Slide 1

From Saint to Sinner and Back Again: The Confounding Effect of Linking Error on Gains Estimated from Value-Added Models

Harold C. Doran
American Institutes for Research
Computer and Statistical Sciences Center
hdoran@air.org
http://www.air.org/hdoran/personal.html

Slide 2

Purpose of My Talk

- To illustrate the bias introduced via linking error
- Provide empirical evidence of its confounding effect on TVAAS data

Slide 3

The Increasing Popularity of VAM

- No Child Left Behind illustrates national interest in test-based accountability
- However, AYP is a cross-sectional model
- Letter from 16 State Chief School Officers
- VAMs better align with the notion of student learning

Slide 4

The Increasing Popularity of VAM

- Assessment systems in place to support longitudinal models
- Unique IDs are present in many jurisdictions
- Statistical advances have been tremendous
- Software programs have made the statistical models more accessible

Slide 5

Characterizing Uncertainty

- The (in)accuracy of scores is paramount in statistical applications
- Multiple sources of error are present
- However, sampling error tends to zero with larger samples
- Linking function is based on a sample of potential items

Slide 6

Linking Error Literature

- Cohen, J., Johnson, E., & Angeles, J. (2000). Variance estimation when sampling dimensions via the jackknife with application to the national assessment of educational progress (Tech. Rep.). Washington, DC: American Institute for Research.
- Sheehan, K. M., & Mislevy, R. J. (1988, July). Some consequences of the uncertainty in IRT linking procedures (Tech. Rep.). Educational Testing Service.
- Haertel, E. H. (2004, May). The behavior of linking items in test equating (Tech. Rep.). CRESST/Stanford University.
- Hedges, L. V., & Vevea, J. L. (1997, December). A study of equating in NAEP (Tech. Rep.). http://www.air.org/publications/publications-set.htm.

Slide 7

The Nature of Linking Error

- The process of constructing vertical scales introduces an additional variance component that is currently ignored in VAM applications
- Consequently, standard errors are too small and gains may appear to fluctuate over time due to noise, not instructional quality
- If this occurs, teacher and school effect indices may appear to fluctuate over time

Slide 8

Linking Scales: A Small Example

- Items are embedded across test forms
- $\boldsymbol{b_{3\rightarrow 3}}$ with mean $\mu_{3,3}$ and variance $\sigma^2_{3,3}$
- $\boldsymbol{b_{3\rightarrow 4}}$ with mean $\mu_{3,4}$ and variance $\sigma^2_{3,4}$
- $\boldsymbol{b_{4\rightarrow 3}}$ with mean $\mu_{4,3}$ and variance $\sigma^2_{4,3}$
- $\boldsymbol{b_{4\rightarrow 4}}$ with mean $\mu_{4,4}$ and variance $\sigma^2_{4,4}$

Slide 9

Linking Constants

Obtain the forward, backward, and overall linking constants:

$$\begin{aligned} \beta_{3\rightarrow 4} &= \mu_{3,3} - \mu_{3,4} \\ \beta_{4\rightarrow 3} &= \mu_{4,3} - \mu_{4,4} \\ \beta_{3\leftrightarrow 4} &= \frac{(\beta_{3\rightarrow 4} + \beta_{4\rightarrow 3})}{2} \end{aligned} \tag{1}$$

Slide 10

Calculate Error Variance

Because the linking constants were obtained using a sample of test items from a sample of students, they too are subject to error which can be estimated as:

$$\begin{aligned} Var(\beta_{3\rightarrow 4}) &= \sigma^2_{\beta,3\rightarrow 4} = \sigma^2_{3,3} + \sigma^2_{3,4} - 2\sigma_{(3,3)(3,4)} \\ Var(\beta_{4\rightarrow 3}) &= \sigma^2_{\beta,4\rightarrow 3} = \sigma^2_{4,4} + \sigma^2_{4,3} - 2\sigma_{(4,4)(4,3)} \end{aligned} \tag{2}$$

The variance of the overall linking constant is therefore:

$$Var(\beta_{3\leftrightarrow 4}) = \sigma^2_{3\leftrightarrow 4} = \frac{\sigma^2_{\beta,3\rightarrow 4} + \sigma^2_{\beta,4\rightarrow 3}}{4} \tag{3}$$

Slide 11

Obtain Proficiency Estimates

With test scales linked, the proficiency score of student i is:

$$\theta^*_{4i} = \theta_{4i} + \beta_{3\leftrightarrow 4} \tag{4}$$

The variance of the score is:

$$Var(\theta^*_{4i}) = Var(\theta_{4i}) + Var(\beta_{4\leftrightarrow 3}) \tag{5}$$

Slide 12

The Confounding Effect of Linking Error

With the scales now linked, we can estimate mean scores as follows: Let

- $\bar{\theta}_{4j}$ = mean of Grade 4 students in school j with variance $\sigma^2_{4j} + \sigma^2_{\beta,3\leftrightarrow 4}$
- $\bar{\theta}_{3j}$ = mean of this same cohort in the previous school year with variance σ^2_{3j}

Slide 13

Calculate Difference Score

Using the Grade 4 and Grade 3 means on the vertical scale, it is common to difference the scores as follows to obtain the gain score:

$$D_{3\to 4} = \bar{\theta}_{4j} - \bar{\theta}_{3j} \tag{6}$$

Which has the following variance:

$$Var(D_{3\to 4}) = (\sigma^2_{4j} + \sigma^2_{3j} - 2\sigma_{4j,3j}) + \sigma^2_{\beta,3\leftrightarrow 4} \tag{7}$$

For within grade comparison (e.g, 4 to 4) the linking bias subtracts out.

Slide 14

The Remaining Variance Component

- As sample sizes increase, the sampling variance of the group means tends to zero.
- However linking error is invariant to examinee sample size, but is sensitive to number of embedded items
- If other items were embedded, we might obtain a slightly different linking function
- As a result, part of what appears to be gain is error that should be characterized as uncertainty.

Slide 15

An Empirical Review of TVAAS Math

	2000	2001	2002	2003
Grade 3	C4	C3	C2	C1
Grade 4	C5	C4	C3	C2
Grade 5	C6	C5	C4	C3
Grade 6	C7	C6	C5	C4
Grade 7	C8	C7	C6	C5
Grade 8	C9	C8	C7	C6

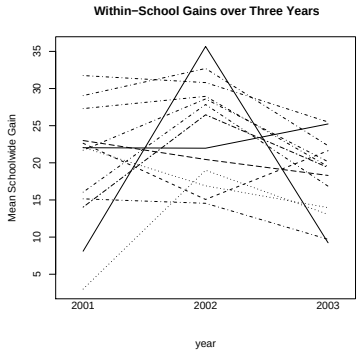
Slide 16

Methodology

- Compare each school's mean at time t to the state mean using TVAAS standard errors.
- Classify schools contingent upon their location in t distribution as "1", "2", or "3".
- We refer to 3's as Saints, 2's as Stable, and 3's as Sinners.
- For example, "111" is consistently low and "333" is consistently high

Slide 17

Random Sample of 15 Elementary Schools



Slide 18

School Frequency Distribution

Pattern	Frequency
113	7
123	18
131	14
132	18
133	16
213	16
231	16
311	5
312	23
313	5
321	14
331	6

Slide 19

School Fluctuation Patterns

	schid	t1	t2	t3	pattern
682	1900040	−7.79	1.98	−2.22	131
2365	3500005	−2.61	6.30	−6.30	131
3793	5800021	6.13	−1.98	2.95	313
5050	7510015	4.16	−4.91	3.33	313
5206	7800045	6.15	−5.18	2.10	313
5362	7900107	−6.36	3.03	−3.25	131
5750	7910220	−7.35	6.46	−5.53	131
6020	7910435	−6.55	8.73	−12.87	131
6140	7910530	−2.07	3.61	−2.53	131
6272	7910620	−2.70	5.19	−9.06	131
6677	7910805	−4.84	6.52	−2.19	131

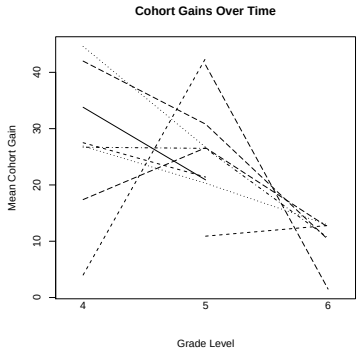
Slide 20

Proportion of School Reversals

Saint to Sinner	14%
Sinner to Saint	25%

Slide 21

Random Sample of 15 Cohorts



Slide 22

Cohort Methodology

- Followed Cohort 4 for 3 years
- Compared estimated gain to TVAAS "expected gain"

Slide 23

Cohort Frequency Distribution

Pattern	Frequency
113	1
131	15
132	11
13NA	40
213	3
231	34
311	6
312	20
313	5
31NA	18
321	21
331	6
NA31	15

Slide 24

Fluctuating Cohort Patterns

	schid	t.2001	t.2002	t.2003	pattern
1666	1000013	−2.43	2.85	−5.18	131
4518	2300035	−4.67	3.75	−7.07	131
4710	2400060	−5.03	4.00	−2.20	131
5390	2900005	−2.76	6.27	−4.48	131
9170	5000050	4.31	−3.06	4.11	313
9410	5200050	7.65	−7.35	3.32	313
9786	5400060	4.29	−1.97	4.18	313
12402	7500077	−4.61	4.23	−3.77	131
12890	7800015	2.71	−3.29	2.10	313
13718	7910118	−3.80	6.52	−10.10	131
13822	7910133	−4.22	3.47	−3.06	131
14198	7910210	−7.60	4.21	−2.64	131

Slide 25

Proportion of Cohort Reversals

Saint to Sinner	22%
Sinner to Saint	49%

Slide 26

Plausible Explanations

- Instructional effects: schools use data to improve
- Cohort effects: school populations change
- Sampling error: should capture instability
- Linking error: is currently ignored.

Slide 27

Implications for Test and Software Developers

- Report linking statistics
- Include algorithms for estimating the linking variance

Slide 28

Implications for Value-Added Models

- Accurately report the uncertainty in the estimated gains by including linking error
- Consider not using value-added models to make causal inferences, the data are too noisy.

Slide 29

Our Next Challenge

- To better understand nature of linking error
- To incorporate this error into VAM estimation
- Distribute *AM* IRT package