Designing Incentive Systems for Schools

By Derek Neal

University of Chicago and NBER

THREE ISSUES

SOCIAL VALUE

AGREGATING INFORMATION

CORRECTING MISTAKES

THE VALUE OF SCHOOL OUTPUT

Skills are indexed by k = 1, 2,.. K.

Schools indexed s = 1,2, … S.

Each school has N students that are indexed by n = 1, 2, ….N.

X^s **is the output of school s with elements** x_{nk}^s**,**

Define $X = \{X^1, X^2 .... X^S\}$ **as the collection of all skill measurements for all students in all schools at a point in time.**

Define $X^* = \{X^{*1}, X^{*2} .... X^{*S}\}$ **as the next set of measurements taken over the same students**

KEY QUESTIONS

If society began the school year at X, how does society evaluate the relative values of ending the year at any one of the infinite possible X^* outcomes?

Further, if we take the matrices of test scores from the beginning and end of the school year for any two schools, how do we use these four matrices to decide which school performed better?

MISSING PRICES ARE A KEY PROBLEM

If policy makers knew not only how to measure but also how to value the output of schools, state departments of education could operate much like the Department of the Interior, which auctions off the right to harvest trees on public land and then allows timber companies to sell the trees that they harvest and transport at the market price for delivered timber.

Because educational policy makers cannot express school output in dollar terms, they have no rational basis for constructing pay for performance systems that look like piece rates systems. Ranking schools or teachers is likely the best that one can do.

THREE DESIRABLE PROPERTIES

The documents describing any accountability or incentive system should spell out the priorities of policy makers.

The mapping between the policy priorities that define the system and the procedures used to create performance rankings for schools and teachers should be clear and precise.

Accountability systems should group schools according to the types of students and families they serve, and then rank schools either relative to other schools that serve similar students or to a performance standard designed for such schools.

1) CLARITY OF CONCERNING PRIORITIES

EXAMPLES OF ISSUES THAT ARISE:

Is progress in reading more or less valuable than progress in math or civics, and if so, how much?

Is it more socially valuable to bring a disadvantaged student closer to grade level or to bring a gifted student closer to her full potential, and if so, how much?

What are the relative values of non-cognitive traits like persistence versus cognitive skills?

2) COHERENCE BETWEEN PRIORITIES AND PROCEDURES

EXAMPLES OF LESS THAN IDEAL IMPLEMENTATION

NCLB

FLORIDA

CALIFORNIA

Figure 1

Elementary Reading Value Table

Elementary Reading							
	Year 2 Level - 2006						
Year 1 Level 2005	1a	1b	2	3	4	5	Average Score
1a	0	100	455	550	675	725	100.0
1b	-50	50	145	265	340	500	100.0
2	-100	-50	125	205	245	350	100.1
3	-175	-100	-90	170	210	250	100.2
4	-200	-150	-140	-75	195	215	100.0
5	-250	-200	-160	-125	25	210	100.2
All Levels							100.1

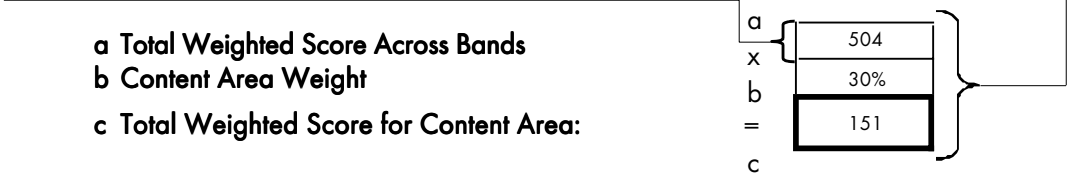
- **Step 3:** Repeat Steps 1 through 4 for each remaining content area.

Stanford 9		Language		Spelling		Mathematics	
A	B	E	F	G	H	I	J
Performance Bands	Weighting Factors	Percent of Pupils in Each Band	Weighted Score in Each Band (B x E)	Percent of Pupils in Each Band	Weighted Score in Each Band (B x G)	Percent of Pupils in Each Band	Weighted Score in Each Band (B x I)
5 80-99th NPR	1000	10%	100	5%	50	5%	50
4 60-79th NPR	875	10%	88	10%	88	10%	88
3 40-59th NPR	700	30%	210	25%	175	25%	175
2 20-39th NPR	500	30%	150	35%	175	35%	175
1 1-19th NPR	200	20%	40	25%	50	25%	50

- **Step 4:** Sum the weighted scores across performance bands. The Total Weighted Score Across Bands for Reading is 504.

- **Step 5:** Multiply the Total Weighted Score Across Bands by its Content Area Weight to obtain the Total Weighted Score for Content Area (a x b = c). In this example, the Total Weighted Score for the Content Area of Reading is 151.

Stanford 9		Reading	
A	B	C	D
Performance Bands	Weighting Factors	Percent of Pupils in Each Band	Weighted Score in Each Band (B x C)
5 80-99th NPR	1000	5%	50
4 60-79th NPR	875	5%	44
3 40-59th NPR	700	25%	175
2 20-39th NPR	500	35%	175
1 1-19th NPR	200	30%	60



NPR = National Percentile Rank

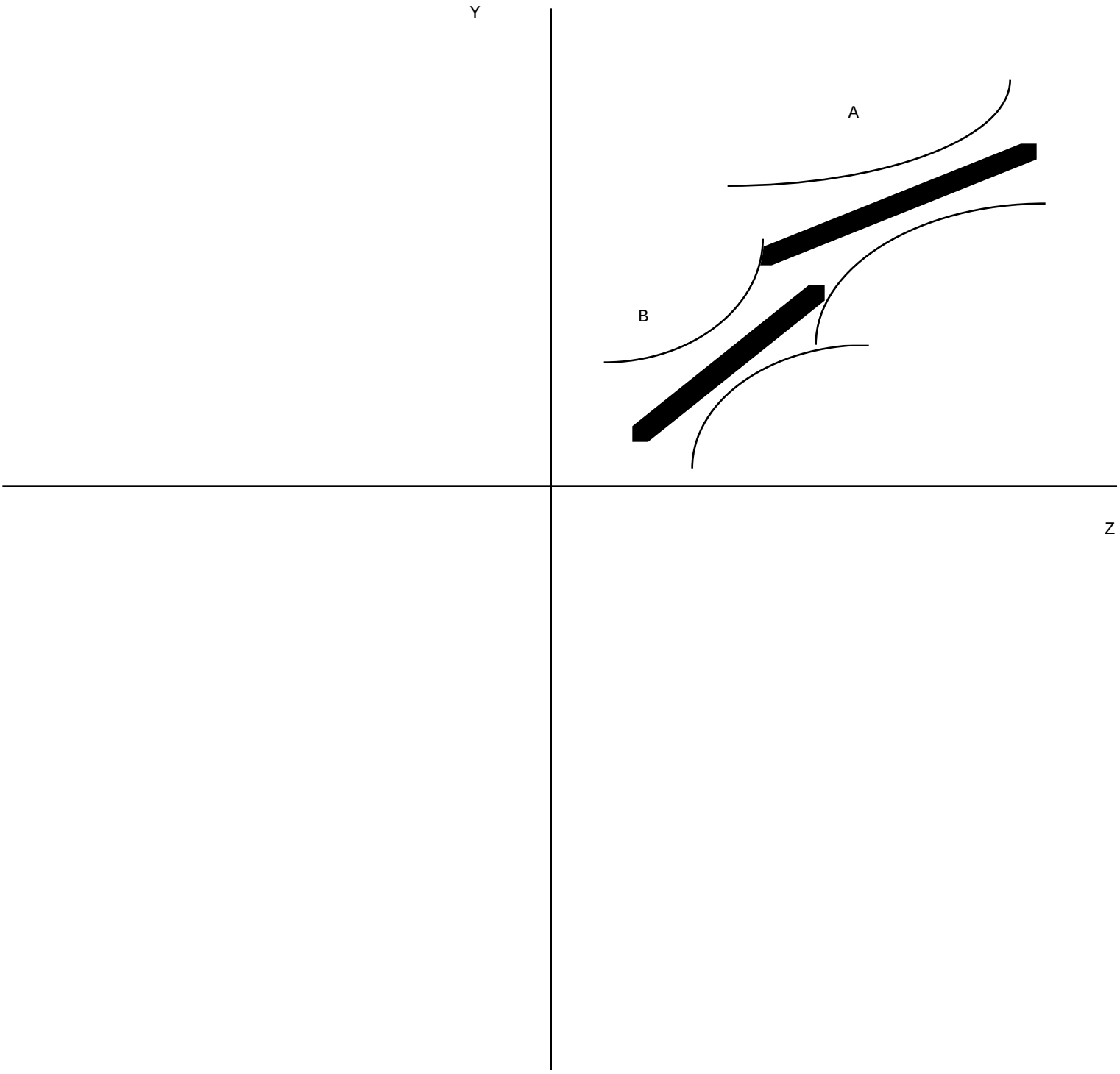
BECAUSE THE RELEVANT PRICES ARE MISSING, ANY SYSTEM IS BOUND TO HAVE ARIBITRARY ELEMENTS.

IS OUR INABILITY TO AVOID OUTCOMES THAT APPEAR ARBITRARY AND CAPRICIOUS EX POST A PRIMA FACIE REASON TO LOOK FOR OTHER MECHANISMS?

3) COMPARISONS AMONG AGENTS WORKING IN HOMOGENEOUS ENVIRONMENTS

FIGURE 2

Figure 2



WHAT IS THE COUNTERFACTUAL?

The USGA collects the data that drives their handicap system by sending out sets of golfers to play multiple courses. If the system works perfectly, one can use a player's handicap, the course rating, and the slope rating for a given course to predict how the player's expected score on a given course compares to the expected score of a scratch golfer on that course.

However, what do we learn from knowing that school A received a higher accountability rating than school B if we know that school A and school B serve extremely different populations of students? When a state accountability system tells us that some school in an affluent suburb performed better or worse than some school in an economically disadvantaged inner-city neighborhood or another school in a less affluent rural town, what exactly is it telling us? And, why is the answer to this question useful in designing systems that reward schools and teachers for performance on the jobs they actually have?

IF GOD GAVE US THE TRUE VAM EFFECTS FOR BOTH SCHOOLS WHY WOULD WE CARE?

CONSIDER MATRICES OF THE FOLLOWING FORM AS A BASIS FOR RANKING SCHOOLS

	School 1	School 2	School 3	CUTOFF
90th percentile				
80th percentile				
70th percentile				
60th percentile				
50th percentile				
40th percentile				
30th percentile				
20th percentile				
10th percentile				
TOTAL SCORE				

WORKING WITH RANKS WITHIN LEAGUES

1) **LESS ROOM FOR ARBITRARY CHOICES CONCERING SCALES AND WEIGHTS TO AFFECT RANKINGS.**

2) **COMPETITION REVEALS FRONTIER**

3) **LEAGUES MAKE COMPARISIONS MEANINGFUL**

4) **MATCHING OF SCHOOLS AND TEACHERS**

WHEN CONSIDERING HOW DIFFERENT SYSTEMS AGGREGATE INFORMATION: DO WE GAIN INFORMATIONAL ADVANTAGES FROM USING SYSTEMS BASED ON SCHOOL RATHER THAN TEACHER PERFORMANCE?

1) ***Technical problems associated with measuring the performance of individual teachers are only half the story. Incentive systems require not only performance statistics but also a mapping between these statistics and the rewards and penalties that workers receive. In other professional labor markets, competition among firms for talented workers determines the form of incentive pay we observe in equilibrium given available measures of job performance. In the absence of this competition, it is hard to see what mechanism will reveal that optimal way to link rewards and sanctions to any set of performance statistics regardless of the quality of the statistics in question.***

Lazear and Rosen (1981)

2) Principals should compete with each other not only in determining the educational practices used within their schools but also in terms of developing and implementing the personnel policies and procedures that identify and retain the best teachers.

3) Recent work by Jacob and Lefgren (2007) indicates that principals can predict which classrooms in their schools are likely to experience the largest achievement gains during a year

It seems reasonable to assume that not only principals but also other teachers in a school possess a great deal of information, not fully captured by any single set of test scores, concerning the performance of other teachers and how their performances could be improved.

4) When accountability systems involve school-level reward pay, individual teachers have incentives to help their peers improve.

When principals have discretion over pay, they can build reputations for rewarding cooperation and punishing destructive forms of competition among teachers. They have strong incentives to pursue this course of action if their pay and job security depend on their schools' overall performances rankings.

5) Team incentive systems are quite common in the private sector, and many professional partnerships operate in a manner that fits roughly within this framework if one views school principals as managing partners, tenured teachers as partners, and new teachers as associates.

Identifying, training, and retaining talented teachers is key to running an effective school, and these tasks are too difficulty to accomplish within systems that only make use of information that is contained in objective student assessments.

POTENTIAL OBJECTION:

A common system for mapping objective assessment outcomes into rewards and punishments for teachers is the only feasible policy option because public employees unions have historically opposed granting such discretion to administrators.

THERE ARE STILL BEENFITS from bonus payments delivered at the school level and shared equally by all teachers.

IT IS EASIER TO FORM RANKINGS OF SCHOOLS THAN TEACHERS

1) NOISE

2) SELECTION WITHIN SCHOOLS --

See Rothstein (2007) for recent evidence.

FURTHER: ***If one asserts that the assignment of teachers to students in any given school is ignorable given a standard set of student characteristics, one is really asserting that the principal of the school in question should be fired.***

ECONOMISTS HAVE SPENT THE LAST 30 YEARS LEARNING THAT YOU CANNOT ESTIMATE HOW TYPES OF JOBS EFFECT PAY PER SE WITHOUT AN EXPLICIT MODEL OF HOW PEOPLE'S CAREER DECISIONS EVOLVE AS THEY GAIN MORE INFORMATION. NONE OF THE VAM MODELS INCLUDE AN EXPLICIT MODEL OF HOW A STUDENT'S TEACHERS AND SCHOOLS ARE CHOSEN THIS YEAR BASED ON INFORMATION THAT PARENTS AND OTHERS LEARNED LAST YEAR.

3) AMONG SCHOOLS -- GEOGRAPHY CAN MITIGATE SELECTION CONCERNS

SELF-CORRECTING MECHANISMS

THE VALUE OF COMPETITION AMONG SCHOOLS IS ILLUSTRATED BY THE FOLLOWING QUESTION:

ARE WE CERTAIN THAT HIGH-STAKES INCENTIVES ARE DESIREABLE, AND HOW WILL WE KNOW IF WE ARE WRONG?

THIS QUESTION IS IMPORTANT BECAUSE IT IS NOT OBVIOUS THAT HIGH-POWERED INCENTIVES ARE THE OPTIMAL POLICY IN EDUCATION?

Holmstrom and Milgrom (1991)

Everyone knows the old saying that "you get what you pay for," but HM take this line of reasoning a step further and argue that, because effort is costly, if you pay for one type of effort, you may get less effort of other types. Does "slack in the system" make sense?

Baker (2002) presents an instructive case of the HM model with only two tasks.

In the context of teaching math, these tasks might be labeled as (i) reviewing concepts and question formats that will appear on the coming high stakes assessment versus (ii) activities that require students to use these concepts to write research reports, make presentations, or build equipment.

If the goal of the school is to increase the value of math skills that students bring to adulthood, should the school be willing to attach high-stakes for teachers to the results of the coming math assessment?

Maybe Yes … but Maybe No

THREE ISSUES ARISE:

SCALE

ALIGNMENT

MEASUREMENT ERROR

NOTE THESE ARE NOT ADDITIVE EFFECTS IN ANY SENSE. WEAK INCENTIVES ARE OPTIMAL WITH NOT MEASUREMENT ERROR IF ALIGNMENT PROBLEMS ARE SEVERE.

FURTHER, THERE ARE OTHER REASONS THAT HIGH STAKES MAY NOT BE OPTIMAL

Besley and Ghatak (2005) note that many non-profit organizations that provide education, health, or related services choose personnel policies that include relatively little incentive pay.

Besley and Ghatak argue that this outcome may imply that it is efficient to devote considerable resources to screening potential hires, and then only hire those with high levels of personal commitment to the mission of the organization. If it is possible to identify such individuals, incentive pay is no longer necessary.

The most important task for those who run schools may be to identify and retain talented persons who love to help children learn, and real competition among schools may simply reveal which principals know how to identify these teachers.

TWO FACTORS SHOULD CREATE PAUSE:

The EMPIRICAL absence of high-powered incentives tied to performance statistics in most professional labor markets.

The POLITICAL reality that we cannot count on competition as a force for weeding out ill-framed accountability systems.

AT A MINIMUM WE NEED MORE REALITY (ACCURACY) IN OUR RHETORIC

Accountability systems do not bring "business practices" or "competitive pressures" to public education.

In the traditional public school model, teachers and principals, as public employees, are accountable to the appointed agents of elected officials. In accountability systems, teachers and principals are accountable to formulas and procedures that are created by these same agents.

If a state or large district implements a seriously flawed accountability system, what competitive pressures will force them to abandon it before it has done great damage?

IN CONTRAST: Professionals in the private sector are almost always given reward pay based on wide-ranging assessments by "principals" or the agents of "principals" who really are "principals" in that they lose THEIR OWN money if they or their agents make the wrong decisions.

THE GREATEST DANGER IS THAT WE WILL SEE WIDE ADOPTION OF SOME RULES-BASED PAY FOR PERFORMANCE SYSTEM IN AN ENVIRONMENT WHERE THERE ARE NO COMPETITIVE PRESSURES THAT FORCE THE ABANDOMENT OF A BAD APPROACH.

WE NEED TO WORRY SERIOUSLY ABOUT A SPECIFIC AND FORUMLAIC APPROACH TO ACCOUNTABILITY BECOMING THE "MASTER'S DEGREE" POLICY OF THE 21 CENTURY