




*Assessment, Accountability, and
Instruction for ELLs Under NCLB*

*David J. Francis
University of Houston*

 **TIMES** Texas Institute for Measurement,
Evaluation, and Statistics

Purpose of this Talk

- Establish a framework for understanding the role of language development in content area knowledge acquisition
- To demonstrate that the achievement gap is not a simple testing gap that can be solved through improved tests and test accommodations
- In this way, this talk is intended to help motivate our attention to developing students' knowledge of academic language in order to close achievement gaps for EL learners



Overview

- Background on ELLs
- Examine the unique challenges posed by Limited English Proficiency as a subgroup under NCLB
- Examine the link between language proficiency and content mastery
- Examine the effectiveness of test accommodations for ELLs
- Conclude by suggesting that the key to success on content area assessments is to develop children's knowledge of academic English.



Who Are English Language Learners?

- **Language Minority Student (LM)**
 - a child who hears and/or speaks a language other than English in the home (see August & Shanahan, 2006 for review of literature)
- **English Language Learner (ELL)**
 - an LM student designated locally (i.e., by the state) as limited English Proficient
- **Limited English Proficient (LEP)**
 - an LM student whose limited command of English prevents independent participation in instruction



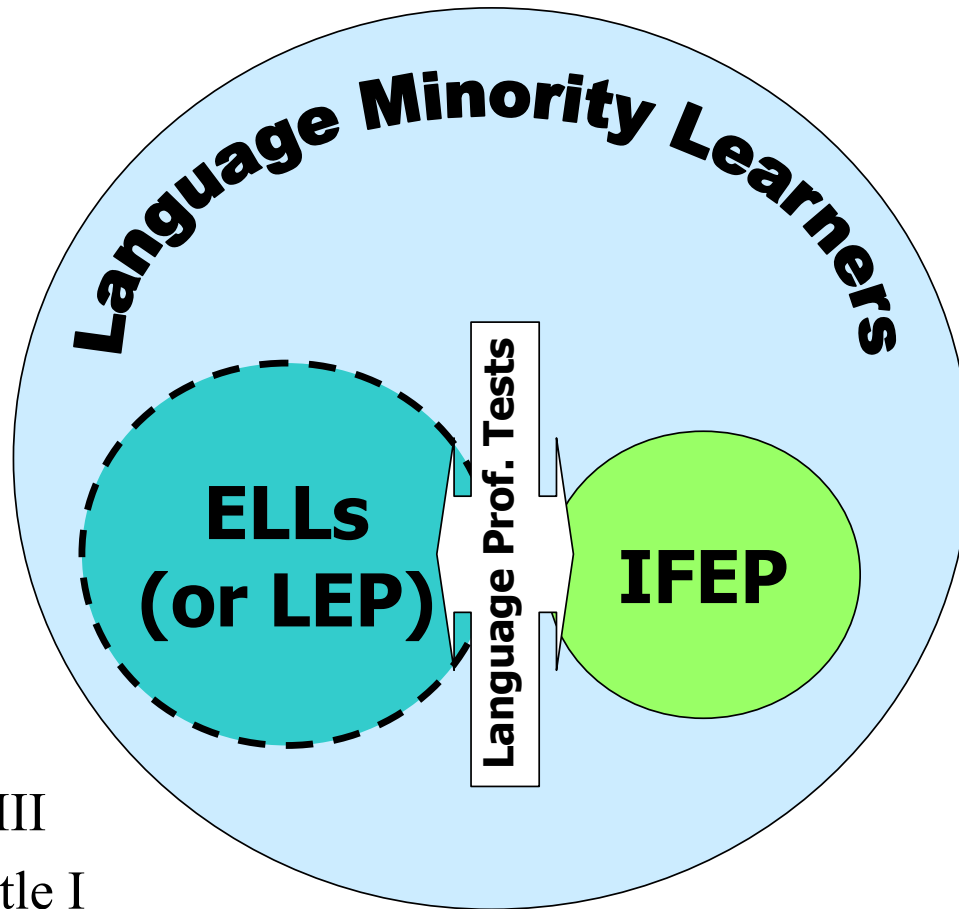
Definitions: At school entry

Identification

- Home survey
- Language proficiency tests
- Other input (e.g., teachers)

Monitoring

- Language – Title III
- Achievement – Title I

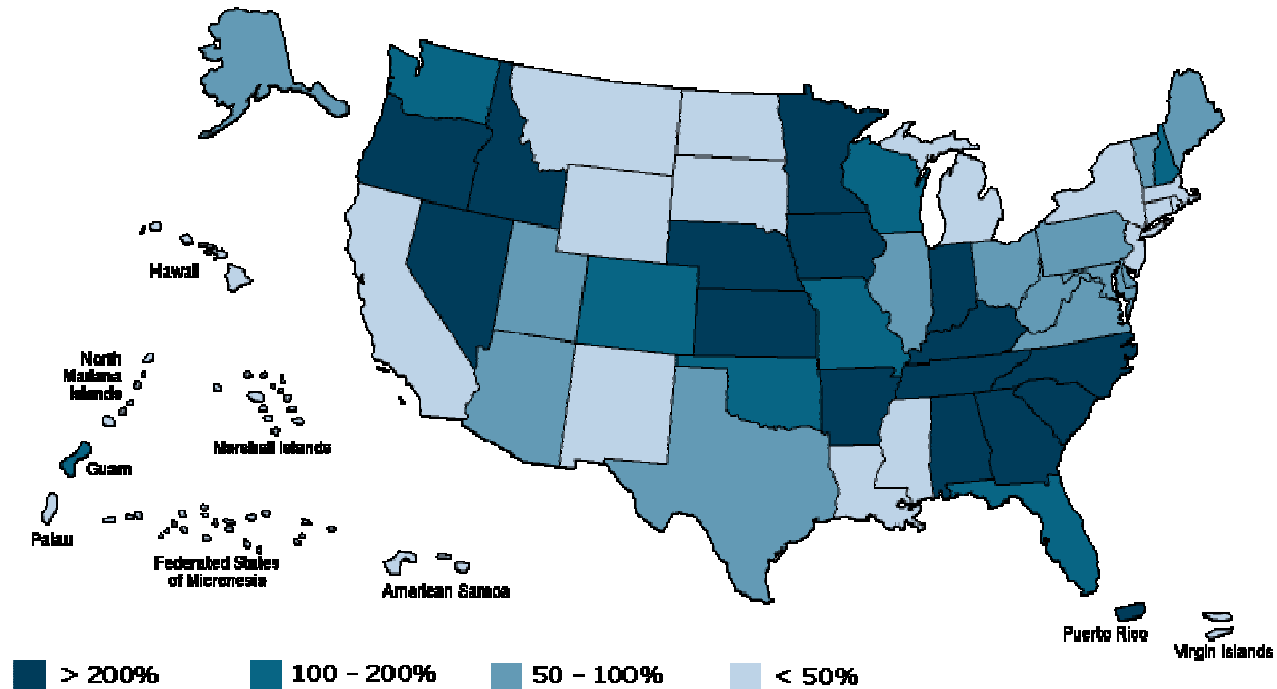


Who Are English Language Learners?

- Over 9M LM students, roughly 5.5M classified as LEP
- Comprise one of the fastest-growing groups among the school-aged population in this nation
 - 169% from 1979 to 2003 (vs. 12% growth in general)
 - Expected to be 30% of school-aged population in 2015
 - 19 states have reported 10-year growth in excess of 200%
- Largest and fastest growing segment of ELL population is
 - Students who immigrated before Kindergarten, and
 - U.S. born children of immigrants



ELL Growth



States with over 200% growth:

Alabama	Indiana	Minnesota	Oregon
Arkansas	Iowa	Nebraska	South Carolina
Georgia	Kansas	Nevada	Tennessee
Idaho	Kentucky	North Carolina	Puerto Rico

Source: U.S. Department of Education's Survey of the States, Limited English Proficient Students and Available Educational Programs and Services, 1991-1992 through 2000-2001. Slide courtesy of Kathleen Leos, USDOE.



Who Are English Language Learners?

- Heterogeneous population
 - Time / age of arrival
 - Prior school experience
 - Parental education
 - Degree of economic and social advantage/disadvantage
 - Home Language
- Well over 400 different home languages among LMs in US
 - Spanish (over 70%); Vietnamese (roughly 4%)
 - Specific geographic regions have significant numbers of children speaking a particular language (Chinese; Russian; Arabic, etc.)

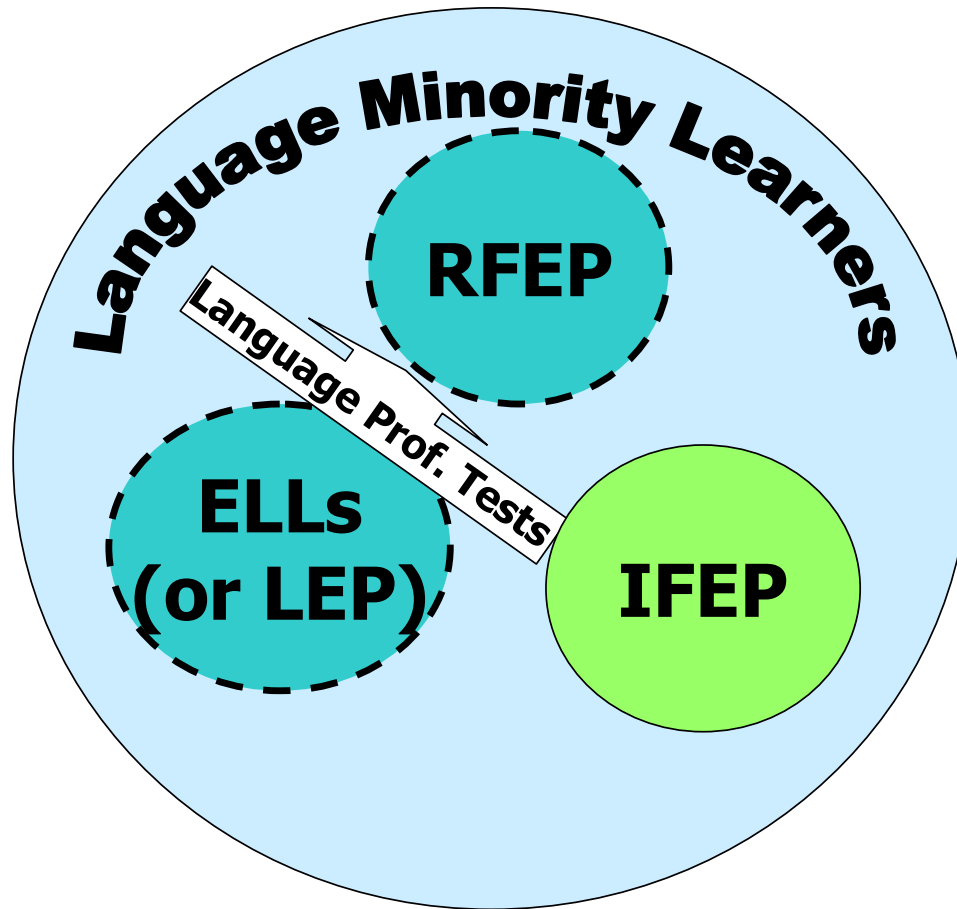


Academic performance indicators for ELLs

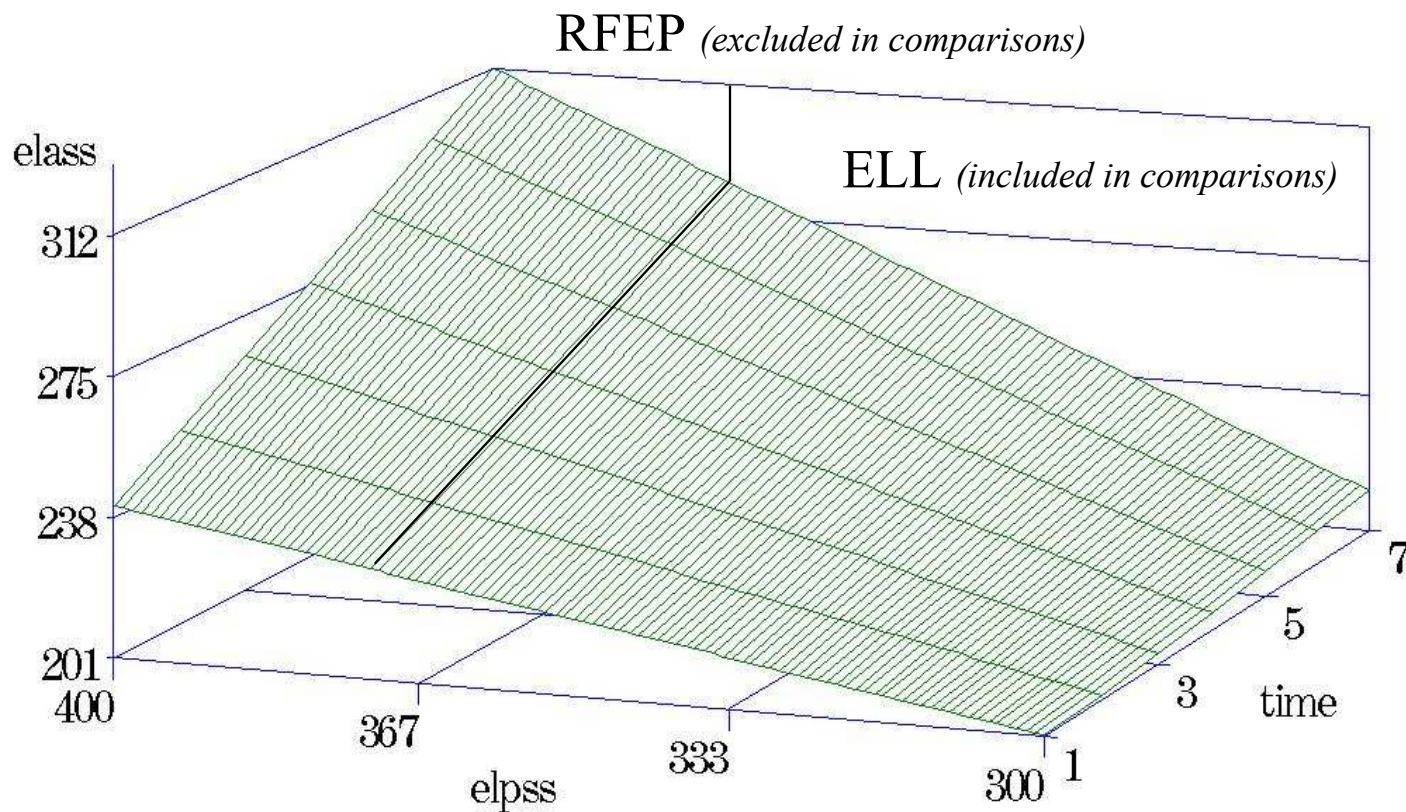
- Compared to native English-speaking peers on Grade 4 NAEP, ELLs were
 - 1/4th as likely to score proficient or above in Reading
 - 1/3rd as likely in Math
- ELLs also perform more poorly on State tests
 - For example, in 2002, only 18.7% of designated ELLs scored proficient in reading on state tests (9 states did not report)
- However, state and federal accountability systems may bias such comparisons against ELLs



Definitions: Over time



Surface Plot of Predicted Values for ELA



$ELA > 240 = \text{Proficient} / ELP > 375 = \text{Proficient}$



Comparison of ELLs and former ELLs on State Reading Test in Texas 2002

Grade	Level of Language Proficiency for ELL Groups			
	Beginning	Intermediate	Advanced (2002)	Advanced (2000)
3	13.9	38.3	90.6	90.0
4	13.1	37.4	84.1	93.6
5	16.5	24.1	69.5	96.1
6	14.5	12.8	46.0	86.8
7	15.0	12.4	43.9	85.0
8	23.2	19.2	55.3	90.2
10	21.3	28.5	66.4	85.8
Overall	15.8	30.4	76.4	89.6

<http://www.tea.state.tx.us/student.assessment/reporting/results/rpteanalysis/2002/reading/statewide.html>



Comparison of Graduation Rates among ELL, Former ELLs, and Never ELLs in New York City¹

Group	After four Years of High School	After 7 years of High School
Current ELLs	32.6	49.5
Former ELLs	60.1	76.5
Never ELLs	54.5	70.5

¹Cohort entering grade 9 in 1996.

<http://www.regents.nysed.gov/2005Meetings/March2005/0305emscvesidd4.html>

The logo for The Times, featuring the word "TIMES" in a bold, serif font with a stylized red and black graphic element to the left.

Current Law

- Allows retention of ELL label for accountability purposes for two years after achieving FEP status
 - Improvement over past practice, but ...
 - ▲ Does not allow reporting of long term outcomes for students who began school as ELL
 - ▲ Fails to take into account the developmental nature of language acquisition
 - ▲ Fails to take into account the causal role of language in content area performance
- Requires development of Language Proficiency standards that are closely aligned to content standards



Assessment of Language Proficiency under NCLB

- Must include Listening, Speaking, Reading, and Writing
- Used for Title 3 accountability, but...
- Also commonly used for placement, re-designation, decisions about interventions, etc.
- Must be uniform within state for all students covered by Title III, but is not required for ELL students not receiving Title III services
- Language proficiency assessments vary greatly across states in content and format



High Stakes Assessments and ELLs

- The effect of language proficiency on ELL performance on content area assessments is a matter of considerable interest to educators, as well as to students and their families.
- At the same time, this question has not been widely studied in the contexts of today's Title I and Title III assessments
- Title I assessments - content area assessments
- Title III assessments - measure English language acquisition
 - Reading,
 - Writing,
 - Speaking,
 - and Listening



Sample Sizes for state-level data base

Cluster Type	N	Number with 10 or Fewer Students	Maximum N
Schools	948	573	304
Districts	170	67	2,590



Student Demographics

Demographic	ALL	Grade				
		4	5	6	7	8
%Female	47%	48%	47%	46%	48%	55%
%Asian ^a	18%	19%	17%	17%	17%	17%
%Black ^a	11%	10%	11%	10%	11%	12%
%Hispanic ^a	58%	57%	58%	59%	60%	60%
%Native American ^a	<1%	<1%	<1%	<1%	<1%	<1%
%Caucasian ^a	13%	14%	14%	14%	12%	12%
%Title I	71%	78%	75%	71%	62%	63%
%Free/Reduced Lunch	81%	81%	80%	81%	81%	82%
%Special Education	20%	19%	22%	21%	20%	17%
%Proficient ELP ^c	46%	60%	40%	44%	36%	44%
Sample Size ^b	17,767	4,663	3,566	3,303	3,172	3,063



How are Years in the US, and ELP performance related to performance on the ELA and Math content assessment?

○ ELP Assessment

- Subtests of Reading (0-30), Writing (0-30), Speaking (0-30), Listening (0-5)
- Composite Scaled Score (300-400)
- Performance Levels¹
 - ▲ Beginning (300-324)
 - ▲ Early Intermediate (325-348)
 - ▲ Late Intermediate (349-374)
 - ▲ Transitioning (375-400)

¹Proficiency cut points vary slightly by grade



3-Level Model for ELA and Math

○ Conditional Models

- Model A
 - ▲ Years in US
 - ▲ ELP measured as Performance Levels
 - ▲ Years in US + ELP measured as Performance Levels
- Model B
 - ▲ Years in US +
 - ✦ ELP measured as Performance Levels
 - ✦ ELP measured as a composite Scaled Score
 - ✦ ELP measured as Domain Scores (R, W, S, L)
 - ✦ ELP measured as Domain Scores (Reading, Writing)



3-Level Model for ELA and Math

○ Unconditional Model (within grade)

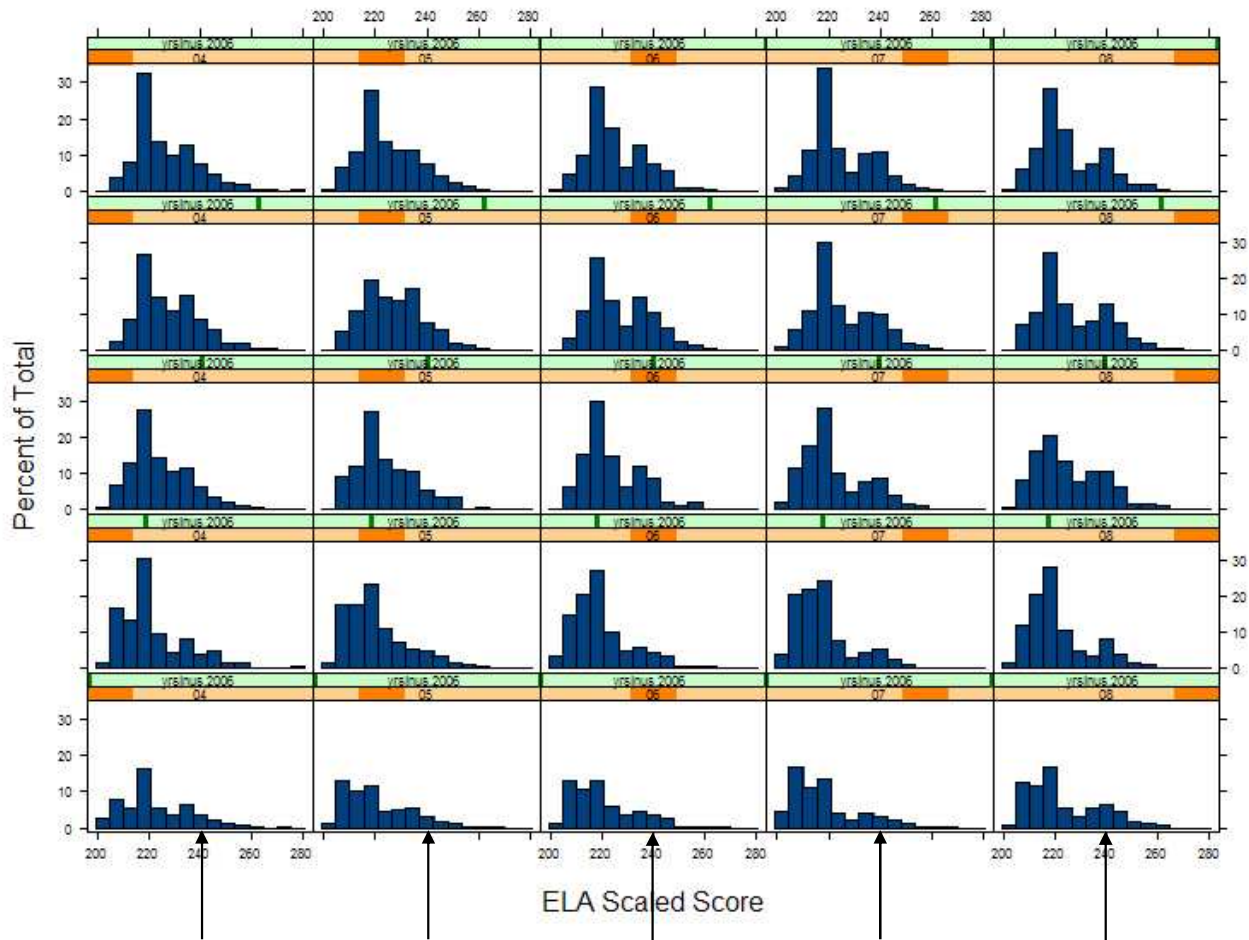
- $V(\text{Students}(\text{schools}))$
- $V(\text{Schools}(\text{Districts}))$
- $V(\text{Districts})$

○ Conditional Models

- Years in US
- ELP
- Years in US and ELP

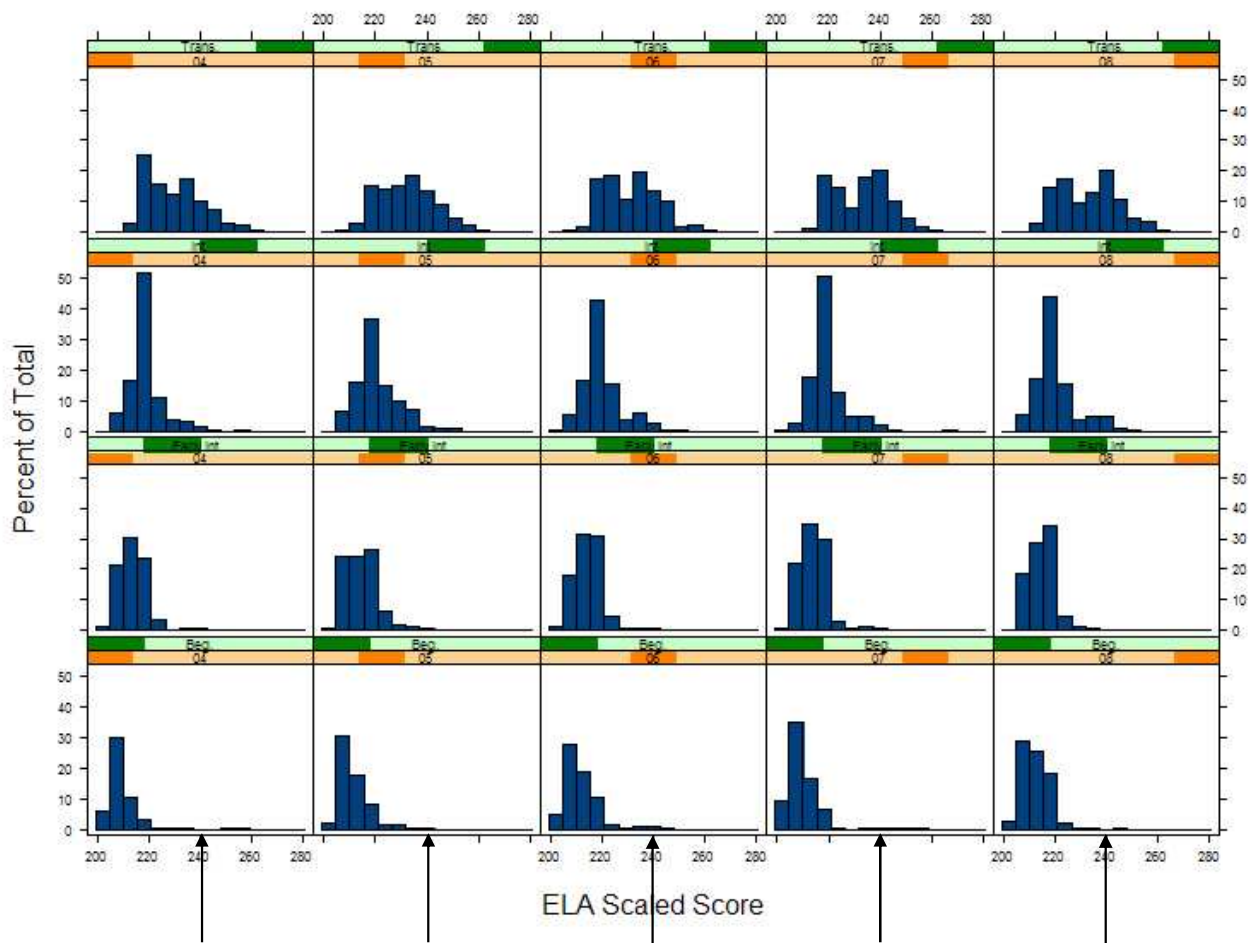


Distribution of ELA by Grade and Years in US

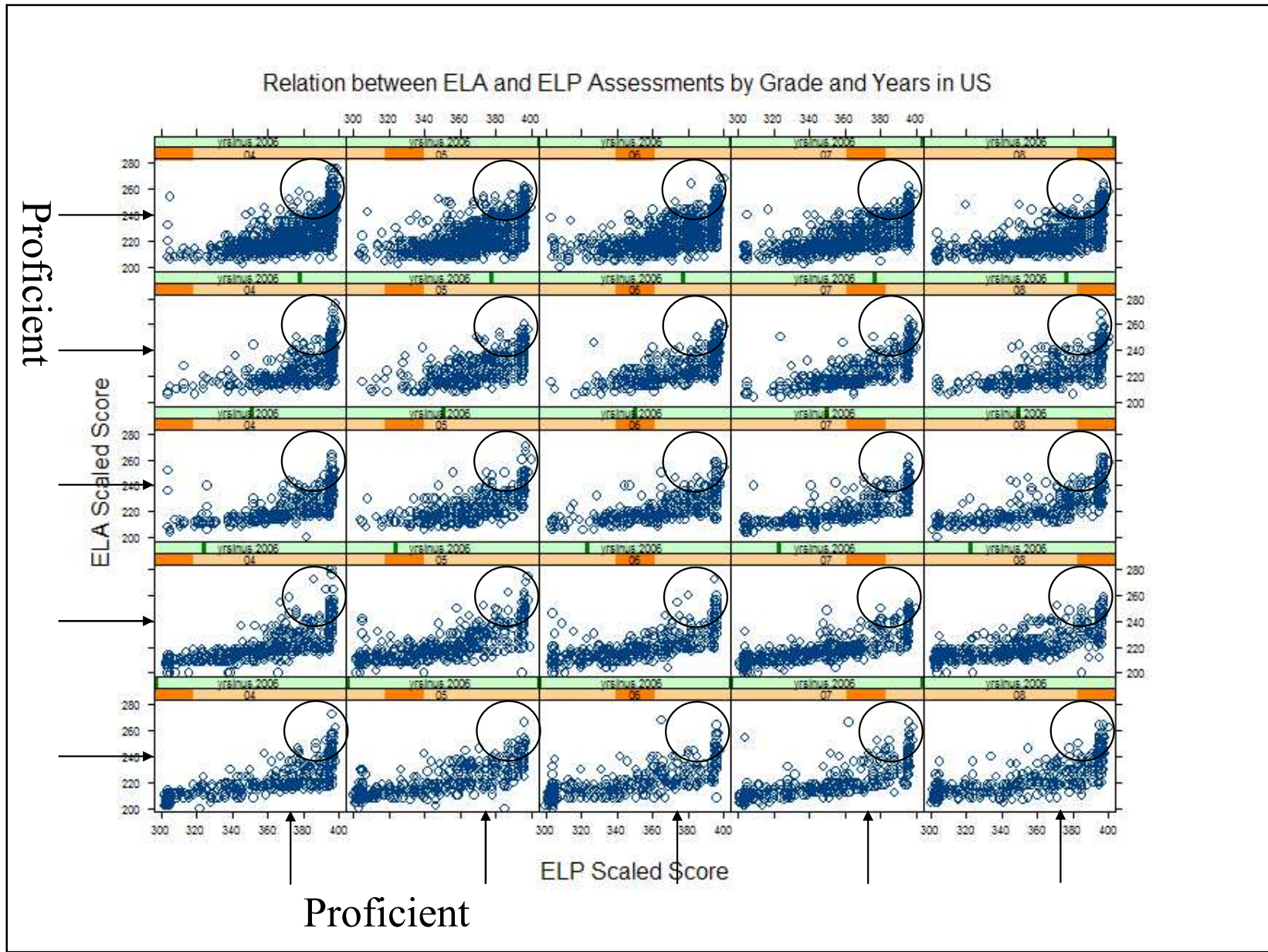


Proficient

Distribution of ELA by Grade and ELP Proficiency Level



Proficient



Conditional Random Effects for ELA and MATH predicted from Years in US, ELP, and Years + ELP

Grade	Source	ELA						MATH					
		Years in US	ΔR^2	ELP-Perf.	ΔR^2	Years and ELP	ΔR^2	Years in US	ΔR^2	ELP-Perf.	ΔR^2	Years and ELP	ΔR^2
4	District	27.21	-0.13	15.13	0.37	14.73	0.39	41.11	-0.12	29.10	0.21	26.51	0.28
	Schools	25.04	0.02	15.66	0.39	15.72	0.39	32.74	0.04	22.62	0.34	22.85	0.33
	Students	108.37	0.03	81.83	0.27	81.67	0.27	145.14	0.03	119.72	0.20	118.84	0.21
5	District	25.73	-0.02	11.62	0.54	11.11	0.56	45.24	-0.04	36.52	0.16	35.45	0.19
	Schools	14.83	0.05	9.25	0.41	9.53	0.39	33.28	0.05	23.34	0.33	22.88	0.35
	Students	104.37	0.03	70.30	0.35	69.65	0.35	149.57	0.01	120.02	0.21	117.86	0.22
6	District	22.15	-0.05	9.16	0.56	8.59	0.59	49.56	-0.02	35.88	0.26	31.27	0.36
	Schools	18.24	0.11	12.68	0.38	12.90	0.37	23.81	0.01	20.05	0.16	20.34	0.15
	Students	97.03	0.04	66.38	0.34	66.07	0.34	133.72	0.01	111.82	0.18	109.45	0.19
7	District	27.88	-0.08	11.20	0.57	11.05	0.57	61.72	-0.05	47.32	0.20	43.68	0.26
	Schools	13.08	0.26	4.53	0.74	4.63	0.74	19.42	0.03	14.44	0.28	15.03	0.25
	Students	104.51	0.03	60.65	0.44	60.68	0.44	119.63	0.01	97.70	0.19	95.91	0.21
8	District	26.70	-0.02	10.87	0.58	9.54	0.63	51.31	0.02	42.76	0.18	37.47	0.28
	Schools	22.99	0.05	7.58	0.69	8.46	0.65	30.17	-0.02	22.09	0.26	21.84	0.26
	Students	113.83	0.01	73.83	0.36	72.40	0.37	109.00	0.01	92.13	0.16	89.03	0.19

^a ΔR^2 computed as change in variance component from unconditional model (Table 5) relative to magnitude of variance component in unconditional model (Table 5-Table 6)/(Table 5).

Conditional Random Effects for ELA and MATH predicted from Years in US + ELP measured as (1) Performance Levels, (2) Scaled Score, or (3) Domain Scores

Grade	Source	ELA						MATH					
		Years + ELP-PL	ΔR^2	Years + ELP-SS	ΔR^2	Years + ELP-DS	ΔR^2	Years ELP-PL	ΔR^2	Years + ELP-SS	ΔR^2	Years and ELP-DS	ΔR^2
4	District	14.73	0.39	14.00	0.42	10.92	0.54	26.51	0.28	25.27	0.31	20.12	0.45
	Schools	15.72	0.39	14.90	0.42	12.81	0.50	22.85	0.33	22.08	0.35	18.65	0.45
	Students	81.67	0.27	74.28	0.33	60.20	0.46	118.84	0.21	112.80	0.25	100.78	0.33
5	District	11.11	0.56	10.72	0.57	7.66	0.70	35.45	0.19	33.94	0.22	25.35	0.42
	Schools	9.53	0.39	8.27	0.47	7.21	0.54	22.88	0.35	22.44	0.36	20.03	0.43
	Students	69.65	0.35	65.37	0.39	60.02	0.44	117.86	0.22	112.85	0.25	105.12	0.31
6	District	8.59	0.59	7.02	0.67	7.60	0.64	31.27	0.36	28.26	0.42	26.63	0.45
	Schools	12.90	0.37	10.78	0.47	6.67	0.67	20.34	0.15	18.99	0.21	17.36	0.27
	Students	66.07	0.34	61.97	0.39	56.49	0.44	109.45	0.19	104.52	0.23	97.72	0.28
7	District	11.05	0.57	11.09	0.57	8.04	0.69	43.68	0.26	42.19	0.28	34.16	0.42
	Schools	4.63	0.74	4.05	0.77	3.06	0.83	15.03	0.25	14.21	0.29	12.25	0.39
	Students	60.68	0.44	57.85	0.47	53.16	0.51	95.91	0.21	93.23	0.23	85.99	0.29
8	District	9.54	0.63	8.36	0.68	3.32	0.87	37.47	0.28	35.03	0.33	27.24	0.48
	Schools	8.46	0.65	7.14	0.70	5.52	0.77	21.84	0.26	20.94	0.29	19.99	0.33
	Students	72.40	0.37	69.24	0.40	60.64	0.47	89.03	0.19	85.41	0.22	75.99	0.31

^a ΔR^2 computed as change in variance component from unconditional model (Table 5) relative to magnitude of variance component in unconditional model (Table 5-Table 7)/(Table 5).

Analysis Summary

- Years in the US predicted ELA and MATH performance at the district, school, and student levels
- However, Years in the US was a relatively weak predictor compared with ELP
- When ELP was included with Years in US, the effects of Years in the US were unsystematic and small;
- Effects of ELP remained strong and consistent (i.e., outcomes increased with increases in ELP)



Analysis Summary

- How ELP was measured made some difference in its value as a predictor; Domain Scores predicted best
- Using Domain Scores for Reading and Writing only was almost as good as using Reading, Writing, Speaking, and Listening
- These results suggest that the academic components of the language assessment are the most important predictors of content area achievement
- It is noteworthy that ELP performance explained so much of the school and district variability in ELA and MATH

The logo for 'TIMES' features the word in a bold, serif font. A red circular graphic element is positioned behind the letter 'I', partially overlapping it.

Including ELLs in Accountability Assessments

- Raises educators' awareness of ELLs' academic needs
- But also raises questions about the appropriateness and validity of content area assessments...



Validity of Test Scores for ELLs

- Threatened if scores reflect irrelevant language abilities rather than content knowledge
(AERA, APA, NCME, 1999)
- Strong link between language proficiency & performance in content areas
(Abedi & Leon, 1999; Bailey, 2005; Butler & Castellon-Wellington, 2005)
- Gaps between ELLs and non-ELLs vary as a function of language demands of the test
(Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Leon, & Mirocha, 2003; Abedi, Lord, & Plummer, 1997).



Test Accommodations

- Alterations to the test or testing conditions
- Address specific needs (e.g., limited English proficiency) but do not change the construct
- Dual criteria for appropriateness:
 - Effectiveness: Accommodation should improve the scores of students who need it.
 - Validity: (in part) Accommodation should not improve the scores of students who do not need it.

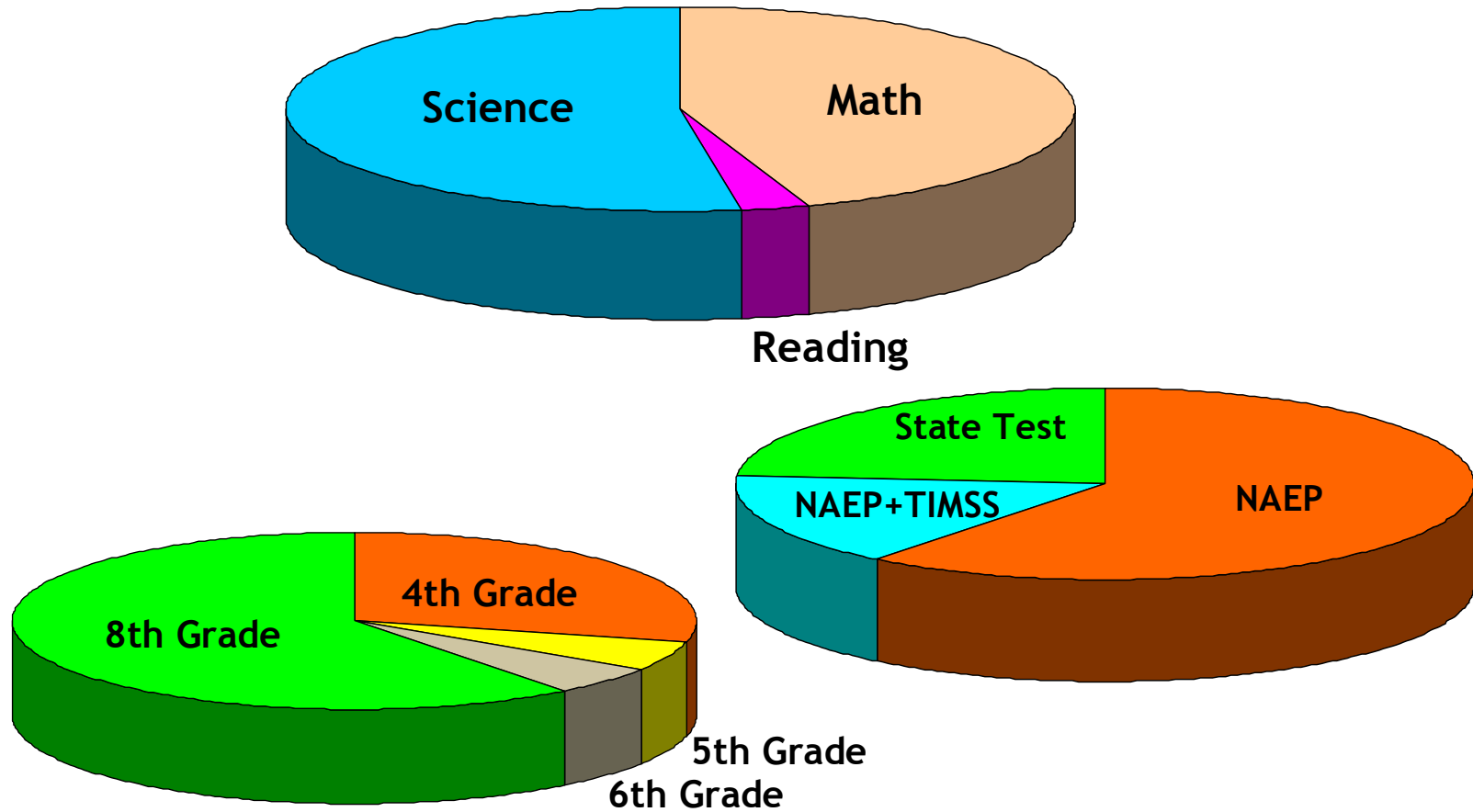


Meta-Analytic Review of Accommodations for ELLs

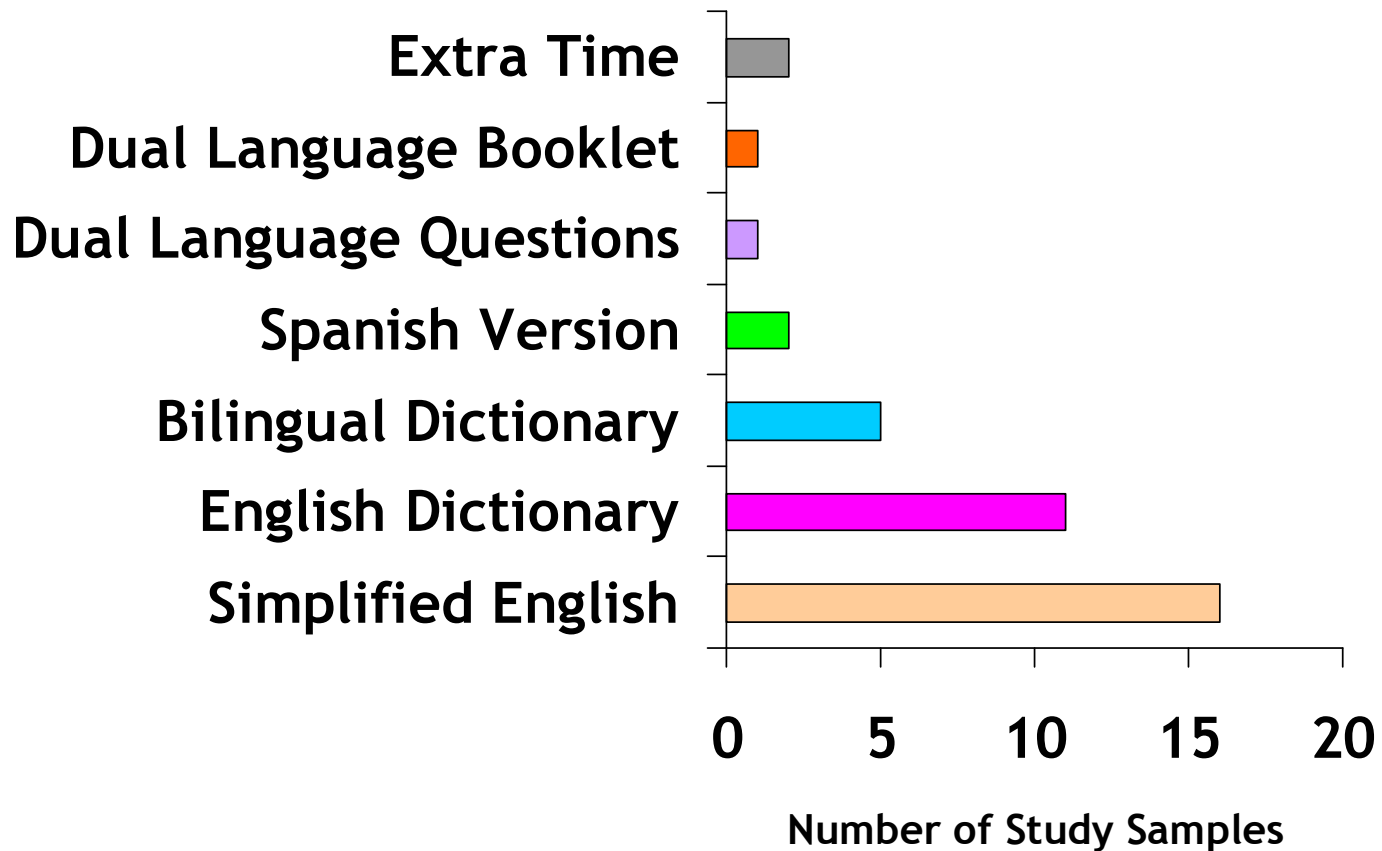
- Inclusion criteria
 - Experimental & quasi-experimental studies focused on the effects of accommodations for ELLs
 - Journal articles & technical reports
 - Appropriate data reported
- Meta-analysis
 - Effectiveness: average effect for ELLs
 - Moderators: Interactions with grade, domain, extra time, language of instruction, study design (experimental vs. quasi-experimental)
 - Validity: average effect for native English speakers



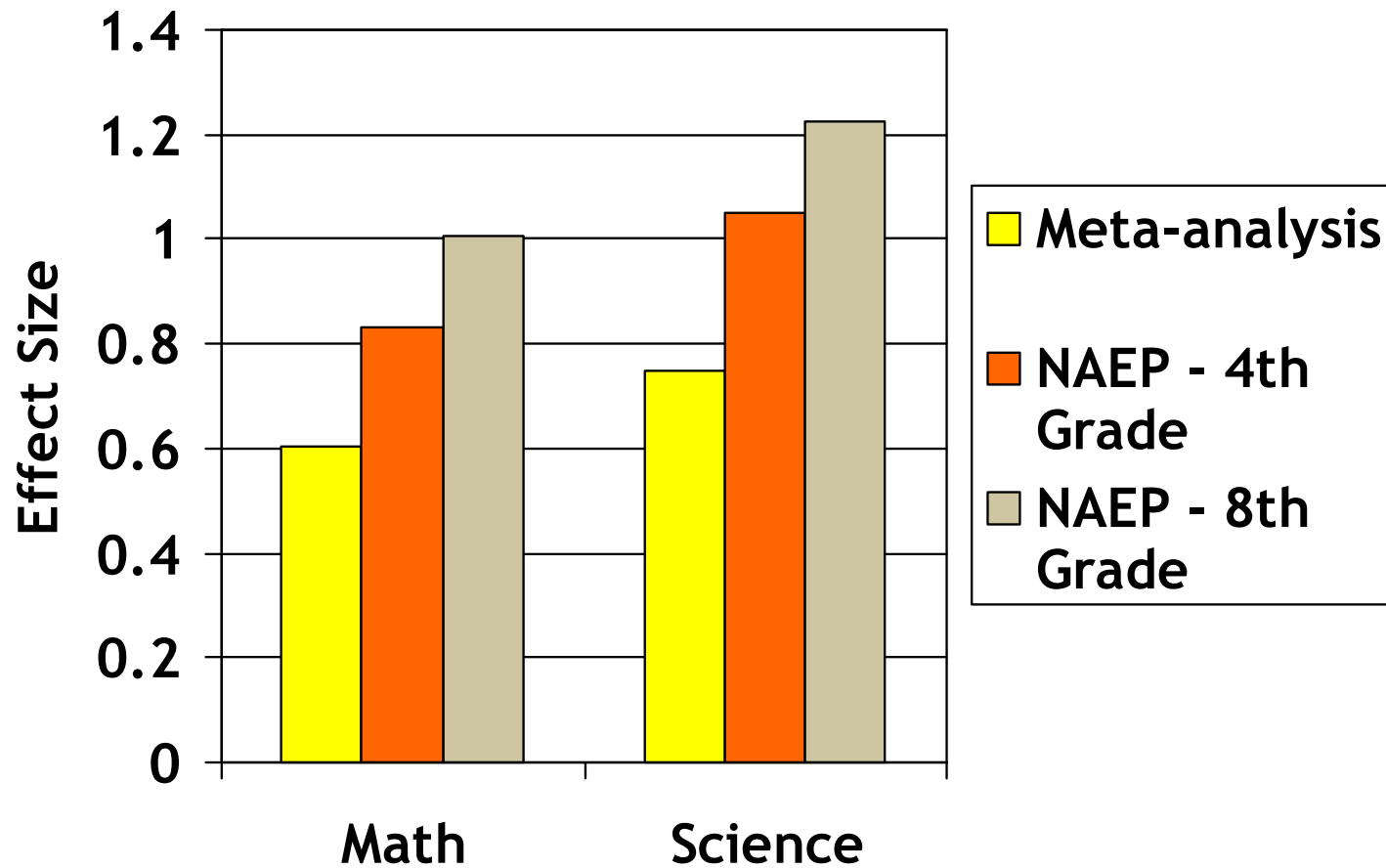
11 Studies yielding 38 Effect Sizes



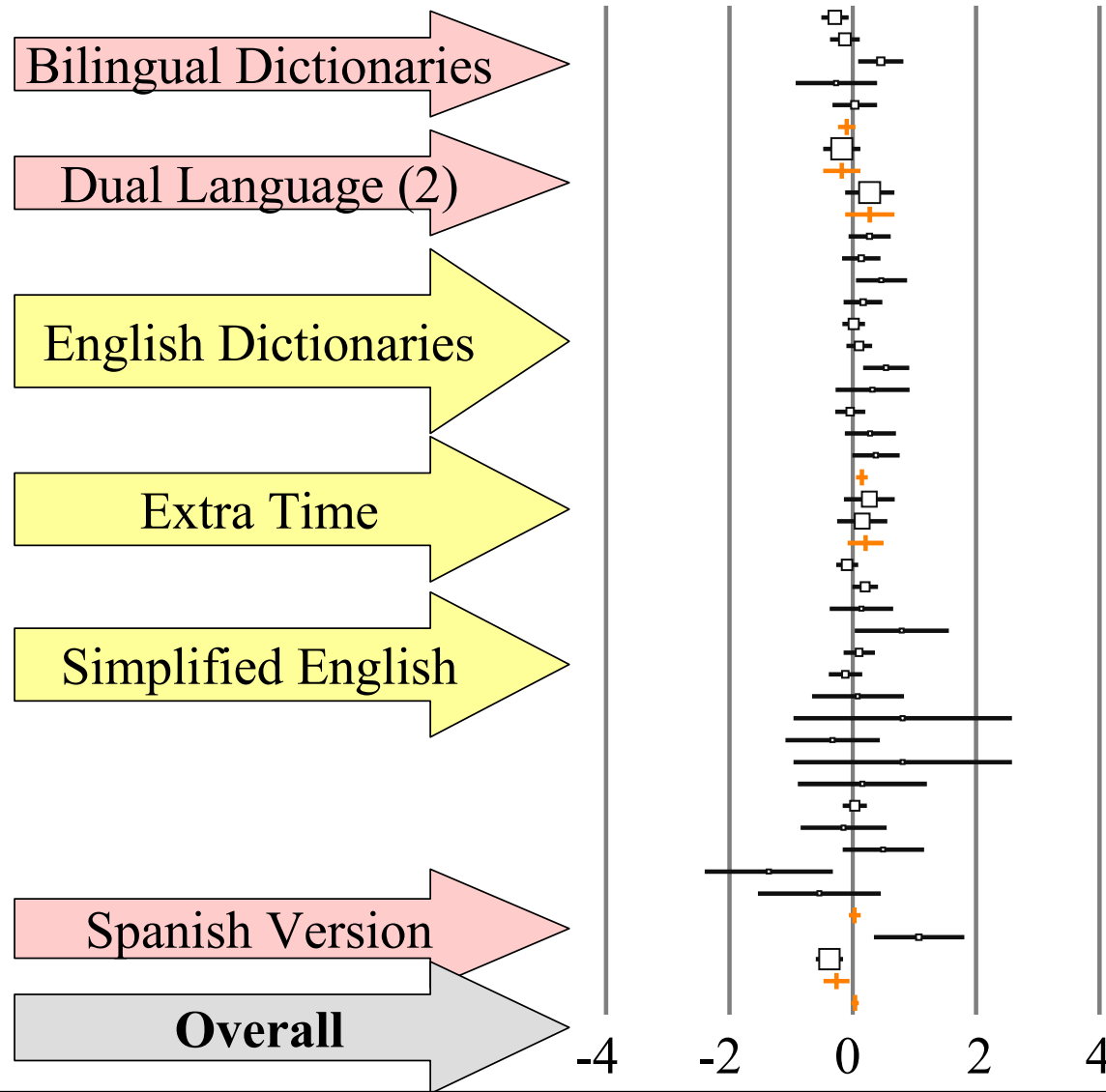
Types of Accommodations



How large are the achievement gaps between ELLs tested without accommodations and non-ELLs?

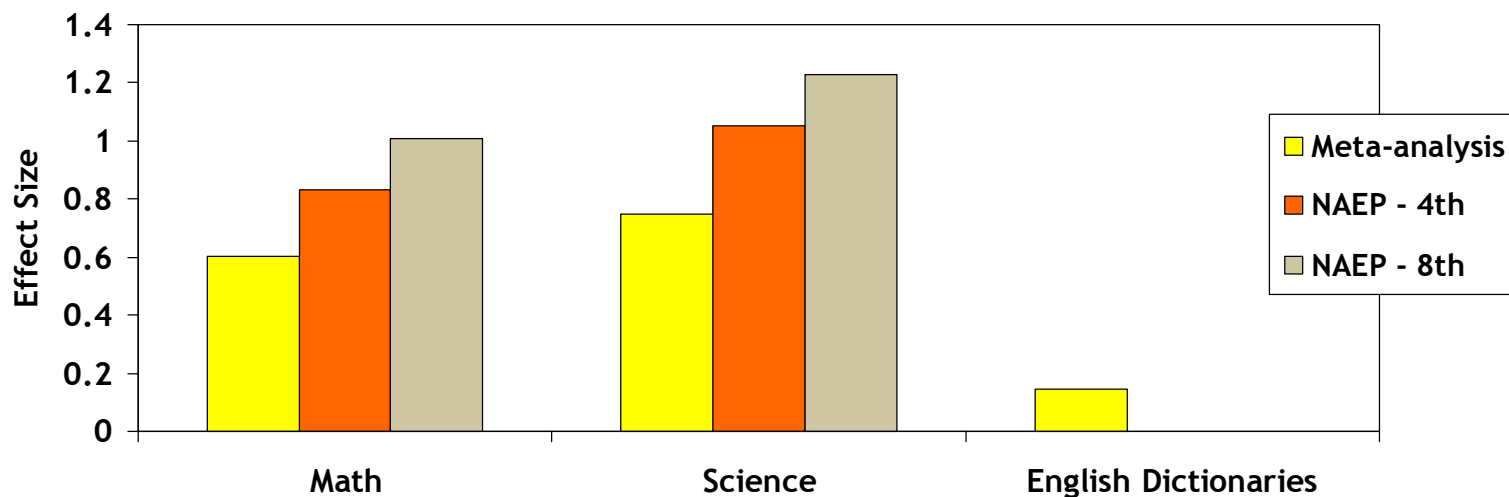


Forest Plot of Effect Sizes Grouped by Accommodation Type



Findings: Effectiveness

- Of the accommodations studied, only providing English dictionaries had a significant effect.
 - Hedges' $g^u = .15$ ($p = .001$)
 - Approximately 10% – 25% of the difference between ELLs & native English speakers



Conclusions

- Future research should investigate accommodations on state tests.
- Question the assumption that accommodations will significantly improve pass-rates for ELLs.
- Taken together, the results highlight the importance of instruction in the academic language which is at the heart of developing content area knowledge.
- They further highlight that children need to be taught in order to close achievement gaps.

The logo for 'TIMES' features the word in a bold, serif font. A red circular graphic element is positioned behind the letter 'T', partially overlapping it.

Broader Implications for Assessment and Accountability

- We clearly have a reporting problem that is fueled by the dynamic nature of the ELL category
- Why not...?
 - Report achievement results within ELP proficiency levels –
 - ▲ Beginner,
 - ▲ Intermediate,
 - ▲ Advanced Intermediate,
 - ▲ Fluent English Proficient



Broader Implications for Assessment and Accountability

○ Why not...?

- Report acquisition of language proficiency as a function of Years in US (or Years in State)
- Integrate Title III and Title I assessment by taking into account the developmental nature of language, and the central role of language in content acquisition
 - ▲ Index – weighted average of ELP and Content Tests
 - ▲ Weights vary with Years in State
 - ▲ Weight for ELP declines with increasing years
 - ▲ Weight for Content Test increases with increasing years

The logo for 'TIMES' features the word in a bold, serif font. A red graphic element, resembling a stylized 'T' or a checkmark, is positioned to the left of the letters 'I' and 'M', partially overlapping them.

Broader Implications for Assessment and Accountability

- An accountability model that addresses these issues will provide more accurate information to teachers, principals, and other stakeholders about the performance of ELLs
- Place emphasis on integration of language instruction into content area instruction, and
- Increase the emphasis on teaching content when ELLs first reach school
- Increase the demand for language tests that will serve as better barometers of ELL students' acquisition of the academic language skills needed to master content domains.





Thank You

dfrancis@uh.edu