The Connie Belin & Jacqueline N. Blank International Center
for Gifted Education and Talent Development

The University of Iowa College of Education

Searching more successfully for academic talent

Finding the right measures and using the right norm groups

David Lohman
Institute for Research & Policy on Acceleration

The University of Iowa
Be remarkable

Nurturing **Potential** | *Inspiring* **Excellence**

Overview

- A potpourri of recent research
  - Sex differences in Quantitative Reasoning
  - Stability of scores on ability & achievement tests
  - Frequency of extreme discrepancies in score profiles for the most (and least) able
- An aptitude *perspective* on talent
- Getting serious about Opportunity to Learn
- New developments

Sex differences in Quantitative Reasoning

Lohman, D. F. & Lakin, J. (under review). Consistencies in Sex Differences on the Cognitive Abilities Test across Countries, Grades, Test Forms, and Cohorts

Comparing Variability

- Sample size
- Ceiling effects
- Scaling & distributional assumptions
- Differential Item Functioning
  - Distinguish from sex differences
  - Magnified effects on extreme scores

Strand, Deary, & Smith (2006)

- Over 320,000 11-12 year old UK students
- Level D of the UK version of the Cognitive Abilities Test (CAT)
  - Verbal Battery (3 tests)
  - Quantitative Battery (3 tests)
  - Nonverbal Battery (3 tests)

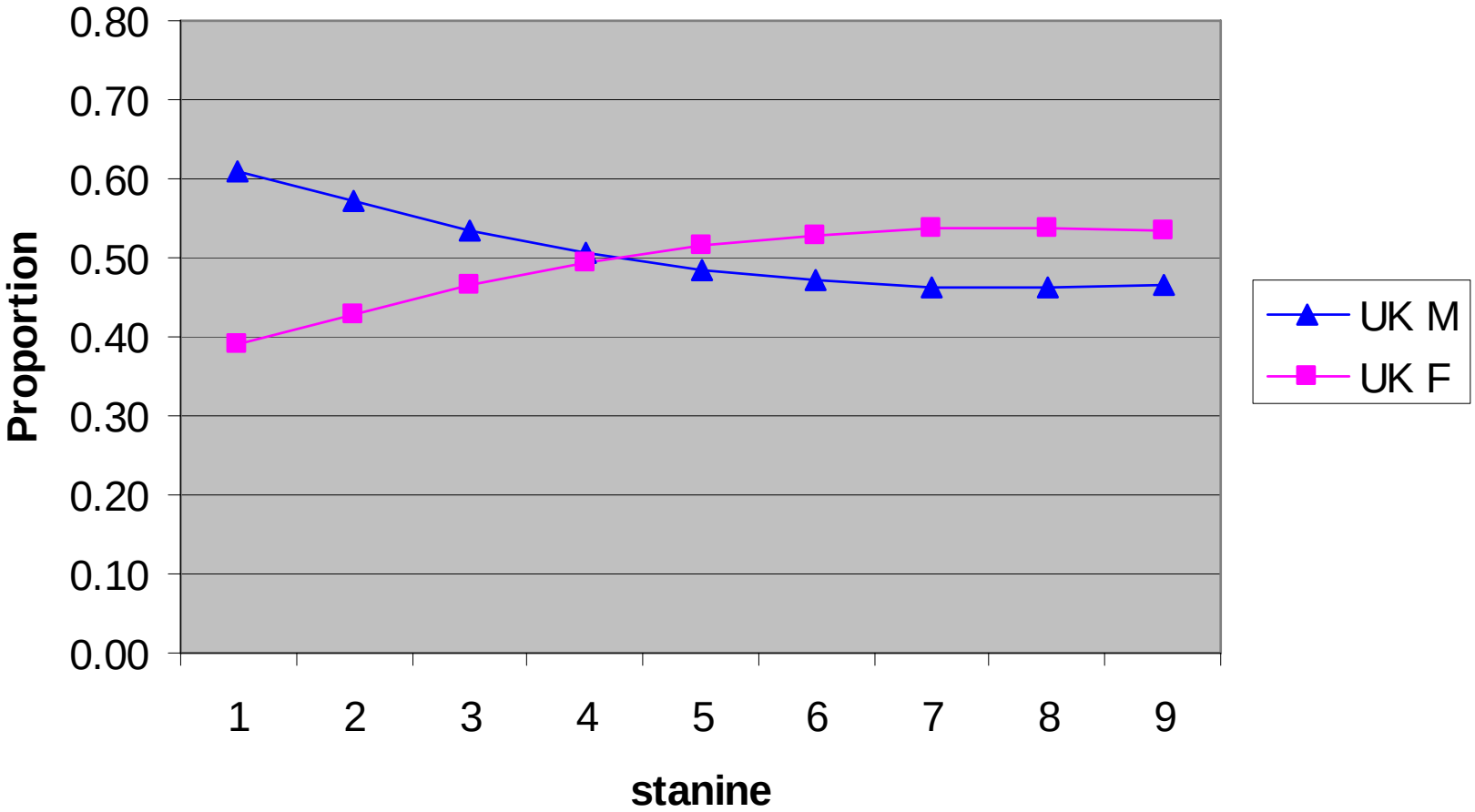
Cognitive Abilities Test

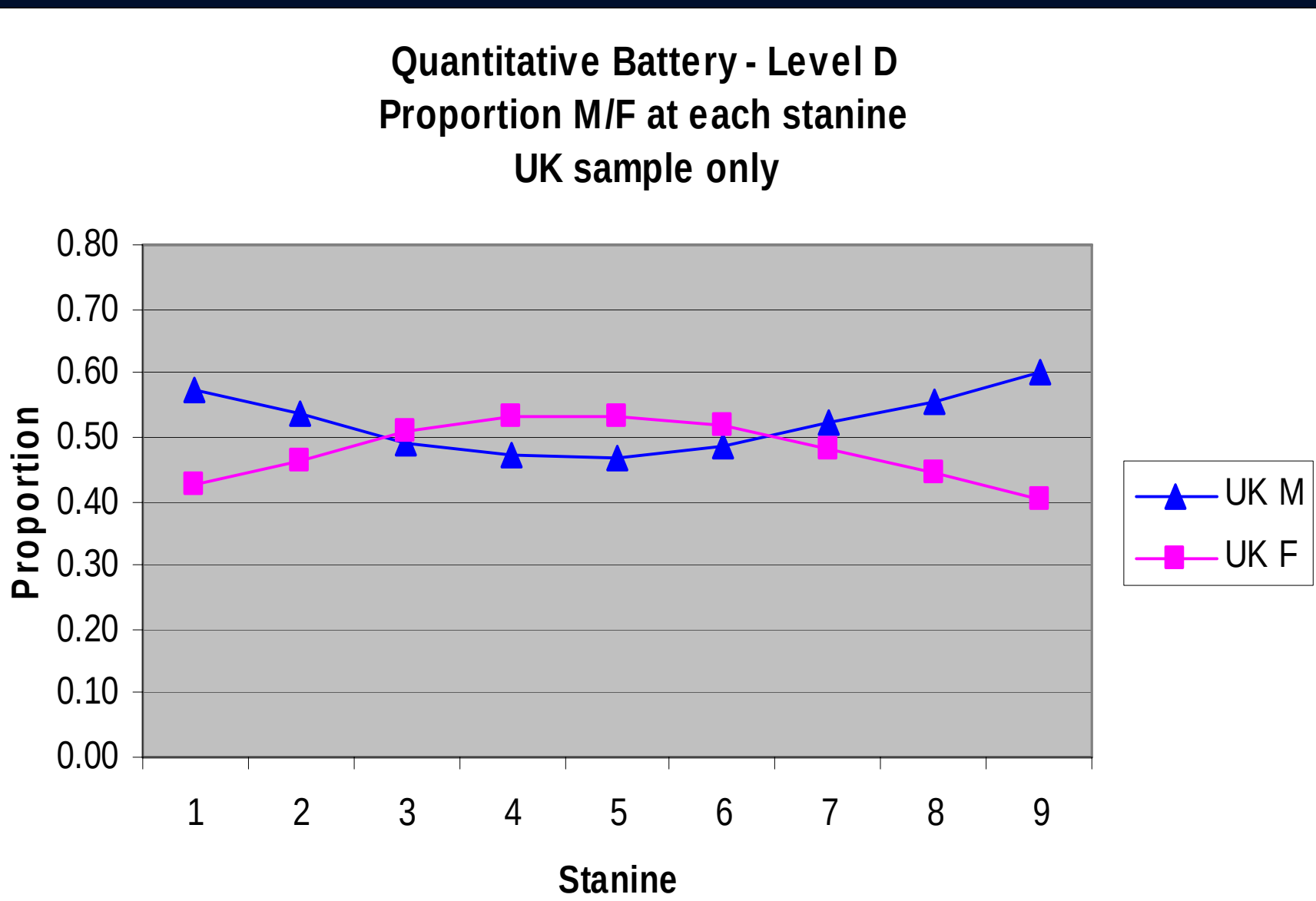
Verbal	Quantitative	Nonverbal
Verbal Classification Verbal Analogies Sentence Completion	Number Series Quantitative Relations Equation Building	Figure Analogies Figure Classification Figure Analysis

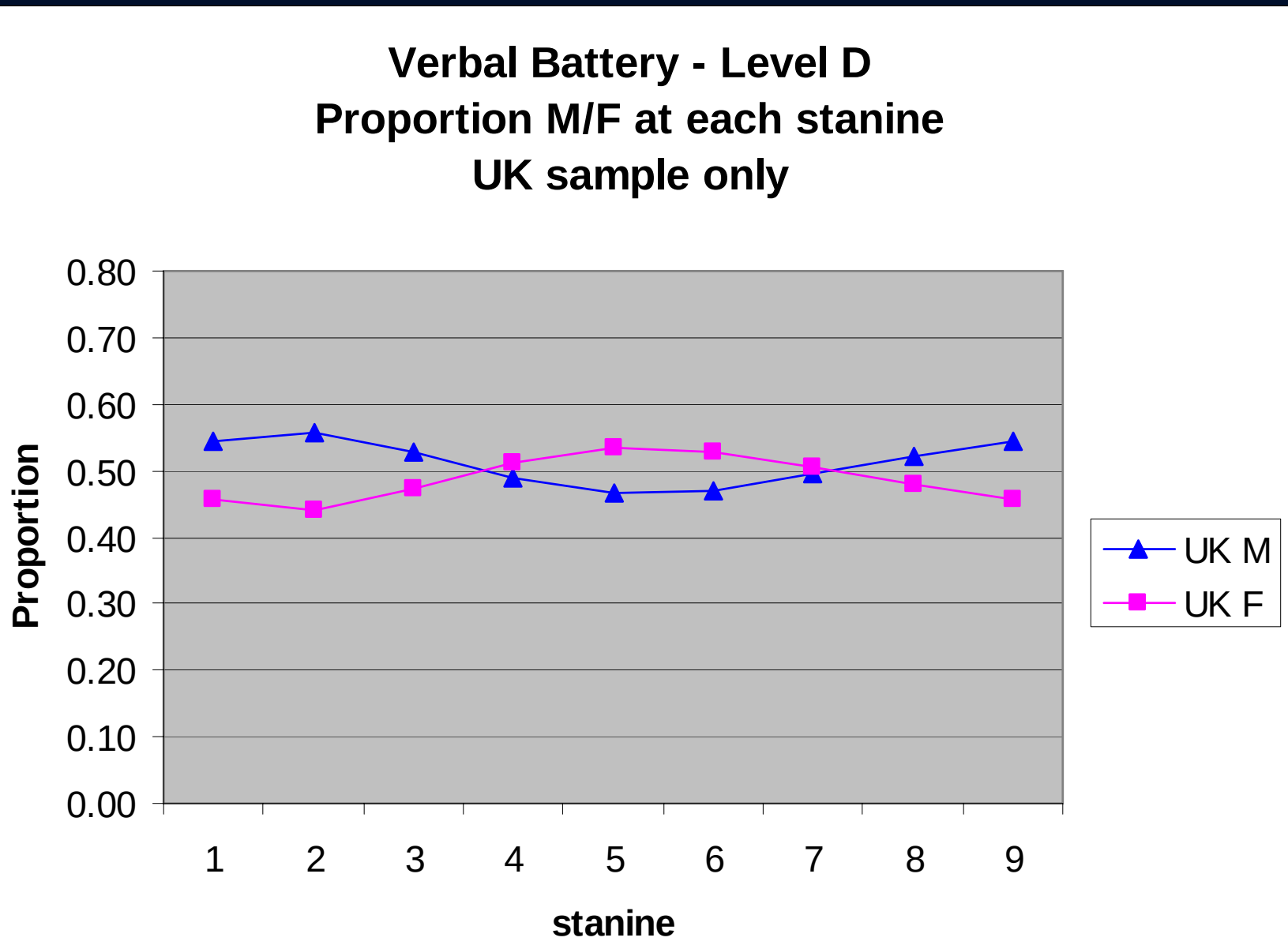
Number Analogies

3 Separate Test Batteries (not one)

Verbal Battery
Proportion M/F at each stanine
UK sample only




Quantitative Battery - Level D
Proportion M/F at each stanine
UK sample only
Proportion
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
2
3
4
5
6
7
8
9
Stanine
UK M
UK F


Verbal Battery - Level D
Proportion M/F at each stanine
UK sample only
Proportion
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
2
3
4
5
6
7
8
9
stanine
UK M
UK F

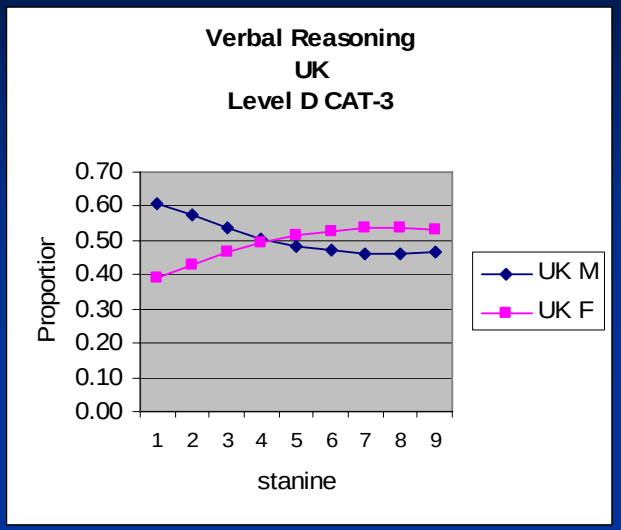
Do these results generalize?

- To 6th grade students in the US?
- Across grades/test levels in 2000 US sample?
  - Developmental changes
  - Impact of a few items on extreme scores
- Across cohorts (1984, 1992, 2000)?

U.S. Data

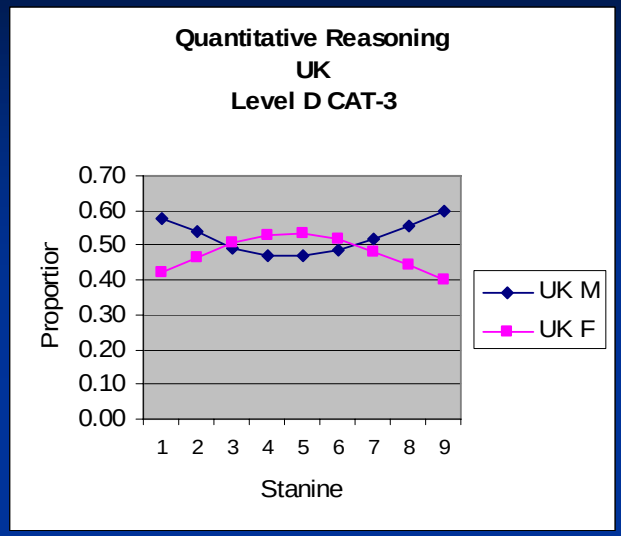
VERBAL

UK



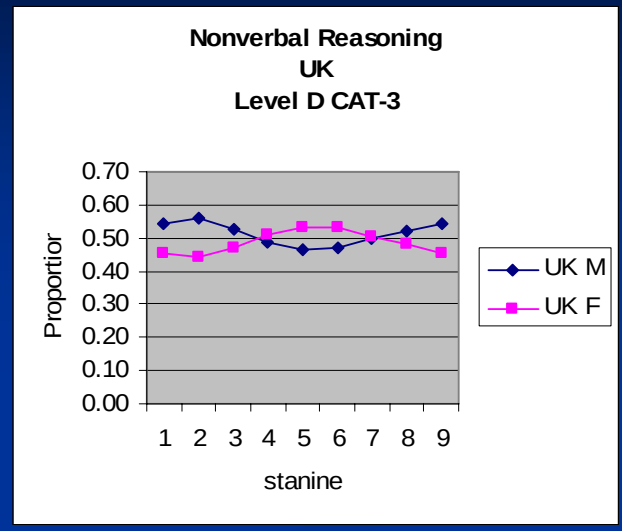
QUANTITATIVE

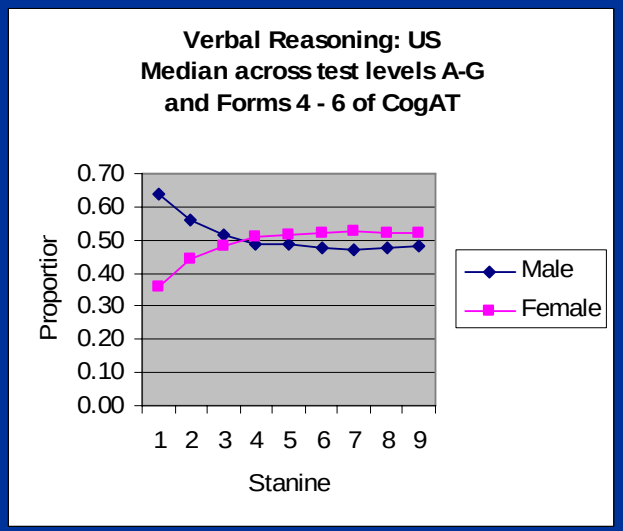
UK

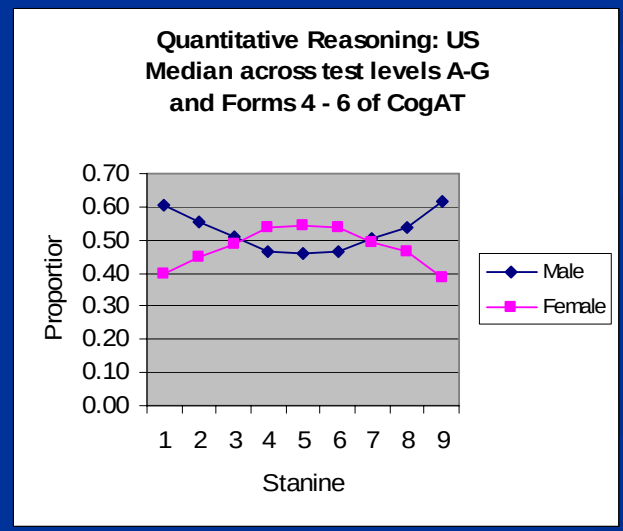


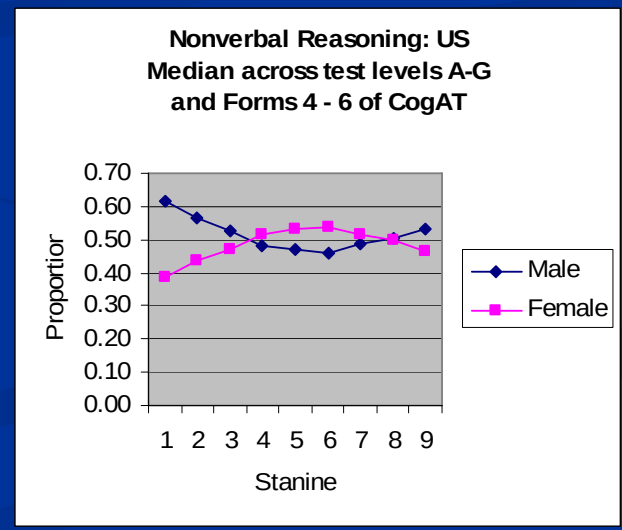
NONVERBAL

UK









US

US

US

Conclusions

- Do results of Strand et al. (2006) generalize?
  - UK to US
  - Across grades (3 – 11)
  - Across assessments/cohorts
- Advantages of not changing the measure in a fundamental way
  - Versus NAEP & TIMMS
- Implications

Talent identification requires:

- A focus on *potential* rather than on current *accomplishment*
- Control for *opportunity to learn*
- The *prediction* of an unusual level future competence from present and past *behavior*
- Assumptions about the *stability of individual differences* in estimates of that competence

Correlations Between IQ Scores across grades (Lohman & Korb, 2006)

Grade	3	4	5	6	7
4	86				
5	83	86			
6	80	83	86		
7	75	80	83	86	
8	70	75	80	83	86

Correlations Between IQ Scores across grades (Lohman & K0rb, 2006)

Grade	3	4	5	6	7
4	86				
5	83	86			
6	80	83	86		
7	75	80	83	86	
8	70	75	80	83	86

Proportion of students identified by one test also identified by the second test

	Correlation between tests				
Cut score	0.50	0.60	0.70	0.80	0.90
Top 1%	0.13	0.19	0.27	0.38	0.54
Top 2%	0.17	0.23	0.31	0.42	0.58
Top 3%	0.20	0.26	0.35	0.45	0.60

Individual differences will be stable to the extent that

- Individuals do not acquire new knowledge and skills at different rates.
- The opportunities that individuals have had to develop the measured competencies do not change
- The aspects of competence that are assessed also do not change.

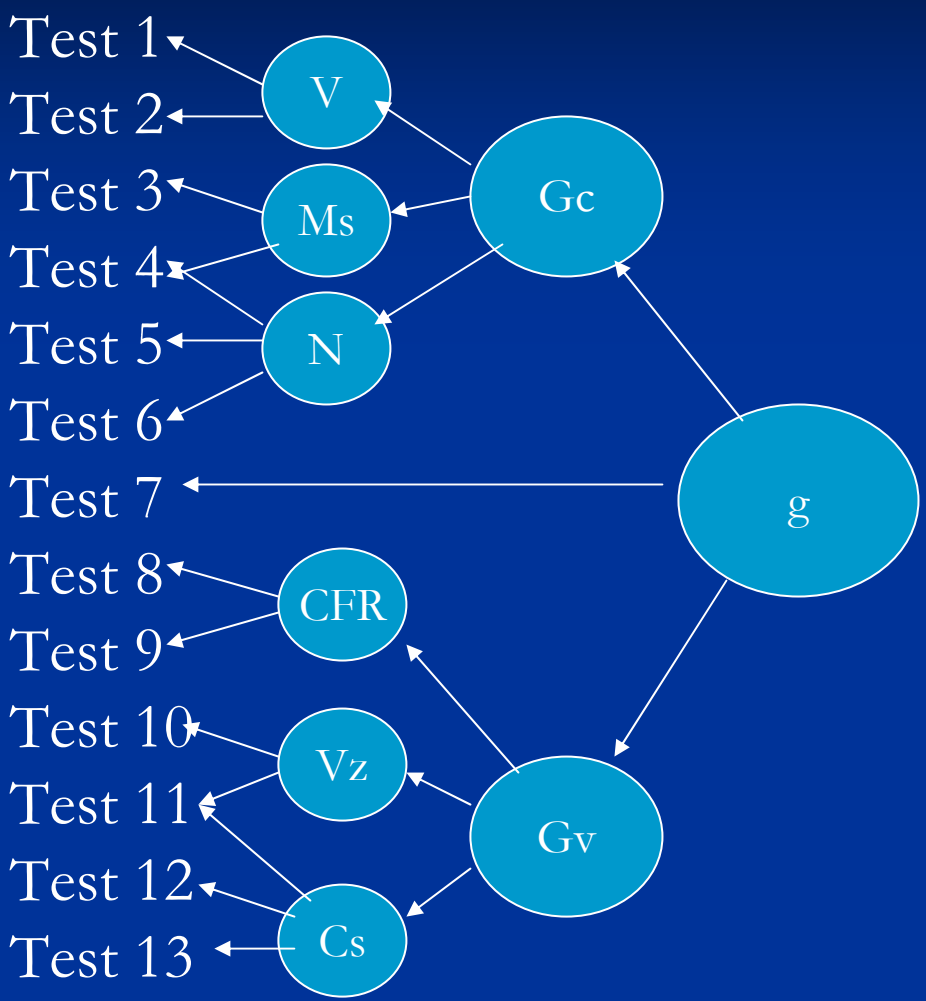
- Implications?
  - Assess cognitive development early and often
  - Revolving door

Beyond a single rank order

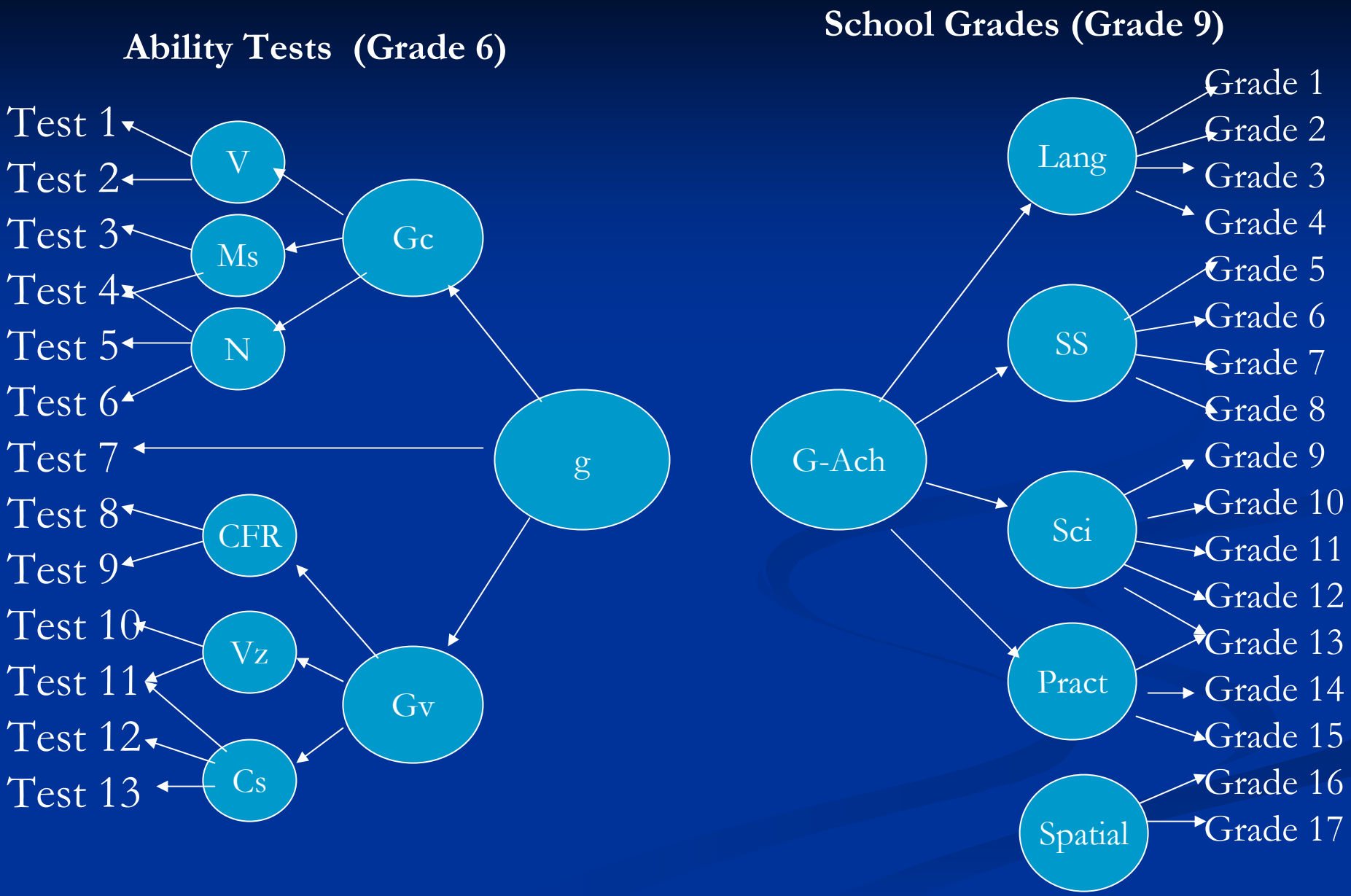
- Brunswikian symmetry – picking the right predictors for the criteria of interest

N = 866

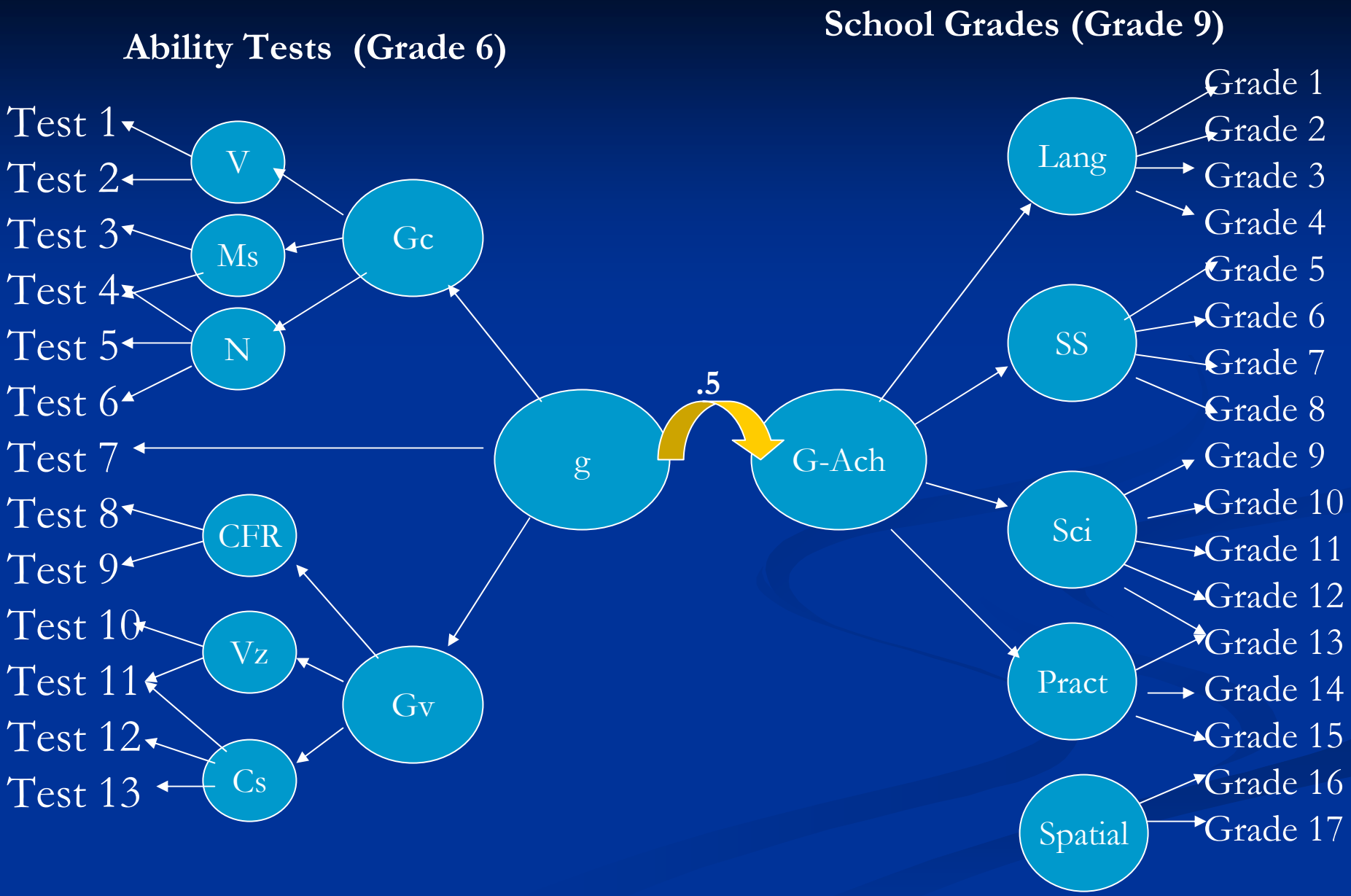
Ability Tests (Grade 6)



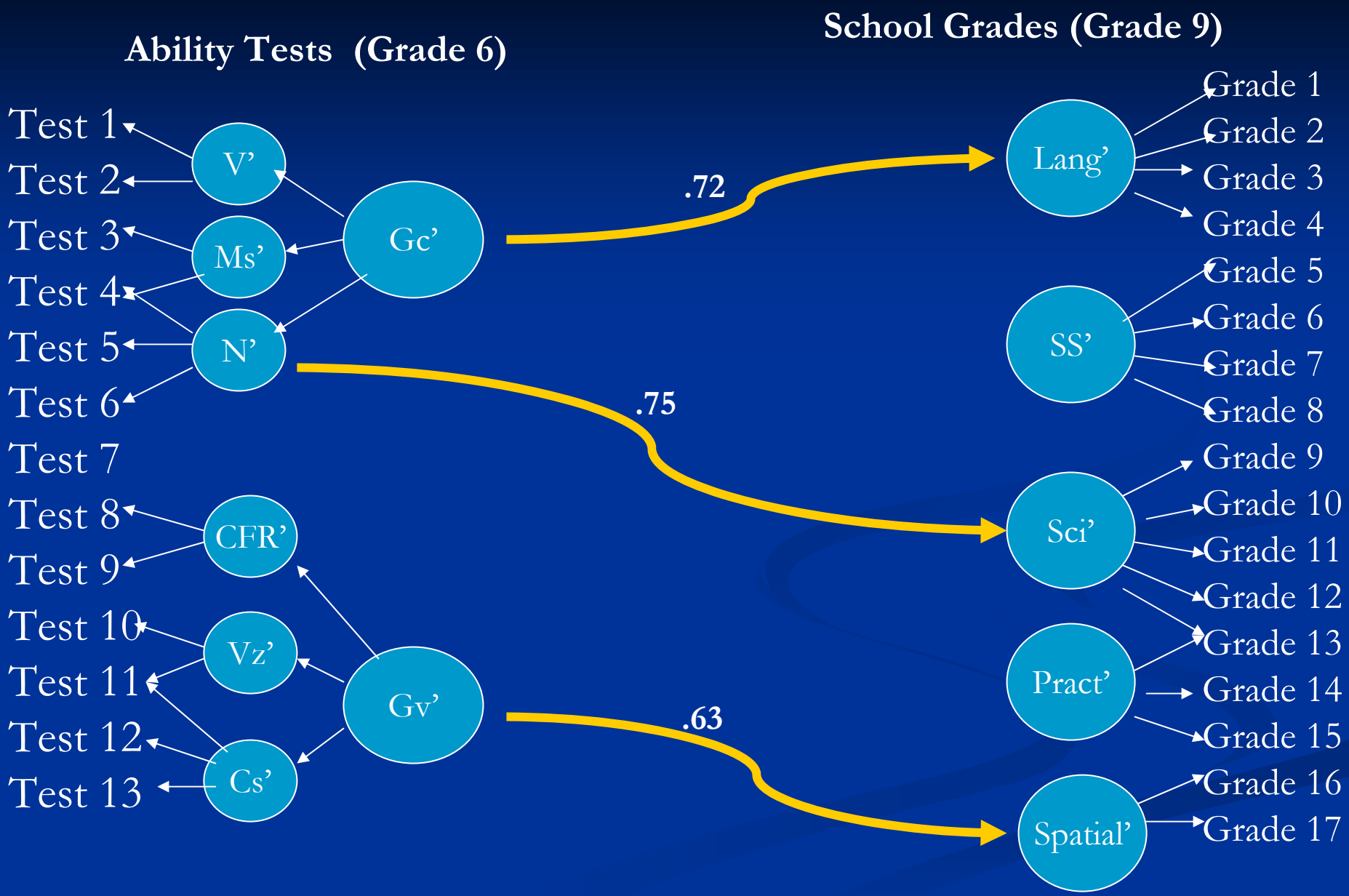
Gustafsson & Balke, MBR, 1993



Gustafsson & Balke, MBR, 1993



Gustafsson & Balke, MBR, 1993



Gustafsson & Balke, MBR, 1993

- g mattered most when predicting GPA
- Specific abilities mattered even more when predicting success in many domains
- Obtaining useful & reliable score profiles

Score Profiles on CogAT

Profiles for the most (and least) able

Lohman, D. F., Gambrell, J., & Lakin, J. (in press). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychological Science*

CogAT6 Profile frequencies for students in K-12 pop. and for students with two stanine scores of 9

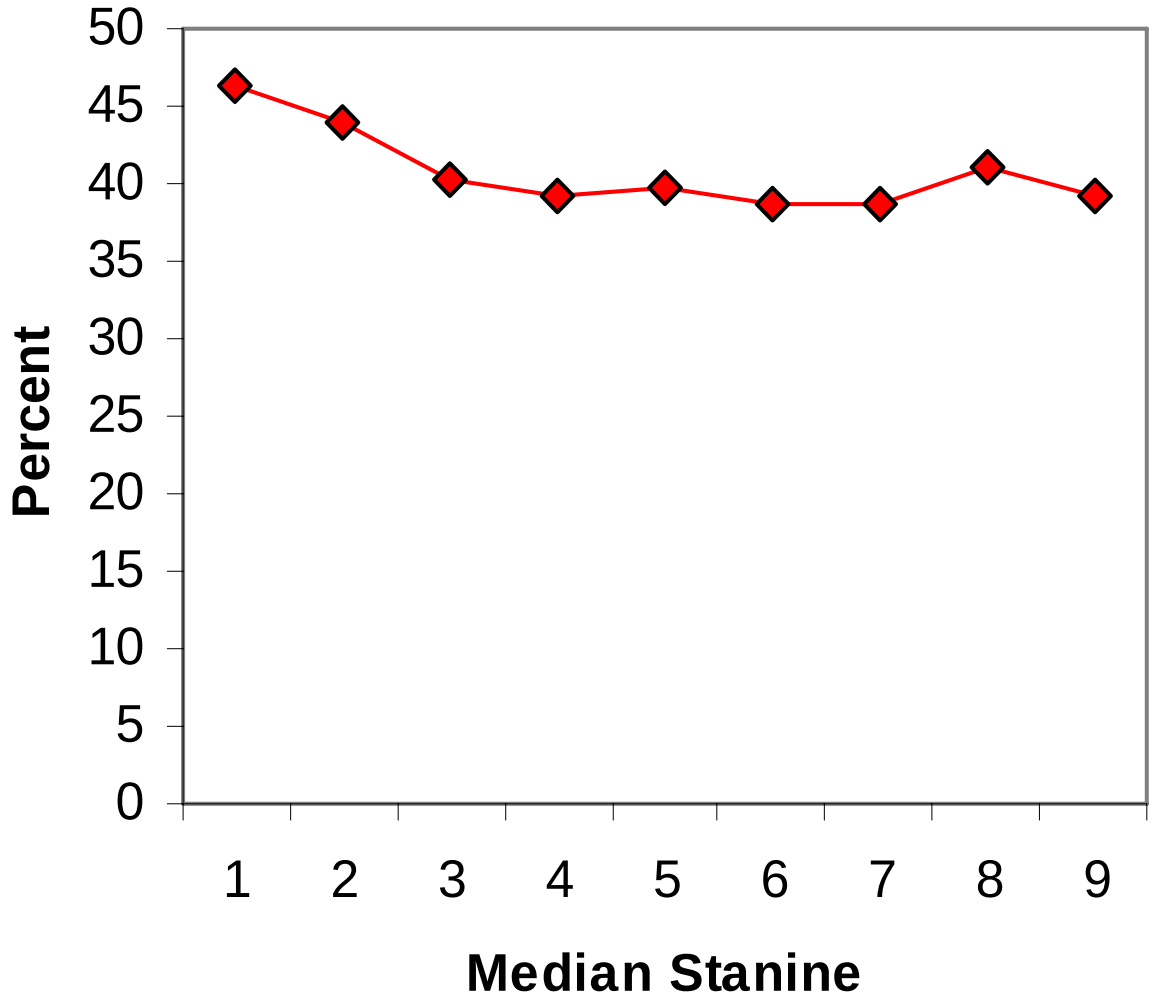
Profile	Percent in K-12 population	
Flat	**33**	
Significant (10 - 23)		
Strength	**21**	
Weakness	**22**	
Extreme (24 +)		
Strength	**4**	
Weakness	**3**	

CogAT6 Profile frequencies for students in K-12 pop. and for students with two stanine scores of 9

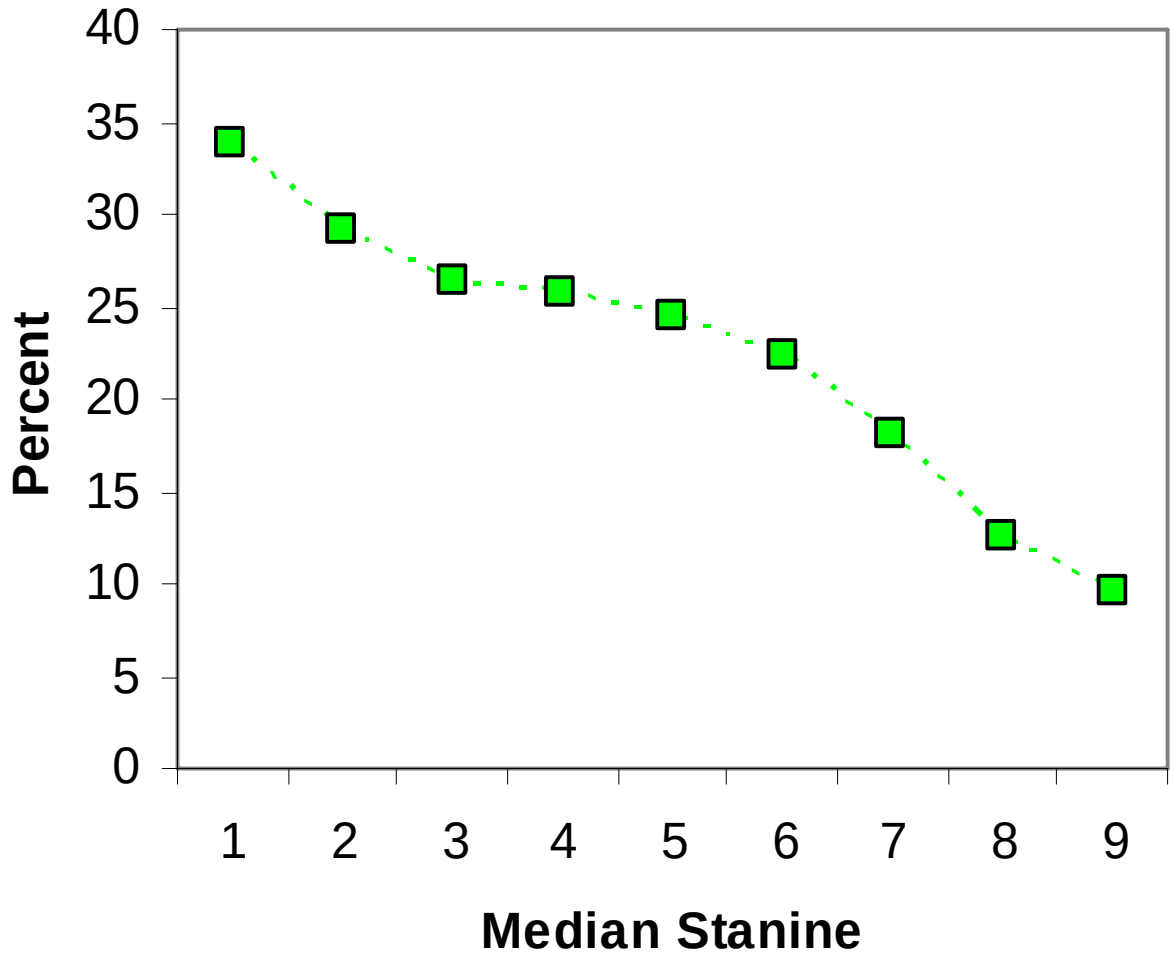
Profile	Percent in K-12 population	Percent in Stanine=9 group
Flat	33	37
Significant		
Strength	21	6
Weakness	22	21
Extreme		
Strength	4	3
Weakness	3	16

37%

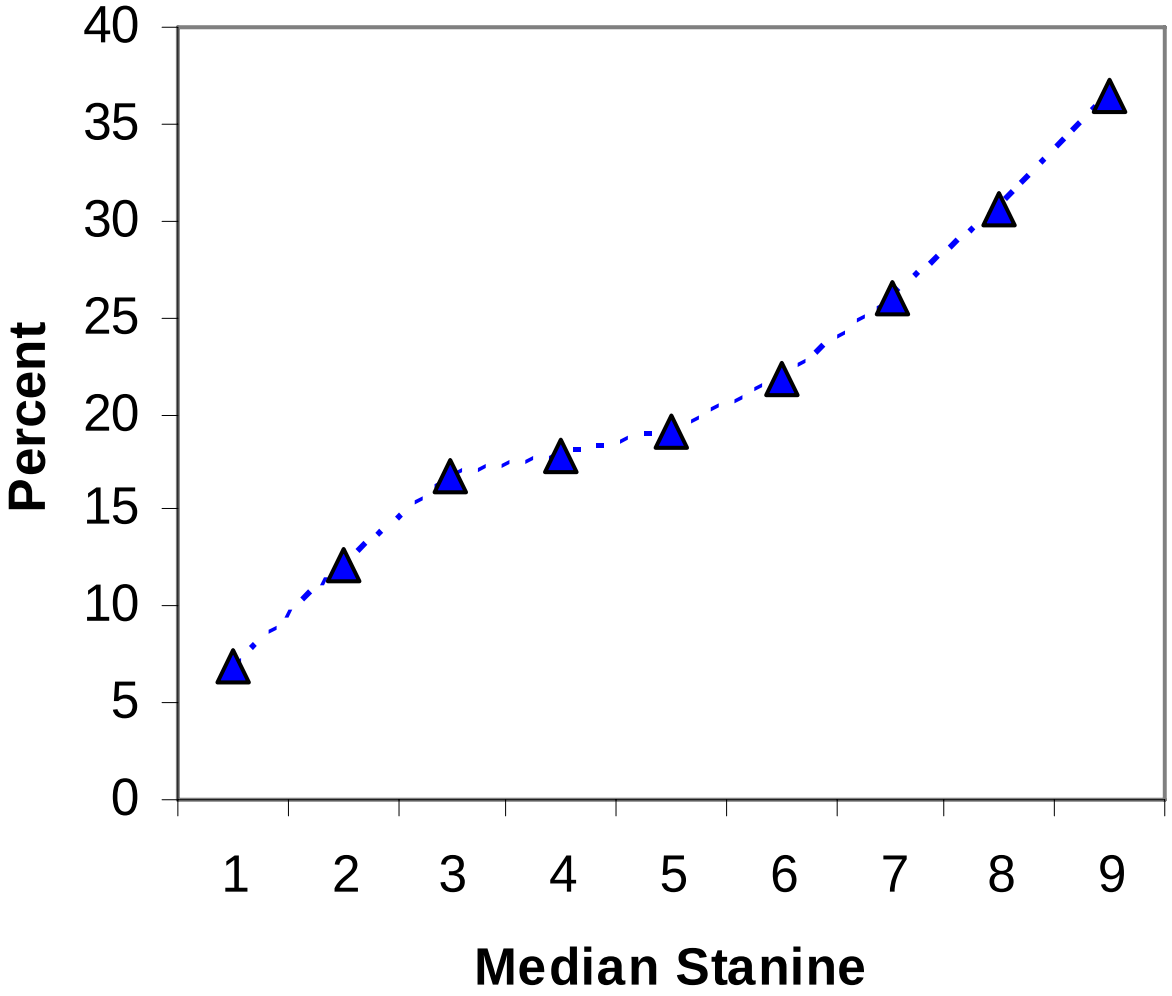
Frequency of Flat Profiles



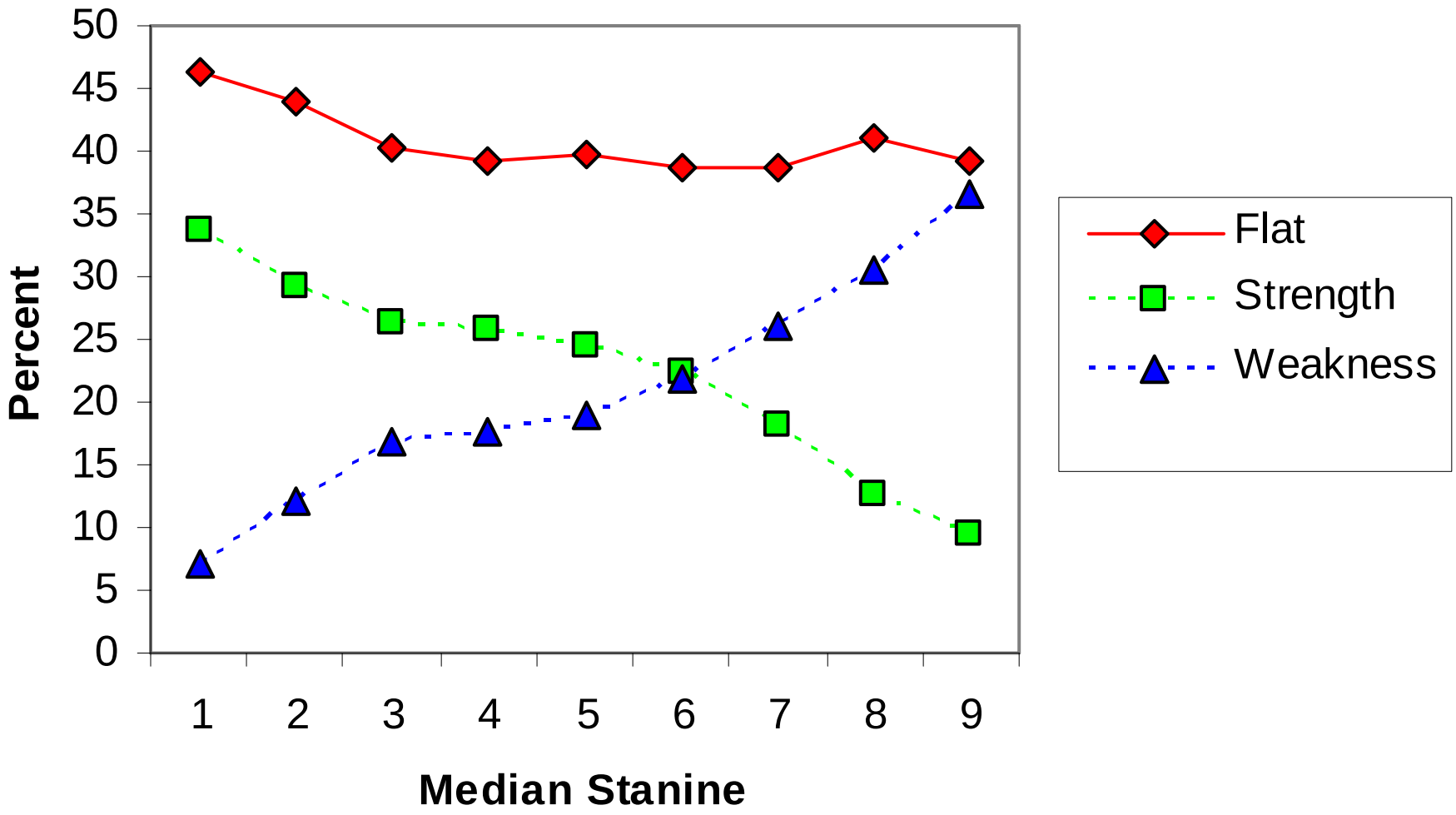
Profiles with a significant or extreme STRENGTH



Profiles with a significant or extreme WEAKNESS



Profile Frequency by Median Stanine



Lohman, D. F., Gambrell, J., & Lakin, J. (in press). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychological Science*

A brief introduction to an aptitude perspective on talent identification

Sources

- Foundation in IO psychology
  - Bingham (1937)
  - *Remaking the concept of Aptitude* (2000)
  - Lohman (2005, J. Ed. Gifted) An aptitude perspective
- Dawes theory of work adjustment (Lubinski)
  - Satisfactoriness versus Satisfaction
- Aptitude/trait complexes (Ackerman)
  - Constellation of traits that combine in a non-compensatory way
- Gagne's DMGT theory

Talent identification and development

- Goal is to identify those who exhibit talent for learning in particular domains
- And then to provide appropriately challenging instruction to develop that talent
- Talent as a constellation of aptitudes

Aptitude: What it is

- Aptitude is the degree of *readiness* to learn & perform well in a *particular situation* or *domain*.

Important Aptitudes for Academic Learning

- Cognition (knowing)
  - Domain knowledge & skill
  - Reasoning abilities in the symbol systems used to communicate knowledge (Verbal, Quant., Spatial)
- Affection (feeling)
  - anxiety, interests, working alone/with others
- Conation (willing)
  - persistence, impulsivity

Ways of measuring aptitude

1. Direct:
   - Learns in a few trials what takes others many trials to learn.
   - Opportunity to learn essential
2. Indirect:
   - Brings required (or helpful) cognitive, affective, & conative resources to the situation
   - Opportunity to learn essential

Opportunity to Learn

- On ability tests:
  - Approximate by **age** (years + months)
- On achievement tests
  - Approximate by grade (and week within grade)
- What about students who are poor, ELL, or have markedly different educational preparation?

Non-normative Experiences

Alternative assessments (common norms)

Non-normative Experiences

Advantages of nonverbal tests

- English Language Learners are disadvantaged on tests that use English
  - Allow one to use common norms for all
- Often good measures of g
- Appearance of measuring something more innate than tests that use other symbols
- Claims that they will solve the problem

Disadvantages of nonverbal tests

- Construct under-representation
- Predict academic learning less well than measures of verbal & quant reasoning
  - Specific variance negatively related to success in verbal domains
- Sensitive to environmental changes:
  - Large Flynn, practice effects, and schooling effects
- (P.S. Not tests of spatial ability)

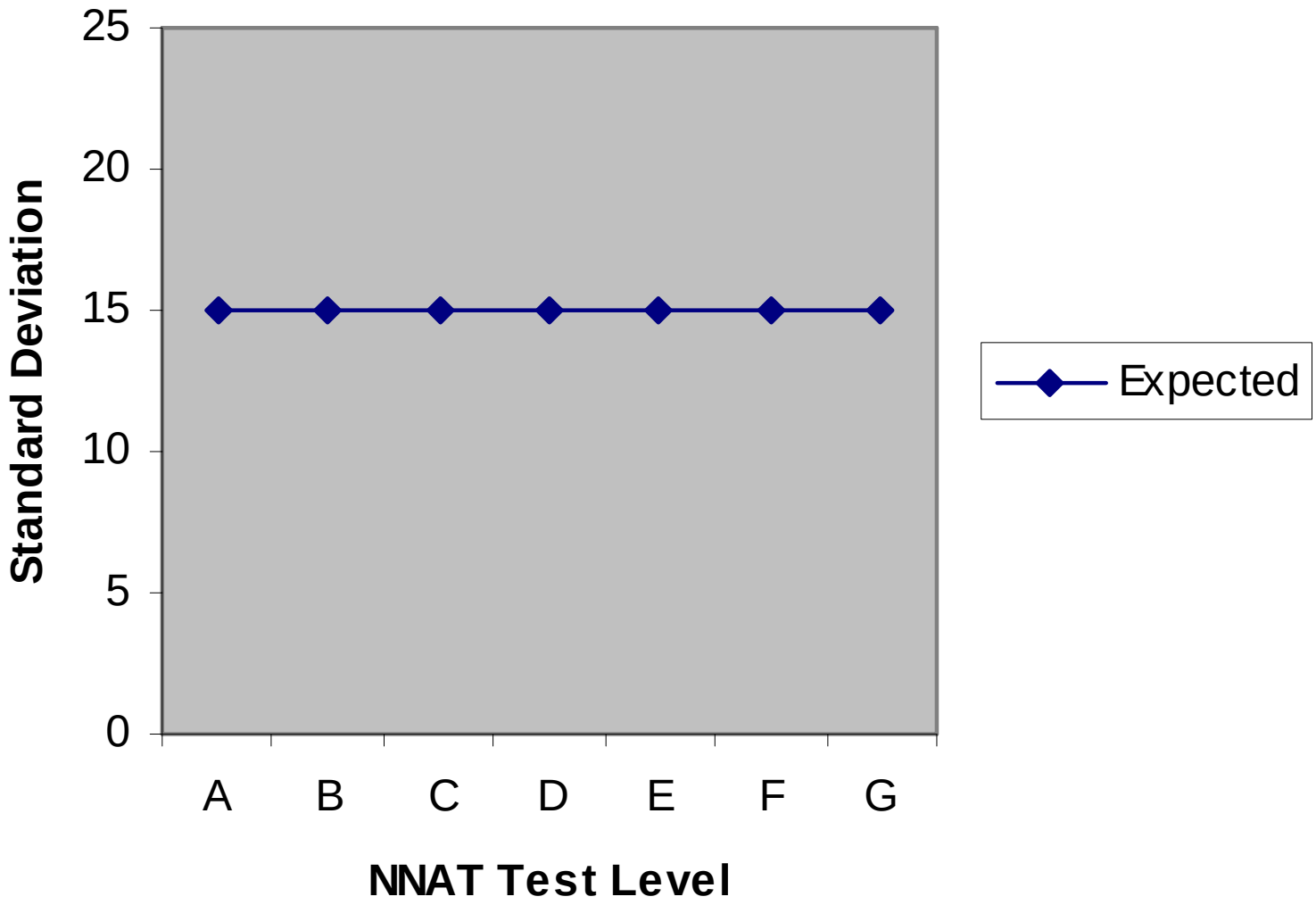
Level the playing field?

- 371 ELL & 332 Non-ELL K-6 children
  - All Hispanic
  - All free/reduced lunch
- ELL – Non-ELL score differences
  - CogAT Nonverbal 7.5 (.47 SD)
  - Raven SPM 7.3 (.46 SD)
  - NNAT 10.1 (.63 SD)

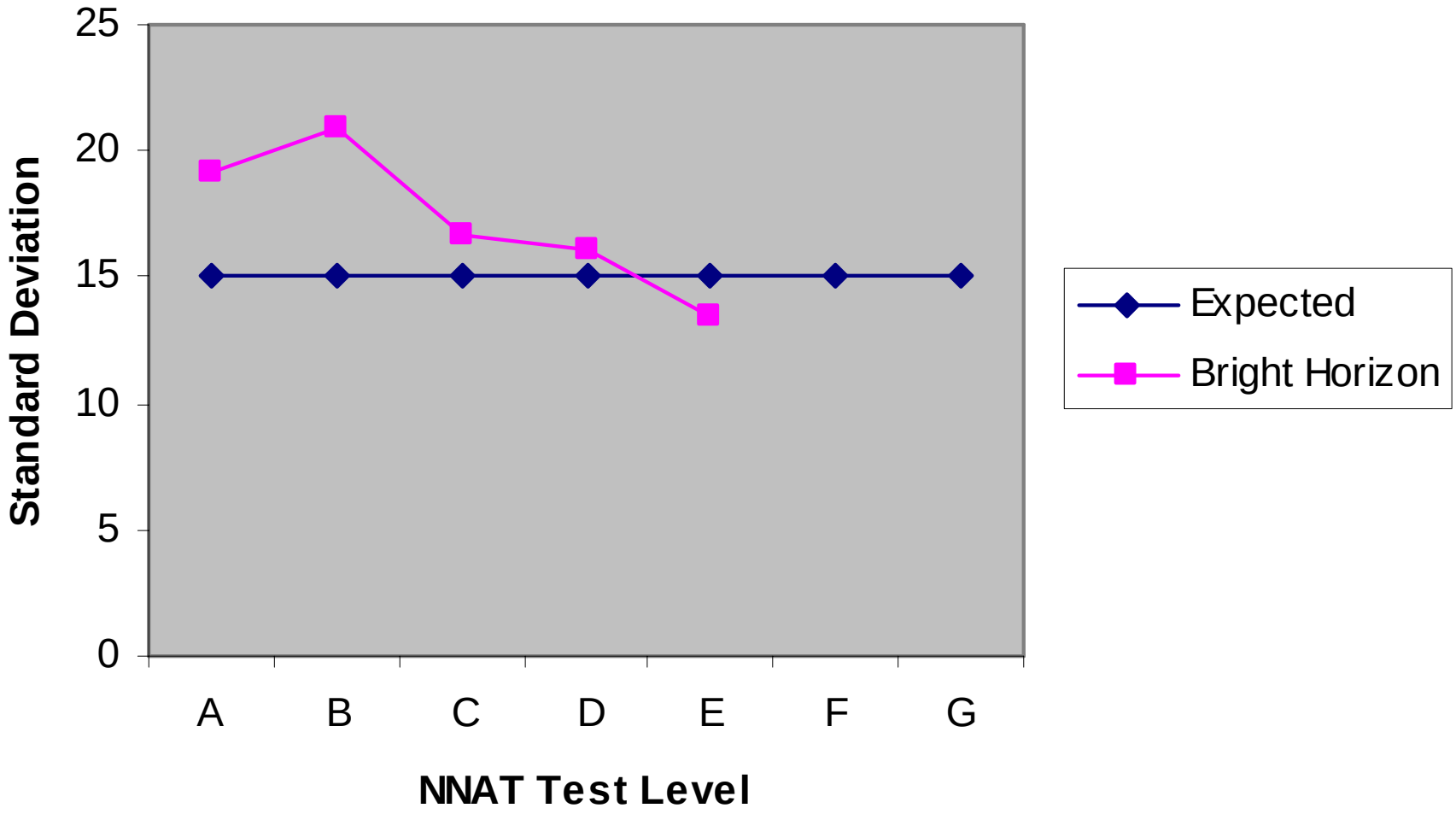
Trusting the Norms

- Judgments of exceptionality depend critically on the appropriateness & quality of the norms
- Raven Norms – about 10 points too high
  - 100.7 versus 111.5 (n = 733)
- Cattell Culture Fair – about 17 points too high
- CogAT Quantitative Battery
  - Shift in educational practice
- NNAT?

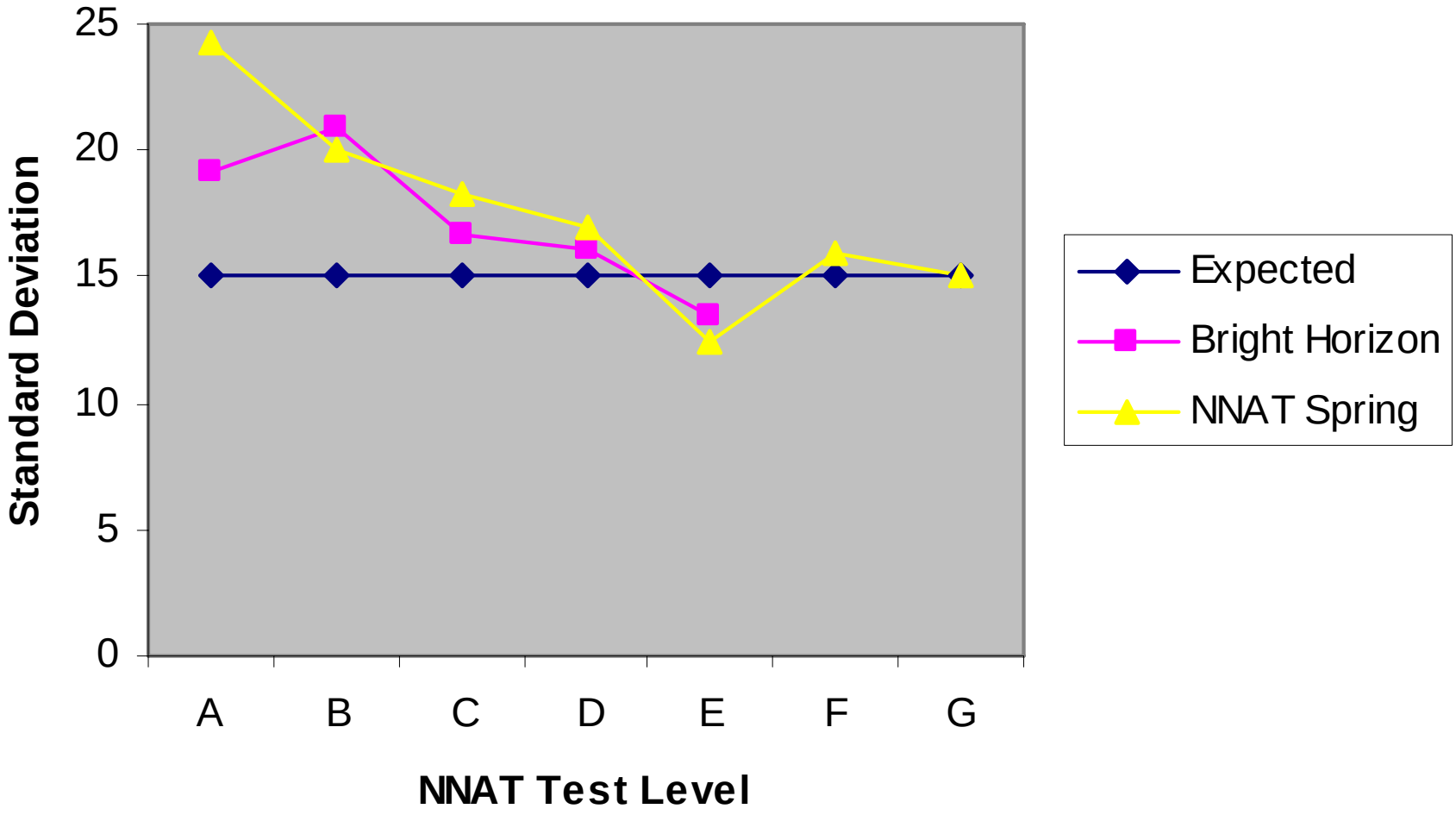
SD's of Nonverbal Ability Index (NAI) scores on NNAT by Test Level



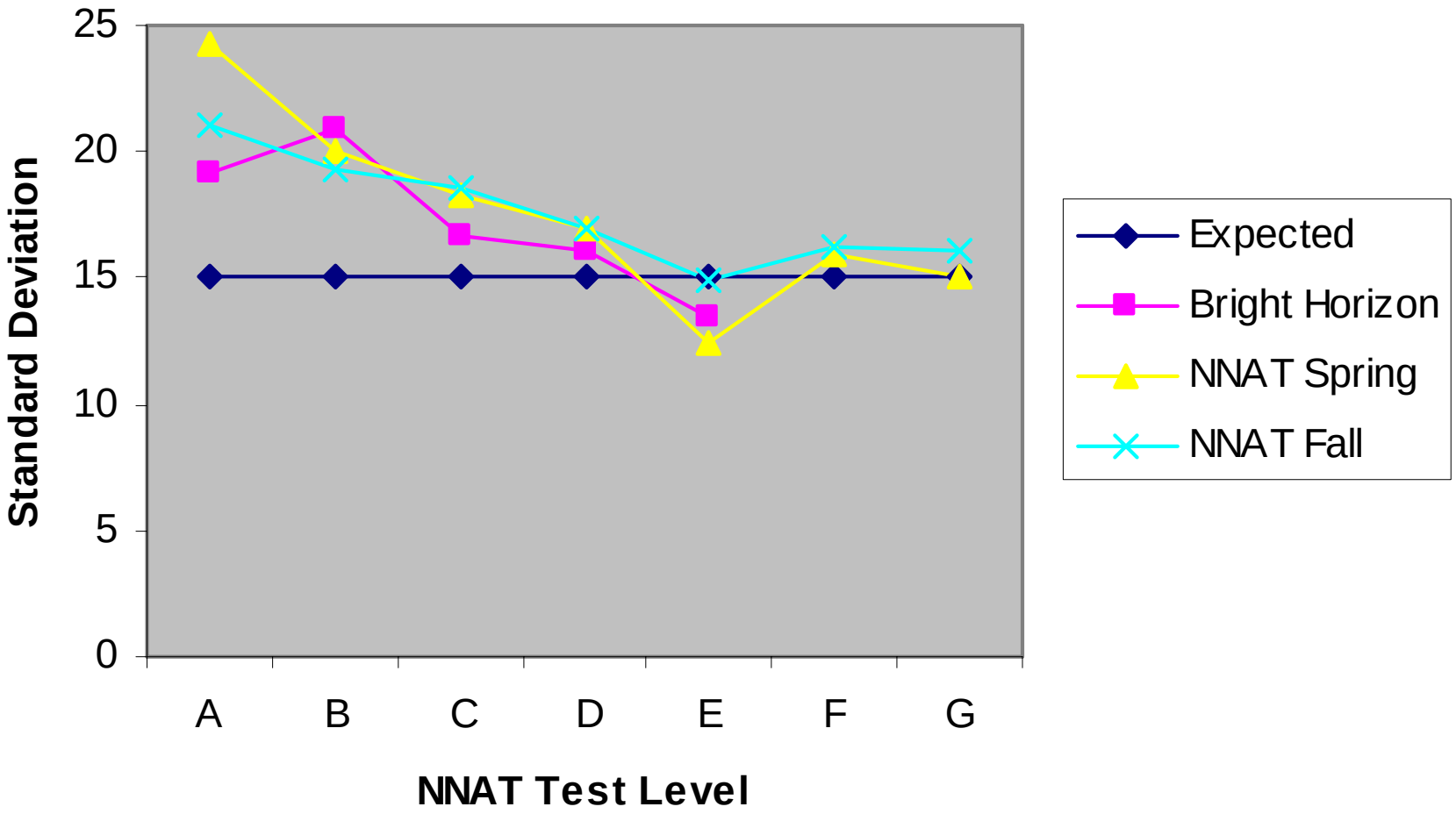
SD's of Nonverbal Ability Index (NAI) scores on NNAT by Test Level



SD's of Nonverbal Ability Index (NAI) scores on NNAT by Test Level



SD's of Nonverbal Ability Index (NAI) scores on NNAT by Test Level



Over-identification Rates for NNAT

	True NAI Score		
Level	**115**	**130**	**145**
A	1.5	3.4	11.9
B	1.4	2.6	7.3
C	1.3	2.3	5.8
D	1.2	1.7	2.9
E	1.0	1.0	1.0
F	1.1	1.4	2.0
G	1.1	1.4	1.9

Over-identification Rates for NNAT

	True NAI Score		
Level	**115**	**130**	**145**
A	1.5	3.4	11.9
B	1.4	2.6	7.3
C	1.3	2.3	5.8
D	1.2	1.7	2.9
E	**1.0**	**1.0**	**1.0**
F	1.1	1.4	2.0
G	1.1	1.4	1.9

Over-identification Rates for NNAT

	True NAI Score		
Level	**115**	**130**	**145**
A	1.5	3.4	11.9
B	1.4	2.6	7.3
C	1.3	2.3	5.8
D	1.2	1.7	2.9
E	1.0	1.0	1.0
F	1.1	1.4	2.0
G	1.1	1.4	1.9

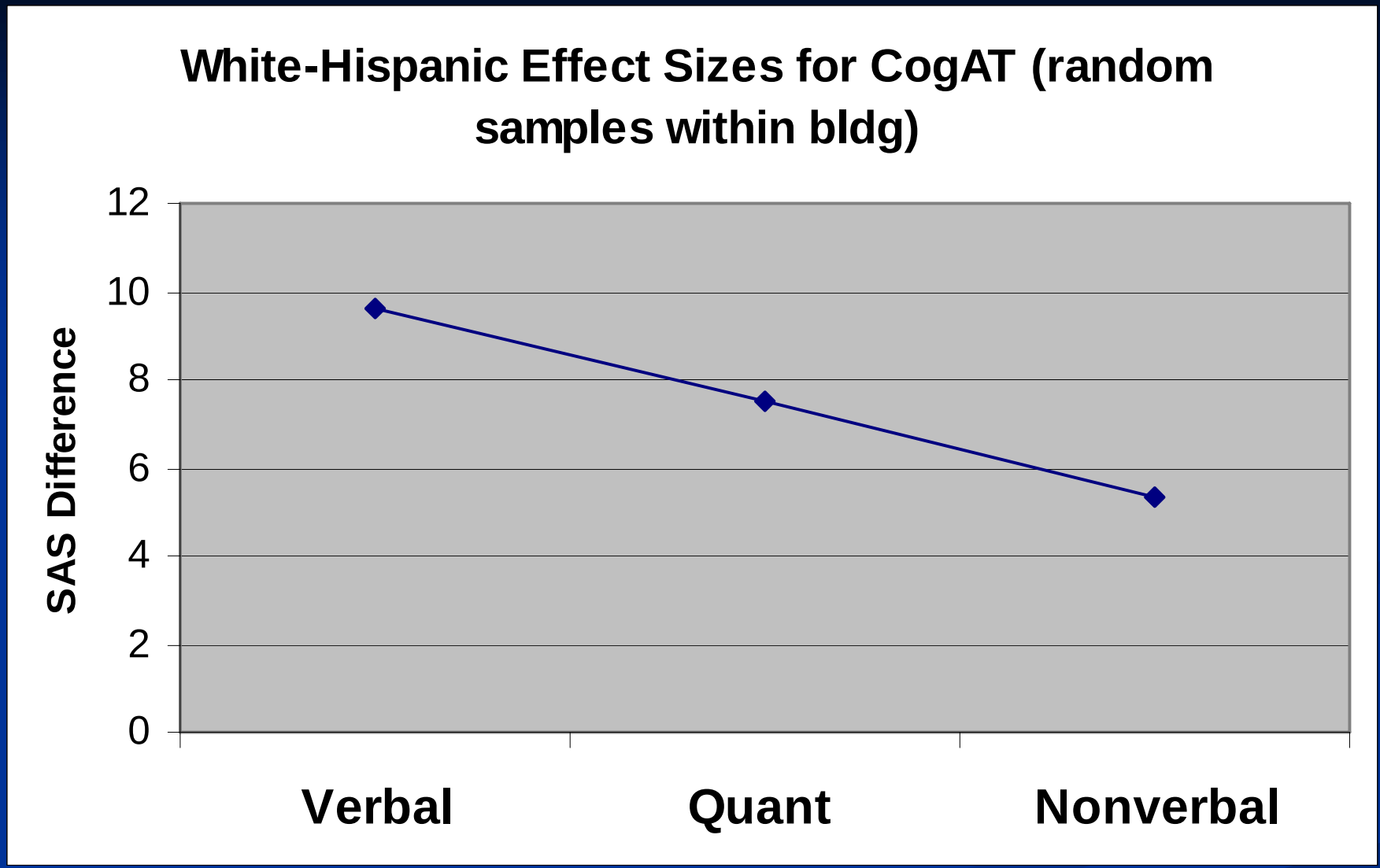
- Lohman, D. F., Korb, K., & Lakin, J. (in press). Identifying academically gifted English language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly*.
- *Cognitively Speaking*, (**6**, Winter 2008) "Comparing CogAT, NNAT, and Raven"

Caveat emptor

- Critically examine your data
- If possible, always use local norms as well

Means versus correlations

On the whole, academic success for ELL and minority children requires the same knowledge, skills, abilities, interests, motivation, perseverance as their non-minority peers



Nonverbal the fairest test?

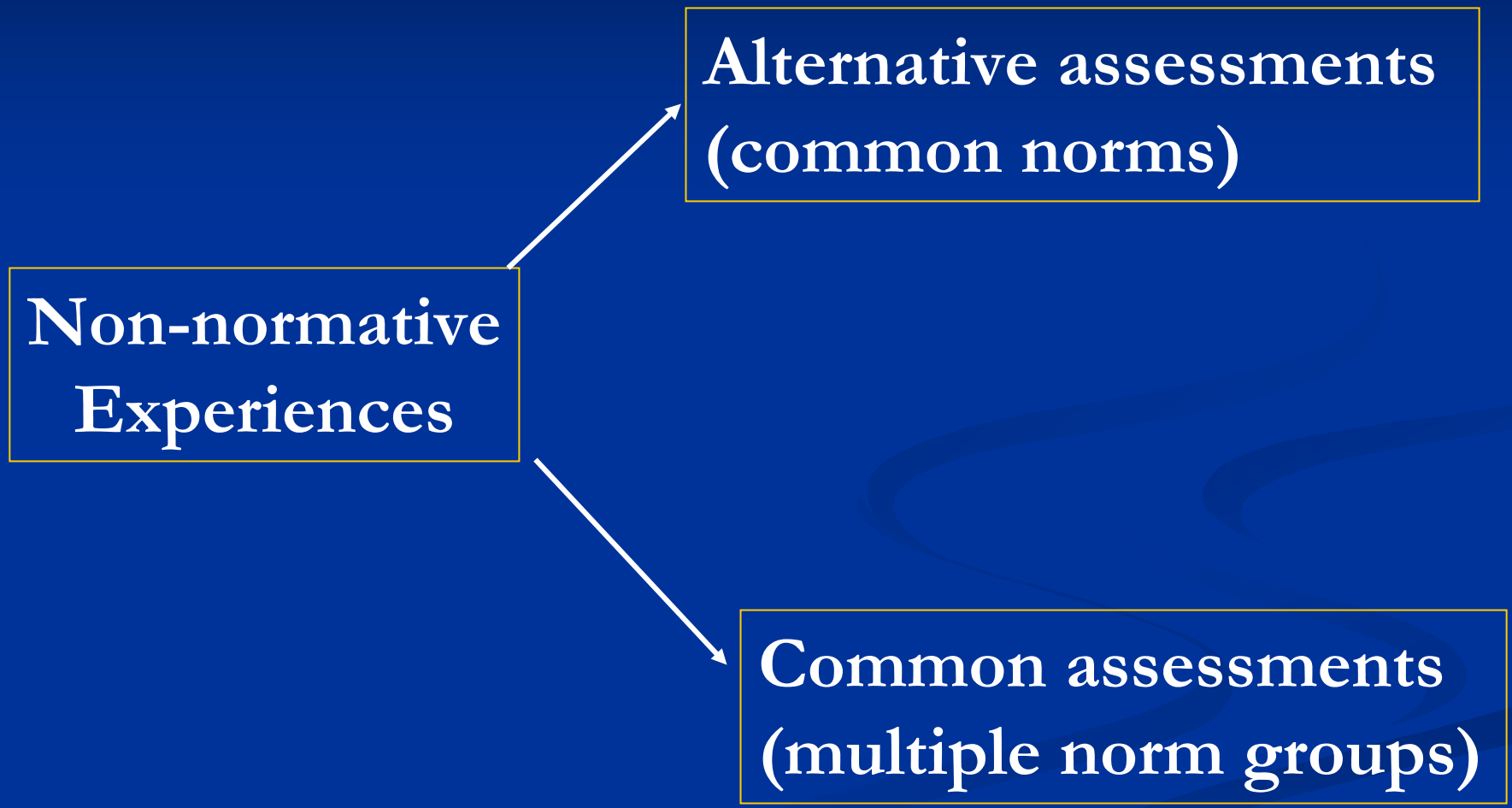
Project Bright Horizon Data

ELL Students Grade **3** (N = 74 - 81)

	CogAT V	CogAT Q	CogAT N	NNAT	Raven
Dev. Reading Assess.	0.63	0.43	0.24	0.11	0.32
Reading - Terra Nova	0.56	0.36	0.30	0.33	0.41
Reading - AIMS	0.73	0.45	0.34	0.39	0.43
Math - Terra Nova	0.57	0.61	0.51	0.42	0.43
Math - AIMS	0.61	0.70	0.57	0.50	0.47

Is the problem the test?

Or the norm group?


Alternative assessments
(common norms)
Non-normative
Experiences
Common assessments
(multiple norm groups)

Multiple Perspectives

- For ELL students in grade 3, compare scores to:
  - Other grade 3 students in the nation (common norms)
  - Other students in grade 3 in the district/school
  - Other ELL students in grade 3 in the district
- Not a single statistical adjustment of scores (Mercer)
- Not an assertion that all are ready for the same level of instruction

Tradeoff

Measuring the right things approximately for ELL students

or

the wrong things with greater accuracy

Improvements in Measurement for ELL children

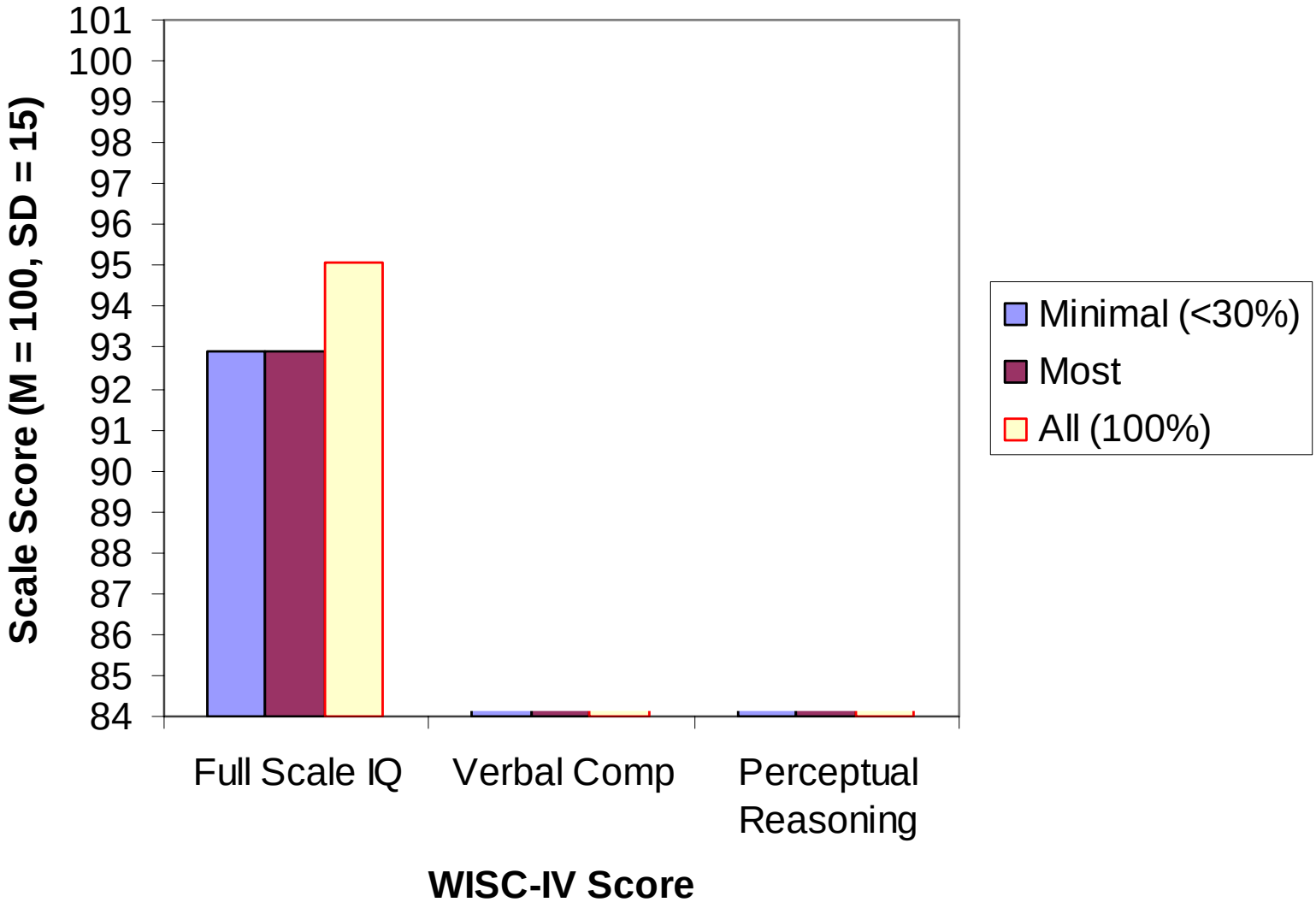
1. One test for all (e.g., early Binet)
2. "Nonverbal" tests (Army Beta)

   Reduced difference between mean scores of ELL and native speakers of English
3. Translated/adapted English-language tests (Woodcock-Munoz)
4. Translated tests with contextualized norms (WISC-IV Spanish)
5. Simultaneous development in (both) languages with contextualized norms (Current work)

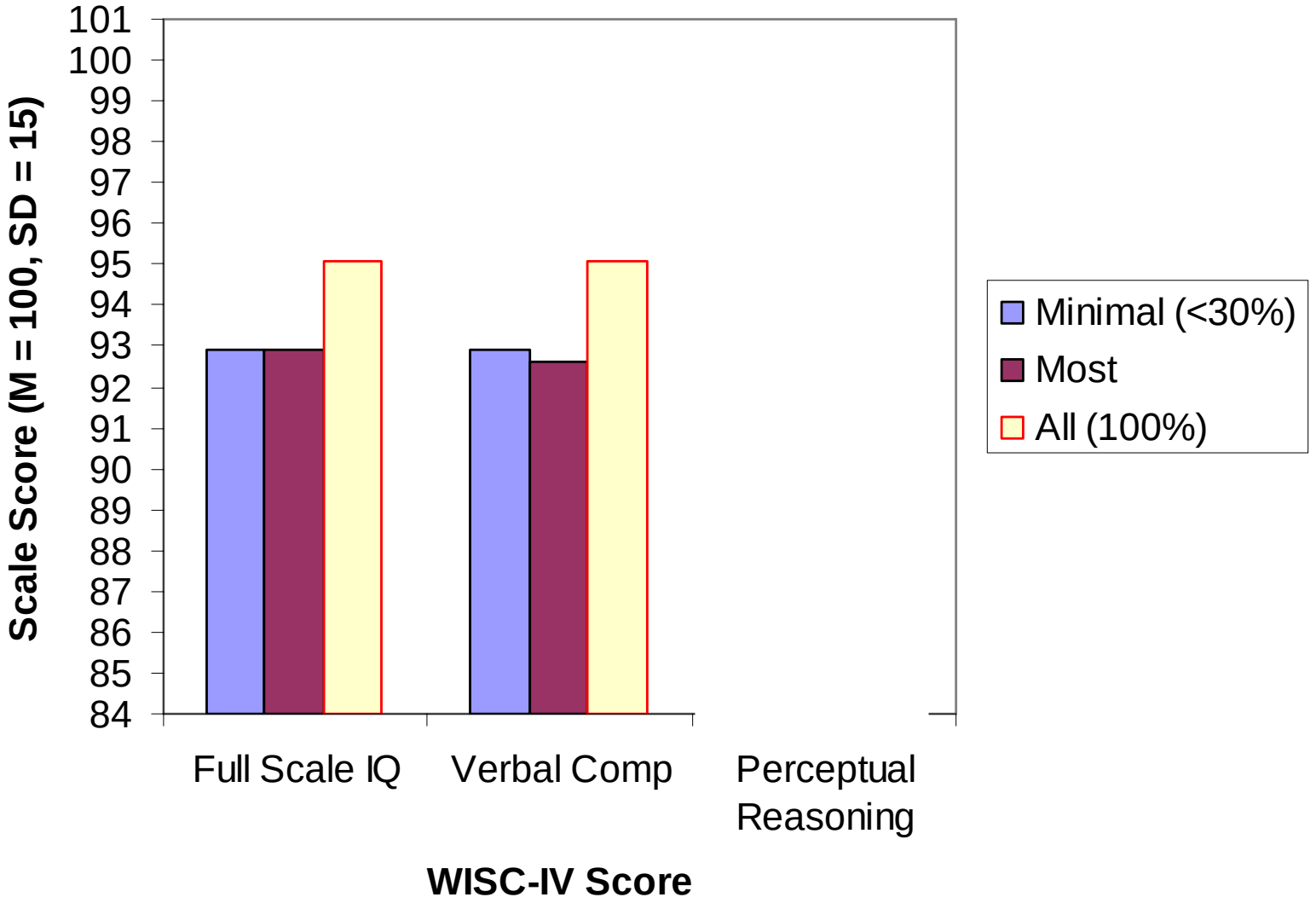
WISC-IV *Spanish*

- Target population: *Bilingual Spanish-speaking children in the U.S.* with no more than 5 years of U.S. schooling
- Translated & adapted the verbal tests
- Calibrated the Spanish Verbal Scale with the normative (English) Verbal Scale
- Indexed opportunity to learn by
  - % child's education in U.S.
  - Parental educational level

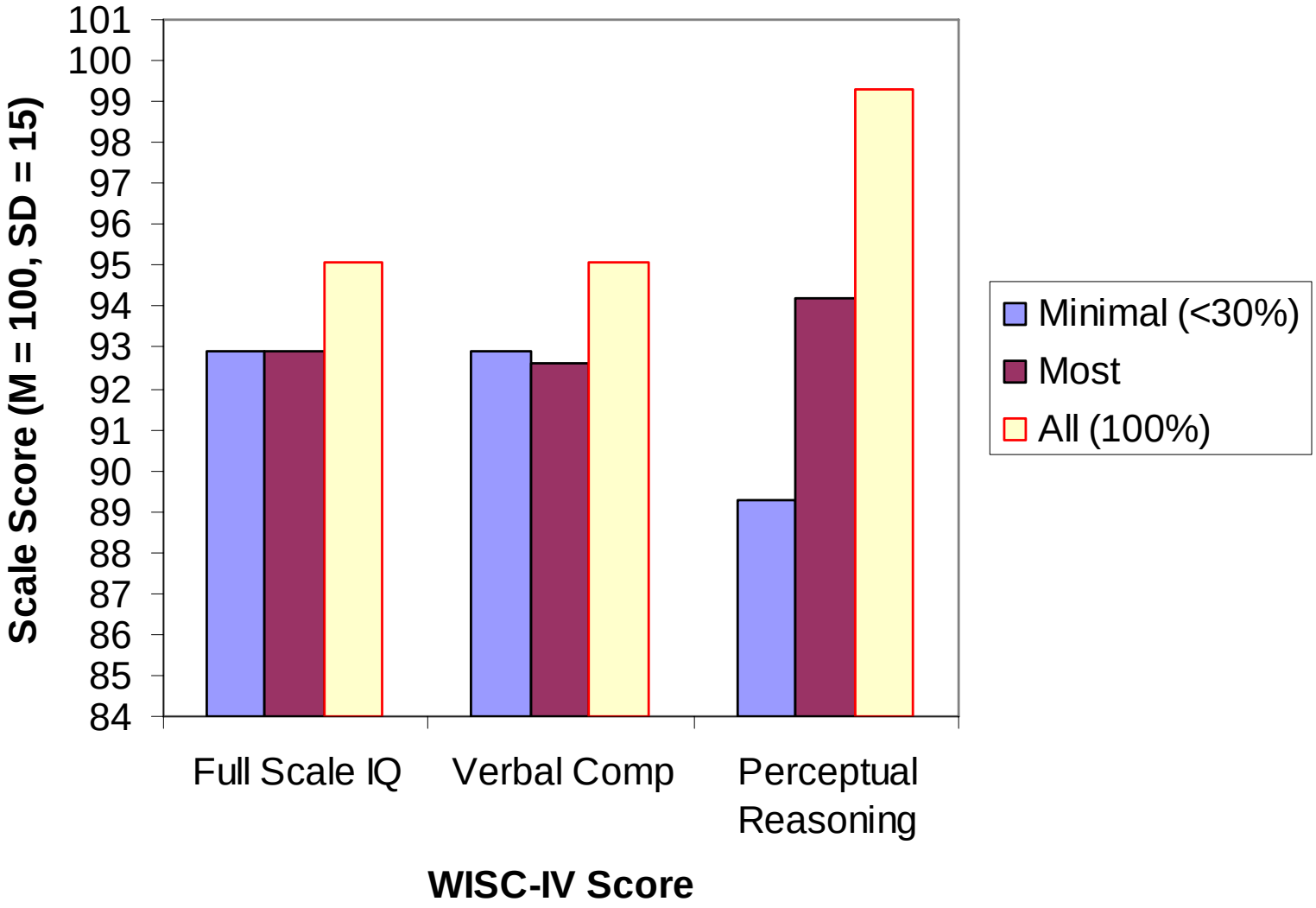
WISC-IV Spanish Scores by Percent Education in the U.S.



WISC-IV Spanish Scores by Percent Education in the U.S.



WISC-IV Spanish Scores by Percent Education in the U.S.



“nonlanguage tests may be more culturally loaded than language tests”

Anastasi & Urbina, 1997, p. 344

WISC-IV Spanish Scores: Multiple Perspectives

- Index scores (Mean 100, SD 15)
  - Verbal Comprehension (Spanish Calibration)
  - Perceptual Reasoning
  - Working memory
  - Processing Speed
- PR (by Pct Education in U.S.)
- PR (by Pct Education in U.S & Parental Education)
- Professional judgment required

Main Points

- Focus on talent identification & development
- Instability of observed scores on the best tests
- Importance of affective & conative traits
- Assumptions about opportunity to learn
- Consider alternative norms rather than alternative assessments
- Develop tests for bilingual children that are not simply translated versions of the English language test

The Connie Belin & Jacqueline N. Blank International Center
for Gifted Education and Talent Development

The University of Iowa College of Education

Thank you.

faculty.education.uiowa.edu/dlohman

The University of Iowa
Be remarkable

Nurturing **Potential** | *Inspiring* **Excellence**