

# Testing and Schooling Outcomes

October 24<sup>th</sup>, EdChoice Academy

*C. Kirabo Jackson, Northwestern University and NBER*

# Open Questions

- Why do we want to measure student outcomes at all?
  - To identify those who are at risk of failure or success?
  - To evaluate teachers and schools?
    - Do we want to attach stakes to it?
  - To best tailor interventions?
- Do we really want to measure skills, or simply predictors of success?
- How should we deal with gaming and cheating?
  - Can this be overcome?
  - Are some outcomes more susceptible to this than others?
  - Can a gameable outcome be useful?

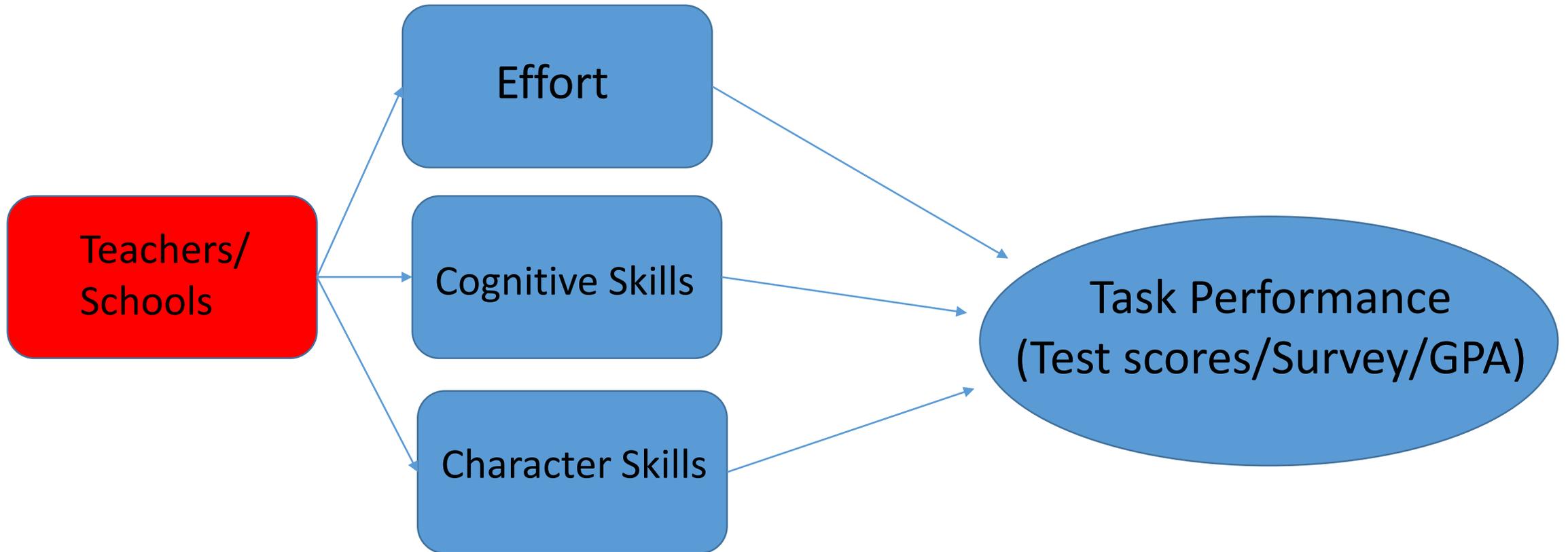
# How Do We Define “What is a good”?

- The concept of a good teacher or school “*as one who is best able to get learners to master the content of the school's subject is becoming rather generally accepted*” (Harl Douglas 1958).
- “*A good teacher is one who thinks properly in terms of the preparation of young people for life- preparation of young people to perform well in the situations which they will face as citizens, as workers, as members of homes, in their leisure life, and in their attainment of physical and mental health*” (Harl Douglas 1958).
- “we conceive of teacher (or school) quality as the ability to increase students’ stock of human capital, however that may be achieved—better communication to students, classroom management, encouragement of greater effort by students or parents, etc.” (Jackson, Rockoff and Staiger 2013)

**Often, we cannot observe the long run outcomes we care about, so we rely on short run outcomes to serve as proxies. What is a good proxy?**

# Framework for Thinking About Measures

(Everything is just a performance of a task!)



# Paper: What Do Test Scores Miss

1. Data: All 9<sup>th</sup> grade public school students in North Carolina from 2005 - 2012.
2. Use non-test score skill measures (behaviors) to form a proxy for skills not well measured by standardized tests
  1. I use 9<sup>th</sup> grade GPA, whether they enrolled in 10<sup>th</sup> grade on time (i.e. retained) , the log of absences in 9<sup>th</sup> grade, if suspended during 9<sup>th</sup> grade.  
**G-R-A-D index!**
3. Estimate 9<sup>th</sup> grade Math and English teacher effects on both test-scores and behaviors (value-added models).
4. Investigate how well test-score measures and non-test score measures of teacher quality predict teacher effects on longer-run outcomes.

# Using Averages of the Skill Measures as Predictors

**Table 2: Predicting Long Run Effect Using 9<sup>th</sup> Grade Outcomes (short version)**

	7
	Intend 4yr
Average Test Score	0.0341** [0.00115]
Behaviors Index z-score	0.0743** [0.000645]
Observations	468,015
Robust standard errors in parentheses	

If any one skill measure is missing, the observation is dropped with the individual skill measures because observations with any single missing skill measure are dropped when all are included.

OK...this seems reasonable. But these are simply correlations!

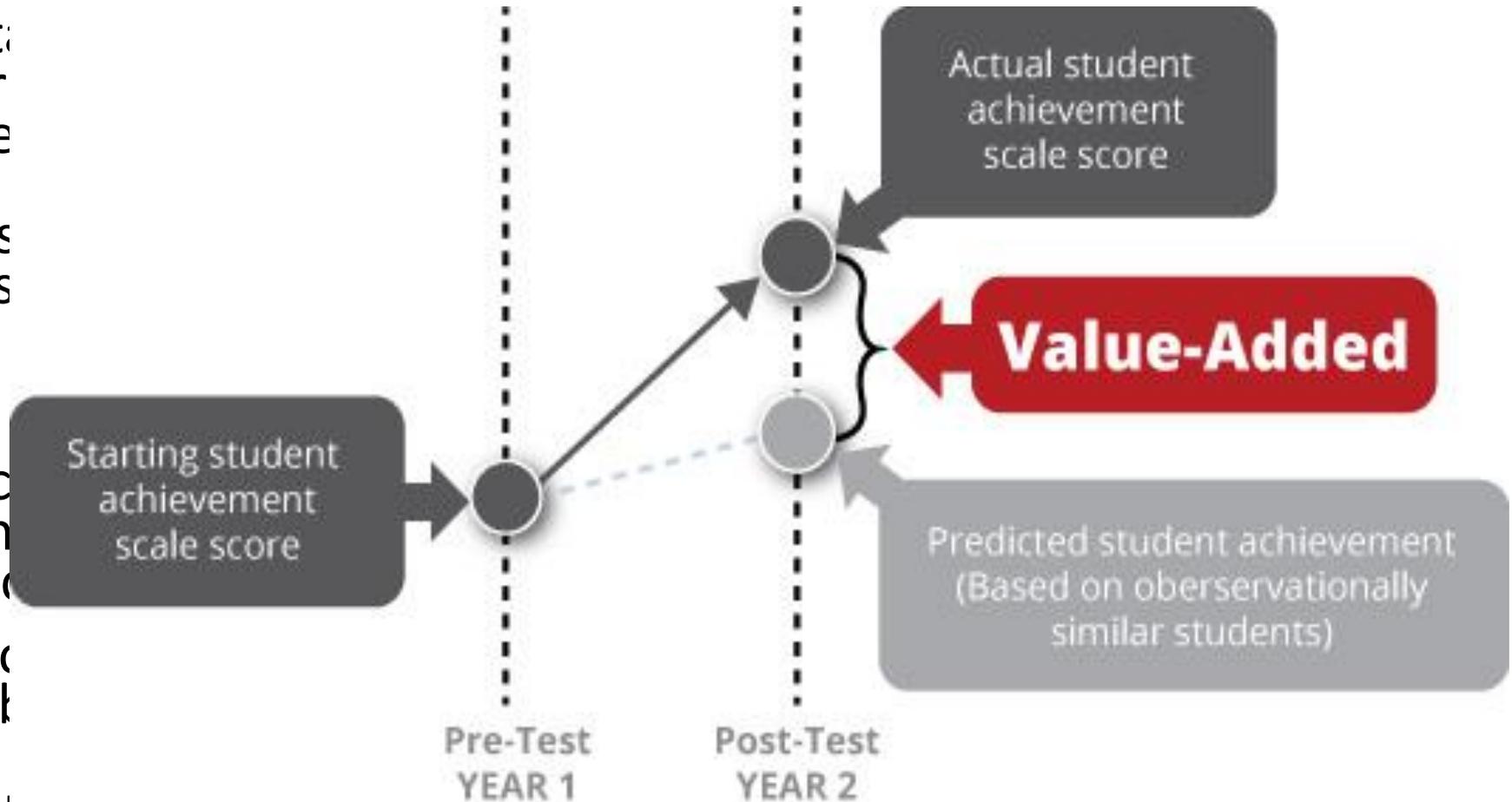
This does not mean that teachers who *improve* these measures will actually *improve* longer-run outcomes.....

# The Logic of Value-Added

- With enough data from previous year (or education, gender, or year scores of a relatively precise any student will score next year.

- If ONE student score is expected, it might be that student did participate.

- However, if ALL (or most) classroom scores can be confidently predicted, then we can do with the classroom.



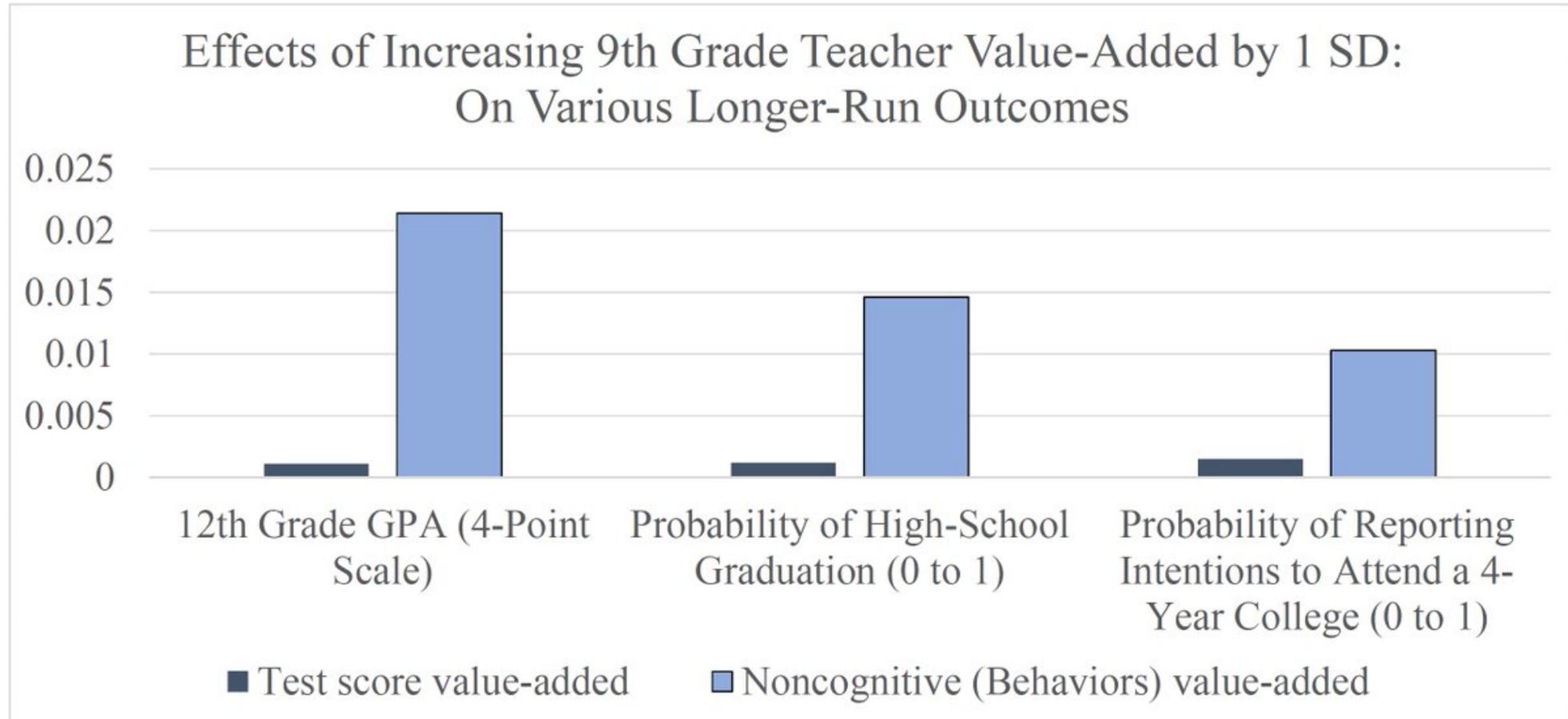
# Do Teacher Effects (value-added) on Test Scores and Effects on Behaviors Measure the Same Thing?

**TABLE 4**  
CORRELATIONS BETWEEN ESTIMATED TEACHER EFFECTS

	Teacher Effect: Test Score	Teacher Effect: Suspended	Teacher Effect: Absences	Teacher Effect: GPA	Teacher Effect: In 10 <sup>th</sup> Grade On time	Teacher Effect: Behavior index	Teacher Effect: 10 <sup>th</sup> Grade GPA
Teacher Effect: Test Score	1						
Teacher Effect: Suspended	-0.0713	1					
Teacher Effect: Absences	-0.0352	0.0856	1				
Teacher Effect: GPA	0.2206	-0.1361	-0.0924	1			
Teacher Effect: In 10 <sup>th</sup> Grade On time	0.144	-0.1167	-0.052	0.385	1		
Teacher Effect: Behavior index	0.2206	-0.4149	-0.2493	0.7484	0.6498	1	
Teacher Effect: 10 <sup>th</sup> Grade GPA	0.122	-0.0391	-0.0304	0.3971	0.1116	0.2911	1

Note: This table reports the estimated two-way correlation coefficient between the estimated teacher effects ( $\hat{\mu}_{zjt}$ ) on each outcome and their effects on each other outcome.

# So Do Teacher Impacts Predict Longer run Outcomes?



# What do We Learn From This Study?

- We knew that test scores measure some skills that matter.
- We knew that test scores did not capture everything.
- What we didn't know was how much of a teacher's effect was missed by test scores (at least 75% is not captured by test score value-added).
- We didn't know that we can do a much better job of capturing a teacher's impact using non-test score behaviors.
- However, we still don't know why some teachers improve non-test-score outcomes.
- We still don't know how some teachers improve non-test-score outcomes.



# Discussion Questions

- Why do we want to measure student outcomes?
  - To identify those who are at risk of failure/success?
  - To evaluate teachers and schools?
    - Do we want to attach stakes to it?
  - To best tailor interventions?
- Do we really want to measure *skills*, or simply predictors of success?
  - Why not just use survey measures since these are designed to measure skills?
  - How do surveys and behaviors compare in predicting longer run success?
  - Why even use surveys at all?
- How do we deal with gaming and cheating?
  - Can this be overcome?
  - Are some outcomes more susceptible to this than others?
  - Can a gameable outcome be useful?