WIDA CONSORTIUM

Delivering ACCESS for ELLs® Data for WIDA Research: Methodology and Artifacts

Rahul Joshi
Kristopher Stewart
Yajie Zhao
H. Gary Cook, Ph.D.

© 2010 Board of Regents of the University of Wisconsin System, on behalf of the WIDA Consortium

www.wida.us

WIDA CONSORTIUM

Section 1: Introduction

Introduction

WIDA CONSORTIUM

- *Who are we?*
  - Rahul Joshi, WIDA Data Warehouse Developer
  - Kris Stewart, WIDA Research Analyst
  - Yajie Zhao ("Grace"), Graduate Student
  - H. Gary Cook, WIDA Research Director

What is WIDA? - 1

WiDA CONSORTIUM

- WIDA stands for World-Class Instructional Design and Assessment
- Located in Madison, WI at the University of Wisconsin's Wisconsin Center for Education Research (WCER)
- Established in 2002 from a federal grant to create standards and assessment for ELL students

What is WIDA? - 2

WiDA CONSORTIUM

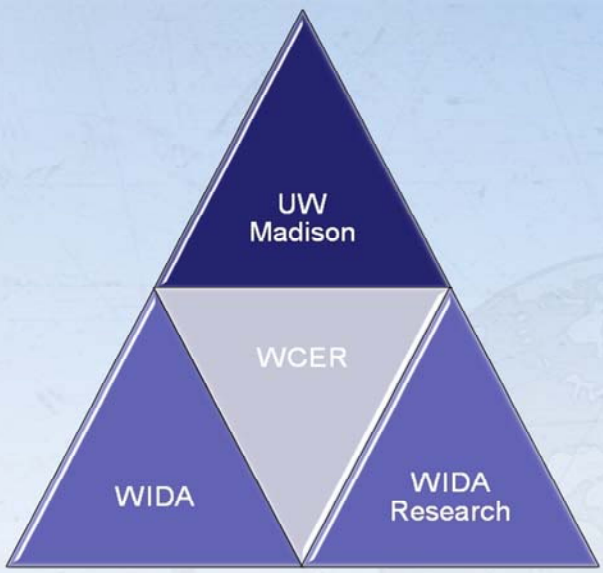
- A consortium that serves 23 states, their districts, and over 1 million ELL students
- Member states use a test for ELL students, namely Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs)
- ACCESS - Large scale test aligned to academic English language proficiency (ELP) standards

WIDA – ACCESS ELP Exam

WIDA CONSORTIUM

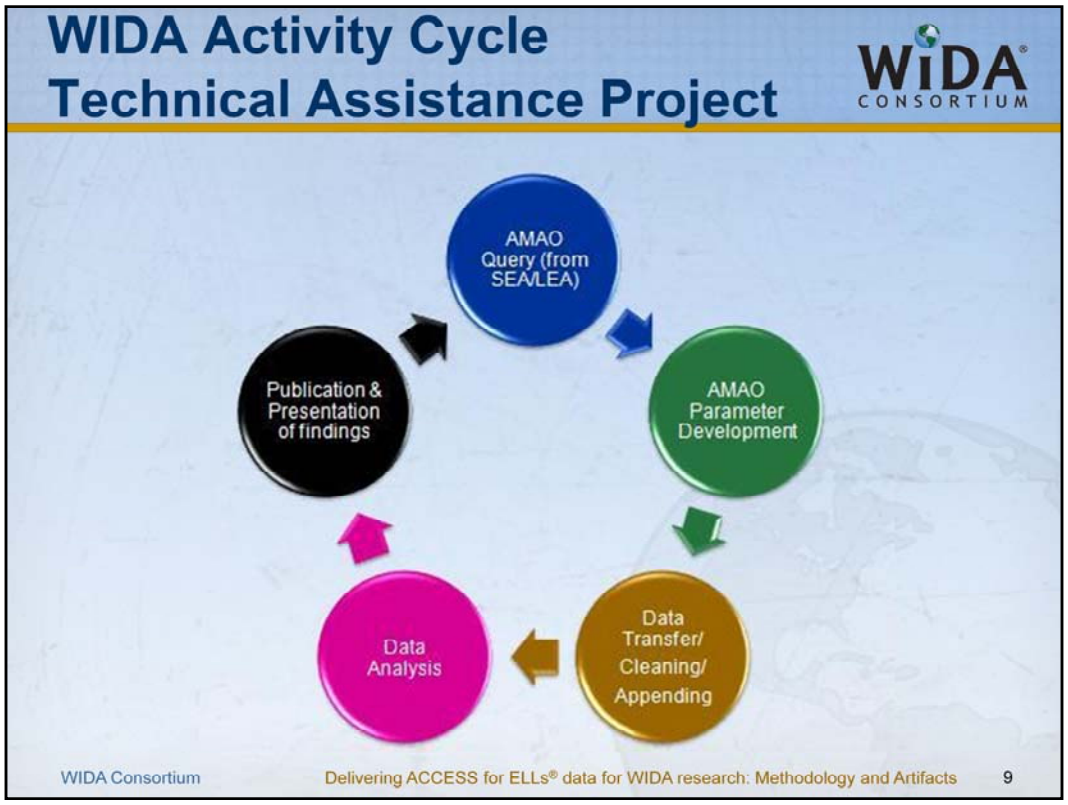
- Examines four language domains of ELP: Reading, Writing, Speaking, and Listening
- Uses five ELP standards
  - Social and instructional language
  - Language of language arts
  - Language of mathematics
  - Language of science
  - Language of social studies
- Provides the raw score, scale score, and proficiency level for a test taker in the language domains

WIDA Organizational Structure - 1



WIDA CONSORTIUM

Section 2:
WIDA Activity Cycle



This slide illustrates the process that WIDA Research undertakes for a typical technical assistance project request that concerns determining "Annual Measurable Achievement Objectives (AMAOs). The process begins when WIDA Research receives either a formal or informal communication from a state's education agency to assist them in an analysis of their WIDA ACCESS for ELL test data. Second, during a formal communication process, WIDA Research and the state education agency have a discussion regarding AMAO 1, which shows growth, AMAO2, which shows proficiency, and at times AMAO 3 which shows adequate yearly progress. The third phase of the activity cycle involves WIDA's data warehouse manager working with the state's technical personnel (and our test vendor) in ensuring that the data is "clean" and useful for reliable analysis. From there, WIDA Research conducts a statistical analysis where the data will illustrate the state's growth, etc. on the ACCESS Test for ELL students. Finally, after the analysis has been done, a report is written and the findings are presented to the state's stakeholders.

WIDA CONSORTIUM

Section 3: Historical Background

WIDA Data Warehouse - History and need

- No data warehouse 'yesterday' but only disconnected datasets
- Quick expansion of WIDA Consortium & commensurate data explosion
- Need to consolidate ACCESS data to be able to extract meaningful information
- Need to effectively manage data
- Need to connect ACCESS data to national educational data collections for multi-faceted research

WiDA CONSORTIUM

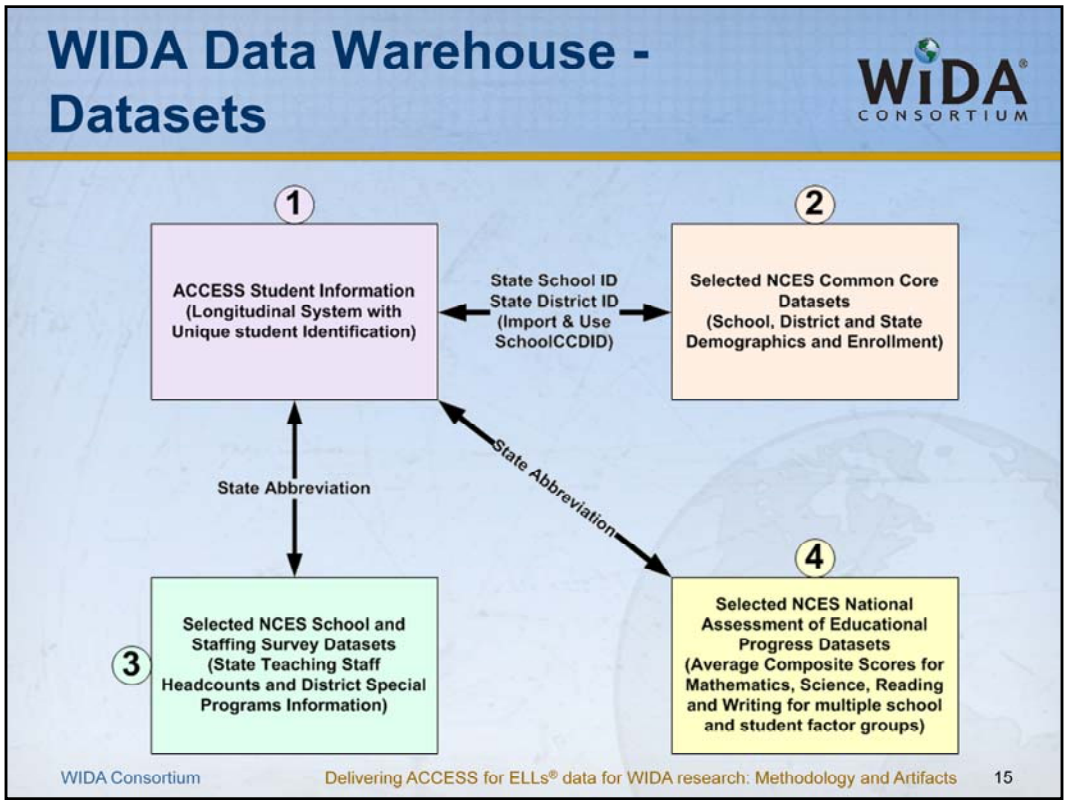
Section 4: Data Warehouse Framework - Tools, Processes and Methodology

Salient Features of WIDA Data Warehouse

- High-performance, scalable SQL Server database design
- Over a million individual ACCESS test takers from 20 states across US
- ACCESS Test Information(test scores, restricted student identification data and demographics)
- Connected to selected NCES Research Data Collections
- Core database for WIDA projects and research initiatives

Data Delivery – Secure and On Demand

- Statistical packages on secure, doubly encrypted Citrix server
- SAS used heavily with SAS/SQL Server connection, plus a few specialized tools
- On-Demand, remote access to data with authentication handshake
- Analysis results are the only static artifacts stored, no source or intermediate data storage
- Data Warehouse is the 'Single Version of Truth'
- On-demand longitudinal dataset generation



This slide indicates the main data components of the WIDA Data Warehouse.

The primary data source is the ACCESS for ELLs student performance data (indicated as 1) with restricted student identification information, for all the WIDA Consortium member states. Every student in this dataset has a unique student identifier and is longitudinally tracked across successive testing cycles.

The data warehouse also houses selected data elements from some of the publicly available research datasets collected by NCES namely: Common Core Datasets (indicated as 2), School and Staffing Survey Datasets (indicated as 3) and National Assessment of Educational Progress Datasets (indicated as 4). The Common Core Datasets are connected with the ACCESS datasets at both the school and district level, using unique State School ID and State District ID's. In this case, nationwide uniquely assigned School CCD Identifier is stored and also used for connecting these datasets.

Since School and Staffing Survey and National Assessment of Educational Progress Datasets are only available at the state level and hence connected accordingly. State level Common Core Data aggregates also form a part of the Common Core datasets.

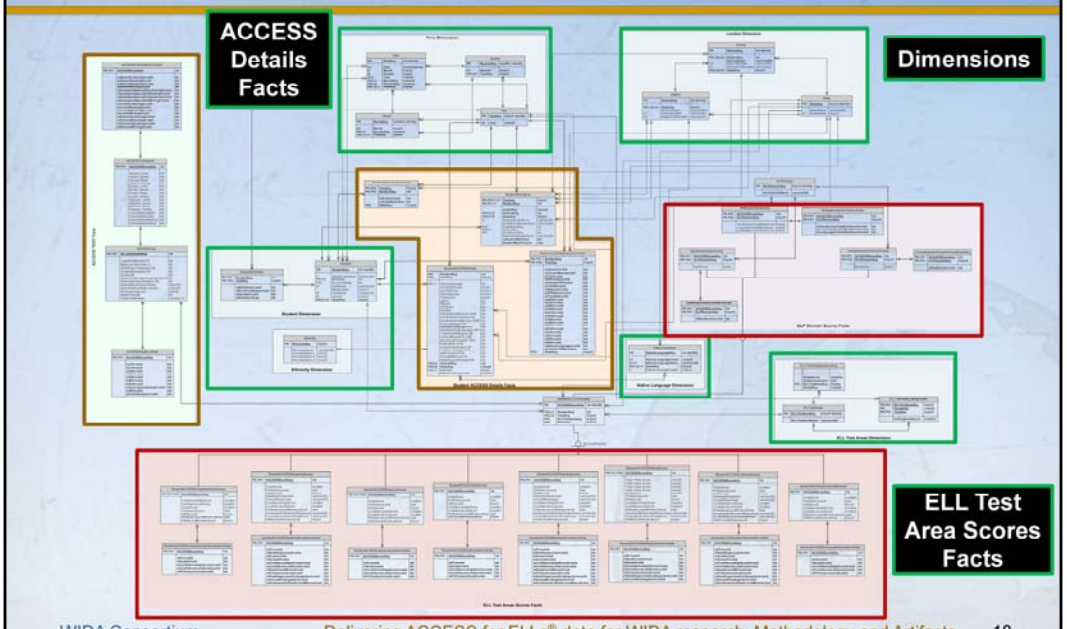
WIDA Data Warehouse - Design Overview

- Important Dimensions - 1 2 3 4
  - Location → State, District, School
  - Time → Date, Month, Quarter, Year
  - Student → Student, Ethnicity, Native Language
  - ACCESS Test → ELPDomain, ELLTestArea
- Important Facts (ACCESS Student Information)
  - Student Attendance 1
  - Student ACCESS Scores (Both for ELPDomain and ELLTestArea dimensions) 1
  - Student ACCESS Test Details 1
  - Data validation information for all above facts 1

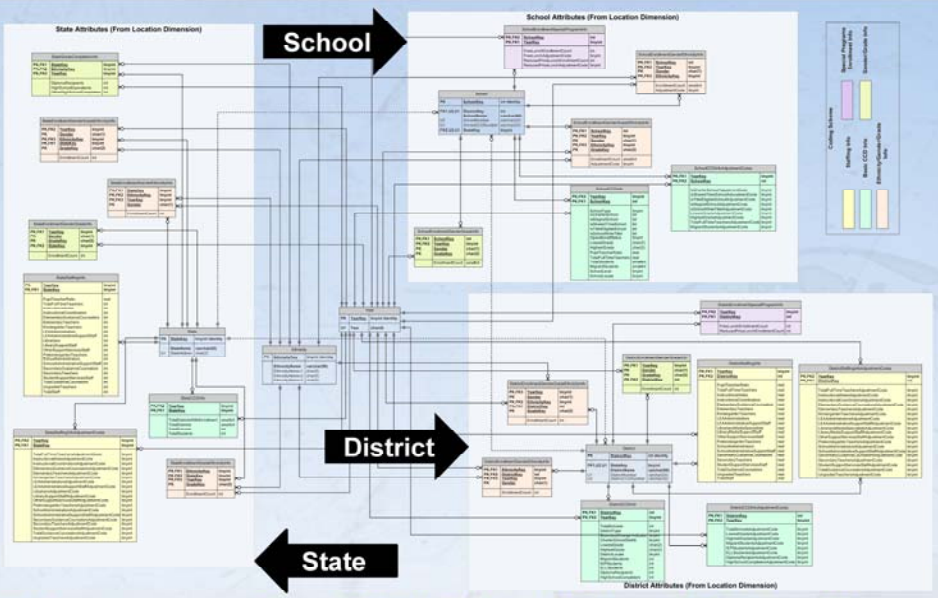
WIDA Data Warehouse - Overview(Contd.)

- Important Facts (NCES Research Datasets)
  - School Enrollment by Grade, Gender and Ethnicity; School Staffing Information (Atomic, aggregates at District and State Levels) 2 3
  - State District Level Special Programs Information 3
  - Average Composite Scale Scores for NAEP Subject Areas (Mathematics, Science, Reading and Writing) for selected School and Student Factor Groups 4
  - Suitable data validation information for above facts 2 3 4
- ACCESS datasets and NCES datasets are available on yearly basis. NCES has specific data collection cycles.

Design - ACCESS Student Information 1

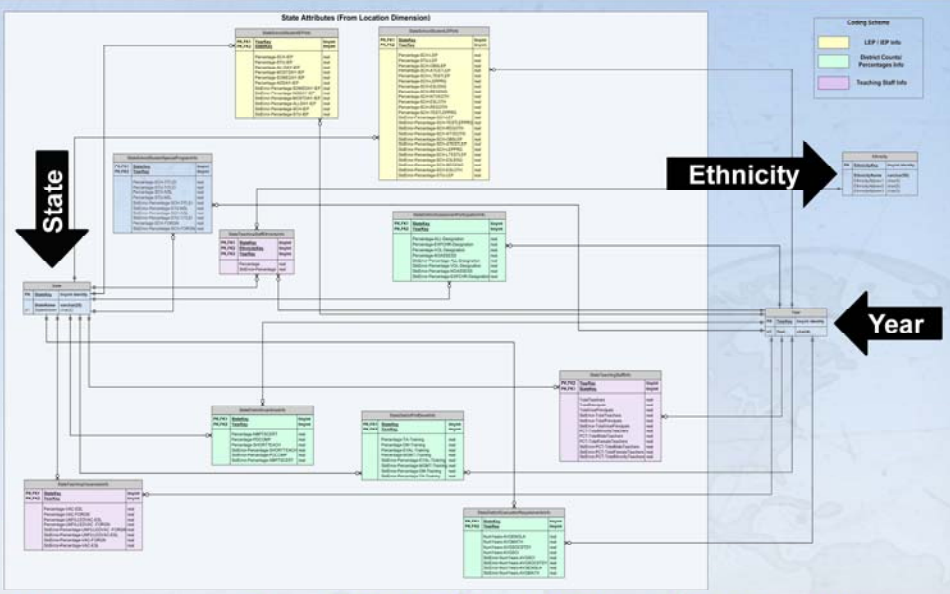


Design - NCES Common Core Datasets (2)

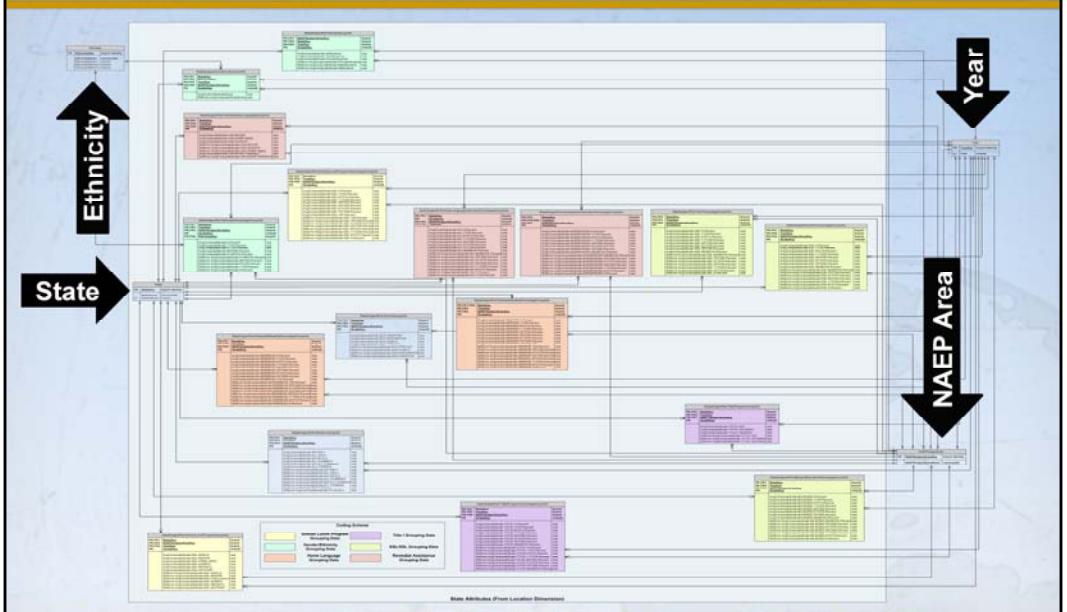


Design - NCES School and Staffing Survey Datasets (3)

WIDA CONSORTIUM



Design - NCES NAEP Datasets 4



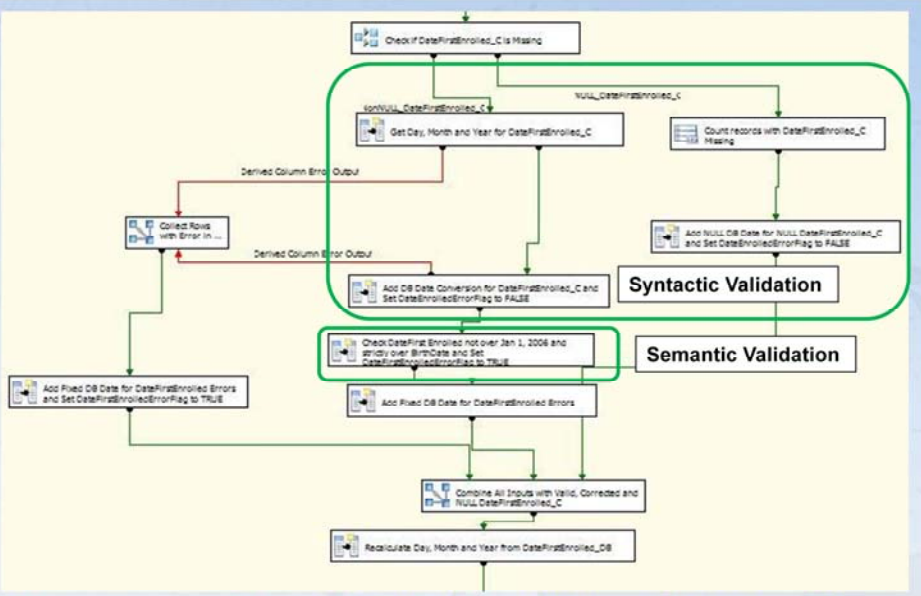
Data Validation – Gauging Data Quality

- Yearly ACCESS raw datasets 150-column wide, with both syntactic and semantic (domain-specific) meanings
- Thorough data validation through certain direct and indirect rules on single/combination of data fields
- Identification of the outliers and reporting relevant counts
- Automated ETL packages developed with SSIS using Visual Studio 2008

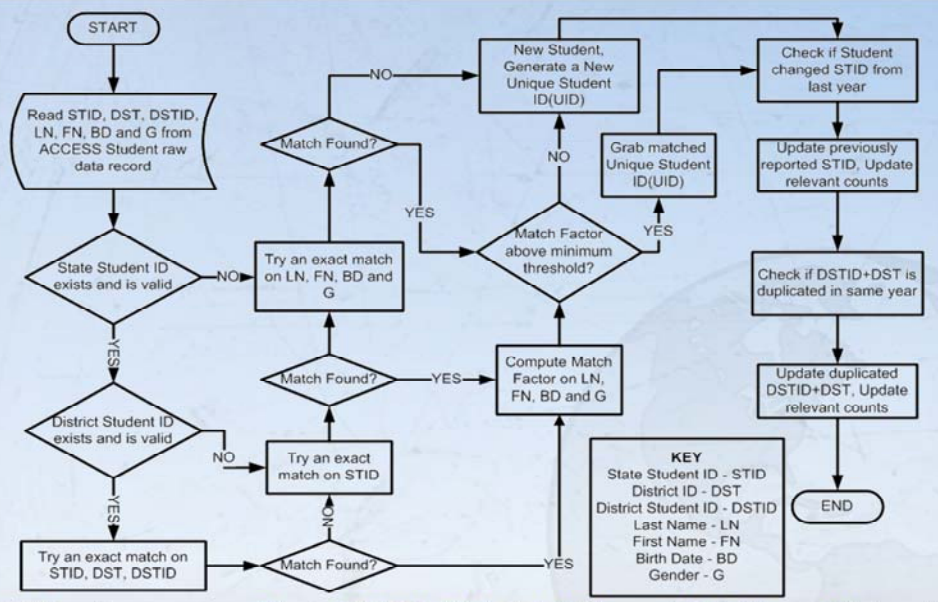
Example – Validation logic for Date of enrollment in ACCESS and Birth Dates

Sample Data Validation Workflow

WIDA CONSORTIUM



Building Longitudinal Student Record System



WIDA CONSORTIUM

Section 5: Uses of Data Warehouse Framework

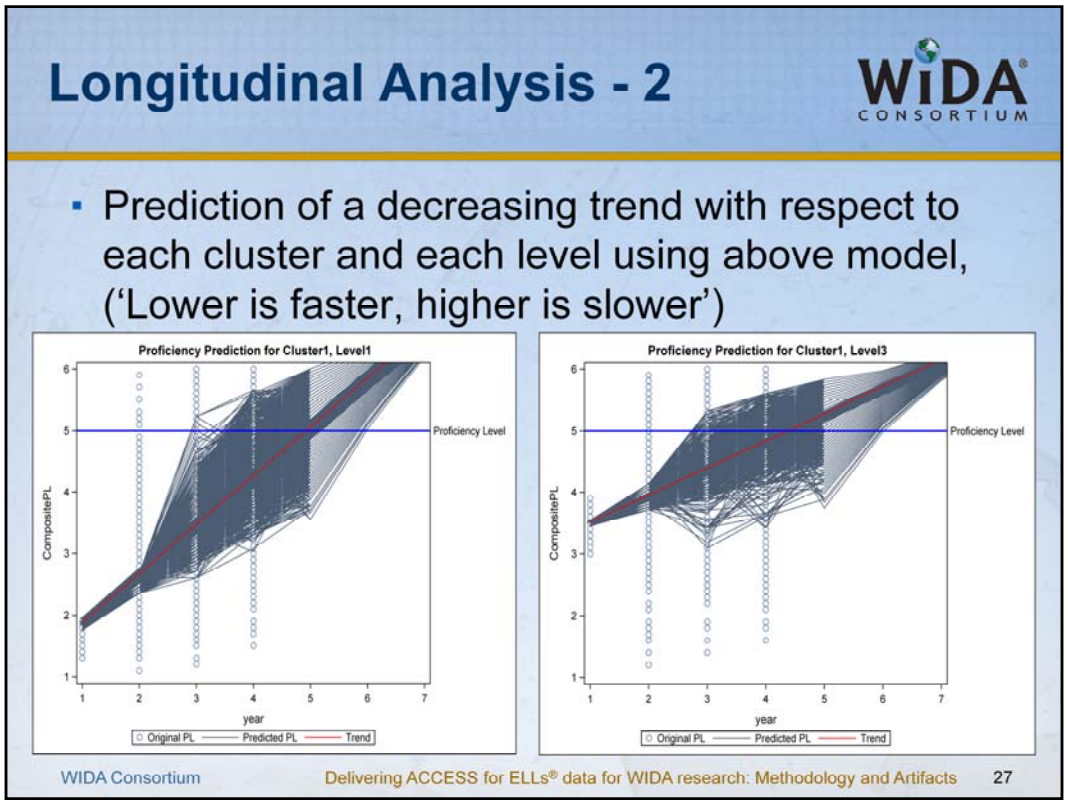
Longitudinal Analysis - 1

WIDA CONSORTIUM

- Research Problem - How many years will it take for students to achieve a proficiency level 5.0 once they enter the program?
- Statistical analysis using a first-order autoregressive model for each starting WIDA level in each cluster

Grade	0	1-2	3-5	6-8	9-12
Cluster	0	1	3	6	9

Proficiency Level	< 2.0	2.0 – 2.9	3.0 – 3.9	4.0 – 4.9	5.0 – 5.9
WIDA Level	1	2	3	4	5



This is a statistical model known as a 'predictive auto-regressive model' which examines how many years it will take an ELL student to achieve a proficiency level of 5.0 once they have entered an ELL/LEP program. The X axis shows the year a student enters, while the Y axis shows the composite proficiency level. This model looks at a student's starting proficiency level (denoted by the small circle), and then the number of years he/she will achieve a 5.0 composite proficiency level. As one examines the predicted proficiency level (denoted by the straight line), this shows where the student will be the following year. Accordingly, this graph is an illustration of the principle: *lower is faster, higher is slower.* That is, ELL students in lower grades (younger students) and at lower proficiency levels acquire language at faster rates.

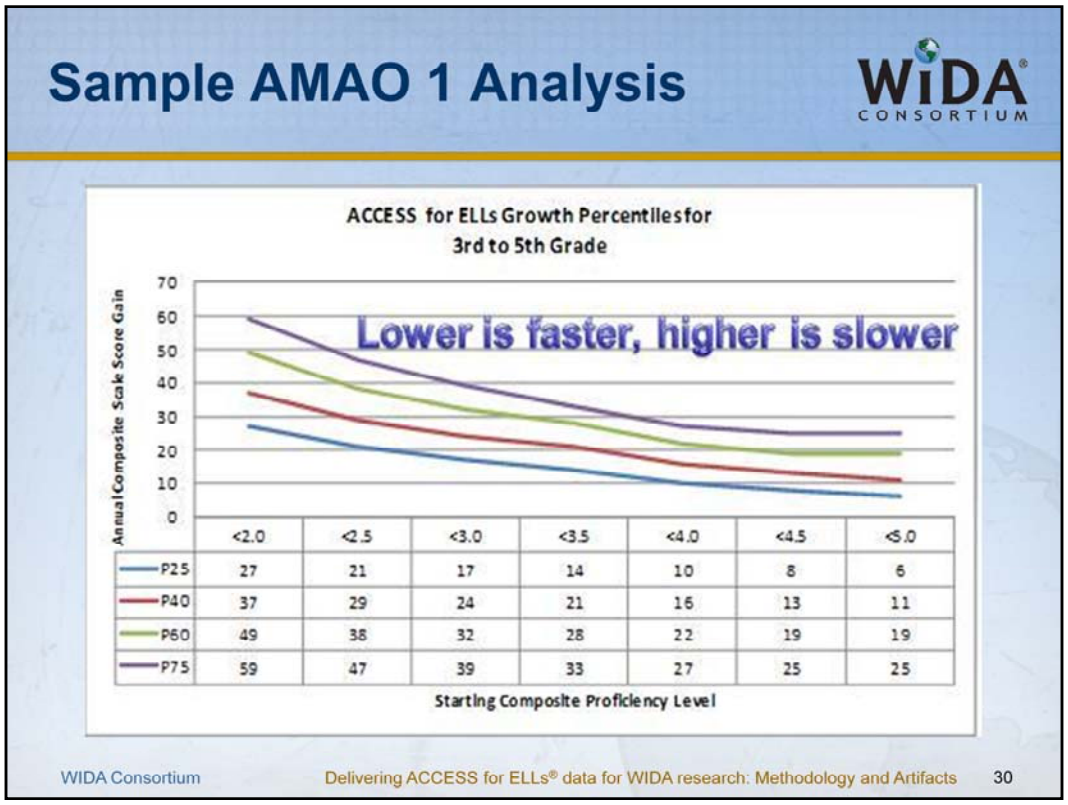
Technical Assistance Projects and Policy Guidance - 1

- SEA/LEAs seek guidance regarding ELL students and assessment accountability:
  - AMAO 1 (Progress)
  - AMAO 2 (Attainment)
  - AMAO 3 (Adequate Yearly Progress - AYP)
- Uses the data warehouse to:
  - Conduct analysis of ELL student progression
  - Determine district performance
  - Comparative analysis among Consortium members

Technical Assistance Projects and Policy Guidance - 2

WIDA CONSORTIUM

- Provide policy guidance on AMAO 1:
  1. Determine the scoring metric used to measure growth
  2. Determine the annual growth target
  3. Set the starting point for AMAO 1 targets
  4. Set the ending point for AMAO 1 targets
  5. Determine the annual rate of growth
- Meet with the state stakeholders to discuss findings
- State stakeholders make *recommendations* to SEA/LEA



This slide is a graphical example related to AMAOs showing growth which accompanies a technical assistance report to a Consortium member. This chart represents four important aspects: grade level cluster, percentile rank, proficiency level, and scale score gain. Cluster being examined includes the third, fourth, and fifth grade. A student's *starting* composite proficiency level (<2.0 - <5.0) accompanied by the percentile rank on the x axis shows the ELL student's growth. For example, when you subtract last year's scale score from the present year, that will give a student growth, let's say for a scale score gain of 32. Then, based on the fourth grade student's composite proficiency level of 3.3, you can see that student's is at the 60th percentile based on a scale score gain of 32 and starting proficiency level of 3.3 within the 3-5 grade cluster.

Technical Assistance Projects and Policy Guidance - 3

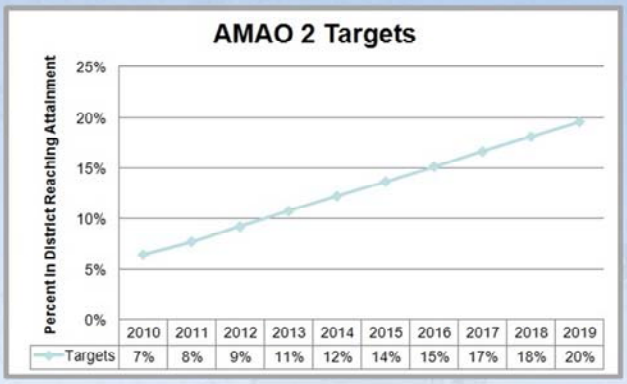
WIDA CONSORTIUM

- Provide policy guidance on AMAO 2:
  1. Define the English proficiency level
  2. Determine the cohort of ELLs for analysis
  3. Set the starting point for AMAO 2 targets
  4. Set the ending point for AMAO 2 targets
  5. Determine the annual rate of growth
- Meet with the state stakeholders to discuss findings
- State stakeholders make *recommendations* to SEA/LEA

Sample AMAO 2 Analysis

WIDA CONSORTIUM

4.8 CPL Gain	
AMAO 2 Targets	
Year	Targets
2010	7%
2011	8%
2012	9%
2013	11%
2014	12%
2015	14%
2016	15%
2017	17%
2018	18%
2019	20%



Average AMAO 2 Percentages (All districts)

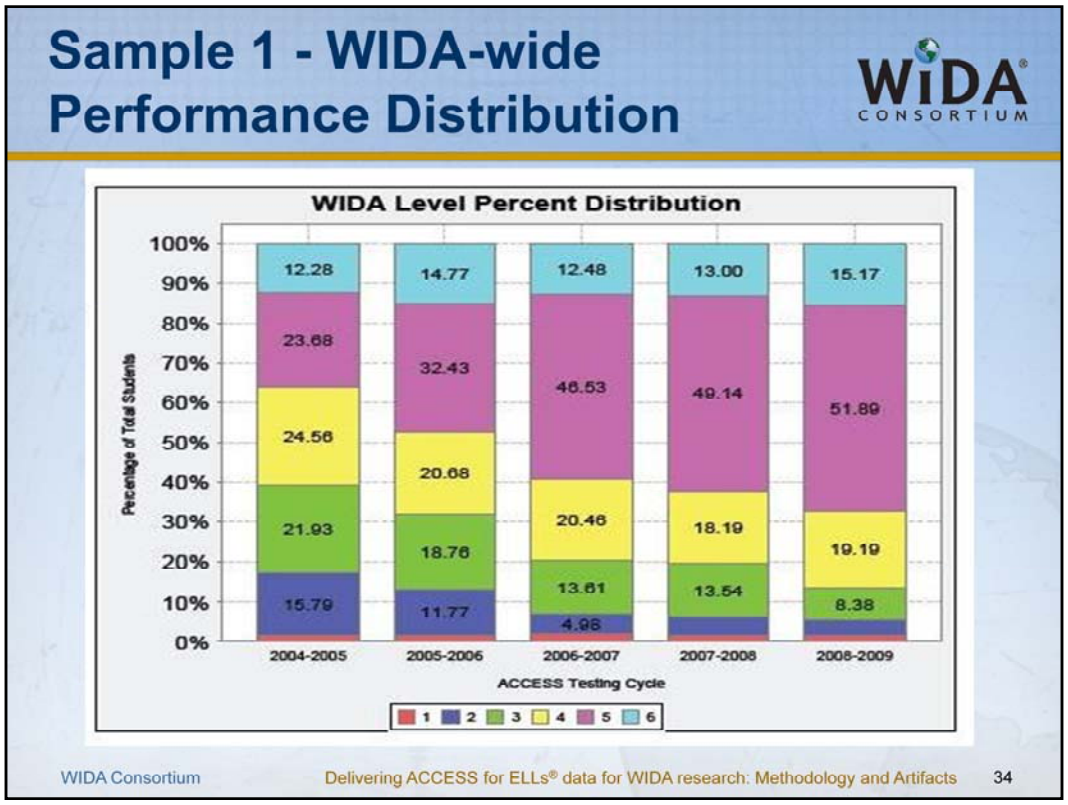
year	P10	P15	P20	P25	P50	P60	P75	P80	P90
2008	4.0%	4.9%	6.0%	6.7%	10.6%	13.4%	18.2%	21.5%	27.0%
2009	4.9%	6.1%	6.8%	8.2%	14.3%	16.7%	20.9%	23.8%	31.9%
Average	4.4%	5.5%	6.4%	7.4%	12.5%	15.0%	19.6%	22.6%	29.5%

WIDA Consortium Delivering ACCESS for ELLs® data for WIDA research: Methodology and Artifacts 32

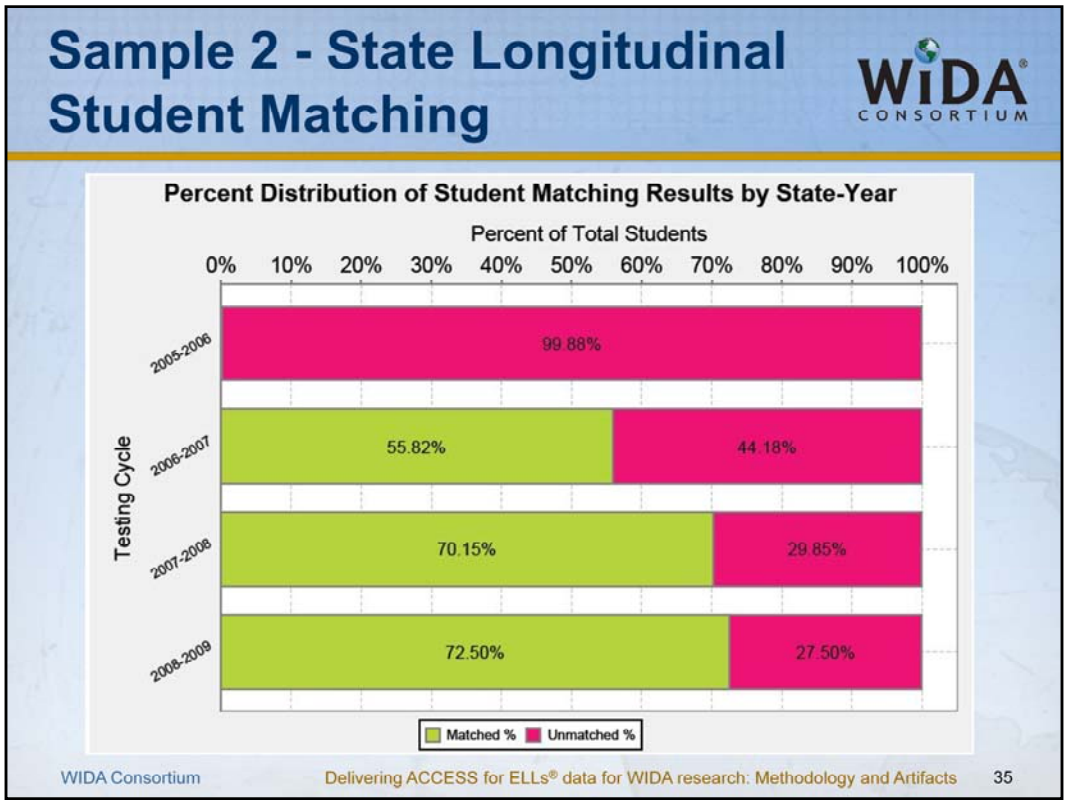
These tables and accompanying chart are also graphical examples related to AMAOs. The table highlighting the year and averages, shows the percentile rank of districts for AMAO 2: Attainment. Once a state education agency has reviewed this initial table showing the distribution of percentiles, they are then tasked with choosing a beginning percentage (target) and ending percentage (target) to show that their districts will achieve that percentage of ELL students reaching attainment. Finally, the chart labeled, “AMAO 2 Targets” simply illustrates the percentage targets by year that a district must show attainment.

Reporting Framework for WIDA Members

- Aggregated statewide and WIDA-wide performance reports for WIDA Member state stakeholders
- Ongoing reporting framework development for more insightful reporting on ACCESS and NCES data repository
- Pilot development in Pentaho, an open source BI Tool, using Mondrian Analysis Engine
- Subsequent migration to SSRS using Visual Studio 2008 and Microsoft Web Server deployment



This is a graphical representation of distribution of students across all WIDA Consortium member states taking ACCESS for ELLs test, based on their composite domain WIDA levels in each of the testing cycle. ACCESS composite domain proficiency levels are directly related to WIDA levels, as defined in slide 26.



This is a graphical representation of results of longitudinal matching among ACCESS for ELLs student population from a characteristic WIDA member state. In each of the testing cycle, the percentage of students from the state who could be matched with an existing student record and the percentage of students who could not be matched are represented as a stacked bar graph.

Sample 3 – State Minimum Proficiency Gain Analysis

WIDA CONSORTIUM

Starting PL Band	No. of Students(State)	No. of Students(WIDA)	Percentage (against yearly total) for State	Percentage (against yearly total) for WIDA	
< 2	533	940	1.89%	2.01%	↓
< 2.5	559	814	1.98%	1.74%	↑
< 3	1,151	1,403	4.08%	3.01%	↑
< 3.5	1,006	1,271	3.57%	2.72%	↑
< 4	1,345	1,987	4.77%	4.26%	↑
< 4.5	612	1,918	2.17%	4.11%	↓
< 5	395	1,415	1.40%	3.03%	↓
< 5.5	16	65	0.06%	0.14%	↓
Total Students with valid 2006-2007 To 2007-2008 scores	28210	46665			

WIDA Consortium Delivering ACCESS for ELLs® data for WIDA research: Methodology and Artifacts 36

This is a tabular report showing a performance comparison of a characteristic WIDA Consortium member state ACCESS for ELLs student population with the entire WIDA Consortium student population. The table indicates number and percentages of students across the state and the entire WIDA, who had at least one point or more composite proficiency level gain in a typical testing cycle, from their proficiency levels measured in the previous testing cycle. The students are grouped by their starting composite proficiency level bands. The green upward arrow indicates that the state percentage is higher than that of the entire WIDA population for a corresponding starting PL band population (State performance better than average WIDA population). The red downward arrow, on the other hand, indicates that the state percentage is lower than (or equal to) that of the entire WIDA population.

WiDA CONSORTIUM

Section 6: What's Next ?

More Data, More Fun….

WIDA CONSORTIUM

- On-line, secure Business Intelligence (BI) reporting framework for WIDA State and District members
  - Development of new metrics for data insights
  - Drilling data with OLAP and MOLAP
  - Predictive analysis using data mining techniques
  - Comprehensive and searchable documentation
  - Data Literacy through guided data visibility experience
- Inclusion of other publicly available data collections to support new research areas
- Capacity expansion to support ELL research external to WIDA

WiDA CONSORTIUM

Thank You !

Questions / Comments ??