

Analysis of Missing Social Security Numbers (SSN) for Accountability Reporting for the Community Colleges (ARCC) Research & Planning, System Office

Introduction

Student Social Security Numbers (SSNs) are used to determine First-Time Freshman (FTF) status in the California Community College (CCC) system. This determination involves identifying prior enrollments outside the CCC system through a match with data from the National Student Clearinghouse (NSC), University of California (UC) and California State University (CSU). SSNs are also used to match student records with four-year institutions to establish subsequent transfers and to link students with wage data from California's Employment Development Department (EDD). The Accountability Reporting for the Community Colleges (ARCC) uses wage data to measure the impact of a community college degree or certificate on student income as one of its performance indicators. The purpose of the analysis reported below is to assess the quality of the SSN data for each college and statewide.

Methodology

The Student Identifier (SB00 in the CCC's Management Information System) is the data element that uniquely identifies a student for all terms and at all colleges. In general, the student's SSN is reported in this data element. If a student does not have or declines to use a SSN, the system requires the college to assign a unique number instead. This locally assigned identifier is distinct for that particular college or district but does not apply statewide. In other words, local identifiers cannot be used to establish FTF status, or match with other datasets from four-year institutions or the EDD. For the purpose of this analysis, missing SSN is defined as the combination of invalid SSNs and the use of local identifiers. The analysis of the Student Identifier data element focused on credit students enrolled between 2000-01 and 2005-06.

Results

Data for this analysis were extracted from the CCC MIS in September 2006. The number and percentage of missing SSNs statewide for six academic years is presented in Table 1. The percentage of missing SSNs statewide ranges from 5.7% to 6.3% during the period from 2000-01 to 2005-06.

Table 1: Missing SSNs Analysis Statewide Credit Students, 2000-01 to 2005-06						
	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06
SSNs¹	911,288	953,995	904,585	748,791	740,342	742,950
Invalid SSNs²	9,990	9,553	8,439	6,856	6,682	7,080
Local Identifiers³	49,451	49,765	45,415	41,952	40,129	61,022
Total Missing SSNs⁴	59,441	59,319	53,854	48,808	46,811	68,102
Percent Missing SSNs⁵	6.12%	5.85%	5.62%	6.12%	5.95%	8.40%

¹ **SSNs:** Valid SSN in the System Office MIS

² **Invalid SSNs:** Identified at the System Office as invalid SSN

³ **Local Identifiers:** Student does not have SSN or declines to give SSN; college assigns a unique local identifier

⁴ **Total Missing SSNs:** Invalid SSNs + Local Identifiers

⁵ **Percent Missing SSNs:** [Total Missing SSNs / (SSNs + Invalid SSNs + Local Identifiers)] * 100

The percentage of missing SSNs by college for each academic year is displayed in Table 2. The colleges are listed in alphabetical order.

College	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006
ALAMEDA	3.9%	4.2%	4.8%	4.2%	3.7%	3.6%
ALLAN HANCOCK	1.3%	1.7%	1.1%	1.0%	2.1%	2.1%
AMERICAN RIVER	1.7%	1.5%	2.2%	2.1%	2.1%	2.9%
ANTELOPE VALLEY	2.5%	2.5%	3.1%	3.0%	3.0%	2.8%
BAKERSFIELD	2.1%	2.2%	2.5%	2.7%	3.0%	3.5%
BARSTOW	2.4%	2.7%	2.7%	1.6%	1.1%	1.5%
BERKELEY	3.1%	3.3%	3.3%	4.3%	3.3%	4.8%
BUTTE	12.0%	16.3%	12.8%	7.6%	5.0%	4.8%
CABRILLO	2.0%	1.7%	1.9%	3.3%	2.6%	3.5%
CANADA	8.0%	15.6%	17.4%	26.5%	27.3%	27.9%
CANYONS	4.9%	4.1%	7.5%	19.1%	29.0%	32.9%
CERRITOS	3.6%	5.1%	6.0%	7.6%	7.4%	16.1%
CERRO COSO	3.0%	4.6%	4.2%	3.1%	3.3%	2.1%
CHABOT	3.7%	3.4%	4.1%	3.2%	4.2%	5.3%
CHAFFEY	3.3%	4.5%	6.7%	4.8%	5.6%	6.7%
CITRUS	4.5%	4.2%	4.5%	6.8%	6.2%	5.7%
COASTLINE	5.4%	6.1%	5.7%	5.2%	5.5%	5.3%
COLUMBIA	1.9%	1.7%	1.6%	1.4%	1.2%	1.7%
COMPTON	45.8%	17.4%	23.6%	22.0%	22.9%	24.5%
CONTRA COSTA	13.6%	15.1%	17.6%	7.1%	7.9%	11.6%
COPPER MOUNTAIN	1.1%	1.2%	1.0%	1.4%	1.0%	1.5%
COSUMNES RIVER	1.8%	1.8%	2.2%	2.7%	2.1%	2.3%
CRAFTON HILLS	1.2%	1.6%	1.6%	1.6%	1.5%	1.9%
CUESTA	2.0%	2.4%	2.7%	3.2%	3.6%	5.3%
CUYAMACA	3.1%	3.1%	4.0%	4.7%	4.2%	4.3%
CYPRESS	8.4%	8.1%	3.8%	4.3%	4.4%	5.6%
DE ANZA	6.6%	6.5%	6.0%	6.8%	6.6%	15.0%
DESERT	6.4%	6.5%	6.6%	5.6%	7.1%	8.3%
DIABLO VALLEY	4.2%	5.7%	5.9%	6.2%	5.9%	8.3%
EAST L.A.	14.1%	10.8%	13.9%	6.5%	6.9%	45.3%
EL CAMINO	4.9%	5.4%	4.9%	5.7%	6.0%	6.7%
EVERGREEN VALLEY	3.9%	4.6%	5.7%	9.4%	7.3%	15.2%
FEATHER RIVER	7.7%	13.4%	8.1%	7.1%	6.8%	8.8%
FOLSOM LAKE					2.8%	4.5%
FOOTHILL	5.4%	5.7%	6.2%	6.6%	7.5%	7.7%
FRESNO CITY	2.0%	2.2%	3.1%	3.3%	3.6%	4.0%
FULLERTON	7.3%	6.8%	4.8%	5.1%	5.5%	6.0%
GAVILAN	2.1%	3.4%	2.4%	2.3%	3.6%	3.9%
GLENDALE	5.0%	5.8%	5.8%	6.8%	7.7%	8.4%
GOLDEN WEST	6.0%	10.9%	10.9%	5.1%	5.8%	5.5%
GROSSMONT	4.7%	4.4%	4.6%	5.1%	5.2%	6.7%
HARTNELL	3.9%	3.9%	4.9%	8.1%	5.8%	7.5%
IMPERIAL VALLEY	2.5%	2.3%	2.0%	2.4%	2.0%	3.0%

Table 2: Percent Missing SSNs by College By Year

College	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006
IRVINE VALLEY	6.6%	7.0%	6.7%	9.0%	8.9%	10.8%
L.A. CITY	5.3%	6.8%	9.1%	9.6%	9.8%	11.8%
L.A. HARBOR	11.7%	11.8%	6.1%	7.7%	8.6%	8.6%
L.A. MISSION	4.0%	5.8%	6.3%	7.1%	7.7%	8.3%
L.A. PIERCE	7.2%	7.7%	8.1%	9.1%	10.5%	11.4%
L.A. TRADE-TECH	34.3%	18.0%	6.0%	6.1%	7.4%	12.8%
L.A. VALLEY	3.4%	4.2%	6.5%	7.3%	7.1%	8.9%
LAKE TAHOE	2.5%	2.2%	3.0%	2.9%	3.4%	4.0%
LANEY	5.5%	5.2%	4.7%	5.0%	4.1%	4.3%
LAS POSITAS	3.4%	3.6%	4.8%	3.3%	3.3%	4.6%
LASSEN	18.3%	17.7%	15.4%	16.4%	8.3%	3.3%
LONG BEACH CITY	2.4%	2.6%	2.4%	5.1%	4.3%	4.7%
LOS MEDANOS	4.9%	3.7%	6.1%	3.8%	5.5%	7.0%
MARIN	7.6%	8.6%	8.4%	7.8%	7.3%	9.1%
MENDOCINO	1.5%	1.9%	2.2%	2.5%	4.2%	3.7%
MERCED	7.2%	7.5%	4.9%	3.6%	6.9%	8.2%
MERRITT	9.7%	9.6%	5.1%	3.9%	3.6%	3.1%
MIRA COSTA	3.6%	4.5%	7.2%	5.2%	5.7%	7.4%
MISSION	5.2%	5.2%	5.0%	6.2%	6.3%	6.9%
MODESTO	2.2%	2.4%	2.2%	2.2%	2.4%	2.9%
MONTEREY	4.4%	3.9%	5.3%	6.0%	5.7%	6.0%
MOORPARK	2.9%	2.9%	2.9%	3.3%	3.6%	3.8%
MT. SAN ANTONIO	4.5%	10.5%	5.4%	5.3%	6.3%	7.7%
MT. SAN JACINTO	5.3%	4.7%	5.1%	4.8%	4.2%	4.4%
NAPA VALLEY	2.1%	2.2%	3.8%	5.7%	6.3%	10.3%
OHLONE	5.0%	10.1%	13.1%	8.6%	9.9%	11.4%
ORANGE COAST	6.4%	6.5%	6.8%	6.6%	6.7%	5.9%
OXNARD	4.1%	5.4%	3.7%	4.0%	3.9%	4.4%
PALO VERDE	8.0%	3.3%	11.0%	7.3%	10.4%	18.9%
PALOMAR	5.5%	5.4%	4.8%	4.6%	4.1%	4.5%
PASADENA CITY	5.3%	6.3%	6.5%	7.3%	7.5%	10.3%
PORTERVILLE	1.8%	2.1%	3.0%	2.3%	2.6%	2.6%
REDWOODS	1.7%	1.5%	7.0%	2.3%	2.1%	2.3%
REEDLEY	2.4%	2.1%	2.2%	3.0%	3.1%	3.8%
RIO HONDO	3.9%	4.8%	5.6%	7.1%	9.4%	18.2%
RIVERSIDE	4.6%	5.6%	4.8%	3.9%	4.4%	6.0%
SACRAMENTO CITY	2.0%	2.0%	2.9%	2.6%	2.9%	3.6%
SADDLEBACK	5.4%	6.2%	6.1%	8.2%	8.3%	9.4%
SAN BERNARDINO	1.4%	1.5%	1.6%	1.5%	1.9%	1.8%
SAN DIEGO CITY	2.3%	2.7%	4.4%	4.4%	4.5%	7.9%
SAN DIEGO MESA	2.7%	3.6%	4.2%	4.7%	6.2%	12.9%
SAN DIEGO MIRAMAR	3.1%	2.9%	3.7%	4.0%	4.9%	10.8%
SAN FRANCISCO CITY	5.5%	5.9%	7.0%	39.4%	12.4%	6.6%
SAN JOAQUIN DELTA	2.0%	1.8%	2.0%	2.0%	2.2%	2.8%
SAN JOSE CITY	5.6%	4.4%	5.3%	3.8%	4.2%	4.8%
SAN MATEO	3.7%	4.0%	4.4%	6.7%	5.9%	8.6%

Table 2: Percent Missing SSNs by College By Year

College	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006
SANTA ANA	5.2%	6.3%	8.7%	9.5%	8.0%	9.9%
SANTA BARBARA CITY	5.2%	4.6%	5.4%	5.6%	7.0%	7.3%
SANTA MONICA CITY	17.4%	17.7%	16.2%	9.6%	9.3%	11.9%
SANTA ROSA	2.6%	2.9%	3.5%	3.6%	4.1%	4.5%
SANTIAGO CANYON	5.0%	4.6%	4.8%	5.2%	4.9%	6.8%
SEQUOIAS	2.5%	2.5%	4.1%	3.7%	3.9%	4.5%
SHASTA	3.0%	3.1%	2.3%	1.8%	1.6%	2.3%
SIERRA	2.0%	2.2%	2.3%	2.5%	2.6%	2.8%
SISKIYOU	7.1%	5.4%	4.5%	11.6%	14.8%	20.0%
SKYLINE	2.7%	3.6%	2.9%	3.8%	3.3%	4.9%
SOLANO	1.4%	1.6%	1.3%	1.5%	1.5%	1.8%
SOUTHWEST L.A.	4.4%	17.3%	3.3%	3.0%	3.5%	3.5%
SOUTHWESTERN	3.3%	2.9%	3.0%	3.2%	3.2%	6.0%
TAFT	1.5%	1.8%	1.9%	2.2%	2.6%	2.8%
VENTURA	2.6%	3.2%	3.0%	4.2%	3.5%	4.4%
VICTOR VALLEY	1.9%	2.0%	1.7%	2.0%	4.8%	3.5%
WEST HILLS	6.9%	10.0%	7.1%	6.5%	5.7%	6.7%
WEST L.A.	27.5%	20.2%	7.0%	5.4%	5.5%	6.4%
WEST VALLEY	4.4%	5.7%	4.1%	4.5%	5.0%	5.1%
YUBA	4.0%	4.9%	4.4%	2.4%	2.6%	3.8%

Statistics related to the percentage of missing SSNs across the colleges are presented in Table 3.

	Percent Missing SSNs 2000-01	Percent Missing SSNs 2001-02	Percent Missing SSNs 2002-03	Percent Missing SSNs 2003-04	Percent Missing SSNs 2004-05	Percent Missing SSNs 2005-06
N	108	108	108	108	109	109
Mean	5.48%	5.60%	5.46%	5.76%	5.74%	7.39%
Median	4.04%	4.33%	4.79%	4.77%	4.94%	5.60%
Std. Dev.	6.13%	4.39%	3.78%	5.07%	4.39%	6.45%
Minimum	1.08%	1.25%	1.04%	1.00%	1.01%	1.50%
Maximum	45.83%	20.16%	23.57%	39.38%	29.03%	45.30%