

The Intersection of Instruction, Assessment & Accountability

David Abrams

**Assistant Commissioner for Standards,
Assessment, and Reporting**

Bilingual/ESL Teachers' Institute

Long Island, NY

March 13, 2008



University of the
State of New York
State Education
Department

3 Cognitive Models in Educational Measurement: Leighton & Gierl 2007

Model 1: The Cognitive Model of Test Specifications

- A model of test specifications is generated using a two-way matrix to establish content and skill groupings for obtaining a *representative sample* of items during test construction from a defined achievement domain.
 - Large-scale standardized tests are often representative of the cognitive model of test specifications.
-

3 Cognitive Models in Educational Measurement: Leighton & Gierl 2007

- Benefit of the cognitive model of test specifications is its simplicity because committees of experts can generate and approve it.
 - Test specifications function as a cognitive model because they not only reflect an explicit demarcation of the domain of achievement, but also the knowledge and skills examinees are *expected to use to answer test items correctly*.
-

3 Cognitive Models in Educational Measurement: Leighton & Gierl 2007

- Large-scale tests, developed from models of test specifications, are the instruments of choice to measure critical but generalized knowledge and skills. Because the model represents a selection of priority content, diagnostic claims about examinees' strengths and weaknesses, either behavioral or cognitive, are not usually compelling.
-

3 Cognitive Models in Educational Measurement: Leighton & Gierl 2007

Model 2: The Cognitive Model of Domain Mastery

- Domain mastery is generated to illustrate the population of knowledge and skills that is believed to conceptualize expertise or mastery within a circumscribed achievement domain.
 - Curriculum-based tests or embedded assessments are representative of the model.
-

3 Cognitive Models in Educational Measurement: Leighton & Gierl 2007

- Curriculum-based tests are used for the purpose of *formative* evaluation and can be described as an “academic thermometer” to assess student progress on a broad set of knowledge and skill component.
 - As a formative evaluation tool, curriculum-based tests are standardized so that changes in test scores can be ascribed to changes in student progress rather than idiosyncrasies with the test itself.
-

3 Cognitive Models in Educational Measurement: Leighton & Gierl 2007

- A benefit of this model is that it provides detailed information about the knowledge and skills required to achieve mastery of curricula.
 - Tests are designed to prove multiple knowledge and skills thoroughly to ensure breadth and depth of mastery at a behavioral level.
-

3 Cognitive Models in Educational Measurement: Leighton & Gierl 2007

Model 3: The Cognitive Model of Task Performance

- This model is generate to illustrate the thinking processes underlying the knowledge and skills students apply *in vivo* when solving educational tasks in a specific domain.
 - This model is the type that educational or cognitive researchers develop to empirically confirm the thinking processes individuals use to answer or solve classes of tests items.
-

3 Cognitive Models in Educational Measurement: Leighton & Gierl 2007

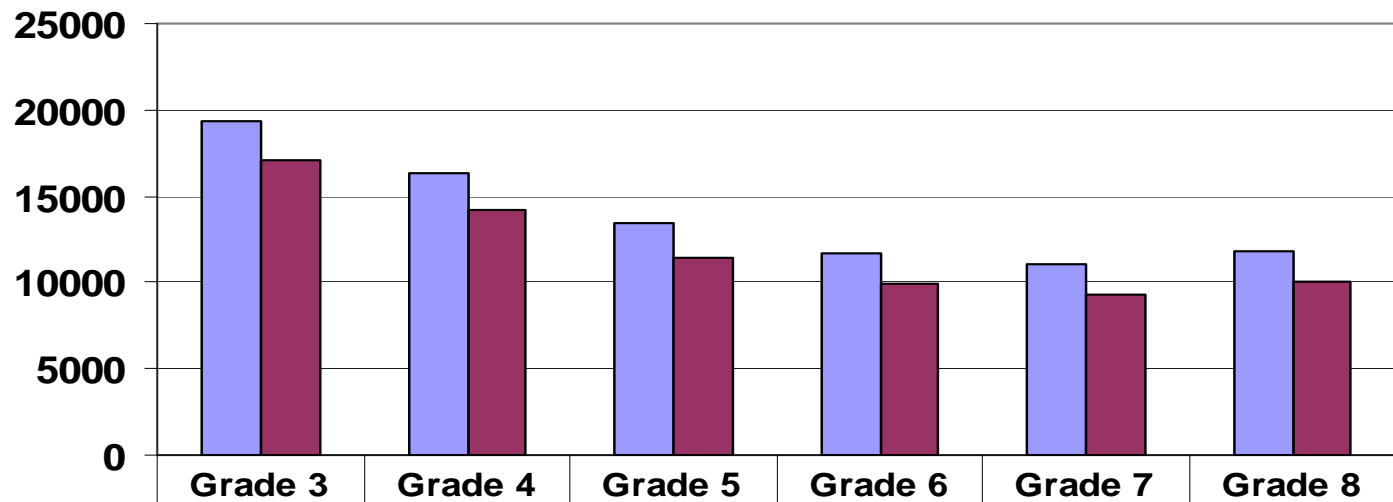
- Tests in this model are relatively uncommon in educational measurement.
 - These are relatively new and include:
 - Embretson's Cognitive Design Systems (1998, 2002, 2005)
 - Mislevy et al. Evidence Centered Design (2003)
 - Tatsuoka's Rule Space Model (2004).
-

ELA 3-8/NYSESLAT 2007

- ❑ In January 2007, all English language learners (ELLs) in New York State public schools, except for those who enrolled in U.S. schools for less than one full year, were administered the Grade 3-8 ELA tests for the first time.
 - ❑ Of the 83,643 Grades 3-8 ELLs enrolled in NYS public schools, 72,089 (86.2%) took the State Grade 3-8 ELA tests in 2007.
 - ❑ The analysis in this presentation are based on the 72,089 ELLs who took the State Grade 3-8 ELA tests in January 2007 and the NYSESLAT in May 2007.
 - ❑ Statewide, 12,939 (18%) ELLs scored proficient on the Grades 3-8 ELA tests, 38,012 (53%) scored at Level 2, and 21,131 (29%) scored at Level 1.
 - ❑ ELLs who scored proficient on the NYSESLAT performed much better on the State Grade 3-8 ELA tests: 52 percent scored at Level 3 and above and 99% scored at Level 2 and above.
-

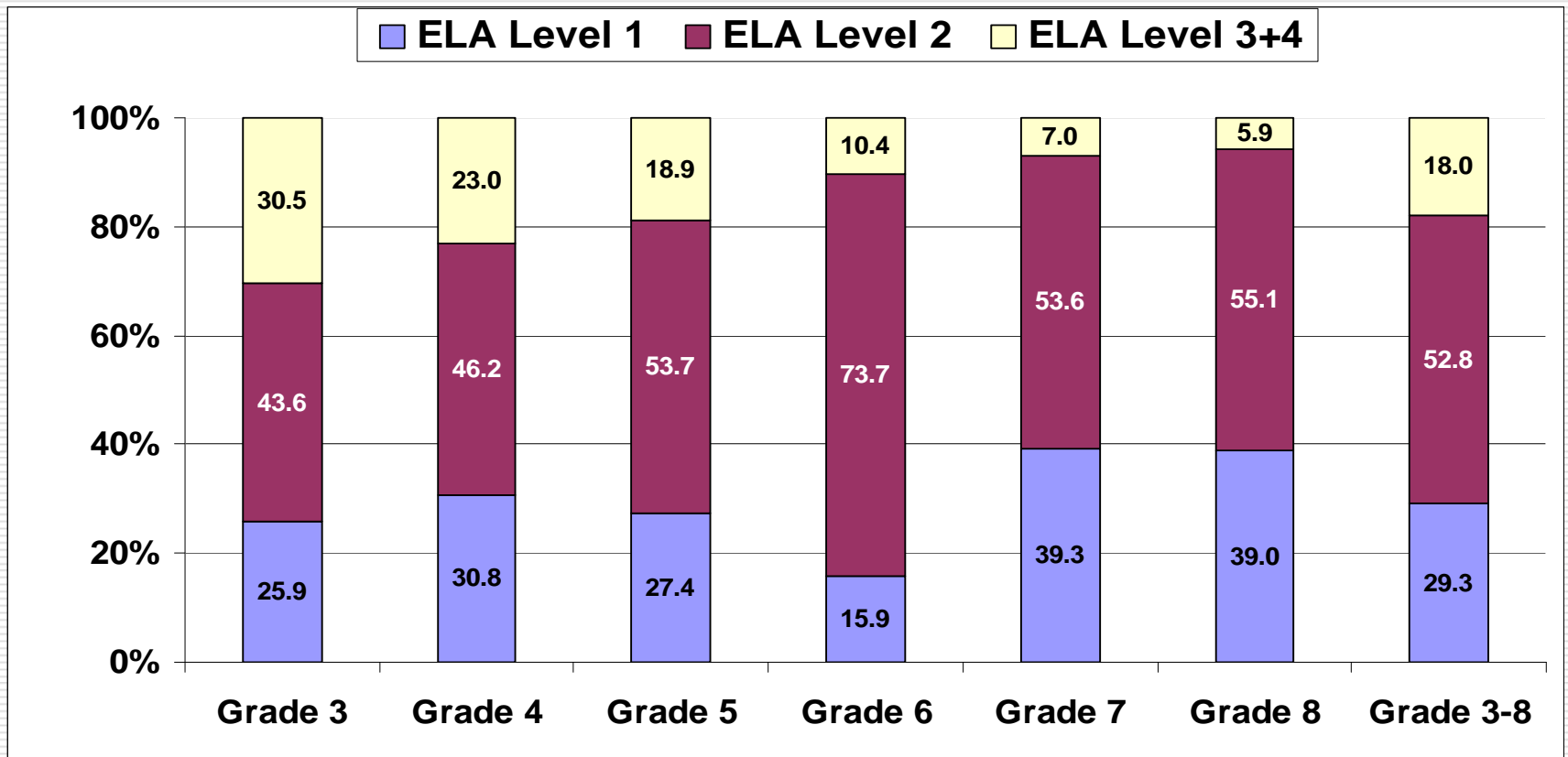
Number of ELL Students Tested On Grade 3-8 ELA Tests

Total ELLs = 83,643; Total ELLs Tested = 72,082

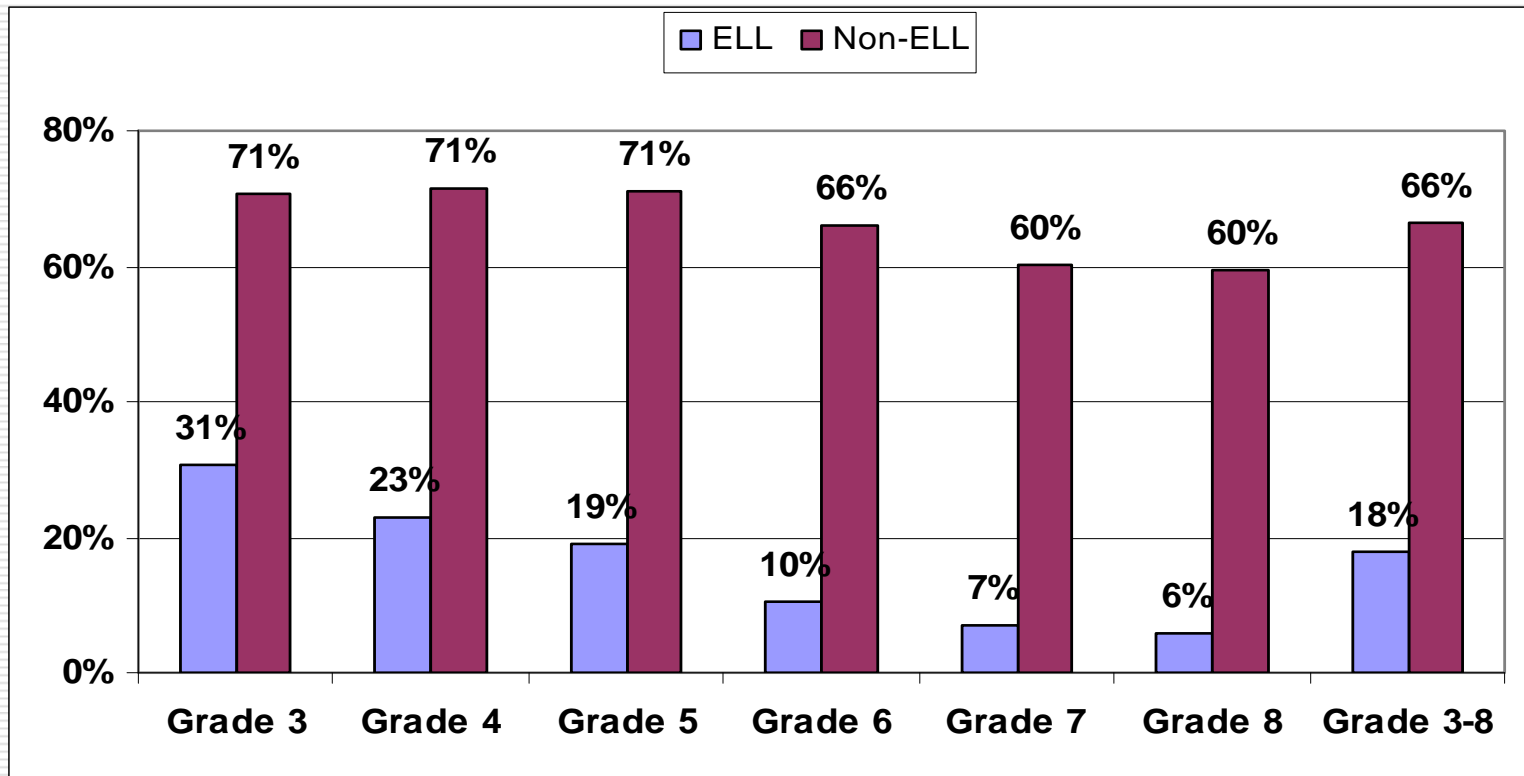


■ Total ELLs in NYS	19376	16276	13388	11695	11095	11813
■ Total ELLs Tested	17093	14200	11480	9934	9299	10076

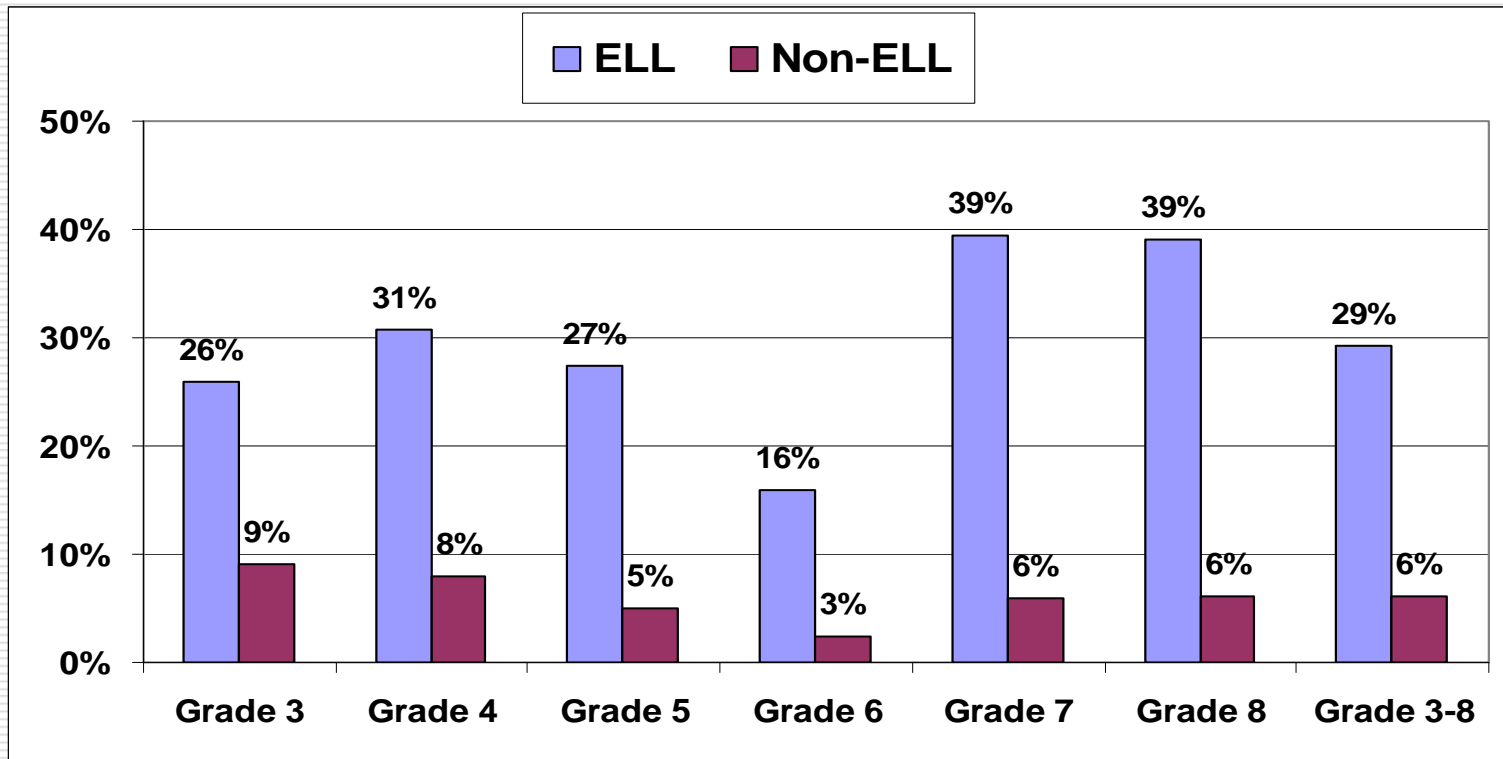
Performance of ELL Students On 2007 Grade 3-8 ELA Tests



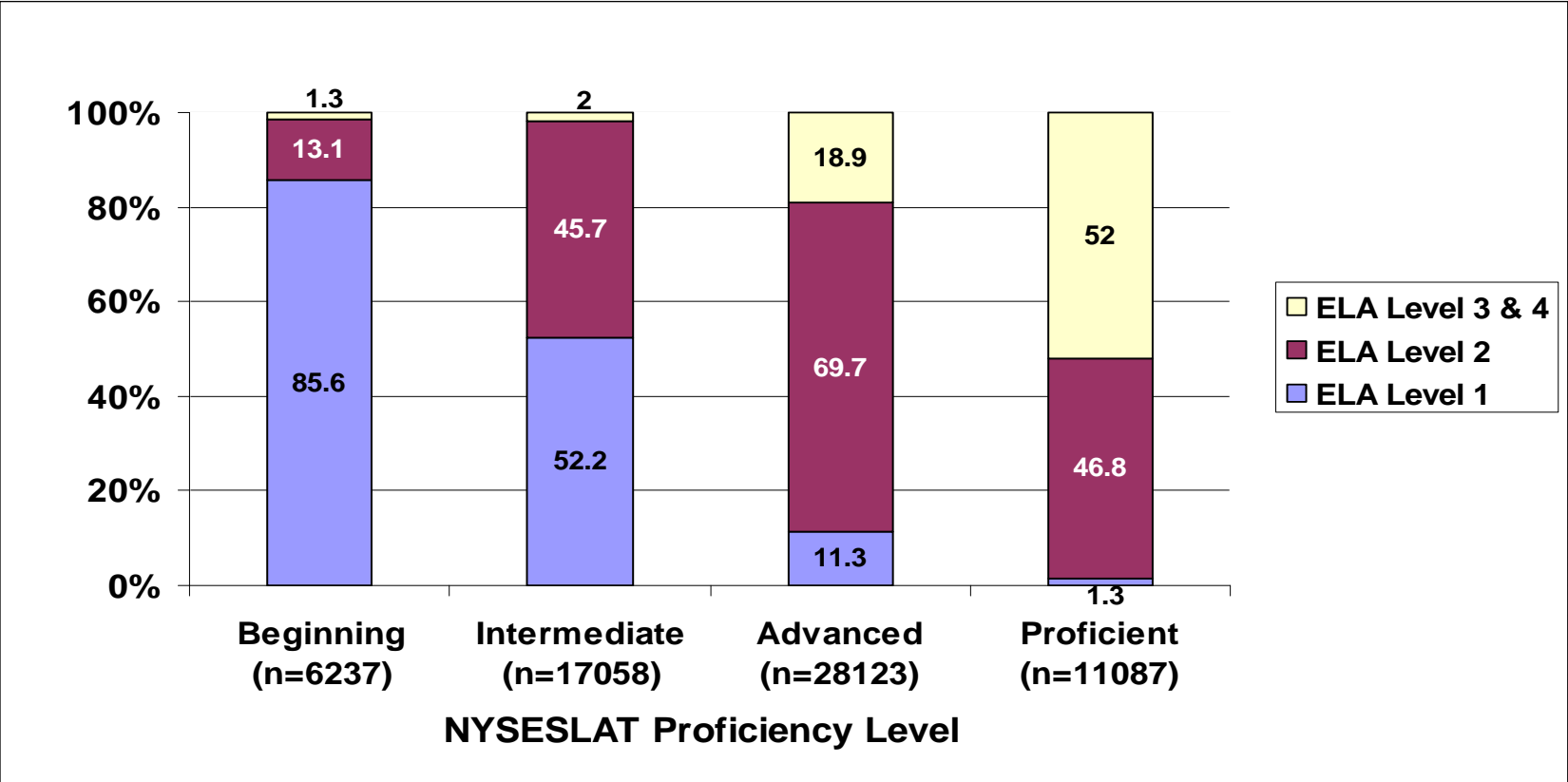
Percentage of All ELLs and Non-ELLs Scoring at Levels 3 and 4



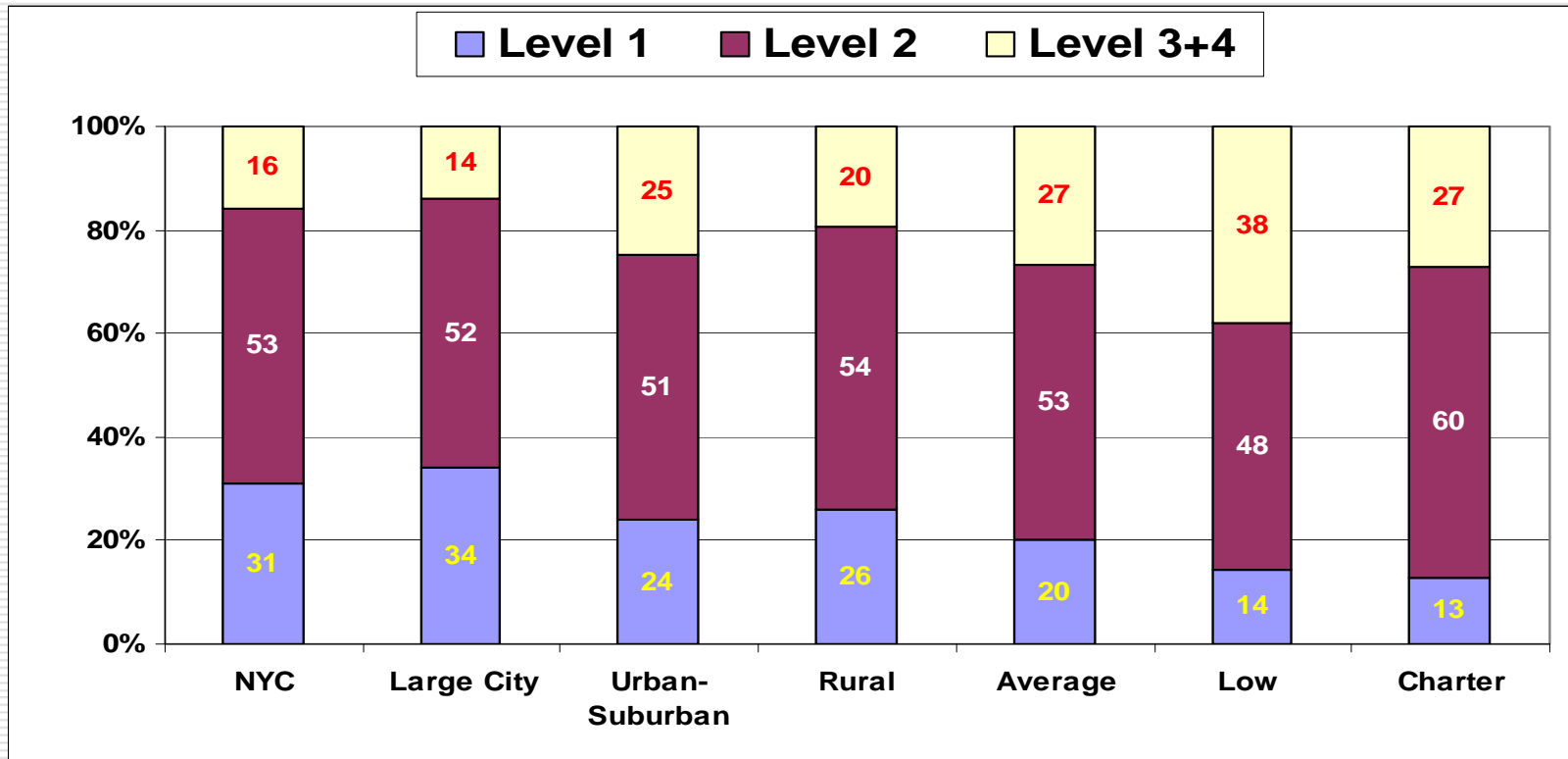
Percentage of ELLs and Non-ELLs Scoring at Levels 1



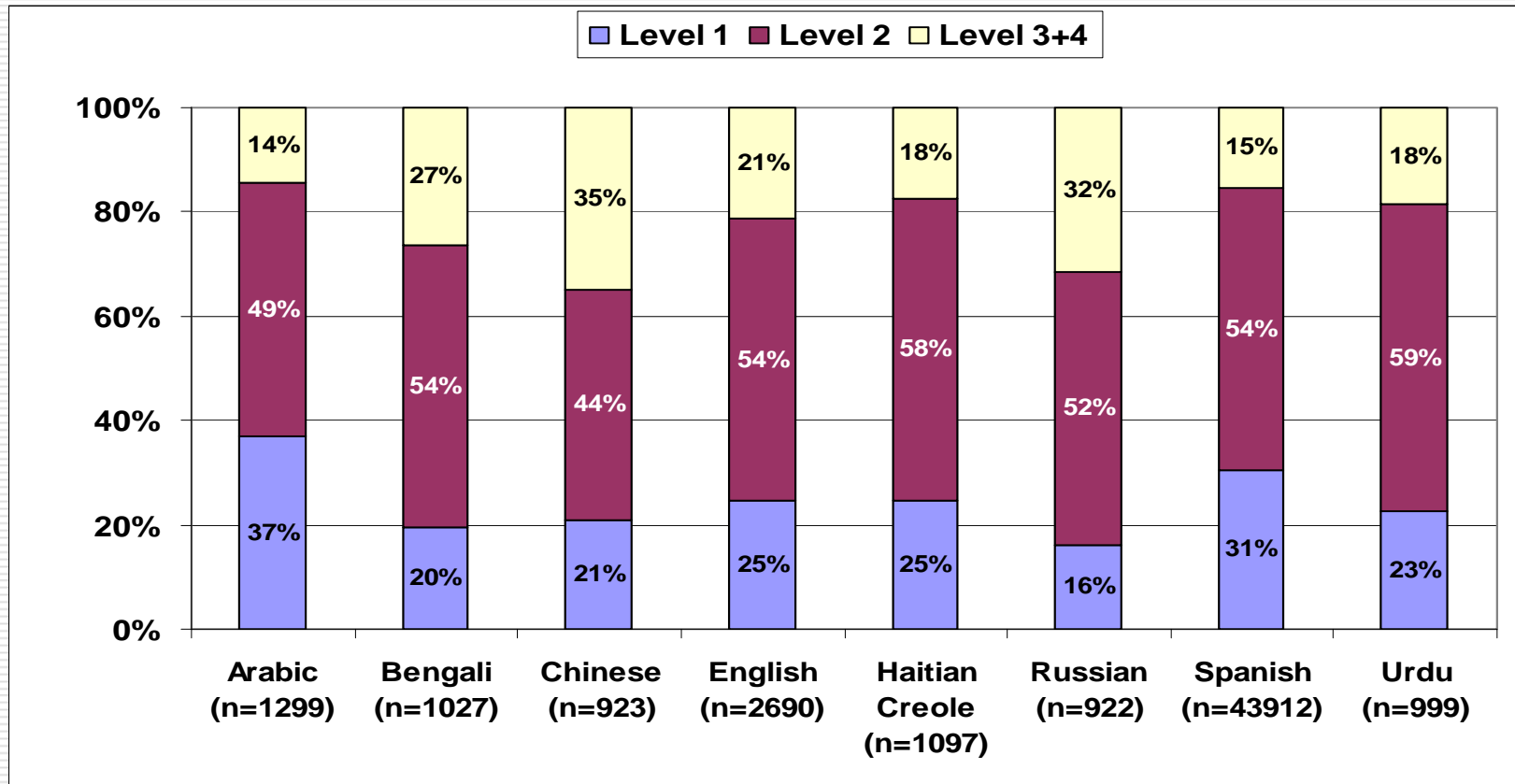
Performance of ELLs on Grade 3-8 ELA Tests by NYSESLAT Overall Proficiency Level



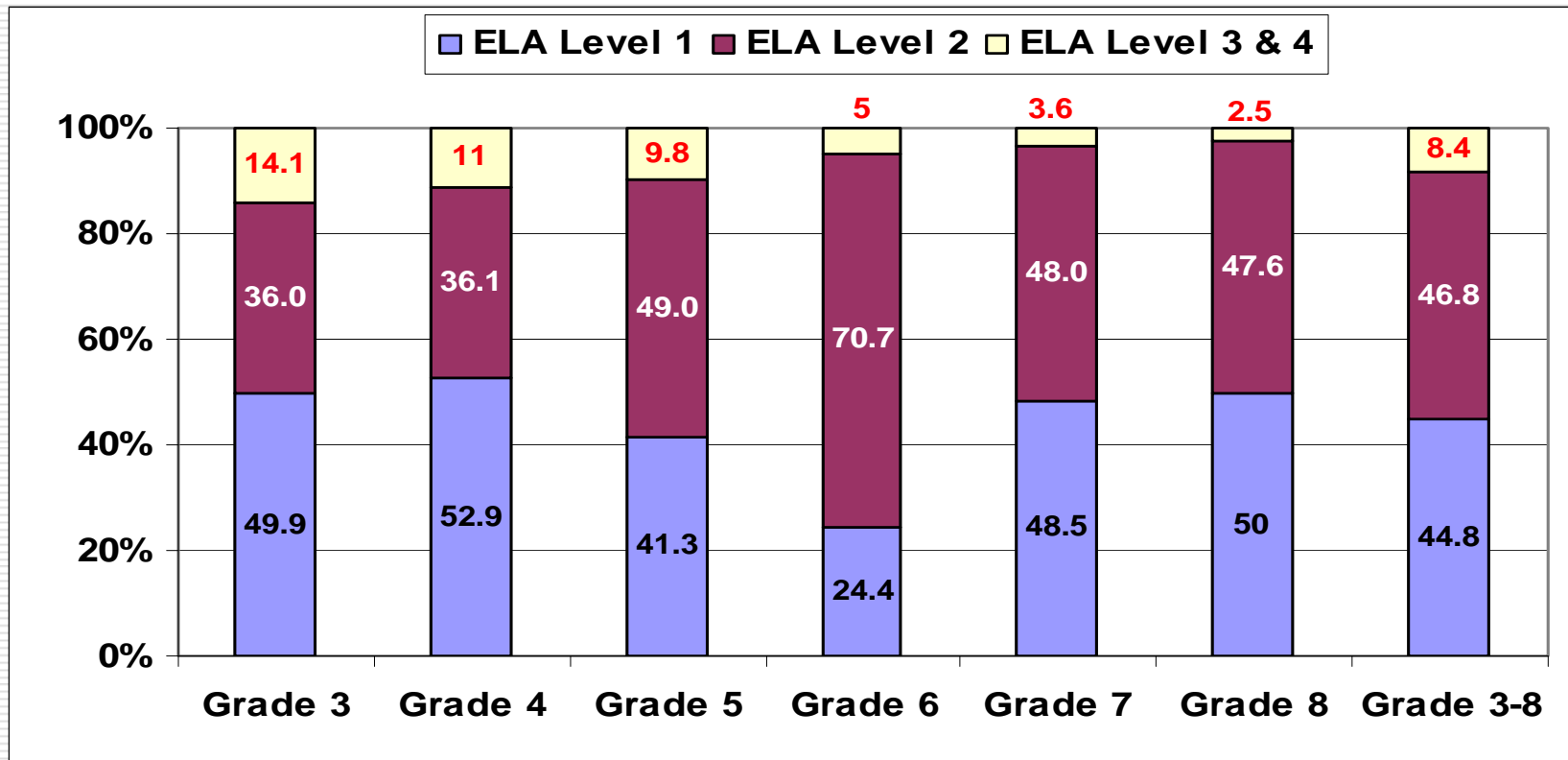
ELL Performance on Grade 3-8 ELA Tests by Need/Resource Capacity Category



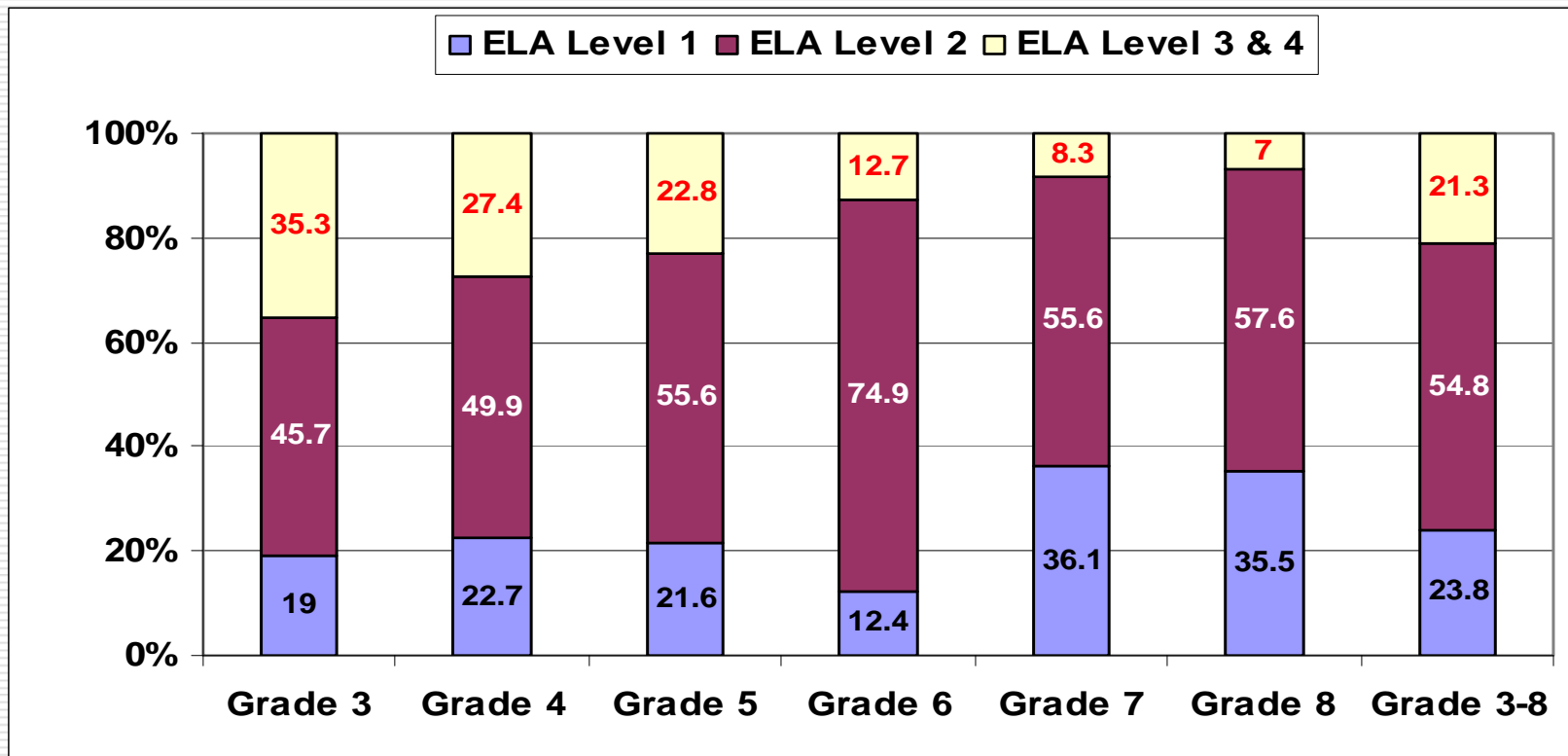
ELL Performance On Grade 3-8 ELA Tests by Major Home Language Group



Performance of ELLs With Disability on Grade 3-8 ELA Tests



Performance of ELLs Without Disability on Grade 3-8 ELA Tests



Summary

- ELLs performed better on the 3-8 ELA tests than expected. Of the 72,082 ELLs who took the 2007 Grade 3-8 ELA tests, 18% scored proficient and 72% scored at Level 2 and above.
 - ELP as measured by the NYSESLAT is a good predictor for ELA performance. Those who scored proficient on the NYSESLAT had a much better chance to score at Level 3 and above on the ELA tests than those who scored below proficient; advanced level ELLs had much better chance to score at Level 2 and above on the ELA tests.
 - ELLs who scored at the beginning level on the NYSESLAT had little chance to score beyond Level 1 on the ELA tests.
 - ELLs in the elementary grades, with 3-5 years of services, and those from schools outside the Big 5 cities were more likely than their counterparts to score proficient on the Grade 3-8 ELA tests.
 - ELLs who scored at Level 1 on the 3-8 ELA tests were more likely to be in the middle school grades, with less than 3 years services, ELLs with disability, and those who scored at the beginning level on the NYSESLAT.
-

Title III AMAO* Review & Update

Title III AMAO 1: Making Progress

- ❑ Definition: Advancement from one proficiency level to the next each year until proficient. It is based on the overall proficiency level as measured by the NYSESLAT scale scores.
- ❑ Unit of Analysis: Student level based on two years of matching NYSESLAT test data.
- ❑ Reporting Level: State and Title III districts aggregated from student level data.
- ❑ Target 2005-06: 60% for 05-06 and 5% increment each year thereafter.

*AMAO: Annual Measurable Achievement Objectives

Title III AMAO Review & Update

Title III AMAO 2: Attaining English Proficiency

- ❑ Definition: Scoring at the proficient level (Level 4) on both the Listening/Speaking and Reading/Writing modality combinations on NYSESLAT.
 - ❑ Unit of Analysis: Student Level
 - ❑ Reporting Level: State and Title III district
 - ❑ Target 2005-06: 10% for LEP/ELLs w/ 1 year of service, 14% for LEP/ELLs with 2 years of service, and 84% for LEP/ELLs with 3 or more years of service.
-

Title III AMAO Review & Update

Title III AMAO 3: Making AYP*

- ❑ Definition: Meeting grade level academic achievement standards in ELA and math based on district AYP data imported from Title I LEAs NCLB Testing Program Results, i.e. AYP for ELL subgroup.

*AYP: Adequate Yearly Progress

Title III AMAO Research Update

- Review Empirical Data to best re-evaluate AMAO Targets 1 and 2
 - Formed Internal AMAO Workgroup that worked w/Robert Linqanti (WestEd) & Harcourt for review of data and methodologies, both for accountability and psychometrics
 - Discuss initial first steps of research with Department's Technical Advisory Group
-

Title III AMAO Research Update

3 Prospective Methodologies:

- Re-calibration Method
- Re-scaling Method
- Equal Weighting Method

Next Steps:

- Continuing Review of Methodologies w/Technical Advisory Group & other experts
 - Meeting w/USDE on methods and waiting for their revised guidance due Spring 2008
-

LEP/ELL Research Agenda

- ❑ Internal Research: ELA 3-8 & Comprehensive English Regents cross analysis w/NYSESLAT Performance
 - ❑ External Validity Study: NYSESLAT Technical Manual 2006
 - ❑ Review Data Trends & Test Structure
 - ❑ Language Threshold Measure
 - ❑ Inform next steps for strengthening the NYSTP for LEP/ELL inclusion
 - ❑ Policy Implications-Test Accommodations & Inclusion
-

LEP/ELL Research Agenda: LEP Partnership

Jamal Abedi: University of California- Davis

- ❑ Contributed testing information for his research on State English Language Development Exams (NYSESLAT)

Charlene Rivera: George Washington University
Center for Equity & Excellence In Education

- ❑ Project 1: Participating in the Best Practices in State ELL Assessment Policies: Handbook Development Project
 - ❑ Major Research Focus: Efficacy of Test Accommodations
-

LEP/ELL Research Agenda-Rivera Con't

- Project 2: Investigating the Academic Language Demands Implicit in State Content Standards

Aims of Project:

- Reviewing and documenting the research on academic language in mathematics and science;
 - Developing a research-based Content Standards Language Analysis Tool (CSLAT) to support the investigation and identification of the specific academic language demands implicit in algebra and biology (Living Environment) content standards;
-

Call For Expertise

To become involved in test development and standard-setting activities, please download and complete our application found at:

www.emsc.nysed.gov/ciai/call.htm
