

comments on papers by
Dale Ballou and
Dan McCaffrey & J.R. Lockwood

Prepared for the National Academy of Sciences
Committee on Value-Added Methodology for
Instructional Improvement, Program Evaluation,
and Accountability

sean f. reardon

13-15 November 2008

analytic issues in VAM

- .. VAM is a causal inference problem
 - ⊗ and, hence, a modeling problem (in the absence of random assignment)
- .. VAM is a measurement problem (or problems)
 - ⊗ how to measure gain?
 - ⊗ can we precisely measure teacher/school effects?
- .. VAM is a system design problem
 - ⊗ yields signals, intended to create incentives
- .. these are not entirely separable problems

VAM is a causal inference problem

- .. issues raised by B & ML papers:
 - ⊗ non-random (and non-ignorable) sorting of students among teachers, schools, programs
 - n esp. note Rothstein (2008), see also Koedel & Betts (2008): evidence of dynamic sorting
 - n not clear that current models can fix this (pace Kane & Staiger)
 - n can we quantify the size of the bias?
 - ⊗ defining the estimand
 - ⊗ fixed-effects models limit possible comparisons and limit generalization
 - n are they useful for intended purposes of VAMs?

VAM is a causal inference problem (cont.)

- additional analytic/modeling issues
 - ⊗ implicit estimand
 - ⊗ heterogeneity of effects
 - n most current modeling approaches assume homogeneity of effects
 - n need better evidence about validity of this assumption
 - n if invalid, what is the right estimand? and what are the implications for design
 - ⊗ functional form/common support
 - n particularly a concern if effects are heterogenous
- additional problems if test scaling assumptions are not met
 - ⊗ models based on ordinal statistics? very inefficient

VAM is a measurement problem

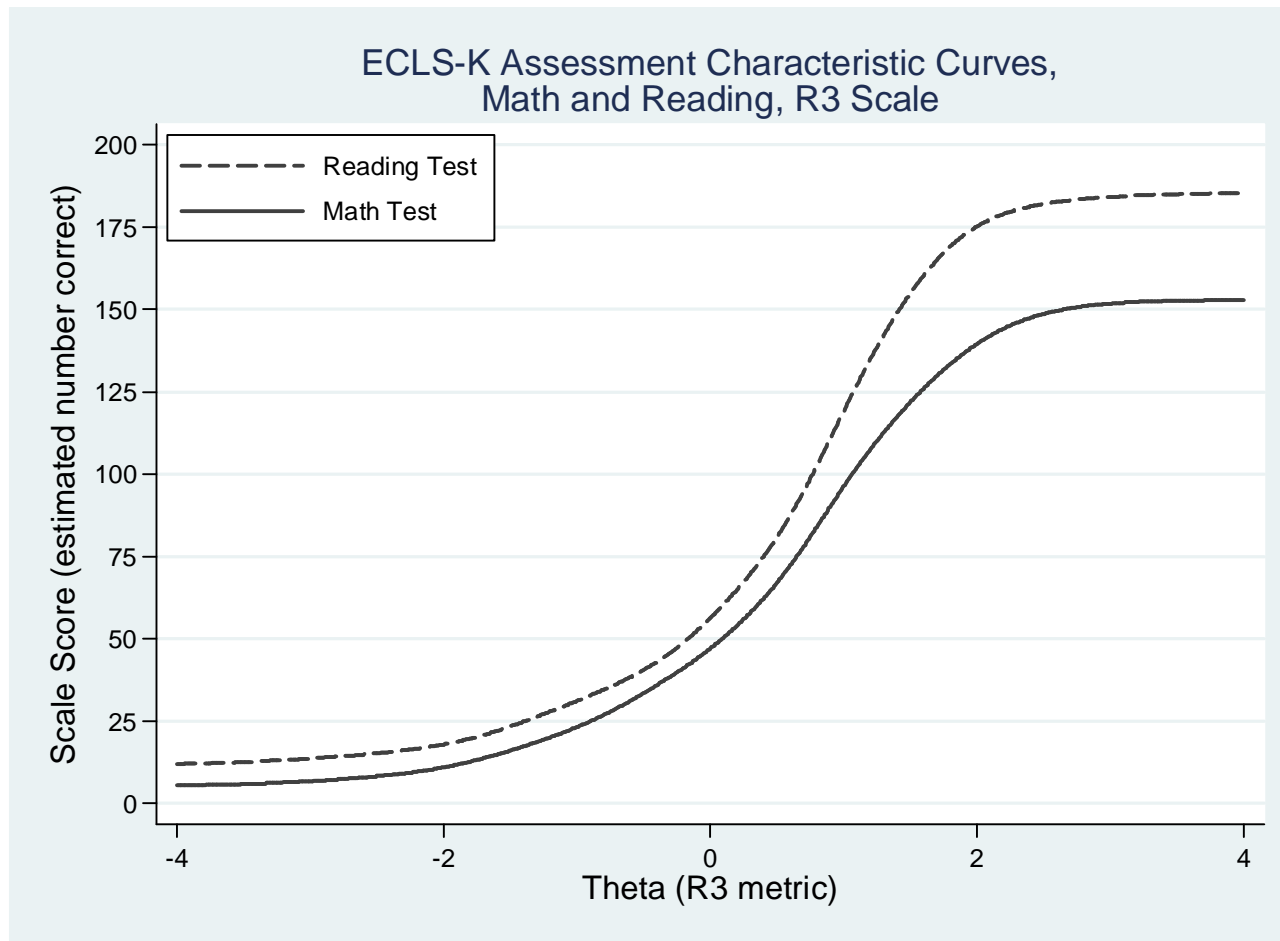


- VAM attempts to provide estimates of teacher or school effects
- even if unbiased, are they precise enough to be useful?
 - ⊗ school estimates are more precise
 - ⊗ teacher estimates more precise when pooled over multiple years
 - ⊗ how precise is precise enough? depends on the use of the results

VAM is a measurement problem (cont.)

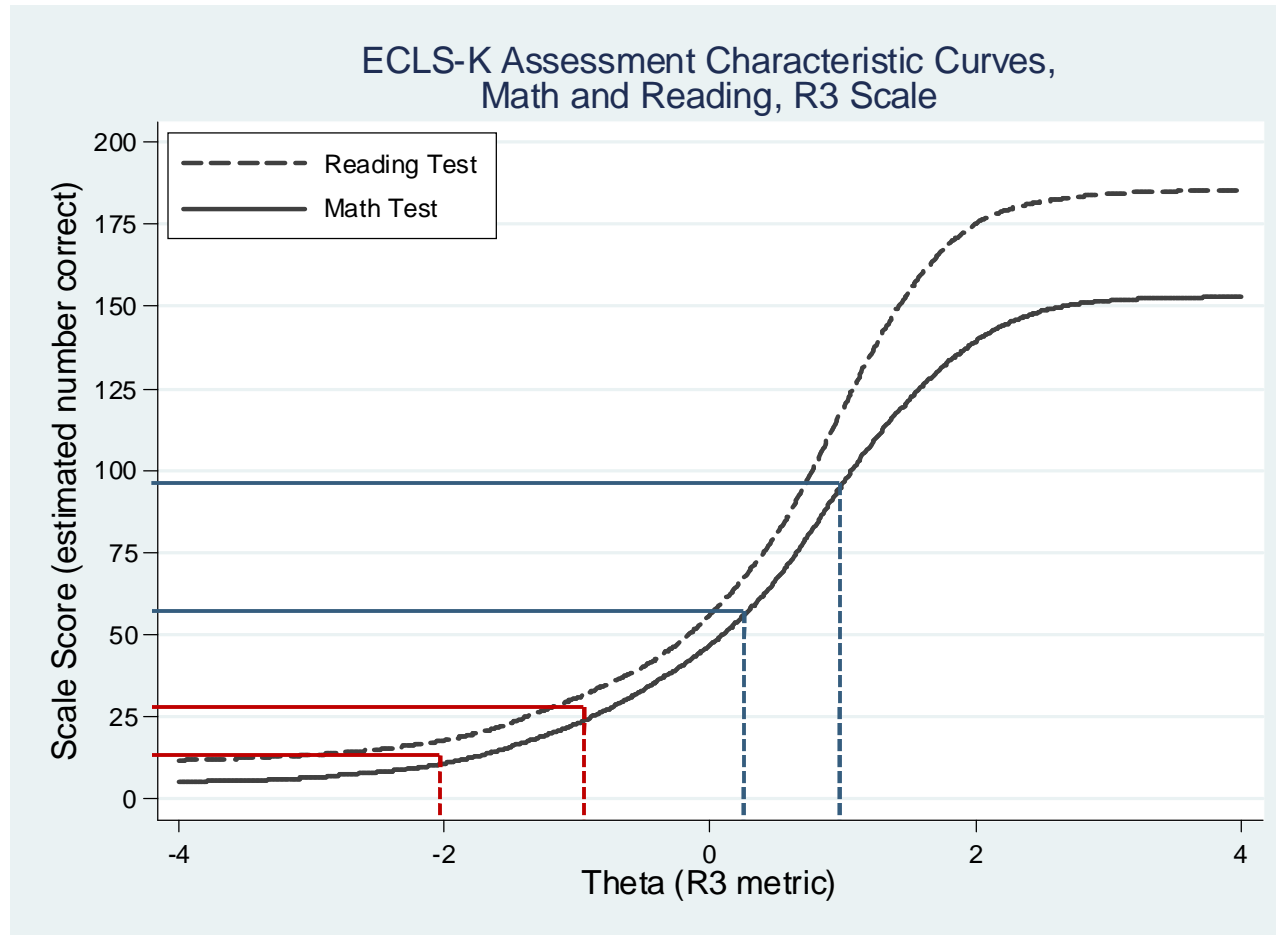
- VAM relies on meaningful measurement of achievement or achievement gains
 - ⊗ all current models rely on interval-scaling
 - ⊗ what does that mean? is it observable?
 - n not the same as vertical-scaling
 - n not the same as IRT-scaled
 - n not the same no test ceiling/floor
 - ⊗ interval-scaled means it is linearly related to social value: we value one-point gain equally regardless of where on the metric it falls
 - n equal average time to gain?
 - n equal effort to move?
 - n equal payoff in wages?

ECLS-K test metric transformation



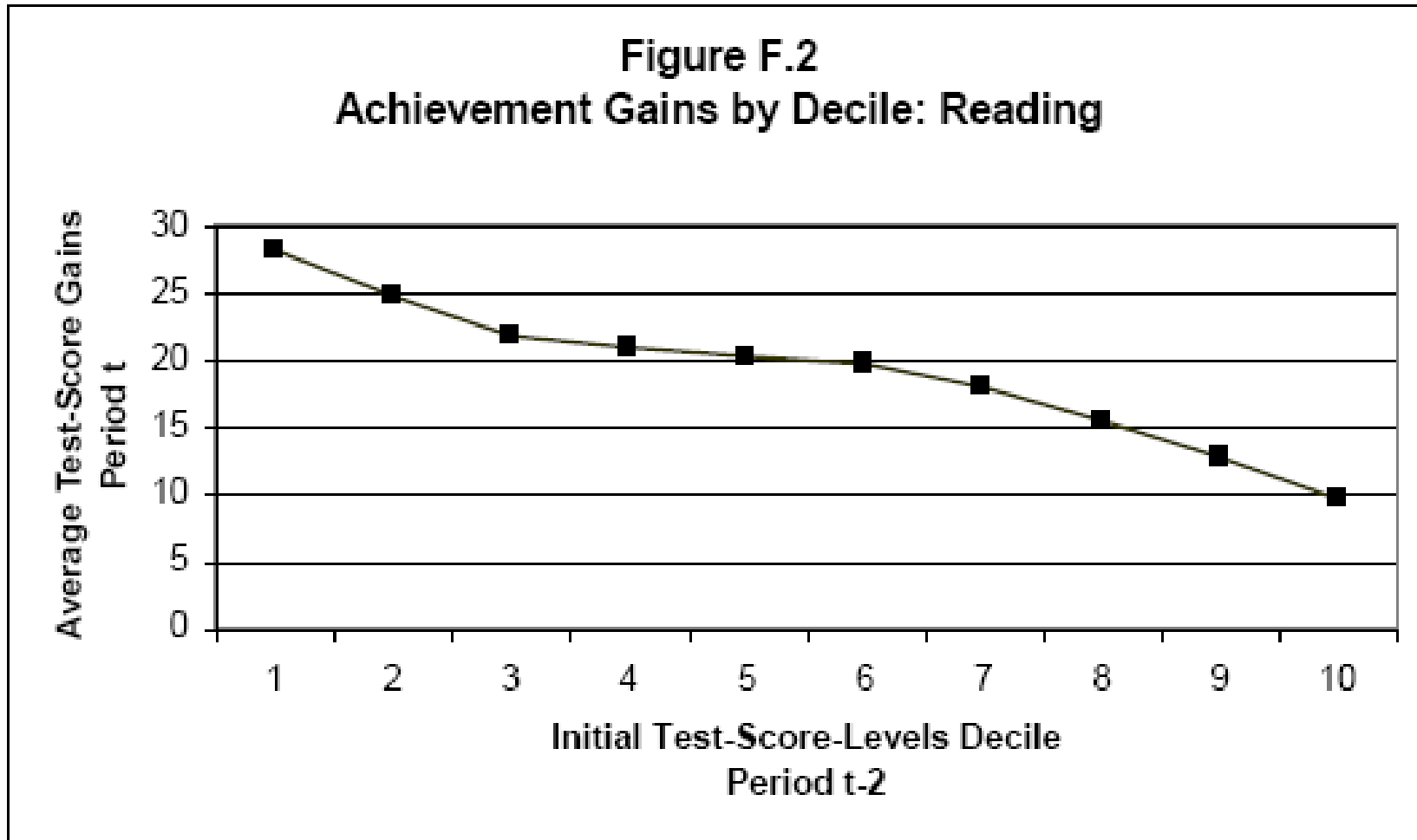
Source: *Reardon (2008)*

ECLS-K test metric transformation



Source: *Reardon (2008)*

non-interval scale or non-random sorting?



Source: Koedel & Betts, 2007

VAM is a system design problem



- design of feedback/incentive/accountability system should drive design of models
 - ⊗ teachers or schools?
 - ⊗ what stakes?
 - ⊗ what teacher/school comparisons?
 - ⊗ each implies incentives for teacher and school behavior
- how to balance accuracy/precision vs. transparency/signaling

VAM is a system design problem (cont.)



- .. an additional, more important causal inference problem:
- .. what is the effect of implementing a given VAM accountability system versus an alternative system?
 - ⊗ the VAM focus has been on estimating effects of teachers/schools, but the underlying theory is that these provide information that will affect educational processes in positive ways

conclusions

- VAM models should be designed with specific system design in mind
- worry about random assignment, because non-random sorting exacerbates all other problems
 - ✧ i.e., scaling, heterogeneity, bias due to measurement error, non-vertical scaling, etc.
 - ✧ ignorability doesn't solve all of these concerns
- worry about test metrics, because non-interval metric exacerbates bias from non-random sorting, heterogeneous effects, etc.
- design studies/policy change to estimate effects of uses of VAM