

TWO FOR THE PRICE OF ONE? PROBLEMS WITH THE DUAL USE OF ASSESSMENTS

DEREK NEAL

UNIVERSITY OF CHICAGO

DEPARTMENT OF ECONOMICS &
THE COMMITTEE ON EDUCATION

THP

9/27/2011

Two Purposes of Tests

STUDENT ACHIEVEMENT

- Consistent scaling over time is required to measure trends
- Requires links between assessments given in different years
 - Repeated items
 - Consistent formats
- => ***PREDICTABILITY***

EDUCATOR PERFORMANCE

- No scaling is required. Barlevy and Neal (2011) show that rank contests are enough.
- Alignment is key
 - i.e. the best strategy for avoiding sanctions and earning rewards is to adopt best practice teaching
 - But, ***predictability*** means that coaching is the best strategy for teachers

- The predictability that makes consistent scaling possible *in theory* invites the coaching that contaminates scale *in practice* given high stakes testing.
- Further coaching diverts class time away from activities that are more beneficial for students.
- Thus by trying to create **both** measures of student achievement and educator performance using one assessment system, policy makers doom themselves to fail at **both** tasks.

CAMPBELL'S LAW (1979)

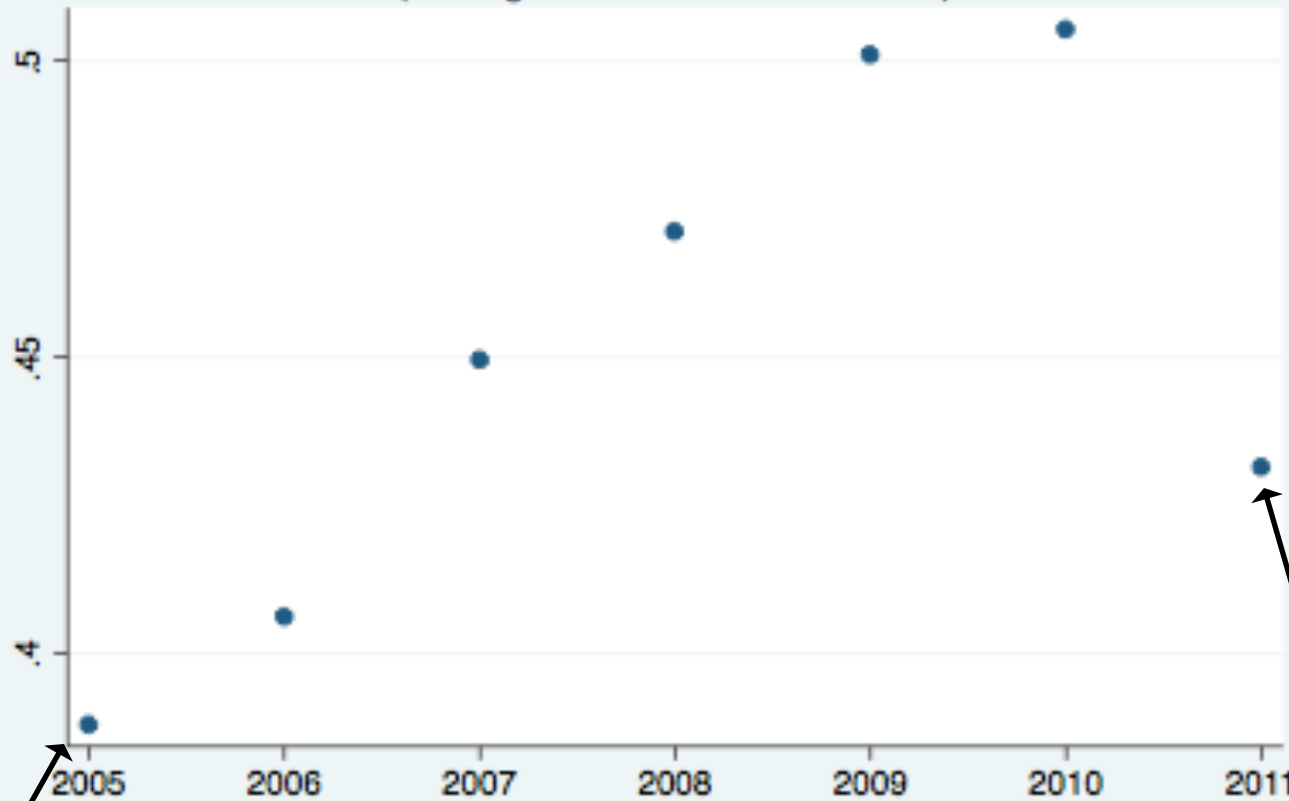
“I come to the following pessimistic laws (at least for the U.S. scene): The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”

(See also Heckman et al (2011), Neal (2011) & Rothstein (2009))

- **THE SMARTER (SBAC) AND PARCC CONSORTIA UNDER RACE TO THE TOP ARE TRYING TO FIGHT “THE LAW.”**
- **BASED ON RECENT DATA FROM THE UNIFORM CPA EXAM, I DO NOT LIKE THEIR CHANCES.**

Case Study: The Uniform CPA Exam

Figure 2. CPA Pass Rates - REG
(Average of 1st and 2nd Quarters)



Introduction of
computer-based
test format

New computer-
based test format

Designing Assessments to Avoid Coaching

- **Do not repeat questions:** When questions are repeated from one year to the next teachers face strong incentives to drill students repeatedly on the exact set of questions used on previous tests.
- **Vary the formats:** There is clear evidence that when students learn material in one format, their command of the material can be limited to the format in question.
- **Avoid or limit multiple-choice questions:** Any scoring rule for multiple-choice questions includes penalties for incorrect answers. Then there is an optimal test-taking strategy that specifies when students should guess and when students should leave questions blank, which could lead teachers to spend class time on these strategies.

Better Way to Provide Incentives for Educators

- *Step One:* Place each student in a comparison set based on expected achievement.
- *Step Two:* At the end of the year, rank all students based on their test scores and assign a percentile.
- *Step Three:* Form the average percentile over all students in a school that are taking a particular subject (e.g., fifth grade math). This is a percentile performance index, PPI.

**ECONOMICS MODELS OF CONTESTS PROVIDE GUIDANCE
CONCERNING HOW TO ATTACH REWARDS OF PENALTIES
TO PPI.**