

***Including Students with Disabilities  
in Large Scale Testing: Traveling  
Blue Highways and the Many Issues  
Encountered In the U.S***

Gerald Tindal – UO

# Abstract

- Key federal legislation served as the primary catalyst for states to develop options for including students with disabilities in their large scale testing programs. For the past 15 years, the sheer diversity and creativity of these options has been stunning. The presentation provides a sample of the issues and options, focusing on many elements that comprise any large scale testing enterprise: Standards and Standardization, Constructs and Test Formats, Design and Universal Design, and Reporting and Articulation of Outcomes.

# The Domestication of Mammals

Of the 148 big wild terrestrial herbivorous mammals that could serve as candidates for domestication, only 14 passed the test.

- Diet
- Growth rate
- Problems with Captive Breeding
- Tendency to Panic
- Social Structure

# Guns, Germs, and Steel

## The Ana Karenina Principle

“We tend to seek easy, single-factor explanations of success. For most important things, though, success actually requires avoiding many separate possible causes of failure” (p. 15).

Advocacy • Legislation • Practice • Technology •  
Training • Systems • Empirical basis

# Five Legislations & Regulations

- IDEA 1997: Students with disabilities included in large-scale testing programs and provided access to the general education curriculum.
- No Child Left Behind Act (2001): 95% participation of specific populations: students with disabilities, students of color, students from disadvantaged backgrounds, and students for whom English is a second language.
- December 9, 2003, regulatory guidance: 1% of students with the most significant cognitive disabilities could be judged proficient on large-scale tests.
- May 10, 2005: Assessments judged against modified achievement standards for students with persistent academic disabilities and served under the IDEA (2% judged proficient).
- 2005-2006 fiscal year: States submit large-scale testing program to a peer review in which a number of criteria would be invoked to document various components of technical adequacy.

# Standardization

- Test administration of tests –most public and visible aspect of testing.
- Much of the standardization of testing conditions relates to the quality of test administration...
- Standardization is a common method of experimental control for all tests.
- Every test (and each question or stimulus within each test) can be considered a mini experiment (van der Linden & Hambleton, 1997).
- The test administration conditions – standard time limits, procedures to ensure no irregularities, **environmental conditions conducive to test taking**, and so on – all seek to control extraneous variables in the "experiment" and make conditions uniform and identical for all examinees.
- **Without adequate control of all relevant variables affecting test performance**, it would be difficult to interpret examinee test scores uniformly and meaningfully (Downing, 2006, p.15).

# The Measurement Conundrum

- Fixing a condition of measurement reduces error and increases the precision of measurements, but it does so at the expense of narrowing interpretations of measurements” (Brennon, 2001, p. 2).
- The reliability-validity paradox: Attempts to increase reliability through standardization can actually lead to a decrease in the validity of interpretations.

# Distinguishing between Nouns and Verbs

- Constructs
  - Meaning and Interpretation
  - Construct Irrelevant Variance
  - Construct Under or Misrepresentation
- To Construct: The Test Environment
  - Contexts and Settings
  - Expected Routines
  - Enacted Behaviors



# Standards – Math Grade 3

- 1. Understand a fraction  $1/b$  as the quantity formed by 1 part when a whole is partitioned into  $b$  equal parts; understand a fraction  $a/b$  as the quantity formed by  $a$  parts of size  $1/b$ .
- 2. Understand a fraction as a number on the number line; represent fractions on a number line diagram.
  - Represent a fraction  $1/b$  on a number line diagram by defining the interval from 0 to 1 as the whole and partitioning it into  $b$  equal parts. Recognize that each part has size  $1/b$  and that the endpoint of the part based at 0 locates the number  $1/b$  on the number line.
  - Represent a fraction  $a/b$  on a number line diagram by marking off  $a$  lengths  $1/b$  from 0. Recognize that the resulting interval has size  $a/b$  and that its endpoint locates the number  $a/b$  on the number line.

# Benefit from Taking a Math Test

Statement	Agree (PA-SC)
16. Read problems and directions aloud.	78% and 43%
17. Simplify language in problems and directions.	64% and 53%
18. Present problems written in a language other than English.	98% and 97%
19. Extend length of testing sessions	75% and 24%
20. Administer test in multiple short testing sessions.	48% and 27%
21. Allow student to work alone in a separate testing location.	59% and 30%
22. Allow student to respond to questions in alternate formats such as typing, pointing, or with the use of a scribe.	44% and 31%
23. Magnify text of problems and directions.	83% and 28%
24. Allow student to respond to questions in an open-ended format where he/she provides the answer in writing.	52% and 37%
25. Allow student to respond to questions in a multiple choice response format where he/she selects the best answer from a list of choices.	29% and 31%
26. Use a variety of manipulatives.	59% and 25%
27. Use a calculator.	69% and 33%
28. Reformat the test to include fewer numbers of questions per page.	45% and 33%

*•Do not know, No benefit, Minimal benefit, Some benefit, Strong benefit*

# How Often are the Following Accommodations Provided ?

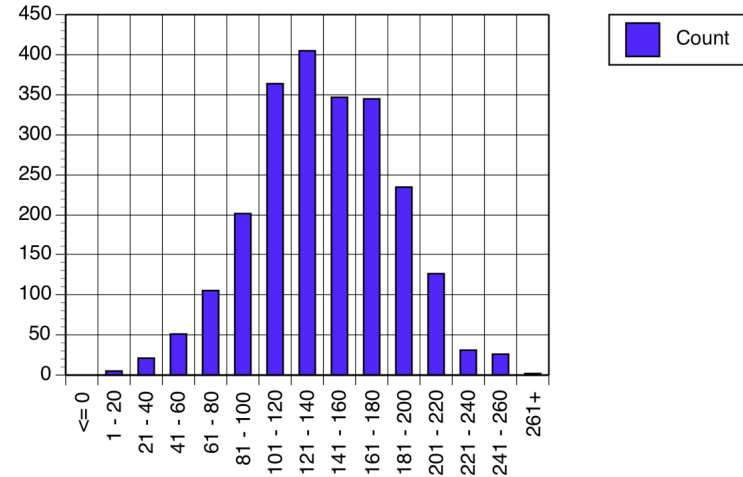
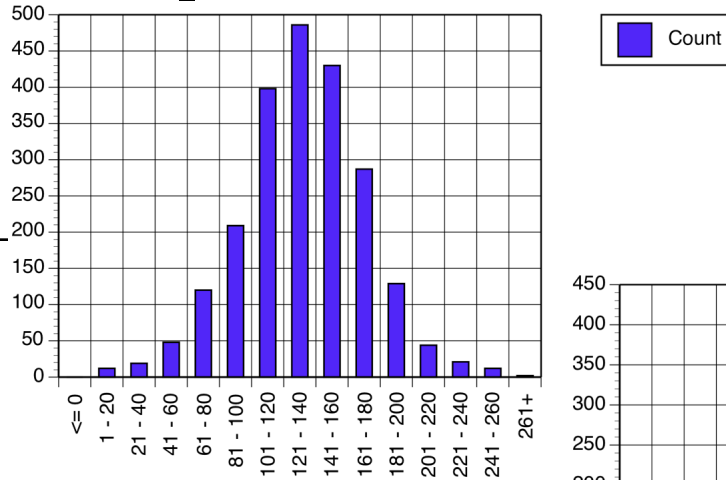
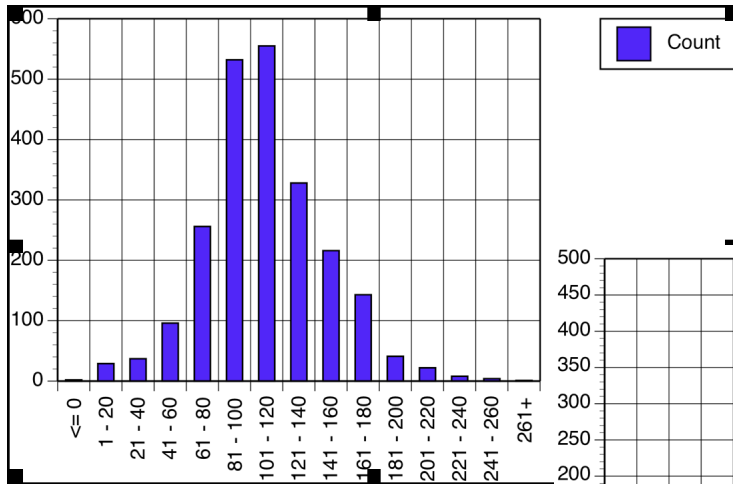
29. Read problems and directions aloud.	56% and 33%
30. Simplify language in problems and directions.	59% and 27%
31. Present problems written in a language other than English.	100% and 100%
32. Extend length of testing sessions.	53% and 32%
33. Administer test in multiple short testing sessions.	63% and 28%
34. Allow student to work alone in a separate testing location.	63% and 38%
35. Allow student to respond to questions in alternate formats such as typing, pointing, or with the use of a scribe.	55% and 44%
36. Magnify text of problems and directions.	65% and 48%
37. Allow student to respond to questions in an open-ended format where he/she provides the answer in writing.	47% and 34%
38. Allow student to respond to questions in a multiple choice response format where he/she selects the best answer from a list of choices.	53% and 35%
39. Use a variety of manipulatives.	52% and 23%
40. Use a calculator.	67% and 28%
41. Reformat the test to include fewer numbers of questions per page.	62% and 32%

•Do Not Know, Never, Sometimes, Often, Always

# Standards – Reading Grade 5

- Demonstrate understanding of figurative language, word relationships, and nuances in word meanings.
  - Interpret figurative language, including similes and metaphors, in context.
  - Recognize and explain the meaning of common idioms, adages, and proverbs.
  - Use the relationship between particular words (e.g., synonyms, antonyms, homographs) to better understand each of the words.

# Oral Reading Fluency



# The Measurement Conundrum Revisited

## Participation and Accommodation

- Keeping Score for ALL
  - The effects of inclusion and accommodation policies on large-scale educational assessment
  - National Research Council, 2004
- “The increased use of accommodations with NAEP assessments has corresponded to increased participation rates for students with disabilities and English Language Learners.”
- So what?

# What We Don't Know

- What accommodations have been used in NAEP (singly and bundled).
- The SKILL of the student (versus the disability).
- Use of the accommodation in instruction.
- Teacher recommendation for accommodation (on NAEP or state test)
- Performance levels on accommodated versus non-accommodated items.

# No Right Way to Do a Wrong Thing

- NAEP data base as it is structured can never address the question of accommodations.
- Research designs are lacking.
- Data are too global to answer any serious question.
- Construct validity at item level is lacking.



# What We Need to Know

- Need to consider accommodations as complex packages
- Need different research designs than randomized experiments (because of low sample size and inappropriate use of group statistics)
- We need to study populations and items more carefully
  - Smart about **items**
  - Smart about **students**

# The Unit of Analysis

- Test Level
  - Bundled Items
  - Variation in Skills
  - Reporting Categories
- Item Level
  - Specific Skills
  - Difficulty and Discrimination
  - Differential Item Functioning

# Design<sup>1</sup> of Research: Smart about Items

- Student is presented a standard item
- Can I solve the problem as presented?

**Yes**

**Incorrect**



**Accommodated**

**No**

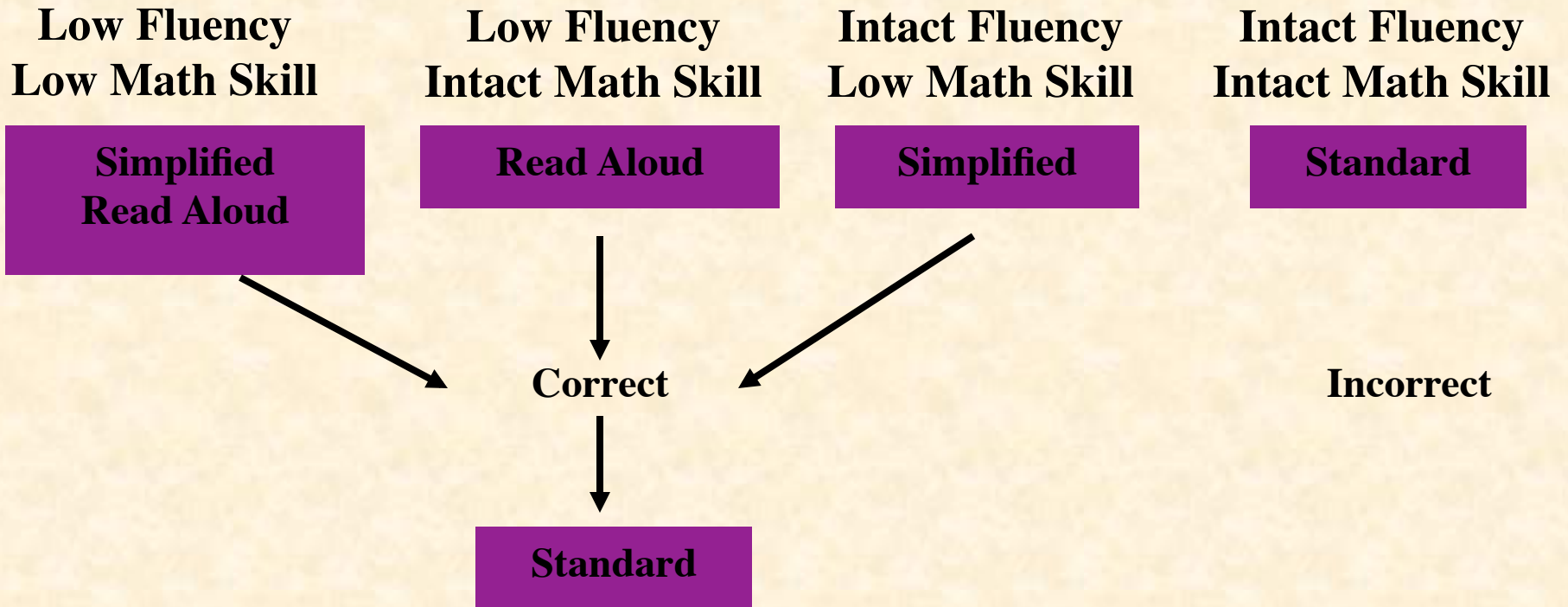
**Accommodated**



**Standard**

# Design<sup>2</sup> of Research: Smart about People

- Pre Measure Student Reading Fluency
- Pre Measure Student Basic Math Skill



# Meta Analysis to Support 'Smart Systems'

## Application of Internal Validity to ELLs

- Research on Accommodations:
  - Dictionaries with terms in English
  - Picture dictionaries
  - Read aloud
  - Plain, simplified, or modified English
  - Bilingual glossary
  - Items and directions in Spanish
  - Dual language or side-by-side in English and Spanish
  - Extra time
  - Small group testing configurations

# Empirical Findings

- Need for treatments in both those with and those without a ‘pre-existing’ condition.
- Effectiveness of accommodations varies by student characteristics (e.g. English proficiency, native language proficiency, language of instruction, etc.)

# Category of Accommodations

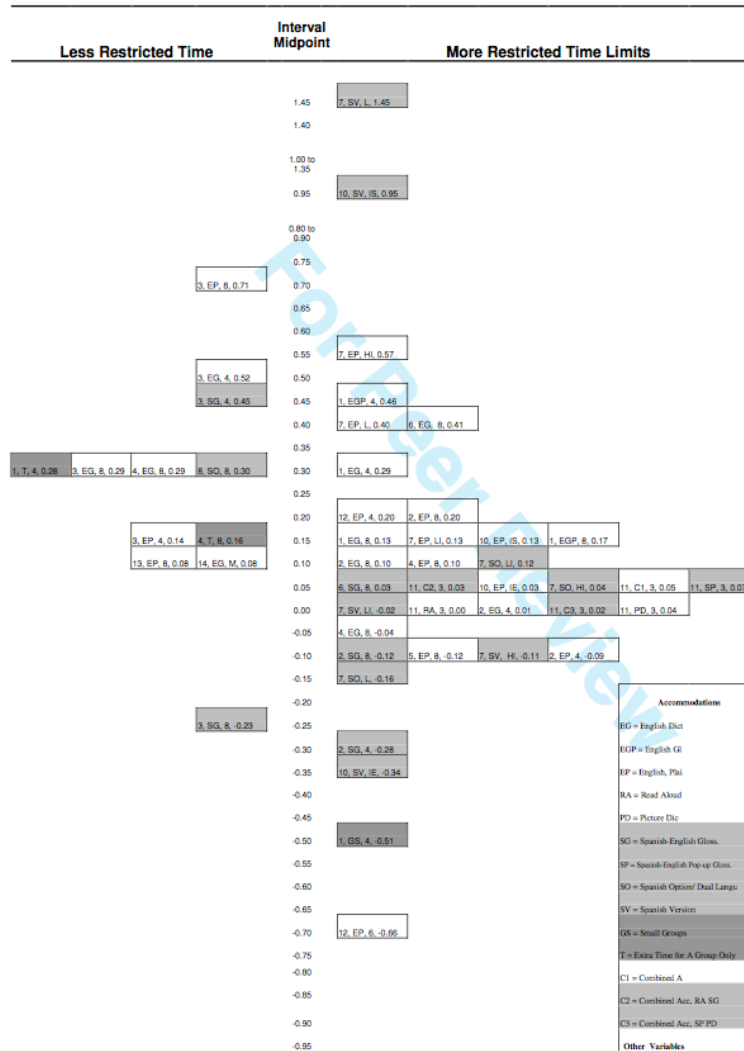
- Type
- Test format
- Time Constraints (equally restricted, equally generous, specifically bundled only for the accommodated group)
- Separated by
  - Proficiency
  - Language of Instruction
- ELL and non-ELL
- Under what conditions and for which students are particular accommodations effective?

# Results

- 94% of the 50 subsamples of ELLs received direct linguistic support accommodations in English (58%), or their native language (36%) and 6% received indirect linguistic support



# Inside Out Display



# Effect Sizes

- Most averaged  $-.012$  to  $+.41$  with the range from  $-1.13$  to  $+1.45$
- Standard time:  $0$  to  $.05$  midpoint
- Generous time:  $.30$  midpoint

# Outcomes

- Spanish language versions of the test had the largest effect sizes for students with low ELP and or who had been instructed in Spanish as compared to other accommodations
- For high or intermediate ELP students, Plain English was more effective than either Spanish version or option
- Plain English was easier for non-ELLs when administered under generous time conditions

# Outcomes

- Pop up glossary/dictionary (with and without restricted or time) versus paper-pencil (with and without restricted time)
  - Pop up had an effect over other English Dictionary only when administered under restricted time limits

# Outcomes

- Spanish Option/Dual Language effective only with generous time limits
- Major conclusion: Research design features (treatment factors) interact with population characteristics
- Language proficiency interacts with test format and time constraints (e.g., generous time with Spanish option or dictionaries in either Spanish or English OR Plain English only effective for moderate ELP versus native language for low EPL)

# Assessment Adaptations beyond Research Findings

- The ASK Settlement in Oregon
- When the Sidewalk Ends: Practice in the Absence of Research
  - Purpose: What is the construct irrelevant variance?
  - Function: How does it work?
  - Error: What are the false positives and false negatives?
  - Systems: What are the implications for the whole?
- Modifications

# Alternate Assessments

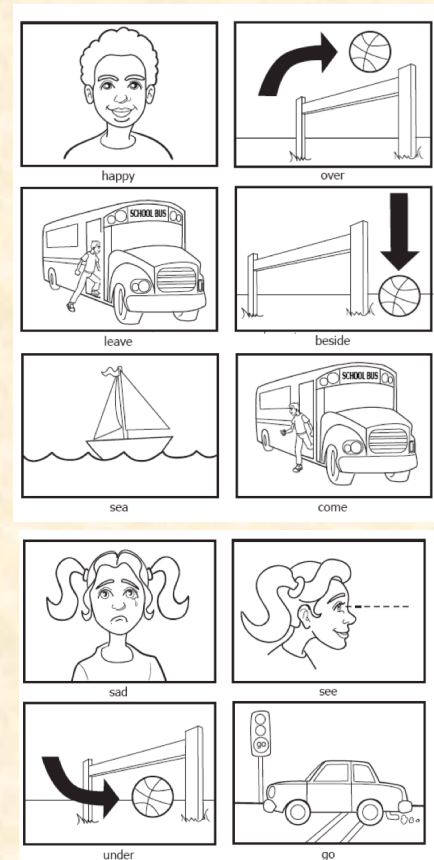
- Extending the population to ALL
- Transitioning toward an interesting nexus of functional versus academic
- Reflecting tension between portfolios and on-demand performance events
- Having considerable success in establishing technical adequacy
- Providing a new look at assistive devices and access

# Type of Administration-Reading

## Standard

happy	over
leave	beside
sea	come
sad	see
under	go

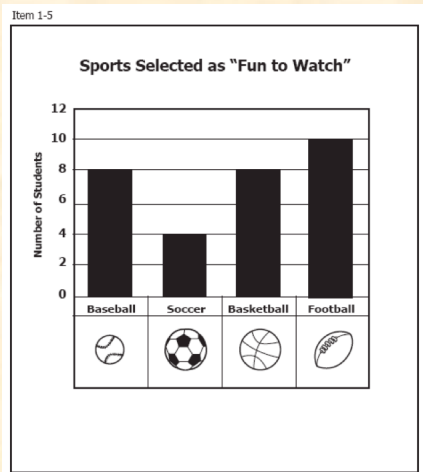
## Scaffold



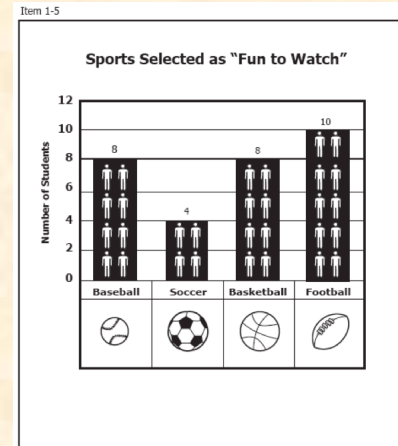


# Type of Administration-Math

## Standard



## Scaffold



# Levels of Independence

Levels of Independence				
I – Inappropriate/ Inaccessible based on the nature of the student's disability	1 - Full Physical Contact for response  <i>(e.g., hand over hand)</i>	2 - Partial Physical Contact for response  <i>(e.g., nudge or adjust body)</i>	3 - Visual: Materials Movement <i>(e.g., move into line of vision)</i>  - Verbal: Auditory Statement <i>(e.g., more than repeat prompt)</i>  - Gesture: Hand Signal <i>(e.g., tap table, pick up card)</i>	4 - Independent: No contact and no prompting
R - Refusal: Student does not complete or participate with or without assistance				
D - Too Difficult for the student				

# Levels of Independence x Disability

	Level of Independence				Total
Code- Disability	LOI-1	LOI-2	LOI-3	LOI-4	
82-ASD	98	37	116	533	784
50-CD	2	3	18	463	486
60-ED	1	0	4	117	122
10-MR <sup>1</sup>	120	27	171	1030	1348
80-OHI <sup>2</sup>	39	7	35	419	500
70-OI	64	8	17	86	175
90-SLD <sup>3</sup>	1	1	22	1135	1159
20-HI	5	1	8	53	67
40-VI	27	0	3	12	42
74-TBI	2	1	5	33	41
Total	359	85	399	3881	4724

# Proficiency Rates-Reading

Grade	Does not meet (%)	Nearly meets (%)	Meets (%)	Exceeds (%)	Meets/exceeds (%)
3	13	21	41	25	66
4	26	15	35	24	59
5	26	20	37	17	54
6	23	15	38	24	62
7	25	20	41	14	55
8	33	22	32	13	45
10	35	28	23	14	37

# Proficiency Rates-Math

Grade	Does not meet (%)	Nearly meets (%)	Meets (%)	Exceeds (%)	Meets/exceeds (%)
@TB:3	22	42	23	13	36
4	51	14	15	20	35
5	54	15	19	12	31
6	56	28	13	3	16
7	65	18	14	3	17
8	77	12	9	2	11
10	73	16	10	1	11

Model	Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Semi-Partial
		B	Std. Error	Beta			
	(Constant)	73.555	1.143		64.376	<.001	
	LOI (centered)	2.145	.042	.541	51.036	<.001	.455
	ADM	-16.448	.694	-.254	-23.705	<.001	-.211
	ELEM	5.370	1.033	.084	5.199	<.001	.046
	MID	2.906	1.056	.044	2.751	<.001	.025
	ASD	-12.886	.955	-.152	-13.495	<.001	-.120
	CD	-3.612	1.048	-.035	-3.447	.001	-.031
	ED	-1.194	1.843	-.006	-.648	.517	-.006
	MR	-13.087	.854	-.187	-15.320	<.001	-.137
	OHI	-6.636	1.048	-.065	-6.330	<.001	-.056
	OI	-12.559	1.670	-.075	-7.523	<.001	-.067
	HI	-10.015	2.444	-.038	-4.097	<.001	-.037
	VI	-9.729	3.136	-.029	-3.102	.002	-.028

<http://www.brtprojects.org>

