

# The Effect of Anchor Test Construction on Scale Drift

Judit Antal  
Gerald Melican  
Thomas Proctor  
Andrew Wiley

The College Board

Annual Meeting of the National Council on Measurement in Education  
Denver, CO  
April 3, 2010



# Anchor Tests

1. Used for Non-Equivalent Anchor Test (NEAT) design
  - to link test forms to be equated
  - *by* measuring differences between populations
  - *when* equivalent groups are not available;
2. Common practice building anchors
  - Miniature version of the total test
  - Representative of content and statistical specification of the full test

# Anchor Tests

- Statistical representativeness
  - Equal mean and equal standard deviation of item difficulties
  - Cover difficulty range of the full test
- Sinharay/Holland Study (2007)
  - Suggests to relax some statistical requirements for external anchors
  - Compared “mini” (same SD) and “midi” (reduced SD) anchor tests
  - Recommendation: SD can be reduced for external anchors

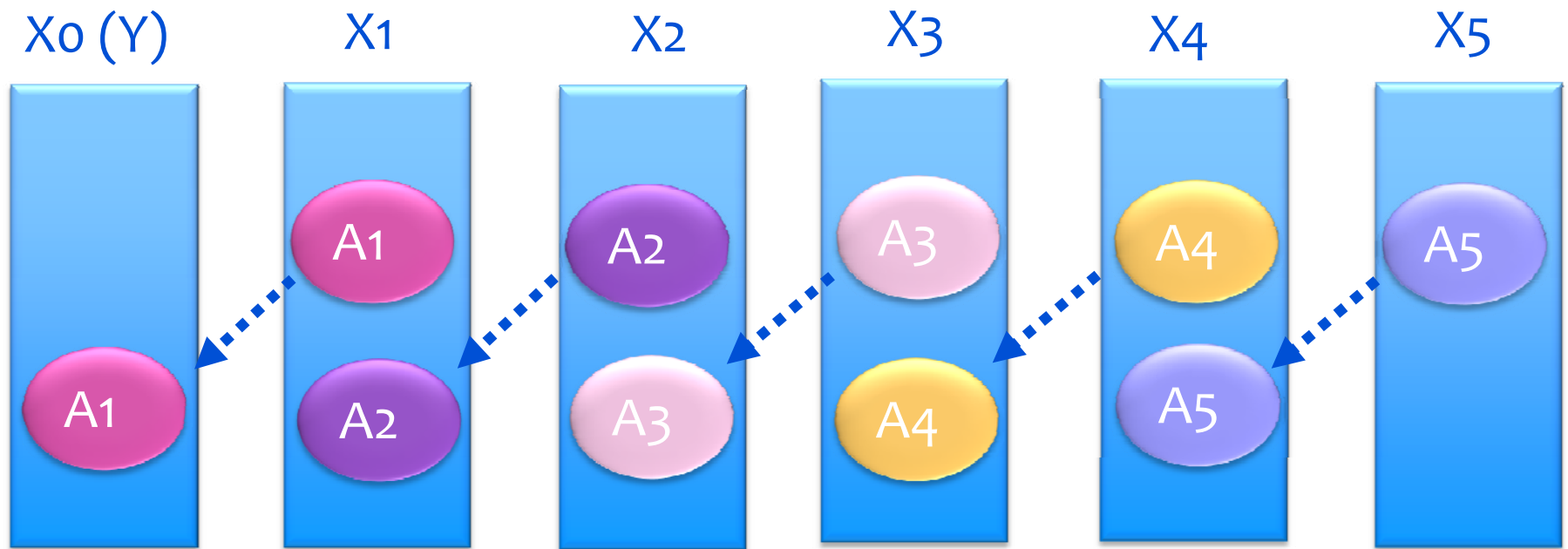
# Study Goals

- Link multiple forms to mimic a multi-year testing program with multiple forms administered in a longer time span – not only two forms
- Evaluation of scale drift across the chain using “mini” vs. “midi” anchor concept proposed by Sinharay/Holland (2007)
  - Scale drift across equating designs (NEAT vs. RG) within each type of anchor
  - Scale drift across anchor types within equating designs
- Current study establishes a baseline for later comparisons

# Design and Data Simulation

- Mathematics Test of a College admission exam was used to obtain initial parameter estimates
  - Main test: 54 items
  - External anchor tests: 20 items
  - All items are dichotomous
- 2PL IRT model (w/ Parscale)
- Two series of six tests are generated with external anchors
- Two anchor tests are embedded into each test form to facilitate common item equating

# Equating Design



# Simulation

1. Built 2 series of tests: a “mini” (equal SD) chain and a “midi” (0.5 SD) chain
2. Anchor tests that link consecutive forms not only needed to be parallel but their item parameters had to be the same as they represent the same item set
3. Means of the item difficulties for all six main forms needed to be the same as the old main test form to create equivalent groups.
4. Item parameters needed to be fixed across the two scenarios to be able to compare the drift across the two anchor types.

# Data Simulation

- 100 replications for each test (600 across the chain)
- Baseline condition:
  - Equal form difficulty across six test forms
  - No difference in ability from year-to-year
  - Fixed test sizes (54, 20, 20)
  - 4,000 simulees for each form
- Simulation condition: anchor type (“mini” vs. “midi”)

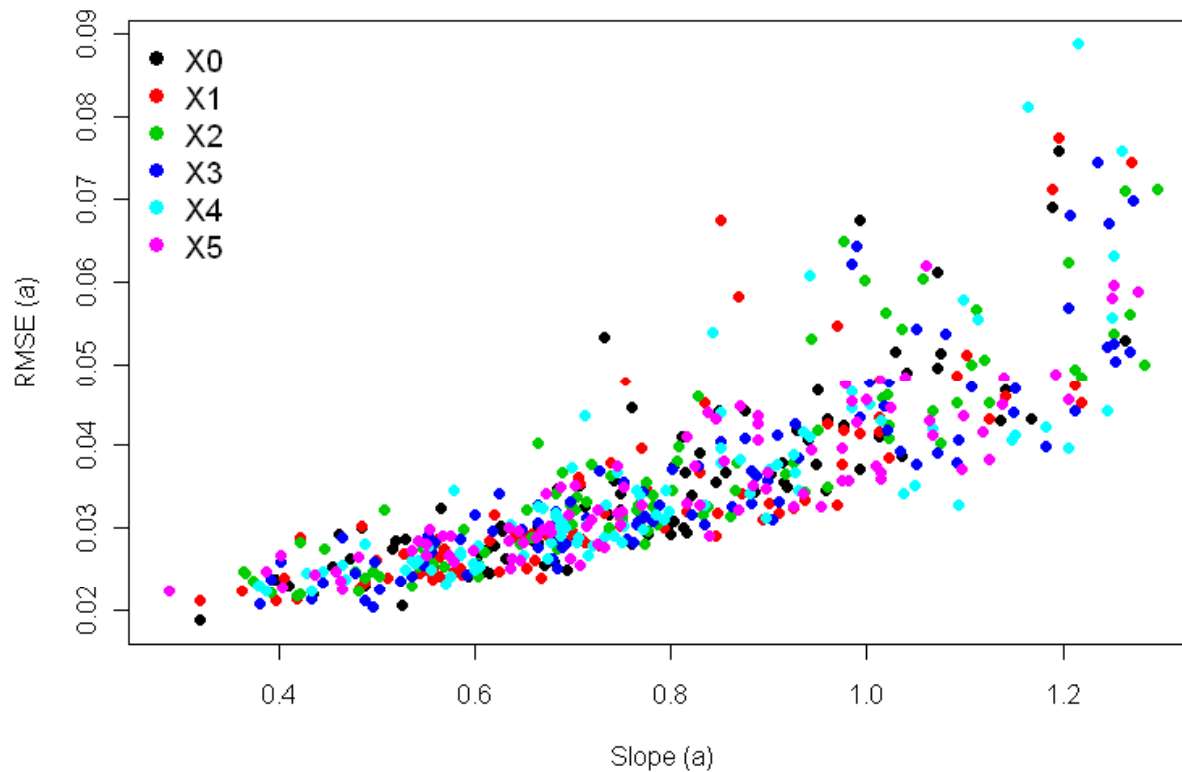


# Simulation: item difficulty statistics

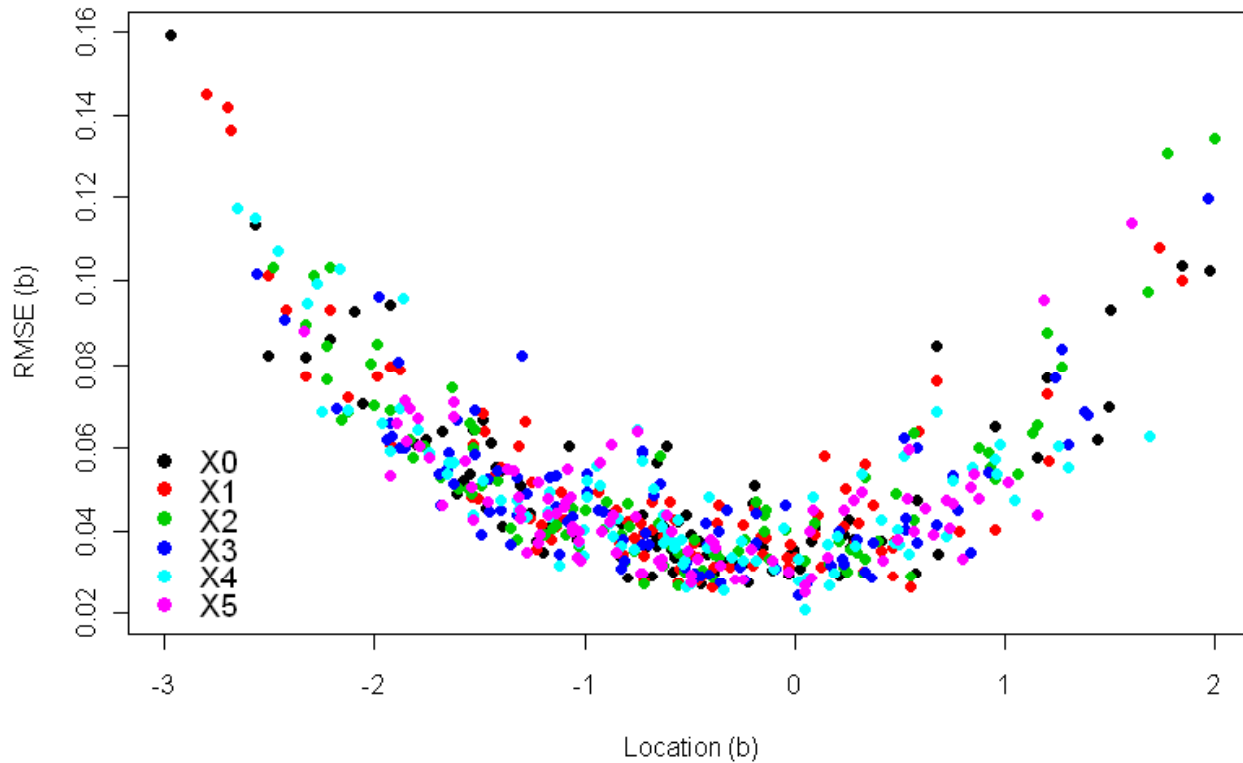
“mini”	Form X0		Form X1		Form X2		Form X3		Form X4		Form X5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Main Test	-0.59	0.89	-0.59	1.00	-0.59	1.21	-0.59	1.07	-0.59	1.10	-0.59	0.96
Anchor 1	--	--	-0.59	<b>1.10</b>	-0.59	<b>0.80</b>	-0.59	<b>0.91</b>	-0.59	<b>0.80</b>	-0.59	<b>0.87</b>
Anchor 2	-0.59	<b>1.10</b>	-0.59	<b>0.80</b>	-0.59	<b>0.91</b>	-0.59	<b>0.80</b>	-0.59	<b>0.87</b>	--	--

“midi”	Form X0		Form X1		Form X2		Form X3		Form X4		Form X5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Main Test	-0.59	0.89	-0.59	1.00	-0.59	1.21	-0.59	1.07	-0.59	1.10	-0.59	0.96
Anchor 1	--	--	-0.59	<b>0.58</b>	-0.59	<b>0.67</b>	-0.59	<b>0.64</b>	-0.59	<b>0.44</b>	-0.59	<b>0.42</b>
Anchor 2	-0.59	<b>0.58</b>	-0.59	<b>0.67</b>	-0.59	<b>0.64</b>	-0.59	<b>0.44</b>	-0.59	<b>0.42</b>	--	--

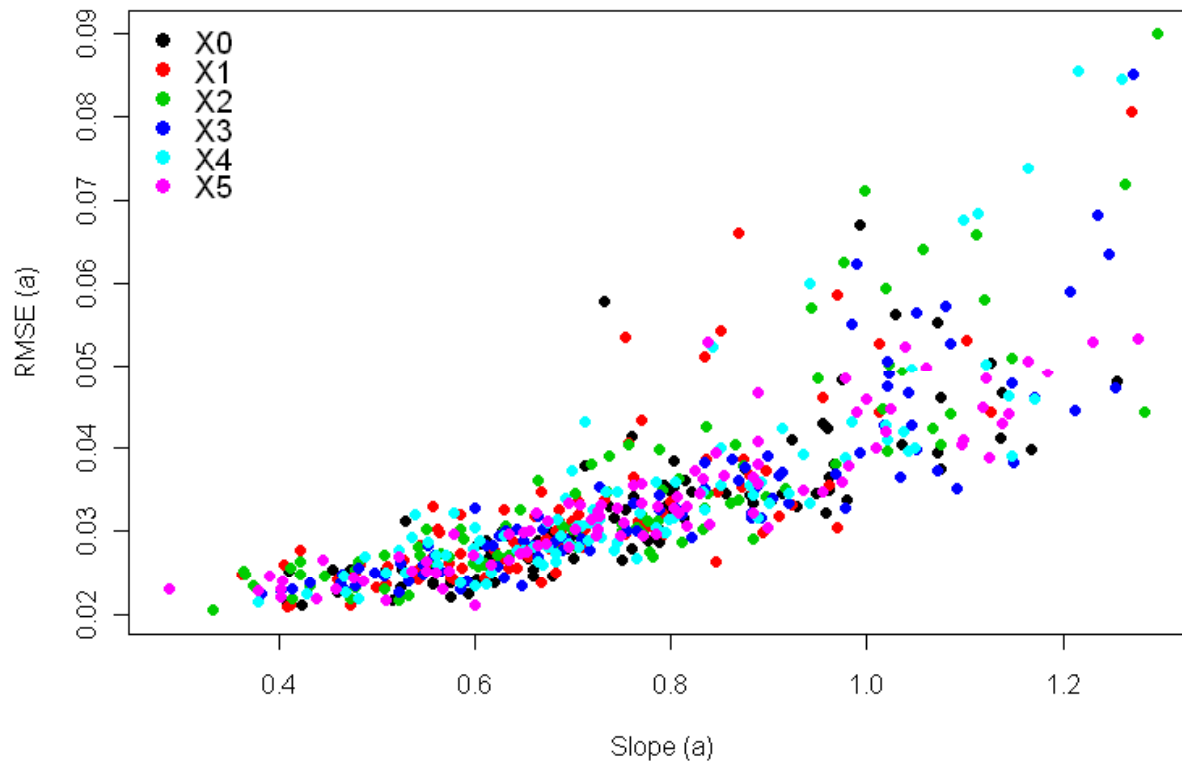
# Simulation: parameter recovery for “a” - equal SD (mini)



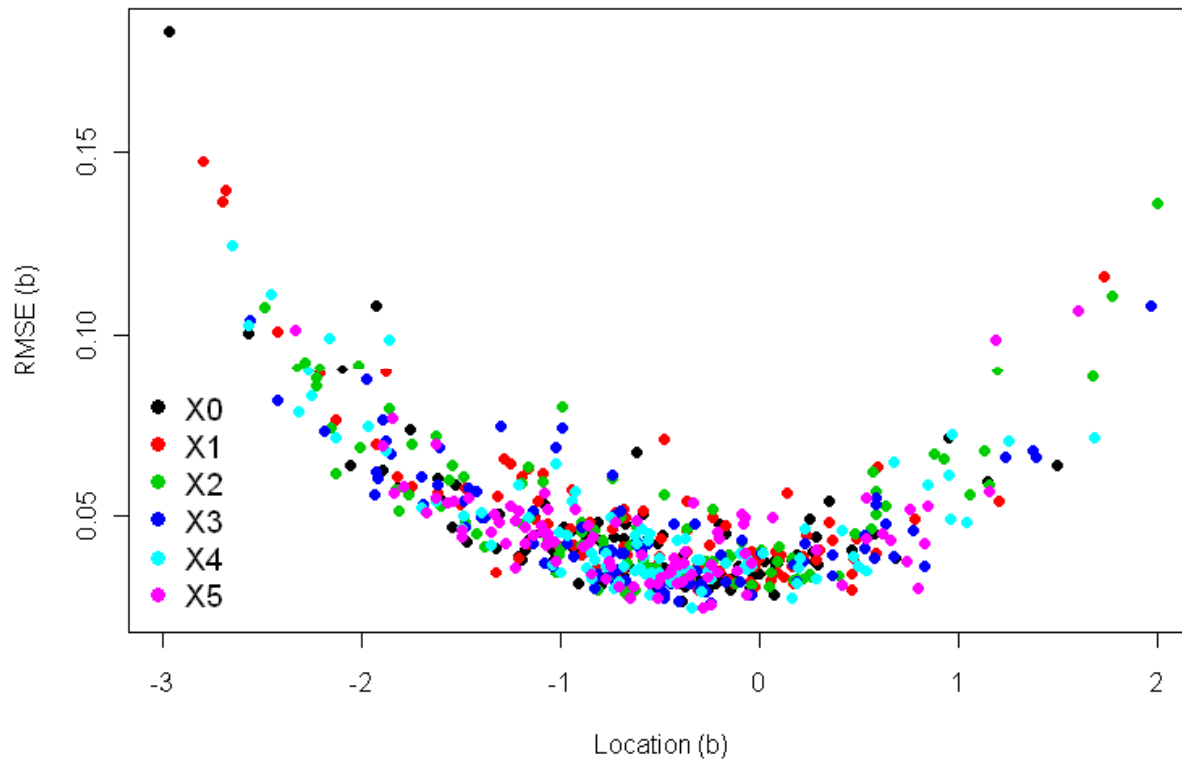
# Simulation: parameter recovery for “b” - equal SD (mini)



# Simulation: parameter recovery for “a” - reduced SD (midi)



# Simulation: parameter recovery for “b” - reduced SD (midi)



# Evaluation Criteria

$$RMSD = \left[ \frac{\sum_i f_i (x_{i.AN} - x_{i.AR})}{\sum_i f_i} \right]^{\frac{1}{2}}$$

$$MAD = \frac{\sum_i f_i |x_{i.AN} - x_{i.AR}|}{\sum_i f_i}$$

$$MSD = \frac{\sum_i f_i (x_{i.AN} - x_{i.AR})}{\sum_i f_i}$$

# Results

	Form 0		Form 1		Form 2		Form 3		Form 4		Form5	
<b>Mini Anchor Condition</b>												
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Main	33.9	11.2	33.5	10.4	34.4	9.8	34.3	10.7	33.7	10.3	34.6	10.9
Anchor 1	--	--	12.9	4.0	12.7	4.6	12.9	4.1	12.8	4.5	12.7	4.3
Anchor 2	12.9	4.0	12.7	4.6	12.9	4.1	12.8	4.5	12.7	4.2	--	--
<b>Midi Anchor Condition</b>												
Main	33.9	11.2	33.5	10.4	34.4	9.7	34.3	10.7	33.7	10.3	34.6	10.9
Anchor 1	--	--	12.7	4.5	12.5	4.3	12.7	4.3	12.8	4.8	12.7	4.8
Anchor 2	12.7	4.5	12.5	4.3	12.7	4.3	12.8	4.8	12.7	4.8	--	--

# Results

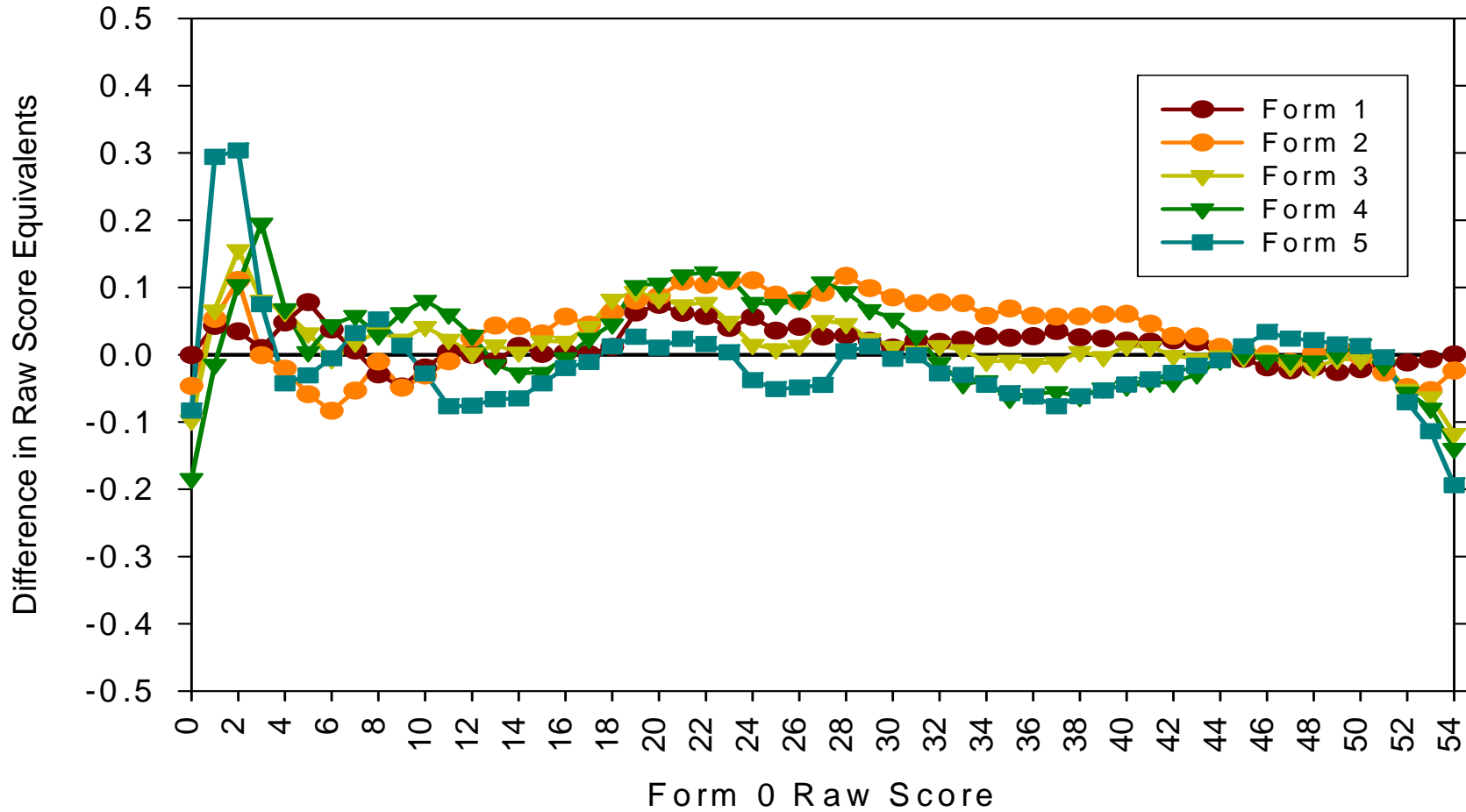
	Form 0	Form 1	Form 2	Form 3	Form 4	Form5
<b>Mini Anchor Condition</b>						
Main – Anchor 1	0.88	0.86	0.89	0.87	0.88	0.88
Main – Anchor 2	0.87	0.89	0.87	0.89	0.87	0.89
Anchor 1 – Anchor 2	0.82	0.83	0.84	0.83	0.84	0.84
<b>Midi Anchor Condition</b>						
Main – Anchor 1	0.89	0.87	0.86	0.87	0.88	0.89
Main – Anchor 2	0.88	0.86	0.87	0.89	0.88	0.87
Anchor 1 – Anchor 2	0.84	0.82	0.81	0.83	0.85	0.83



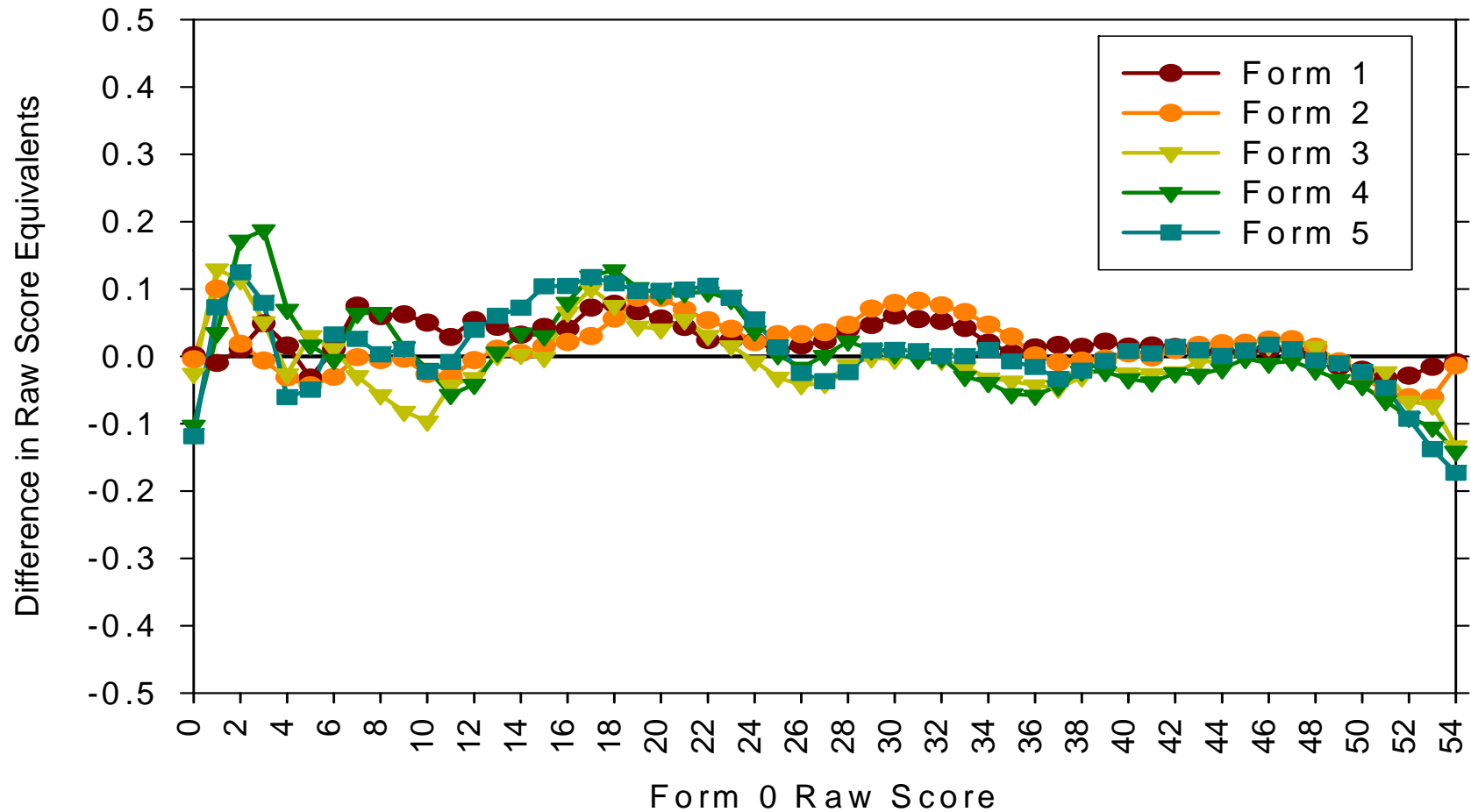
# Results

	Form 1	Form 2	Form 3	Form 4	Form5
<b>Mini Anchor Condition</b>					
RMSD	0.36	0.44	0.47	0.53	0.57
MAD	0.27	0.33	0.36	0.40	0.44
MSD	0.02	0.05	0.01	0.00	-0.02
ewRMSD	0.43	0.55	0.59	0.65	0.72
ewMAD	0.31	0.38	0.43	0.48	0.53
ewMSD	0.02	0.03	0.02	0.01	-0.01
<b>Midi Anchor Condition</b>					
RMSD	0.36	0.46	0.48	0.53	0.56
MAD	0.27	0.34	0.37	0.41	0.44
MSD	0.02	0.03	-0.01	-0.01	0.01
ewRMSD	0.42	0.56	0.61	0.67	0.70
ewMAD	0.30	0.40	0.45	0.49	0.52
ewMSD	0.02	0.02	-0.01	0.01	0.01

# Results: NEAT vs. RG for “mini” chain



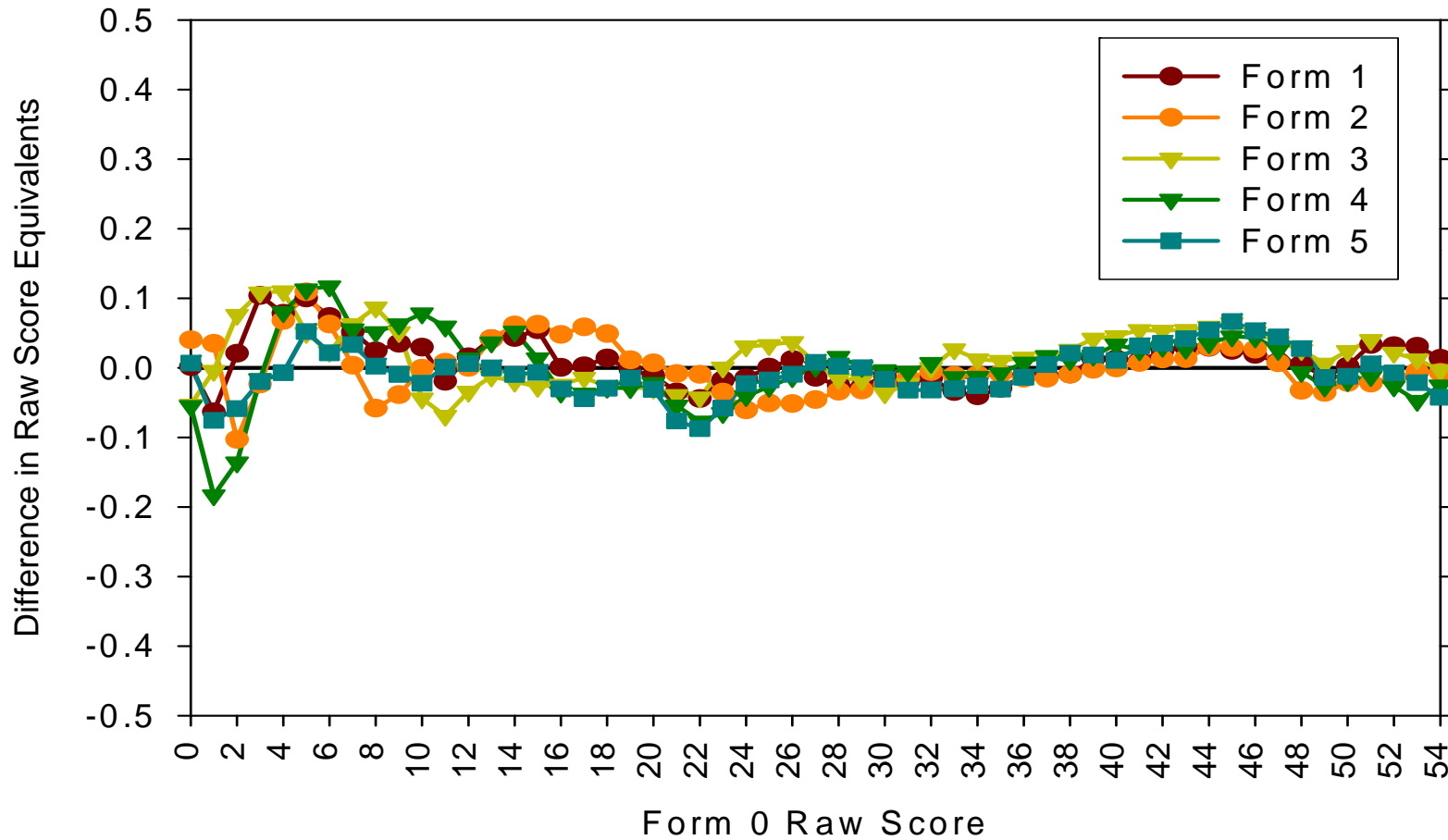
# Results: NEAT vs. RG for “midi” chain



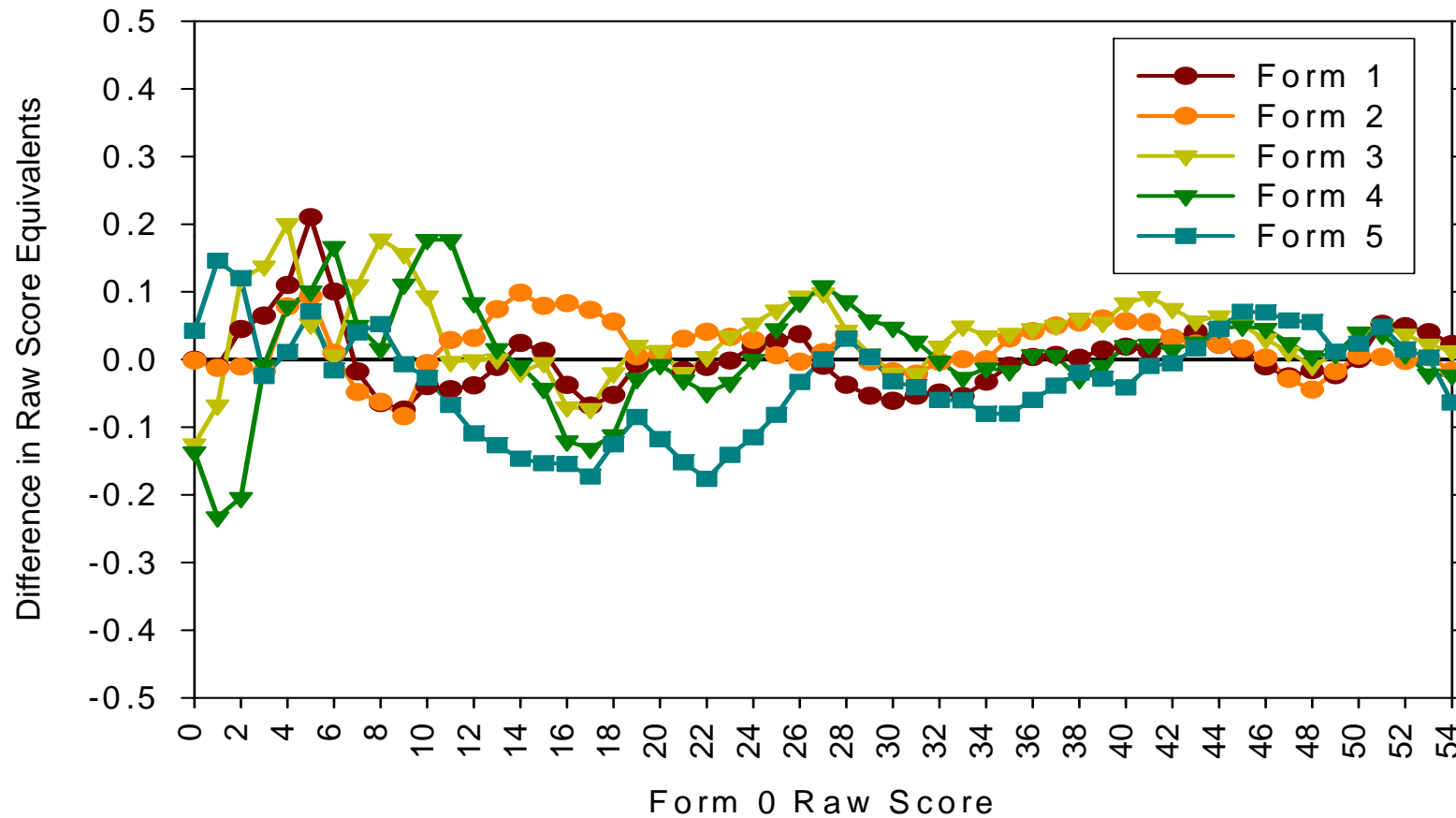
# Results

	Form 1	Form 2	Form 3	Form 4	Form5
<b>RG Equating</b>					
RMSD	0.50	0.52	0.48	0.52	0.48
MAD	0.37	0.39	0.37	0.39	0.36
MSD	0.00	-0.01	0.02	0.01	0.00
ewRMSD	0.62	0.62	0.59	0.62	0.59
ewMAD	0.45	0.45	0.44	0.46	0.43
ewMSD	0.01	0.00	0.01	0.00	-0.01
<b>NEAT Equating</b>					
RMSD	0.43	0.54	0.61	0.68	0.77
MAD	0.32	0.42	0.46	0.52	0.59
MSD	-0.01	0.02	0.04	0.01	-0.04
ewRMSD	0.62	0.74	0.82	0.91	1.01
ewMAD	0.44	0.54	0.61	0.67	0.74
ewMSD	0.00	0.02	0.04	0.01	-0.03

# Results: RG across chains (mini vs. “midi”)



# Results: NEAT across chains (mini vs. “midi”)



# Conclusions & Future Research

- Results support Sinharay/Holland recommendation
  - SD can be reduced when building external anchor tests
- Continuing research efforts:
  - Various ability levels and form difficulties;
  - Total/anchor length ratios;
  - Small sample equating;
  - Correlations between anchor and total test;
  - Effect on IRT equating methods;
  - Real data applications

Thank You!