. [790.115](#), F.S.

FDOC

Obtained from the FDOC were data that included all offenders’ identification information and all sentencing events in its Offender Based Information System (OBIS). To determine if, and when, DJJ releases in the FY2000-01 cohort had been sentenced to prison subsequent to release, it was necessary to match the cohort cases to the FDOC offender identification information. The identifiers used included last name, first name, middle initial, date of birth,

sex, race, and social security number (SSN). Various combinations of these identifiers were tested for matching accuracy, and only in those cases where there was a high degree of confidence that the youth in the cohort was, in fact, the same offender in the FDOC data was a decision made that a valid match had been obtained.

For those cohort cases that matched to the FDOC identification data, the FDOC offender identification number was used to match to the FDOC sentencing data to determine if these youths had a prison sentencing date after their DJJ release date. If so, the DJJ release date was retained as part of the cohort data and used to create indicators to determine whether the youth had been sentenced to prison and the length of time from DJJ release to a prison commitment.

FETPIP

Data from FETPIP consist of an extract provided at JJEEP's request on an annual basis. JJEEP submits a file of student SSNs, names, and dates of birth, which FETPIP matches to its database. The resultant file, which is returned to JJEEP contains the employee number, year and quarter of employment, wages for the quarter in each job held during that quarter, and total wages earned during the quarter for each student. It is important to note, however, that FETPIP only uses SSN to match records, which may result in imprecise matching.

Cleaning the DOE Survey 5 Demographic Format

The first task in this process involves the *grouping* of DOE data in the demographic format in an effort to identify which entries refer to the same individual student, in order to form a complete educational history for each student who may have attended multiple schools within the school year. Getting this "right" is extremely important in the context of tracking individual student outcomes over time.

- There are two possible scenarios that require data "cleaning" and must be considered before records can be successfully *grouped* using a single unique student identifier:
 - a. Two or more *different* students share the same Student ID (SID).
 - b. A single student has records listed under *several different* SIDs.

These issues arise for several different reasons but most frequently occur due to:

- common names
- students, either intentionally or unintentionally, providing inaccurate or inconsistent information to school officials, and
- data entry errors at the school or district level

Correcting these errors requires carefully examining student ID, student alias, name, date of birth, and several other demographic variables for each record. The end result is that all

records referring to the same youth are *grouped* by assigning them a common identifier in the form of a variable derived from SID; this variable is called TRUESID.

Student ID is, in most cases, the student’s SSN; however, it also may be a district-generated identifier. To make matters more difficult, approximately 1/3 of the records in the demographic format for a given year contain both a student ID and an alias variable, which are not the same. For these cases, a duplicate line is created, and the student ID line is recoded to contain the alias so that student ID now contains all possible SSNs and school district IDs present in the Survey 5 demographic format.

TXTID is a concatenation of the first four letters of the student’s last name, the first three letters of the student’s first name, and the month and year of their date of birth. It is used as an additional method for grouping student records in cases where the same student is reported in the demographic format using multiple, different student IDs.

TRUESID is the student’s SSN, whenever present in the demographic file, or the school district identification number if no SSN is present for that student. If multiple SSNs are present then the first one (starting with 592, if possible, since this is a common SSN prefix in Florida) is selected. If no SSN is present then the first district ID is selected. A student is given a TRUESID for every academic year, and the digit that follows the variable title delineates the reference year. For example, TRUESID0 is for the academic year 1999-2000.

The entire demographic format, consisting of nearly 3.9 million records after adding records where alias and SID differ, is assigned a TRUESID. The file is then *unduplicated* (though no records are actually deleted) by SID and again by TXTID. TRUESID is electronically “lagged down” to all records according to scoring criteria. This process is largely automated and compares first name, last name, middle initial, date of birth, race, county, and gender between records sharing Student ID, and again between records sharing TXTID. Using probabilistic record linkage scoring criteria, all but approximately 100,000 records are assigned a TRUESID. Research staff must examine the remainder manually, and a judgment call must be made. Once this process is complete, the cohort(s) may be selected.

Cohorts Produced for the Annual Report

Three student-level cohorts were produced using the “cleaned” DOESurvey 5 Demographic format data for this year’s annual report. These include, by chapter:

Chapter 7 Cohort 1: Incarceration, Educational Opportunity and Community Reintegration

- all youths *released* from any DJJ residential commitment program during FY 2000-01

Chapter 8 Cohort 2: Incarceration, Educational Opportunity and Community Reintegration

- all youths *released* from any DJJ residential commitment program during FY 2001-02

Chapter 9 Volusia County Pilot Project

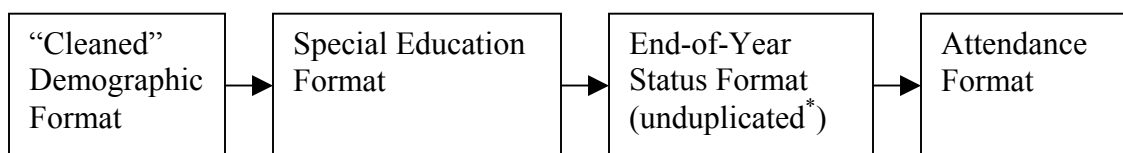
- all youths *released* from either of two Volusia County Alternative Discipline Schools during FY 1999-00
- all youths *released* from either of two Volusia County Alternative Discipline Schools during FY 2000-01
- all youths *released* from either of two Volusia County Alternative Discipline Schools during FY 2001-02

Creating the Cohorts

Data for the three cohorts were selected using the school number from DOE Survey 5 data for a given year. Using the Master School ID list as well as the expertise of JJEPP staff, all residential DJJ Commitment programs were identified by school number and selected from the Survey 5 Demographic Format for FY 2000-01. The process was identical for selecting the Volusia County cohorts, except that instead of DJJ schools, the school numbers for the two Alternative Disciplinary Schools were used. This excludes any students who had already earned diplomas prior to entering the DJJ program since they are not contained in the DOE data, but does not affect the Volusia cohorts. Once identified, the cohorts were further reduced to only those youths who were released from their programs during the school year in question, based on withdrawal code and withdrawal date.

- Data obtained from DOE arrive in separate formats (Student Demographics, Attendance, Disciplinary Referral, End of Year Status, ESE Status, and Transcript), which must be *linked* together and later *matched* to other data sources, such as FDLE, FDOC, FETPIP, and JJEPP's own program-level QA database.
- *Linking* within the DOE Survey 5 formats is done using SID (either an SSN or an alias), District, and School Number.
- *Matching* to data sources outside DOE Survey 5 is done using SSN and TXTID.
- Once data are grouped, linked, and matched, they may be summarized and analyzed.

Data are linked in the following order:



Students may attend, and even be released from, more than one DJJ school within a given school year. In keeping with the notion of longitudinal follow up, the *last* DJJ (or Volusia) school from which the student was released is selected as the cohort record. Because follow-up analyses are calculated using release date from the DJJ (or Volusia) program, records with

no release date are excluded. If a student’s only DJJ (or Volusia) record in the DOE Survey 5 demographic file is missing an exit date, that student cannot be retained in the cohort.*

Widow and Orphan Records

Occasionally, data in the demographic format may not have a corresponding record in the attendance format. Or, conversely, a student who might otherwise be selected for inclusion in the cohort may have a line in the attendance file but not have a corresponding record at the same school in the demographic file. These records are called “widow” and “orphan” records. Widow and orphan records were excluded from the cohorts.

All records with release dates prior to the entry date into the cohort record program were discarded. All subsequent records were used for follow-up analyses.

At this point, the cohort file was matched to subsequent years’ “cleaned” demographic formats to build a placement history spanning the entire period from release to the end of follow-up in order to ascertain short- and long-term outcomes. The matching procedure included three steps. The first used TRUSID, the second used SID and the last used TXTID in an effort to locate students in following years’ data. The cohorts were further refined by examining student withdrawal codes after being linked to the Survey 5 attendance format and matched to subsequent years. Records that could be identified as “rollovers” (i.e., students who appeared in the same school the following year with less than a two week break or who were only gone during the summer semester and did not have any other attendance record at a different school in between) were removed from the analyses since they had not actually been released during the school year. Withdrawal codes also were helpful in making a determination regarding releases; however, since many records did not contain a withdrawal code, it could not be the sole metric used to make the determination.

Tracking Student-Level Data Across Multiple Years

Only about two thirds of cases match from one year to the next in the FLDOE Survey 5 demographic format.

Possible reasons why students may not be found in future Survey 5 data:

- Students obtained a high school diploma or its equivalent while committed to DJJ.
- Students may have left the state after their incarceration.
- Local school district registrar never officially enrolled the student.
- The student’s SSN or SID may have been reported incorrectly.

*Fewer than 200 records in a given year contain duplicate sid disnum1 and school data in the end-of-year status format. These duplicates represent “co-enrollment” where a student simultaneously attends high school and adult education classes during the evening, thereby doubling the number of credits that can be earned in a semester. The result is often graduation or a GED, which only shows up in one of the records. Unduplicating this file involves taking the record with the diploma and discarding the other one.

- Death of the student
- The student dropped out of school.
- The student entered private school.

Educational performance and outcomes are measured using the variables of return to school, arrest, recommitment, attendance rate, employment, diplomas and credits earned. Return to school is defined as whether the youth returned to a secondary, non-DJJ school following release from the DJJ program. There are many possible measures of recidivism. The one used in this report is based on re-arrest using FDLE data. Given that longitudinal recidivism data were not available from DJJ, it was necessary to reach a conclusion regarding recommitment using the data obtained from DOE. The DOE records include youths' placements in juvenile justice schools, but often do not contain the specificity necessary to discern whether such a placement is merely a transfer commitment or an aftercare commitment associated with the original placement resulting in the youth being included in the 2000-2001 cohort, or whether the placement is a continuation of the original placement and re-commitment to the same facility. As such, the most conservative approach was taken by defining a recommitment as only placements in a higher security level program within one year of release from a DJJ program. Individual outcomes also were examined relative to the security levels of the program from which youths were released. DJJ has a four-tier security and restrictiveness level system for its residential programs. In order of restrictiveness, the levels are as follows: low-risk residential, moderate-risk residential, high-risk residential, and maximum-risk residential/juvenile prisons. Day treatment programs often serve a mix of intensive probation, referral, prevention, and conditional release students. Because DOE student level data do not distinguish between these different types of youths served in day treatment programs, day treatment was excluded from the cohort used in Chapter 8.

Measurement of prior arrests and arrests after release from a residential DJJ facility

The FDLE was the source of arrest data for the measurement of both the number of prior arrests and whether and when DJJ youths were arrested subsequent to release from a residential facility. A formal data sharing agreement was first established with FDLE's Statistical Analysis Center (SAC). JJEPP then supplied the SAC with a dataset of the FY2000-01 cohort that contained offender identifiers including; last name, first name, middle initial, social security number, sex, race, and date of birth. Using these identifiers, the SAC matched the cohort to FDLE's Computerized Criminal History (CCH) database to extract all arrest records for any offender who was in both datasets. Only cases that matched on an appropriate number and type of identifiers to ensure they were the same person were retained as legitimate matches. Arrest events with multiple charges were counted as one arrest.

The type of arrest charges reported to FDLE from local law enforcement agencies are those submitted by local law enforcement agencies in accordance with section 943.051, F.S.

Measurement of employment after release from a residential DJJ facility

The data used to determine whether DJJ releases were employed were obtained from the Florida Education and Training Placement Information Program (FETPIP). The SSNs of the

FY2000-01 release cohort were shared with FETPIP as part of a data sharing agreement and were used to match to the quarterly employment data in their repository. Only employment records of those with SSNs that have been verified by the Social Security Administration are retained by FETPIP, therefore, if a youth provided an invalid SSN and was employed, there would be no match between the two datasets. Therefore, the number of employed youths reported for the cohort may be an underrepresentation of the actual number employed.

For those youths who have employment records, FETPIP supplied data on each year and quarter they were employed, from quarter three of 2000 to present. Additionally, the average salary earned during each quarter of employment was part of the data FETPIP shared with JJEPP.

For analysis purposes, the first step was to determine the first quarter after release that the youth was available to work. It was decided that a release during any time in the first half of a quarter made him or her available to work during that quarter and any subsequent quarters. A release in the latter half of a quarter made the youth eligible to be employed during the following quarter and any subsequent quarters. Based on this determination of the quarter of employment eligibility, and which quarters the youth was employed, it was possible to create variables that indicated whether or not the youth was employed at any time during the first six and 12 months after release from a residential facility.

Measurement of academic, vocational, and elective credits earned while in DJJ facilities

The FY2000-01 DJJ release cohort was matched to FLDOE transcript data to capture data on academic, vocational, and elective credits earned while in DJJ facilities. These credits only apply to those earned while in high school because elementary and middle school students do not earn Carnegie credits. These data include a record for each specific type of class taken and the associated number of credits earned. The specific class types were grouped into the three categories of academic, vocational, and electives; the total number of credits earned within each broad category was summed. Additionally, the total number of credits earned while in DJJ facilities was summed across the three types of credits, and the percentage of the total comprised of academic, vocational, and elective credits, was calculated.

In order to then quantify *academic attainment* while in DJJ, a measure was developed which takes into consideration both the total number of academic credits earned and the proportion of all credits earned that were academic. To consider both these indicators of academic attainment, a scale score was developed by first weighting the total number of academic credits earned by the proportion of all credits earned that were academic by multiplying these two values. The scale score after weighting was difficult to interpret. Thus, Z scores for the weighted score were computed by subtracting the mean of the weighted score distribution from every weighted score and then dividing it by standard deviation of the weighted scores. This procedure converted the distribution of the scale score into one that was approximately normal, with a mean of zero and standard deviation of one, such that the deviation from the mean could be interpreted easily in terms of the percentage of the distribution that was above or below a given score.

The final measure of the level of academic attainment was measured based on whether the student was below or above the average on the scale score. A value of zero was used if the student was below the mean on the scale, and a value of one was applied if the student was above the average of all the scale scores.

Measurement of return to school and attendance upon returning to school

The FY2000-01 DJJ release cohort was matched to FLDOE attendance data to determine whether the juvenile returned to public school within one semester after DJJ release and the level of attendance if they returned. The DOE attendance records have the dates of enrollment, the number of days the student was in attendance, and the number of days they were absent. In order to capture the level of commitment to education upon release from DJJ, whether the juvenile returned to school or not was combined with the level of attendance. Whether they returned to school was simply based on whether they were enrolled for at least one day.

The level of school attendance is based on a measure that takes into account both the number of days students attended school and the percentage of enrollment days that they attended. The purpose of this measure is to capture the level of commitment youths have to education. Therefore, if a youth is enrolled in school for a very few days but attends all of those days and then drops out of school, using the percentage of enrolled days attended gives them a value of 100%. Using only the attendance percentage in this case would exaggerate the level of commitment to education. Also, if a student attends for many days (say 180) and has an attendance rate of 90%, his level of commitment to school, based on his attendance, is quite high, but his attendance rate is less than the previous example of low enrollment days with perfect attendance.

To consider both the number of days present in school and the percentage of enrollment days present, a scale score was developed by first weighting the percentage of days present by the number of days present. This was done by multiplying the percentage of days present by the number of days present. The scale score after weighting was difficult to interpret. Thus, Z scores for the weighted score were computed by subtracting the mean of the weighted score distribution from every weighted score and then dividing it by standard deviation of the weighted scores. This procedure converted the distribution of the scale score into one that was approximately normal, with a mean of zero and standard deviation of one, where the deviation from the mean could be interpreted easily in terms of the percentage of the distribution that was above or below a given score .

A variable that combines whether DJJ releases returned to school and their level of attendance was defined with three values. A zero indicated that they did not return to school. If they returned to school and their attendance rate was below the average on the attendance scale score for those who did return, they were given a value of one. If they returned to school and their attendance rate was above average, based on the attendance scale score, they were given a value of two. In other words, the higher the value on this variable, the higher the level of commitment to education. The inclusion of the below or above average attendance provides a more precise and useful indicator of the level of commitment to

education than one that simply indicates if the juvenile returned to school, because many youths return to school but have low rates of attendance.