

# DATA VALIDATION FY 05

---

By Howard Frim

8/22/05

# Index

- Introduction
- Background
- Deciding the Process
- Designing the Database
- Conducting the Review
- Findings by Funding Source
- Missing Records
- Findings by Reporting Item
- Conclusions
- Implications

# Introduction

- **In Fiscal Year 2004 (FY 04) United States Department of Labor (USDOL) Employment and Training Administration (ETA) began a data validation initiative that covers the accuracy of the program activity and performance outcome as well as the accuracy of the individual data elements.**
- **States are now required to ensure report accuracy by reviewing samples of participant records against source documentation to ensure compliance with Federal definitions.**
- **To accomplish this goal, ETA has developed standardized software that utilizes sampling methodology that each state will use in completing the required data validation.**

# Background

- **In Massachusetts, the Division of Career Services (DCS) as the Workforce Investment Act (WIA) Title I Administrator is mandated with the responsibility of conducting data validation.**
- **The Division of Career Service's (DCS) Quality Assurance Unit, part of Career Services Central Operations has the responsibility for conducting data validation.**
- **DCS's Performance and Reporting Unit developed the database, reviewed the information and submitted the reports.**

# Deciding the Process

- **DCS staff met and discussed the various options before deciding upon the strategy that was utilized for conducting the Data Validation. It was immediately agreed that in an effort to decrease workload at the Career Center level DCS would utilize the data validation database.**
- **The major concern was to devise a way to complete the review process and enable the reviewers to utilize laptops rather than the alternative method of using cumbersome print outs which would then necessitate entering all the data from the print outs into the database.**

# Designing the Database

- **When the data validation process began, DCS found there were many issues with both the Workforce Investment Act (WIA and Trade Adjustment Assistance (TAA) databases.**
- **The USDOL database forms (screens) were cumbersome, non user-friendly or intuitive. Users could not find information easy.**
- **Cases could not be accessed by using certain criteria such as last name, or offices. The data could not be viewed or exported easily into other applications such as Microsoft Word and Excel.**
- **The databases lacked certain report features and did not provide a view of completed and non-completed records.**

# Designing the Database Cont.

- **DCS also faced the challenge of the systems inability to allow multiple QA staff to access files at the same time using their laptops.**
- **The DCS Massachusetts Data Element Validation (MDEV) database was designed to meet these needs. The data was imported into both the USDOL TAA and WIA databases in order to produce the sample counts. The MDEV database then was able to communicate with the USDOL TAA and WIA databases by linking to their sample tables.**
- **The database was loaded onto each laptop via CDs, and each night after completion of the days review.**

# Designing the Database Cont.

- **The QA staff were responsible for saving the daily work to a CD and connecting to MDEV server and synchronizing that data.**
- **Prior to the start of the data collection, the staff received training on utilizing the laptop, utilizing the specially designed database, saving the information, and copying it on to the server.**
- **Each reviewer was also provided with copies of USDOL's data validation guides and numerical keys.**
- **Each QA staff member was able to work on his /her own databases which were continuously synchronized on the network with the main database.**

# Designing the Database Cont.

- **The screens were designed to provide a user-friendly environment.**
- **Users were able to view additional description information about each of the data elements.**
- **Users were able to search records by region, office, cohort and user names.**
- **Each user was able to track completed and non-completed cases by name, and office.**
- **Managers were able to track daily and weekly progress of the QA staff using customized reports such as pass/fail statistics.**

# Designing the Database Cont.

- Prior to the start of the data collection process, decisions were made regarding which data elements could be pre-filled via MOSES and/or other methods and did not have to be verified on site.
- Based upon the first round conducted in the field, DCS made significant changes to the MDEV database once it was realized that there were additional items that were available on MOSES and not in the file folders.

# Conducting the Review

- In FY05, DCS Quality Assurance staff conducted the Data Validation Reviews in 15 of the 16 local areas. One local area (South Coastal) did not have customers included in the data validation sample.
- In FY05 the Data Validation sample totaled 1135 customers which included 1004 WIA Dislocated Worker, Adult, Older Youth, and Younger Youth customers and 131 Trade customers.

# Conducting the Review Cont.

## DCS DATA VALIDATION SCHEDULE FY 05

<u>OPERATOR</u>	<u>SCHEDULED DATES</u>	<u>DV SAMPLE</u>
Berkshire	12/13-12/16/04	46
Boston	12/13-12/16, 12/20-12/22/04	97
Bristol	12/13-12/16, 12/20-12/22/04	119
Brockton	12/27-12/29/04	37
Cape Cod & Islands	12/13-12/16, 12/20-12/22/04	75
Central Mass	12/13-12/16, 12/20-12/22/04	73
Franklin/Hampshire	12/20-12/22/04	29
Greater Lowell	12/13-12/16, 12/20-12/22/04	77
Greater New Bedford	12/13-12/16, 12/20-12/22, 12/27-12/29/04	120
Hampden	12/13-12/16, 12/20-12/22/04	160
Lower Merrimack	12/13-12/16/04	48
Metro North	12/20-12/22/04	27
Metro South/West	12/20-12/22/04	29
North Central Mass	12/27-12/29/04	48
South Coastal	N/A (Not included in the Data Validation Sample)	N/A
North Shore	12/13-12/16, 12/20-12/22, 12/27-12/29/04	150

# Findings by Funding Source

- WIA Adult
- WIA Dislocated Worker
- WIA Older Youth
- WIA Younger Youth
- Trade

# WIA Adult

- The sample included 272 records.
- There were 3 missing customer records.
- Date of Birth was the data element with the most significant rate of error.

# WIA Adult

Cont.

- **WIA Data Element Validation Summary and Analytical Report — Adult—**

## DATE OF BIRTH

■ <u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
35	9.92%	10.00%

# WIA Dislocated Worker

- The sample included 290 records.
- There were 5 missing customer records.
- Date of Birth and Dislocation Date were the two data elements with the most significant rates of error.

# WIA Dislocated Worker Cont.

- **WIA Data Element Validation Summary and Analytical Report— Dislocated Worker—**

## DISLOCATION DATE

■ <u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
74	24.10%	25.85%

# WIA Dislocated Worker Cont.

- **WIA Data Element Validation Summary and Analytical Report— Dislocated Worker—**

## DATE OF BIRTH

■ <u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
33	10.91%	11.12%

# WIA Older Youth

- The sample included 186 records.
- There was one missing record.
- Date of Birth, Additional Assistance, Educational Status, Basic Literacy Deficient, and Follow-up were the five data elements with the most significant rates of error.

# WIA Older Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report– Older Youth —**

## DATE OF BIRTH

■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	7	2.23%	2.23.12%

# WIA Older Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report– Older Youth —**

## **ADDITIONAL ASSISTANCE**

■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	10	6.87%	6.89%

# WIA Older Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report— Older Youth —**

## EDUCATIONAL STATUS

■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	14	5.85%	5.87%

# WIA Older Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report– Older Youth**

## BASIC LITERACY DEFICIENT

■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	9	3.25%	5.04%

# WIA Older Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report– Older Youth**

## FOLLOW-UP

<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data</u>
28	25.44%	57.81%

# WIA Younger Youth

- The sample included 256 records.
- There was four missing records.
- Date of Birth, Additional Assistance, Educational Status, Basic Literacy Deficient, Follow-up, Youth Placement Info, and Youth Retention Info were the seven data elements with the most significant rates of error.

# WIA Younger Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report— Younger Youth —**

## DATE OF BIRTH

■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	12	4.84%	4.91%

# WIA Younger Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report– Younger Youth —**

## ADDITIONAL ASSISTANCE

- | <u># of Errors</u> | <u>Overall Error Rate</u> | <u>Reported Data Error Rate</u> |
|--------------------|---------------------------|---------------------------------|
| 9                  | 3.73%                     | 3.79%                           |

# WIA Younger Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report— Younger Youth —**

## EDUCATIONAL STATUS

- | <u># of Errors</u> | <u>Overall Error Rate</u> | <u>Reported Data Error Rate</u> |
|--------------------|---------------------------|---------------------------------|
| 12                 | 4.84%                     | 4.91%                           |

# WIA Younger Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report— Younger Youth —**

## **BASIC LITERACY DEFICIENT**

■ <u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
8	3.29%	4.80%

# WIA Younger Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report– Younger Youth —**

## **FOLLOW-UP**

■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	9	3.73%	6.95%

# WIA Younger Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report— Younger Youth —**

## YOUTH PLACEMENT INFO

■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	20	8.12%	18.97%

# WIA Younger Youth

Cont.

- **WIA Data Element Validation Summary and Analytical Report— Younger Youth —**

## YOUTH RETENTION INFO

■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	23	9.64%	21.77%

# Trade

- The sample included 131 records.
- There were zero (0) missing customer records.
- Dislocation Date was the one data element with the most significant rate of error.

- **TAA Data Element Validation Summary and Analytical Report— TAA—**

## DISLOCATION DATE

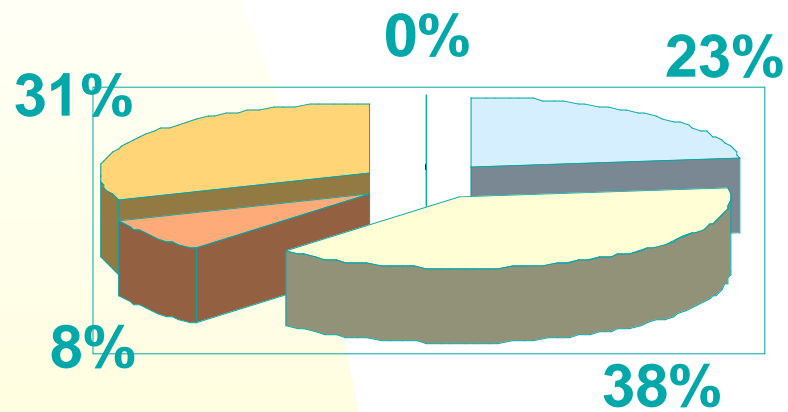
■	<u># of Errors</u>	<u>Overall Error Rate</u>	<u>Reported Data Error Rate</u>
	29	13.6%	13.6%

# Missing Records

- The missing records while statistically minor in nature (one percent of the total sample), demonstrated that certain local areas were not in full compliance with record retention requirements, and had not implemented adequate safeguards to ensure that one hundred percent of the customer records were being maintained.

# Missing Records

Cont.



- WIA Adult
- WIA Dislocated Worker
- WIA Older Youth
- WIA Younger Youth
- TAA

# Findings by Reporting Item

- Date of Birth
- Date of Dislocation
- Education Status
- Additional Assistance
- Basic Literacy Deficient
- Youth Placement Info
- Youth Retention Info
- Follow-up

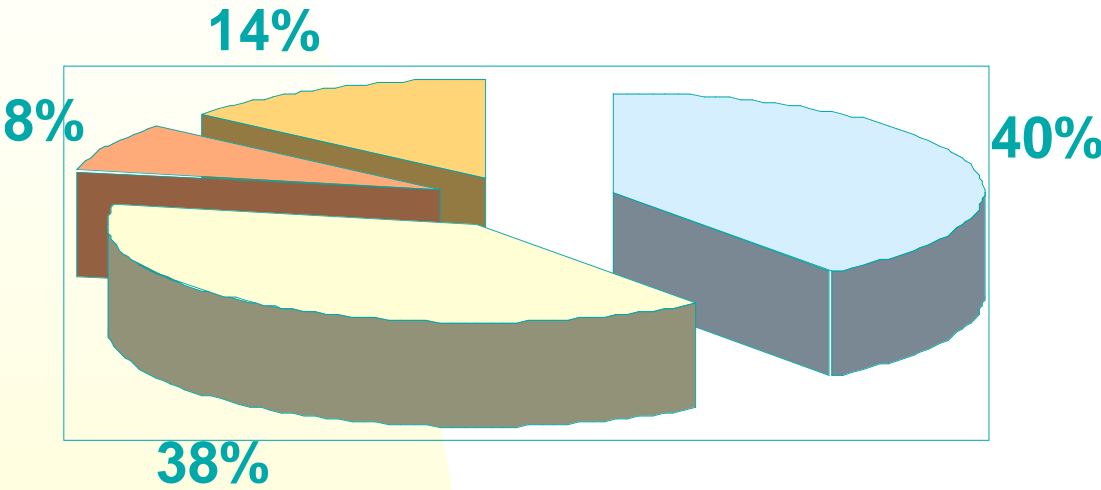
# Findings by Reporting Item Cont.

- Information captured at time of enrollment is sometimes erroneously changed. For example, the employment status of a WIA Title I Dislocated Worker is captured at the time of enrollment as “unemployed.”
- If subsequent to enrollment, the customer becomes employed, the local area needs to record the job placement. The local area should not go back and change the customer’s employment status from “unemployed” to “employed.”

# Findings by Reporting Item Cont.

- Another example is capturing a WIA Title I Youth's educational status at time of enrollment as "less than high school". If subsequent to enrollment, the customer attains their GED the local area needs to record this attainment.
- The local area should not go back and change the customer's education level from less than high school to High School/GED.

# Date of Birth



- WIA Adult
- WIA Dislocated Worker
- WIA Older Youth
- WIA Younger Youth

# Date of Birth

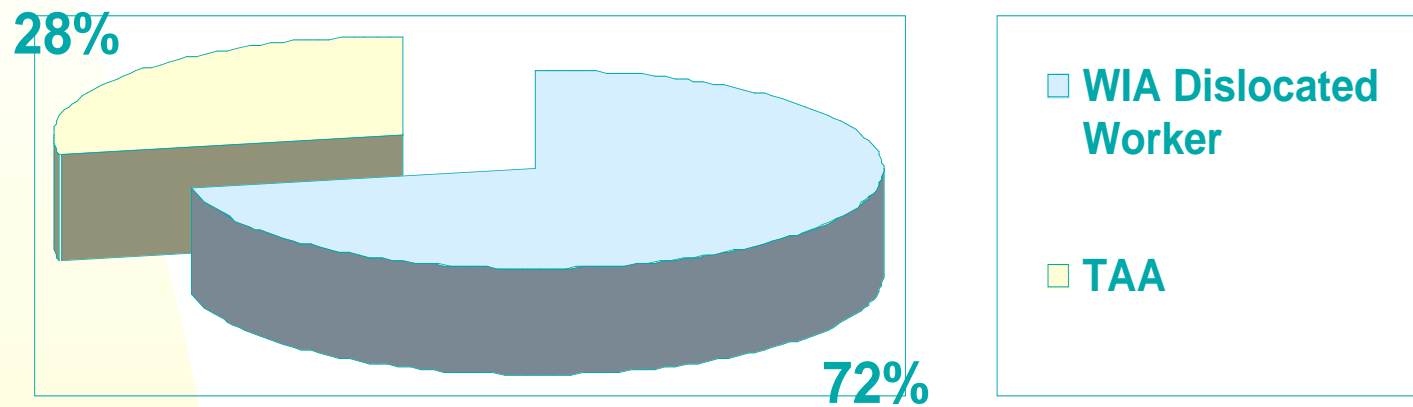
- This data element had one of the more significant error rates.
- Investigation determined that errors included copying the date that the birth was transcribed rather than the actual date of birth.
- Other times the numbers were transcribed from the source document to MOSES.

# Date of Birth

Cont.

- **In certain cases, the date of birth is recorded on documents that do not use the standard month, day of the month and year format. Rather they list the day of the month, month and year in that order. When transcribing the date, the operator has failed to accommodate for this difference.**
- **In other instances, no explanation of the errors could be discerned.**
- **Acceptable sources of documentation would include; copy of birth certificate, passport, driver's license, public assistance records.**

# Date of Dislocation



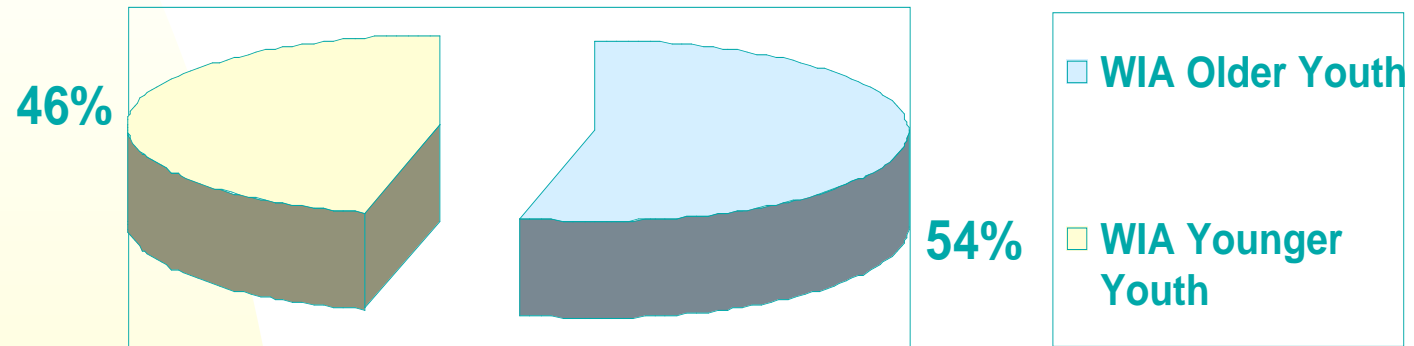
# Date of Dislocation Cont.

- **This data element was one with one of the higher error rate.**
- **Among the most frequently utilized sources for this data element were customer's attestation on registration forms contained in the customer files and Unemployment Insurance (UI) forms contained in the customer files.**
- **In many cases, the data was entered into MOSES with the month and the year without including the actual day of the month, as the customer did not supply this information. MOSES allows the information to be entered without the day of the month but as with certain other data elements MOSES either defaults to the 15th of the specific month. Therefore the information does not match and the data validation fails.**

# Date of Dislocation Cont.

- **In many cases, the information contained on the UI forms did not match the information entered into MOSES. A review of this information seemed to indicate that many customers had experienced multiple dislocations and the UI forms obtained and contained in the files pertained to different dislocations than the ones corresponding to the customers' enrollments.**
- **Acceptable sources of documentation would include; verification from employer, public notice of lay-off, unemployment verification, and rapid response list.**

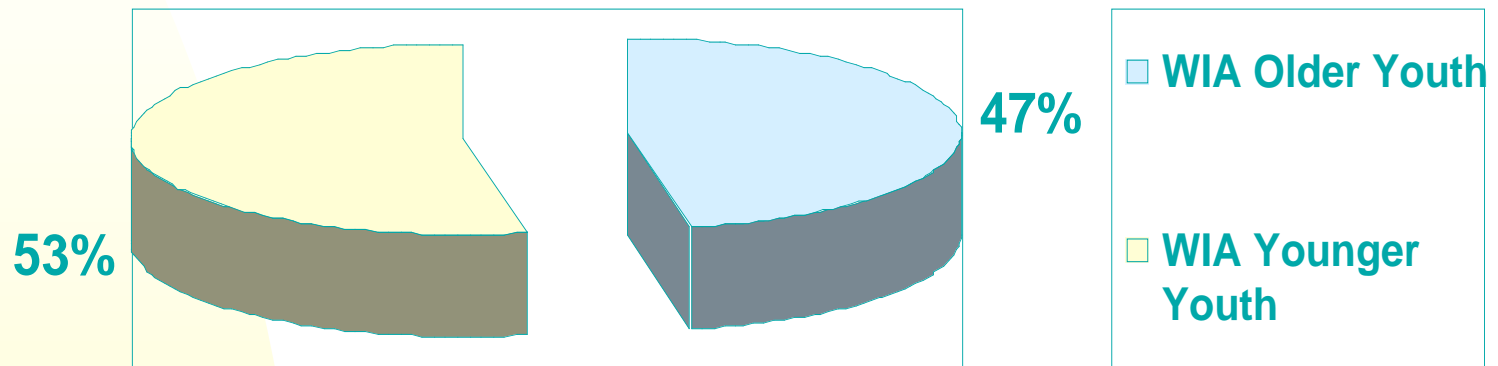
# Educational Status



# Educational Status Cont.

- In many cases, there was no source information contained in the case records. Therefore this data element could not be verified.
- In other cases, the information contained in the case records did not match what was recorded in MOSES.
- Acceptable sources of documentation would include applicable records from educational institutions (GED certificate, diploma, transcripts, drop-out letter), and WIA registration forms.

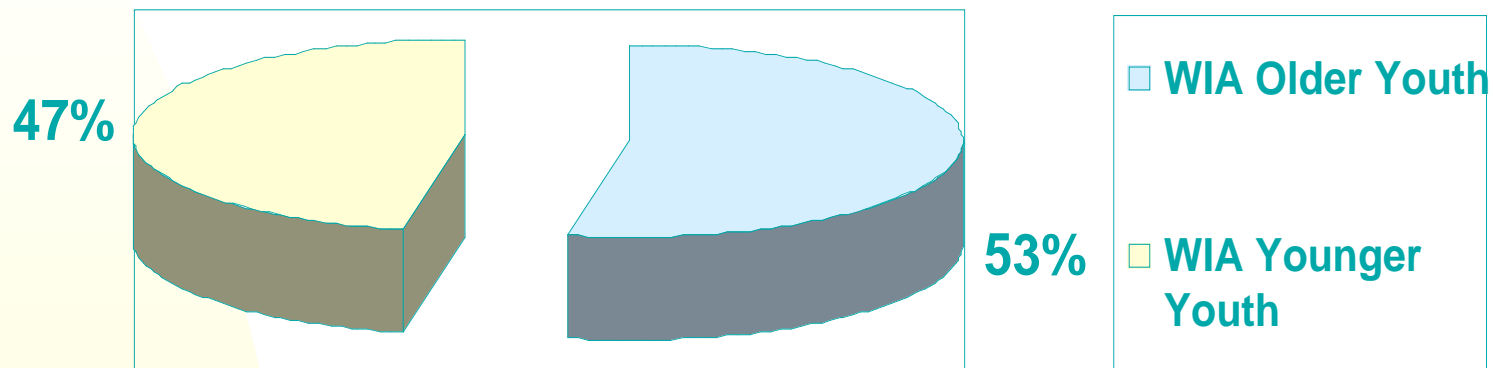
# Additional Assistance



# Additional Assistance Cont.

- **In many cases, there was no source information contained in the case records. Therefore this data element could not be verified.**
- **In other cases, the information contained in the case records did not match what was recorded in MOSES.**
- **Acceptable sources of documentation would include; State Plan and/or local WIB information, individual service strategy (ISS), case notes , and WIA registration forms.**

# Basic Literacy Deficient

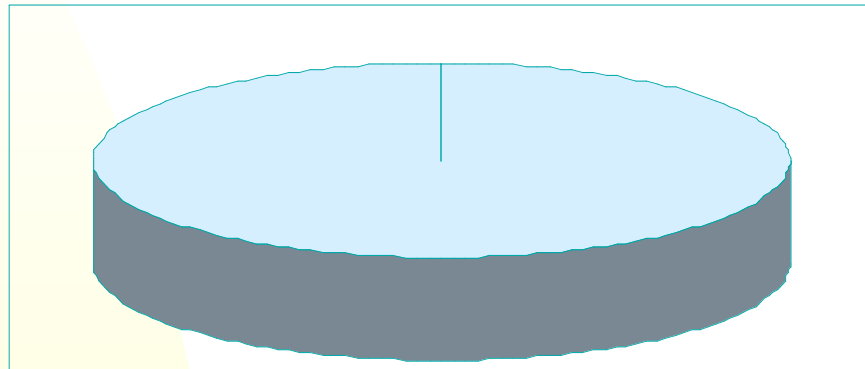


# Basic Literacy Deficient

Cont.

- **In many cases, there was no source information contained in the case records. Therefore this data element could not be verified.**
- **In other cases, the information contained in the case records did not match what was recorded in MOSES.**
- **Acceptable sources of documentation would include; standardized assessment tests, school records, and case notes.**

# Youth Placement Info



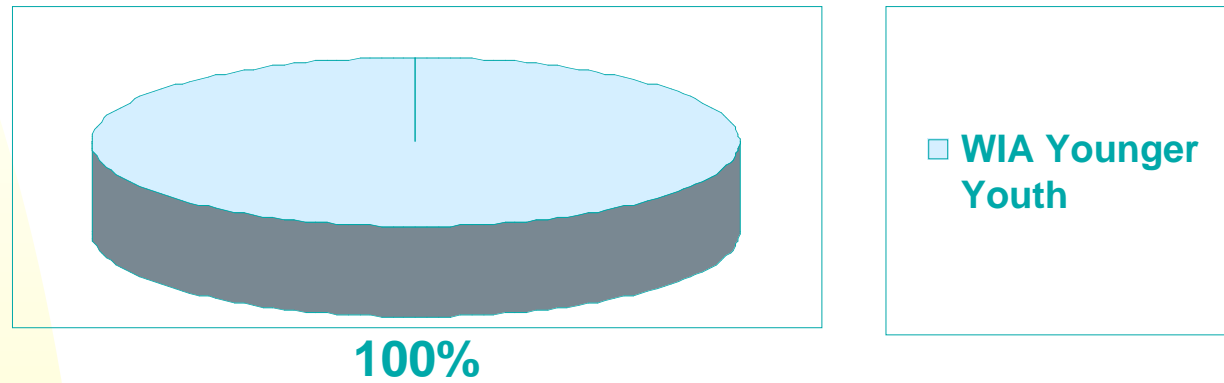
100%

■ WIA Younger Youth

# Youth Placement Info Cont.

- **In many cases, there was no source information contained in the case records. Therefore this data element could not be verified.**
- **In other cases, the information contained in the case records did not match what was recorded in MOSES.**
- **Acceptable sources of documentation would include; employer contact, WIA exit forms, youth vendor reports, and case notes.**

# Youth Retention Info

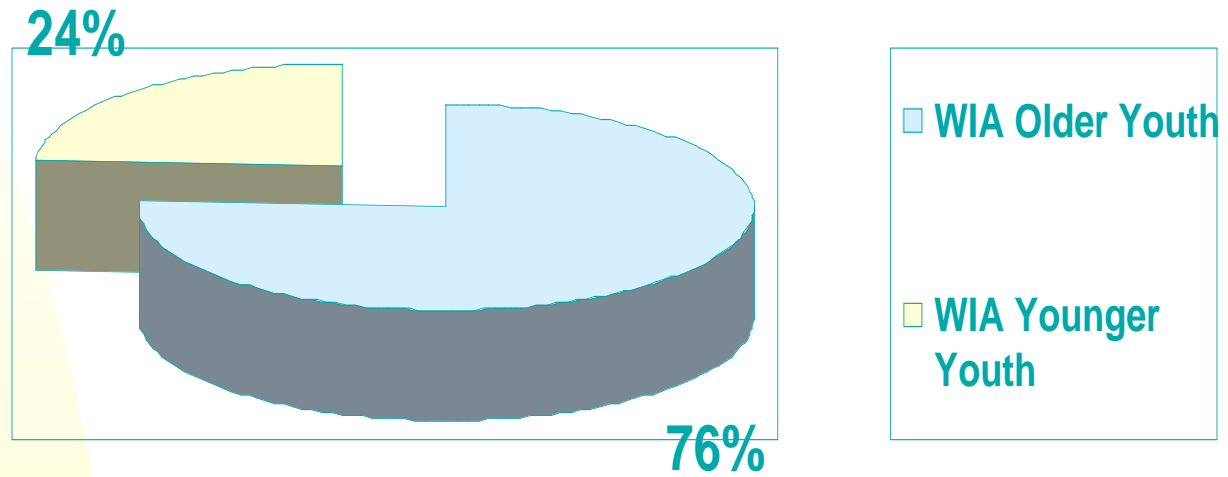


# Youth Retention Info

Cont.

- **In many cases, there was no source information contained in the case records. Therefore this data element could not be verified.**
- **In other cases, the information contained in the case records did not match what was recorded in MOSES.**
- **Acceptable sources of documentation would include; cross match with other agencies, military service verification, apprenticeship verification, employer contacts, post secondary education transcripts, registration forms, community college information, and case notes.**

# Follow-up

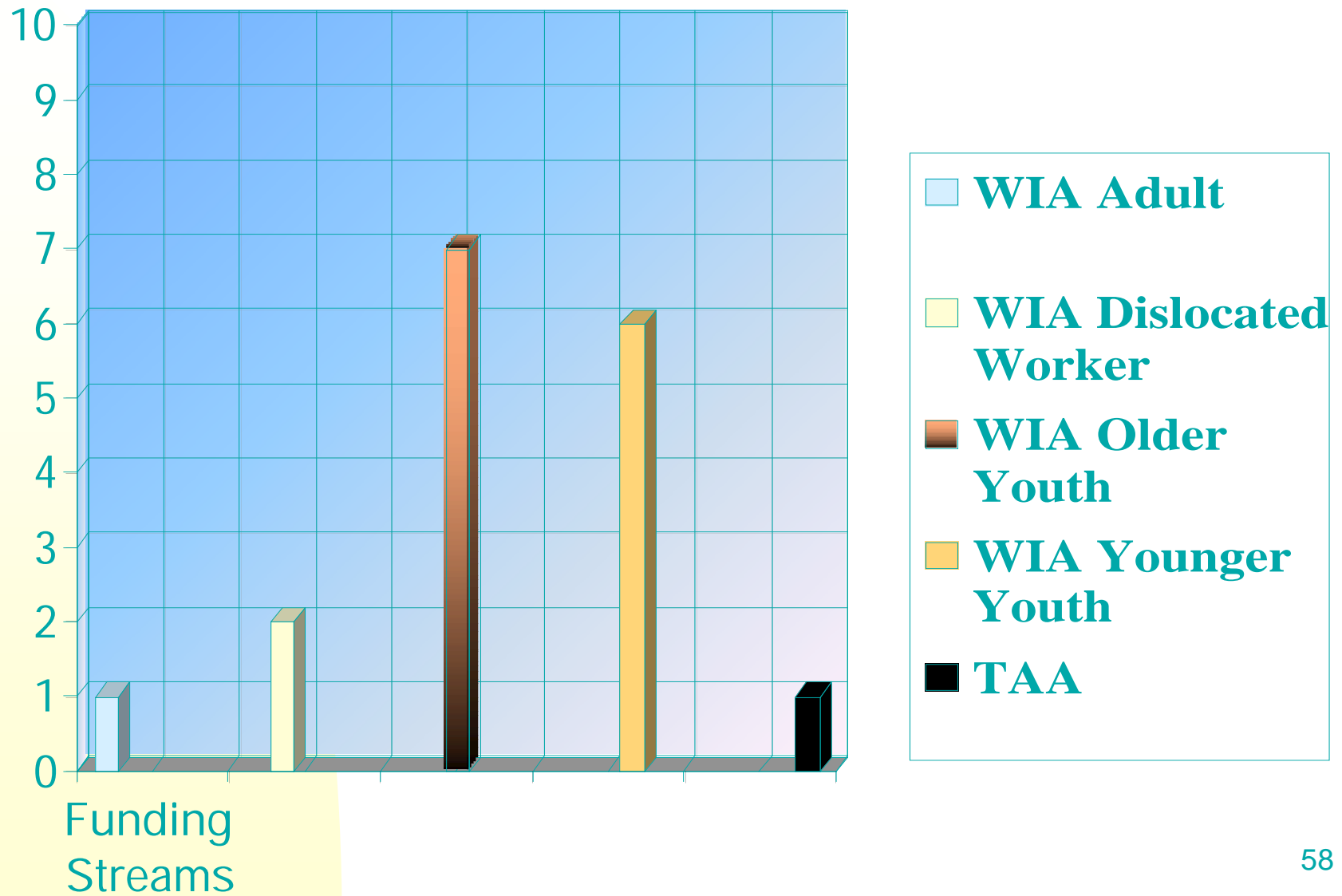


# Follow-up

Cont.

- **In many cases, there was no source information contained in the case records. Therefore this data element could not be verified.**
- **In other cases, the information contained in the case records did not match what was recorded in MOSES.**
- **Acceptable sources of documentation would include; activity sheets, attendance rosters, follow-up surveys, youth vendor reports, and case notes.**

# Snapshot of All Items



# Conclusions

- **The funding streams with the highest level of data elements that could not be validated were WIA Older Youth and WIA Younger Youth.**
- **The problems with the Older and Younger Youth data items not being able to be validated is attributable to how services are provided to Older and Younger Youth. Although the design of services for Older and Younger Youth varies throughout the Commonwealth, there are some commonalities.**
- **Intake and determination of eligibility is usually completed by the framework services provider and subsequent services are provided by youth vendors. The framework services providers maintain the eligibility files at their locations. All additional information may be kept in separate files at each of the youth vendors.**

# Conclusions

Cont.

- **In addition, there remains a problem in getting the proper level of information submitted from the vendors to the framework service providers and then data entered into MOSES, since with one exception, none of the youth service providers across the Commonwealth have access to MOSES.**
- **It is that information, which in most cases does not get submitted to the framework services providers and therefore does not get included in the hard copy files for the Older and Younger Youth customers.**

# Implications

- **The data validation review must be submitted 120 days subsequent to the submission of the validated annual report and the appropriate WIA Standardized Record Data (WIASRD). For PY2004, these reports are due by October 1 and October 15, 2005 respectively.**
- **This means the data validation is associated with “exitors” and “participants” whose activities that are reported upon occurred a minimum of four months prior.**
- **Once the data is reported upon, it is too late for the local areas to correct the information in MOSES and thereby avoid the data items failing data validation. Therefore, all information must be correct prior to it’s inclusion in the reports.**

# Implications

Cont.

- **The quality control by the local areas must be completed in real time and significantly earlier than the conducting of the data validation review.**
- **We are suggesting that local areas increase the level of quality assurance that currently occurs and develop and implement increased quality control over data entry that is incorporated into their Standard Operating Procedures (SOP).**

# Implications

Cont.

- It is anticipated that USDOL may attach some sort of sanctions to error rates in future years. All of this argues the need for increasing levels of quality control over data entry. DCS is suggesting that local areas increase the level of quality assurance that currently occurs within the local areas and develop and implement quality control over data entry that is incorporated into their Standard Operating Procedures (SOP).
- Based upon the potential of sanctions being imposed, and the higher level of data validation elements failure associated with WIA Older and Younger Youth, this is one more incentive to address the pervasive issues around the effective documentation for youth services.