SPEEDING, TERRORISM, AND TEACHING TO THE TEST*

Edward P. Lazear

Educators worry that high-stakes testing will induce teachers and their students to focus only on the test and ignore other, untested aspects of knowledge. Some counter that although this may be true, knowing something is better than knowing nothing and many students would benefit even by learning the material that is to be tested. Using the metaphor of deterring drivers from speeding, it is shown that the optimal rules for high-stakes testing depend on the costs of learning and of monitoring. Incentives need to be concentrated for those whose costs of action are high. For high cost learners this implies announcing the exact requirements of the test. For more able students, a more amorphous standard produces superior results. This is analogous to announcing where the police are when the detection costs are high. Other applications are discussed.

High-stakes testing, where teachers, administrators, or students are punished for failure to pass a particular exam, has become an important policy tool. The "No Child Left Behind" program of the George W. Bush administration makes high-stakes testing a centerpiece of its approach to improving education, especially for the most disadvantaged. Proponents of high-stakes testing argue that testing encourages educators to take proper actions and that testing also identifies those programs that are failing.[1] But critics counter that high-stakes testing induces educators to teach to the test, which has the consequent effect of ignoring important areas of knowledge.[2] Almost every

* This research was supported by CRESST. I am indebted to Edward Glaeser for fundamental suggestions on the framework and for some derivations contained in the paper. In addition, George Akerlof, Gary Becker, Thomas Dohmen, Richard Freeman, Eric Hanushek, Caroline Hoxby, Lawrence Katz, Paul Oyer, Paul Romer, Kathryn Shaw, Andrzej Skrypacz, Michael Spence, and Steven Tadelis were especially helpful in providing comments. Benjamin Ho provided excellent research assistance on the formal analysis, and along with Ron Siegel, provided detailed review and comments.

1. Identification is particularly important if, as Rivkin, Hanushek, and Kain [2001] find, teacher-specific effects go a long way in explaining the performance of their students.

2. See Koretz et al. [1991] discussed in more detail below.

Hoffman, Assaf, and Paris [2001] report on results from Texas Assessment of Academic Skills testing. Using a sample of 200 respondents, they suggest that the Texas exam has negative impacts on the curriculum and on its instructional effectiveness, where eight to ten hours per week on test preparation is typically required of teachers (by their principals) and the curriculum is planned around the test subjects. They also argue that teaching to the test raises test scores without changing underlying knowledge.

Jones et al. [1999] study data from North Carolina and conclude from a survey of 236 participants that the high-stakes test induced two-thirds of teachers

The Quarterly Journal of Economics, August 2006

teacher is familiar with the question, "Will it be on the final?" The implication is that if it will not be on the final, the student will not bother to learn it.

Which argument is correct? The main result of the following analysis is that to maximize the efficiency of learning, high-stakes, predictable testing should be used when learning and monitoring learning are very costly, but should not be used when learning and monitoring are easy.

The best way to focus the question is to examine another problem that is formally equivalent, namely that of deterring speeding.[3] Suppose that the city has available to it a given number of police, who patrol the roads. Should the city announce the exact location of the police or simply allow drivers to guess? At first blush, the answer seems obvious. Of course, their locations should be kept secret. If the locations of the police are announced, then motorists will obey the law only at those locations, and will speed at all other locations. But the answer is not obvious. If police are very few and their locations are unknown, drivers might decide to speed everywhere. If police locations are announced, there is a better chance that speeding will be deterred at least in those places where police are posted. The total amount of speeding could actually be lower when locations are announced.

Tax fraud is virtually identical. The tax authority can announce the items to be audited or just let taxpayers know that there will be random audits. In the absence of announcing specific items to be audited, taxpayers may cheat on all tax items, especially when there are few auditors and audits are unlikely. Instead, the authority can announce those items that will be audited with certainty and likely deter cheating on those items, which is better than failing to deter any cheating.

Terrorism presents a third application of the principle. A country's announcement that it has increased its effort to deter terrorism may have no effect on terrorism if those efforts are spread widely throughout the country. But focusing the increased

---

to spend more time on reading and writing and 56 percent of teachers reported spending more time on math. They also claim that students spend more than 20 percent of instructional time practicing for end-of-grade tests and a significant fraction report a reduction in students' love of learning.

In an early study, Meisels [1989] outlines some of the pitfalls of high-stakes testing and suggests adverse effects of the Gesell School Readiness Test and of Georgia's use of the CAT.

3. Beginning with Becker [1968], there is a large amount of literature on optimal incentives for enforcement of the law.

patrols on airports may at least have the effect of making air travel safer.

Teaching to the test is analogous because the body of knowledge is like all of the roads. Announcing the items to be tested is like telling drivers which miles of road will be patrolled. If the test questions are not announced, but instead some random monitoring is done, students will have to decide whether to study a large amount or very little. When they would choose to study very little or nothing, announcing what is on the test may motivate them to learn at least those items. With the exception of definitions and some other formalities, the problems are the same.

The issue is one of concentrating or spreading incentives. Concentrating incentives provides very strong motivation, but over a limited range of activity. Spreading incentives provides weak motivation, but over a broader range. When agents are responsive to incentives, spreading them over a large range encourages more of the desired action. But if agents are insensitive to incentives, then spreading them too thin might provide too little incentive to do anything, which results in a reduction in the amount of desired action.

The decision to concentrate or spread a given amount of incentive can be put generally and more formally. Suppose that we want to encourage some action A, like obeying the law, and that the decision to obey depends on the expected penalty P, according to $A(P)$. There is a broad space over which we can monitor activity, or we can restrict our attention to some fraction q of that space and let the other $1 - q$ be ignored. In the case of speeding, this is like announcing which q of the roads are patrolled and which $1 - q$ of the roads are not. When the full space is subject to stochastic punishment that has expectation P, then the agent takes action $A(P)$. However, if only q of the space is subject to punishment, then the expected penalty over that subset is $P/q > P$ since $q < 1$ and the action taken there is $A(P/q)$.[4] Incentives are strengthened over that part of the space. However, on the $1 - q$ of the space announced to be ignored, the penalty is 0, so the action is $A(0)$. Incentives are weakened on the part that is clear of patrol. Should incentives be concentrated over some subset q or spread over the entire range? Let there be some social

4. Note that $q > P$. When $q = P$, the probability of detection over this range is one. There is no advantage to having a smaller q than P.

value of the action A, given by $R(A(P))$. Then, it is better to spread the incentives broadly if

$$R(A(P)) > (1 - q)R(A(0)) + qR(A(P/q)). \tag{1}$$

If the inequality in (1) goes the other way, then it is better to concentrate over some subset q.

Expression (1) can be recognized as Jensen's inequality, which states that the left-hand side exceeds the right-hand side if and only if $R(A(x))$ is concave in x. Thus, incentives should be spread out if the R function is concave and concentrated if the R function is convex. Convexity of the R function relates to the responsiveness to incentives. As will be shown in the applications below, when individuals are responsive to incentives, the R function is concave. When they are less responsive, the R function is convex.

I. A Model of Speeding

I.A. *Deterring Speeding*

There are Z miles of road. A driver can either speed or obey the speed limits. Suppose that the extra utility that is derived from speeding is V per mile and that the fine for speeding, if caught, is K. There is a vast literature on optimal fines, but that is not the point of this example, so the fine is assumed to be given exogenously.[5]

Suppose that there are G police and that each policeman can patrol one mile of road. If police are distributed randomly along the road, then on any given mile the probability of being caught speeding is G/Z, and the expected fine from speeding is KG/Z. Thus, if drivers do not know the location of the police, they will speed if

$$KG/Z < V. \tag{2}$$

Since the cost and value of speeding on every mile is the same, if the driver chooses to speed on one mile, he speeds on all.

Now suppose that the location of the police along the roads is announced. A more general approach allows for some miles to be subject to patrol with some probability and others with some

5. In the teaching case analyzed below, the loss may be market determined, and then K is given exogenously to the student or teacher. As such, the model with exogenous fines is more appropriate for the main task of the analysis.

different probability, but to get the basic intuition, let us start with the more extreme version of the model. If roads are either patrolled or not, then drivers are certain to be caught if they speed on a patrolled section. As a result, no speeding occurs on the patrolled section as long as $V < K$, but speeding occurs on all nonpatrolled roads because the drivers know that the probability of detection there is zero. The law will be obeyed on G miles of road, and there will be speeding on the other $Z - G$ miles.

If locations are unannounced, there is either no speeding at all or always speeding, depending on whether the expected fine KG/Z exceeds or falls short of the utility value of speeding, V. But when locations are announced, there is speeding on $Z - G$ miles, but not on G miles as long as $K > V$. More speeding is deterred by announcing the location of police whenever

$$KG/Z < V < K.$$

If $KG/Z < V$, drivers would always speed if locations were secret because the probability of detection is sufficiently low, which makes the speeding gamble worthwhile. But announcing the locations deters speeding on G miles (since $V < K$) so this is the better outcome. If instead, $KG/Z > V$, the expected fine is sufficiently high to deter all speeding when locations are secret, and this dominates revealing locations.[6]

The intuition is simple. If police are few, drivers assume it very unlikely that they will be caught speeding and speed everywhere. Announcing locations of the police strengthens incentives on patrolled roads and at least deters speeding at those locations. If police are abundant and the probability of being caught sufficiently high, no one will speed. With many police, revealing their location induces drivers to speed on all roads except the G miles that are patrolled. So when there are many police, it is better to keep their locations secret; with few police it is better to reveal their locations and at least deter speeding on the few roads that are patrolled.[7]

6. In a related paper, Eeckhout, Persico, and Todd [2004] take the speeding deterrence approach quite literally and test it using data from Belgium. They find that their model, which uses intuition similar to the one in this paper, fits the data on speeding quite well.

7. This logic implies that as long as police are costly, there are an optimal number of police. When police locations are secret, it is never optimal to have more police than

$$G = VZ/K,$$

which makes (2) hold with equality so that cheating is completely deterred.

The structure can be generalized easily. Now allow q of the Z roads to be patrolled and $(1 - q)$ to be unpatrolled. The question can be stated in its most extreme form. Is there any q such that it is better to limit patrol to qZ of the miles and announce it, rather than randomly patrolling all Z miles? Reverting to (1), in this extreme case, $R(A)$ can be thought of as a linear function where A is the number of miles over which the law is obeyed and V is also linear in those miles. For example, if there were a social cost to speeding equal to γ, this would simply require that $V < \gamma$ on all miles. In this context, $A(P)$ is the number of miles on which speeding does not occur if the expected penalty is P. The concavity of $R(A(x))$ depends only on concavity of $A(x)$. If only qZ of the miles were patrolled, then the expected penalty on those qZ miles would be P/q, and the expected penalty on the unpatrolled miles would be zero. By (1) it is better to announce a limitation on patrolled roads when the $R(A(\,))$ function is convex. If it requires strong incentives to induce individuals to do anything, then the $R(A(\,))$ function is convex (see Figure Ia). When individuals are very responsive to incentives, the $R(A(x))$ function is concave in the relevant range between 0 and P/q (see Figure Ib).

In Figure Ia individuals speed when punishment is P, but obey if punishment is P/q. In Figure Ib individuals obey if punishment is P and obviously continue to obey at punishment P/q. The $A(x)$ function is therefore convex over the relevant range in Figure Ia, but concave in Figure Ib.

When the costs of detection are high or penalty is low relative to the gains from speeding, Figure Ia is relevant. Then, $A(P) = 0$ because P, which equals KG/Z, is less than V so the individual

The literature on deterrence that dates back to Becker [1968] notes that it is better to raise the penalty and reduce the number of police in the limit to zero. Since only the expected penalty matters, social costs are saved by reducing the amount of real resources spent on police and increasing the penalty so as to leave the expected fine sufficient to deter all crime.

There have been a number of objections to this rule. One is risk aversion. If false positive errors are made, individuals who are wrongly accused suffer greatly when extreme punishments are doled out. Second, marginal incentives are distorted. An individual who faces the death penalty for a parking violation might take extreme action to avoid arrest, resulting in additional social damage. Third, high penalties create incentive for enforcement officials to engage in extortion. Very high penalties might also encourage extortion by public officials because the penalties to anyone caught are so high that the willingness to pay extortion is increased.

Again, optimal penalties are not the essence of the argument, especially in the context of education where the penalties are likely to be determined exogenously by the market or by some wage structure that is determined in a bargaining context.

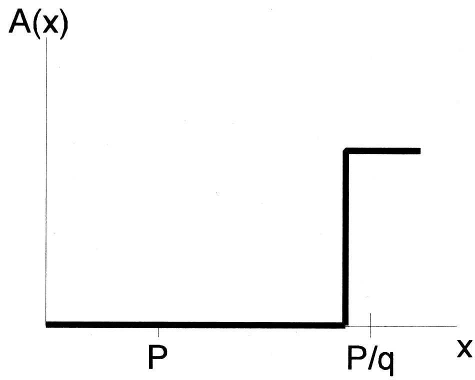

FIGURE Ia

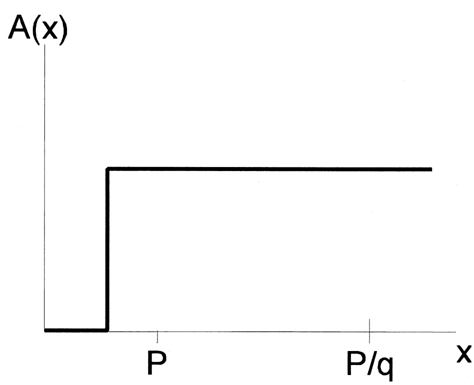

FIGURE Ib

never obeys the law. Since the lower limit of q is zero, there always exists some $q < 1$ for which $KG/qZ > V$, so $A(x)$ equals 1 (obey the law with certainty) over this range. The California Highway Patrol (CHP) uses the strategy of setting $q < 1$ and announcing the roads on which patrolling will occur. For example on the July 4 weekend of 2004, the CHP announced on all TV news stations that the 250 miles of Interstate 80 from San Francisco to the Nevada state line were being singled out to check for intoxicated drivers.

When the cost of detection is low, penalties are high, or benefits to speeding are low, Figure Ib is relevant. Then P, which equals KG/Z, exceeds V, and the individual obeys the law even for $q = 1$.

In this simple case, the optimal q can be calculated to set incentives just sufficient to deter speeding, or using (2), set

$$q = KG/ZV.$$

If $q > 1$, then it is unambiguous that the number of police G should be reduced until $q = 1$ because additional police serve no purpose; all speeding is deterred when $KG/ZV = 1$ or when $G = ZV/K$.

I.B. Tax Fraud

The extension of the idea to tax fraud is straightforward. The tax authority can do random audits, examining taxpayers and items without advance notice, or they can announce that all deductions of a particular kind will be audited. If they announce the items to be audited, taxpayers will report their expenditures honestly on the audited items. If they do not announce, then taxpayers will either cheat profusely or not cheat at all. So, when the cost of auditing is high perhaps because there are very few auditors, announcing the items that will be audited results in less cheating. Announcing audited items ensures that at least some taxes get paid honestly. However, if the cost of auditing were low and auditors were abundant, then keeping audited items secret would result in more taxes being paid. Because auditing is sufficiently likely, taxpayers are honest on all items.

The model is identical. V can be thought of as tax due on each of the Z items. As such, it is the savings on taxes that results from cheating on one of Z reported items on the tax form. K is the fine associated with being caught, which includes repayment of the V dollars initially saved. Thus, $K > V$. Redefine G as the number of items that can be audited (per return), given the number of tax auditors.

As before, when

$$KG/Z < V < K,$$

filers will cheat on every item if monitoring is stochastic and will pay the penalty on those items on which they are caught. If the goal is to deter cheating, then a better system is to announce all of the items that will be audited and to deter cheating at least on those items that are audited with certainty.

When G is low (auditors are very costly), there is a conflict between deterrence and revenue collection. If G is low, more fraud is deterred by announcing the audit rules than by keeping them secret, but more revenue is collected by keeping rules secret. When G is too low to deter cheating if rules are unannounced, individuals cheat on all items, paying zero taxes, but are caught on G items (on average) and so pay GK in total. With announced auditing rules, individuals pay taxes on the G announced items, and revenues are GV. No fines are ever collected because the items on which the individuals cheat go undetected with certainty. Because $V < K$, revenues are highest in the stochastic monitoring regime, even though no cheating is deterred.

Keeping the rules secret induces everyone to cheat, which is like setting a trap for cheaters. Entrapment can be useful for revenue collection because "tricking" people into cheating results in fine collection, which brings more money into the treasury than the paying of taxes without fines.

The difference between the tax auditing problem and the speeding problem is that in speeding, the assumption is that the social cost of speeding is sufficiently high to swamp any distortions associated with reduced fine collection that might be part of an optimal tax structure.[8] Here, if taxes are not collected through fines by the tax authority, the revenues must be raised in other ways, which may create other distortions. The goal of taxing, at least in large part, is revenue collection.

I.C. Terrorism

Speeding and tax fraud are particular forms of crime. Terrorism is another and fits the structure directly. Like speeders, who might speed on a segment of road at an unknown time and place, terrorists can strike randomly at various locations and times. Deterring terrorists from striking targets is conceptually the same problem as deterring drivers from speeding. Let V be the value that terrorists place on hitting a target, γ the social cost of being a victim of terrorism, K is the punishment that the terrorist receives if caught, and Z the total number of potential targets, where each of the G enforcers can patrol one of the Z targets. As before, the size of the expected penalty relative to the

8. Some might argue that speeding fines are part of an optimal tax structure. For example, some very small towns set extremely low speed limits to induce speeding so that they can collect revenue from out-of-town motorists.

terrorist's value of committing the act determines the optimal allocation of enforcers. When terrorists place a very high value on terrorism or the number of enforcers is low relative to the number of potential targets, it is virtually impossible to deter terrorism everywhere, even with very large penalties. (Since many of these individuals make suicide part of the strategy, it is difficult to think of feasible penalties sufficiently large to deter.) This, the most likely situation, corresponds to Figure Ia, where the $R(A(x))$ function is convex. It is better to concentrate incentives, deterring terrorism somewhere rather than not at all. For example, airport screening, reinforced cockpit doors, and other measures make it difficult to conduct terrorism using airports or airplanes. Terrorists know this and avoid those targets so at least airports are relatively secure. If there were infrequent screening spread over all possible targets, none would be safe. The expected penalty would be too low anywhere to deter committed terrorists.

This example points up two issues. The first is heterogeneity; the second is substitution. Heterogeneity is particularly important in the context of terrorism. Some targets produce higher value to terrorists than others. Some create more social damage than others. Some are harder to patrol than others. It has already been shown that even when all targets are equal, it sometimes (when $R(A(x))$ is convex) is better to narrow the patrolled targets to some subset of the total possible and to announce those that are patrolled and those that are not. Heterogeneity makes the case stronger, but one can also say something about which targets should be patrolled.

It is obvious that other things equal (V, K, and γ), it is better to patrol the less costly targets. If there are G enforcers, higher cost of patrol can be thought of as more than one enforcer per target. This simply reduces the number of enforcers available for other targets with no additional gain in deterrence. More interesting is whether the most precious targets, i.e., those for which γ is high, warrant more resources. The analysis, contained in Appendix 1, shows that there are situations in which it pays to concentrate enforcement on high value targets, but the necessary and sufficient conditions are stringent. The reason is that if a target has strategic value, it is reasonable to assume that it has higher value to terrorists as well as to the citizens, who are the owners of the target. Under general circumstances, unless the precious targets are valued more relative to the less precious by the citizens than they are by the terrorists, concentrating on those targets will not be optimal. If Americans value protecting the White House relative to my house by more than terrorists want to take out the White House relative to my house, then it is optimal to guard the White House over my house. If the relative values are reversed, my house should be guarded over the White House. Of course, if enforcement is sufficiently cheap or if the social damage is sufficiently high, it pays to allocate sufficient resources to protect both targets. The idea also relates to substitution. If terrorists regard most targets as close substitutes for one another (having similar V), but citizens do not view all targets similarly, then citizens prefer to concentrate enforcement on those targets on which the citizens place a higher value.

I.D. Announced Inspections

There are many activities in which inspections occur. Government agencies such as the Food and Drug Administration and the Federal Aviation Administration inspect private sector firms that are under their jurisdictions. In the military, drill sergeants inspect the uniforms and gear of their subordinates. Sometimes, inspections are random and unannounced, but at other times they are scheduled. It might seem silly to inform one's charge of the exact time at which an inspection is going to occur, but the logic of previous sections reveals that there are circumstances under which it is better to make the timing known than to keep it secret.

There is nothing in the formalism that distinguishes space from time. Just as Z could reflect geography or items, it can also be thought of as a linear index of time, where a particular Z refers to a point in time. As before, qZ would limit the time during which an inspection would occur to a particular interval; e.g., "an inspection will be held during the week of August 5." Further, just as before, it is better to concentrate incentives and limit inspection possibilities to $q < 1$ when $R(A(x))$ is convex. Convexity is produced by high costs of inspection or by high costs of meeting prescribed requirements.

In the case of checking that shoes are polished, it would seem that drill sergeants could monitor this at relatively low cost and that the benefit from not polishing shoes would be low to soldiers. As a result, $R(A(x))$ is concave, and inspections should be random over time and unannounced. Weak, diffuse incentives are sufficient to motivate.

Consider FAA inspections. There are so many potential parts on an airplane that could be defective that it would be impossible to inspect all on every plane without rebuilding the entire fleet. In the context of the model, Z is interpreted as all the parts that could be inspected at a point in time and G, the number of inspections, is small relative to Z. As a result, $R(A(P))$ is convex, and it is better to concentrate incentives. This appears when the FAA sometimes focuses on a particular part (like a hatch door) and announces that it will inspect all of these. Incentives are thereby concentrated, giving airlines an incentive to increase the attention that they pay to maintenance.

Often, the part that is the subject of attention has been defective on another aircraft, which introduces the kind of heterogeneity discussed in the previous section. Heterogeneity, where one part (target) has higher inspection value than another, reinforces the tendency to concentrate.

I.E. Drug Screening in Sports

In the Tour de France, Lance Armstrong visited a tent after completing the various stages to be tested for drug usage (steroids, etc.). Armstrong knew that he would be tested at the end of the stage.

The argument for announcing that the test would be conducted at the end of the stage rather than at some other random time had little to do with concentrating incentives. Instead, it was based on the technological fact that infractions on one Z indicates an infraction on another Z. In the examples given above, independence across Z is assumed, although primarily for simplicity. In the case of drug testing, independence makes no sense. An individual who uses performance-enhancing drugs before the race will test positively for those drugs after the race.

I.F. Legislature or Courts?

Detailed legislation and court rulings can be viewed as substitutes. When legislation is very detailed and legitimate behavior is prescribed, the courts have less room for interpretation. When legislation is broad, loose, or nonexistent, justice is ex post and less predictable. The distinction between civil and common law has the same flavor. Civil law lays out more specifics than common law. Civil law is an ex ante system, attempting to specify as clearly as possible what is and is not permitted. Common law focuses on the principles of good behavior and to a greater extent determines deviations from good behavior on an ex post basis. Which system is better?

One interpretation of the current model can be used to examine the choice between detailed legislation or court discretion. Define Z as the number of potential actions that can be legislated and G as the number that is checked. Suppose that Z is small relative to G. Such would be the case in a simpler world that existed 500 years ago, in primarily agrarian societies where the variety of interactions were limited compared with those today. The expected penalty P equal to KG/Z was high relative to the benefits of breaking the law. There are three reasons. First, Z is small because of limited interactions. Second, G is large because in small communities, it is difficult to find anonymity. Every neighbor is an enforcer. Third, penalties were high. Individuals who deviated from the common law or social norm and were ostracized could find few other ways to survive outside their small communities. As a result, the $A(P)$ function is concave (and presumably $R(A(P))$ is concave) in the relevant range which means that loosely specified rules work.

In contrast, a complex society with many possible interactions and much anonymity creates conditions where the logic of common decency is insufficient to deter bad behavior. As a result, a detailed system of laws, specified ex ante by the legislature with limited discretion for the courts, concentrates incentives and prevents infractions of those laws that are announced and monitored. In this situation, $R(A(P))$ is convex, so general or less well-specified ex post strategies are inefficient relative to the more concrete structures offered by precise ex ante legislation and civil law.

II. Teaching to the Test

The lesson of the speeding example can be applied in a straightforward way to the issue of high-stakes testing. High-stakes testing as a practical matter places the learning and teaching emphasis on items that are expected to be on the exam. In this sense, it is similar to the idea of announcing where the police are posted. The items on the exam receive special attention, whereas untested items may be neglected by students and teachers. The speeding model can be applied to this problem in an almost direct fashion to obtain some insights. As above, the first result is that high-stakes testing is best used when monitoring is costly or when expenditures on enforcement are low. If expenditures on enforcement are high, then it is better to leave the testing regime more open. Second, high-stakes testing with well-defined exam questions is best used when the distribution is weighted toward high cost learners.[9]

Let us start by defining the knowledge base, which consists of n items. This is analogous to the Z miles of road. Suppose further that there are m questions on a high-stakes exam, analogous to the G policemen. Should the exam questions be announced or not? A more direct way to put the issue is "What comprises a good high-stakes test?" Should it be a test where questions are well-defined and known in advance, or should it be a test where questions are drawn randomly from a larger body of knowledge? Most would say the latter. It will be argued that the former rule is appropriate in some circumstances.

As a policy issue, testing is as much about motivating teachers as it is about motivating students, and the model applies to teachers as well. Initially, however, think of the student as making the choice about learning, and let the teacher be a passive agent. That assumption will be altered below.

To be consistent with the speeding model, the return side is modeled as follows. Think of the test score as an observable signal to employers, or more accurately, to future schools which the student might attend. If a student is asked a question to which he does not know the answer, he bears cost K in the form of lower earnings, most directly reflected as reduced probability of admission into a desirable college. The SAT exam is a high-stakes test with exactly that effect. The "fine" K is taken to be exogenous, but a richer model would allow K to be the solution of an inference problem that colleges or employers make about the individual's ability based on the answers to the exam.

Let us reinterpret V and K from the speeding model as follows: if the student does not learn the item, he does not have to bear cost V of learning the material. The student knows what is on the test, so he opts to avoid learning an item when the extra utility from not learning, V, exceeds the cost of not learning, which is lost earnings K. If the student knows what is on the test, he will choose to learn those items if and only if $V < K$. Since $K = 0$ for items not on the test, he learns nothing that is not to be asked explicitly.

9. The emphasis here is on the incentive aspect of testing. Another role of testing is to provide information to help in modifying the curriculum. To deal with this component of testing, a dynamic model is required.

To generalize the earlier framework, allow there to be a distribution of V that reflects the cost of learning on any given item by any given person. Let that distribution be written as $J(V)$ with corresponding density $j(V)$. The unit of analysis is a person-item so that V can vary for a given individual because some items are more difficult to learn than others. Also, V can vary across people because some people learn more easily than others. Then $j(V)$ is the density across all items potentially learnable by all students. Note that a given student might learn some items and not others and some students might learn everything always and others nothing, depending on the distribution of V across items and people. Further note that the assumption is that all items and students characterized by the distribution $J(V)$ are observationally identical. If items or people are observably different, then separate distributions must be written to characterize each. Finally, note that V is assumed to be independent of whether learning of other items occurs. For the sake of simplicity, such complications are ignored.

Suppose that there is some expected penalty X. A given student learns an item if and only if $V < X$. Let the social value of learning be given by γ. Items for which $V < X$ are learned. Those for which $V > X$ are not learned. Thus, the social damage associated with any expected penalty X is

$$S(X) = \int_X^\infty (\gamma - V) j(V) dV. \tag{3}$$

Also note that

$$S'(X) = (X - \gamma) j(X), \tag{4}$$

and that

$$S''(X) = j'(X)(X - \gamma) + j(X), \tag{5}$$

which will be useful later. In what follows, optimal solutions are found in the more general analysis where a rich structure of strategies is considered.

From (4) and (5) it is clear that social damage is minimized when $X = \gamma$. Setting the expected fine equal to the social cost of the infraction induces the appropriate behavior.

The goal is to choose q so as to maximize social benefit from learning or equivalently to minimize the damage associated with the failure to learn. Items in the knowledge set can be made eligible for testing, and others can be declared off limits. At the extremes, when $q = 1$, all items are fair game. When $q = m/n$, $(m/n)n$ or m items are subject to testing. Since there are m questions, each item subject to testing is identified.[10]

The student learns an item when the cost of learning is less than the expected penalty or when

$$V < Km/qn$$

for items subject to testing. On those items, the fine is K, and the probability that any given item will be tested is m/qn for the items subject to test. The expected penalty is a certain zero for those items not subject to test.

Because $S(X)$ is the social damage on a given item when the expected penalty is X, the expected damage as a function of q is given by

$$\text{(6)} \quad \text{Full social damage} \equiv FSD(q) = n[qS(X) + (1 - q)S(0)].$$

On the qn eligible items, damage is $S(X)$, and on the $(1 - q)n$ ineligible items, the damage is $S(0)$.

The first-order condition for minimizing (6) is

$$\text{(7)} \qquad \frac{\partial}{\partial q} = n\left[S\left(K\frac{m}{qn}\right) - S(0) - K\frac{m}{qn}S'\left(K\frac{m}{qn}\right)\right] = 0.$$

Corner cases are possible where the solutions are $q = m/n$ (reveal the questions) or $q = 1$ (do not reveal anything). Stated formally,

PROPOSITION 1. When $S(X)$ is globally concave, the optimal q is m/n. When $S(X)$ is globally convex, the optimal q is 1.

Proof. See Appendix 2.

In the language used earlier, $S(X)$ being globally concave is equivalent to $R(A(P))$ being globally convex. Under those conditions, it is necessary to concentrate incentives because spreading them too thinly results in no effort.

There is no necessity that q be at either corner. It is possible and perhaps likely that the solution is an interior q with $m/n < q < 1$.

Next, consider what occurs when individuals have higher

10. It is assumed that the same item is never tested twice.

ability. Define ability a, such that individuals choose to learn as long as[11]

$$V + a < X,$$

or equivalently,

$$V < X - a.$$

Ability acts to displace the distribution by the constant a. We can then state the following proposition.

PROPOSITION 2. If an interior solution for q exists such that $m/n < q < 1$, then q is increasing in ability level a.

Proof. See Appendix 2.

The range of material that is subject to testing is optimally larger for high ability students than for low ability students. The intuition is that it is more difficult to motivate low ability students because their costs are high. As a result, failure to concentrate incentives produces little or no learning among low ability individuals. But high ability individuals have low costs of learning. Even mild incentives on any given item are sufficient to induce them to learn a subject. Consequently, a larger number of items can be made subject to test for high ability individuals, and they will still opt to learn those items.

The number of questions available may vary because of the cost associated with testing. When it is cheap to test, m is high. This is analogous to having many police when the cost of police is low. How does the optimal q vary with m? The following proposition shows that it is optimal to choose a higher value of q when m is high.

PROPOSITION 3. Given that q is interior such that $m/n < q < 1$, q varies directly with m.

Proof. See Appendix 2.

11. Since ability is unobservable, as long as we think of the metric of ability as continuous and monotonic, there is also a transformation of some measure of ability into another that allows the functional form above to hold. This formulation shifts only the mean of the distribution and does not alter higher moments.

III. Discussion, Implications, and Extensions

Proposition 1 provides a result on when it is best to announce the specific questions and when it is best to say nothing at all. The intuition is identical to that given in the speeding model. If the distribution of costs is such that little or nothing will be learned when questions are unannounced because incentives are too diffuse, then the only hope is to limit the number of items subject to test. When costs are truly high and concentrated, corresponding to concavity of the social damage function, $S(x)$, (or equivalently convexity of the $R(A(P))$ function), then announcing the questions concentrates incentives and induces at least some learning. When V is concentrated and high, it is optimal to announce the items that are to be tested. When V is concentrated and low, it is optimal to keep the questions secret. High cost of learning requires that the questions be announced to provide sufficient incentive, whereas low cost of learning allows for secrecy because even low expected penalties induce learning.

Another intuitive result is easily derived. If it is very costly to test, it is better to announce the specific questions. If it is very cheap to test, it is better to keep the questions secret. Costly testing is reflected in a low number of questions. Suppose that m is sufficiently small so that Km/n is too low a number to motivate learning; i.e., the minimum value of V exceeds Km/n. Then $S(Km/n) = S(0)$ because failing to limit items subject to testing is equivalent to setting the penalty equal to zero; both result in no incentives. When m is sufficiently low, the optimum cannot be the policy of keeping questions secret.[12] When questions are expensive, failure to announce them results in no learning. Announcing the questions induces learning of the announced items.

Conversely, if it is very cheap to test, then the optimum must be to leave vague the items to be tested. Suppose that m is so

12. To see this, it is sufficient to show that the other extreme, of announcing all questions, dominates complete secrecy. It is better to announce the questions when

$$mS(K) + (n - m)S(0) < nS\left(K\frac{m}{n}\right),$$

which is the same as

$$mS(K) + (n - m)S(0) < nS(0),$$

because $X = Km/n$ is too small to induce learning. The inequality holds because $S(0) > S(K)$.

large that $Km/n > V \ \forall \ V$. Then the optimum is to announce nothing. By announcing nothing, all items in the knowledge base are learned. Any $q < 1$ results in less learning.[13] Questions are so numerous, and the cost of learning is sufficiently low that no student risks leaving any item unlearned.

Interior solutions to q are common throughout education. Study sheets give students strong clues as to the material that will be on the exam without announcing specifically which questions will be asked. "Pop quizzes" on the previous night's reading that are given randomly provide another example of $m/n < q < 1$. If students were told specifically when the exam would be given, then the student could read that night's and only that night's assignment. But the fact that students are told that the quiz will only be on the reading from the previous night limits q to a number much smaller than 1. If nothing else, it rules out material from earlier nights or from future readings.

III.A. What Is a "Good" Test?

One common view is that a good test is one that is not so predictable that students essentially know what is on the exam. It would be possible to create an exam that randomized question selection so as to prevent students from knowing what is on the exam. Educators often view as a goal of testing that test scores generalize to reflect knowledge of material not on the test.[14]

The view that a good test is one that is unpredictable and one where the test score is informative about items not on the test is incorrect as a general proposition. Although it may be optimal to construct a test that draws from a larger body of knowledge, the main theorem of this paper (Proposition 1) is that sometimes it pays to restrict the relevant required material to a specified

13. Formally, $S(Km/n) = S(\infty)$; all items are learned. It is better to keep questions secret when

$$mS(K) + (n - m)S(0) > nS\left(K\frac{m}{n}\right)$$

which is the same as

$$\frac{m}{n}S(K) + \left(1 - \frac{m}{n}\right)S(0) > S(\infty).$$

This must hold because $S(\infty) < S(X) \ \forall \ X$ when $\gamma > \max(V)$.

14. For example, McBee and Barnes [1998] claim that a test would have to test a prohibitively high number of tasks to attain acceptable levels of generalizability.

subset of the entire knowledge set. A "good" test when students have very high costs of learning is a test that announces the questions and sticks to them. Under those circumstances, students at least learn the material that is on the test. The alternative test, which chooses questions from a broader base of knowledge, results in no learning or very little learning.

For low cost learners, the reverse is true. A test that draws from the entire or a larger knowledge base is a better test because it encourages more learning than one that is well-specified and announced. For these students, a "good" test is one that is not completely predictable because it provides more incentives to learn.

Defining a "good test" captures exactly the intuition of both sides of the argument over high-stakes testing. Most agree that imposing high-stakes testing will induce teaching to the test because the incentives are strong to learn what is on the test and then to teach to it. The disagreement is over whether this is good or bad. The concern by critics of such testing is that a strategy that is tantamount to announcing the exam questions will stifle learning of the more general curriculum. These critics are correct if they have in mind students who would be sufficiently motivated to learn all the material. But for those who are less able, choosing to keep questions secret will result in little or no learning because diffuse incentives are too weak to provide any motivation at all for high cost learners. For them, the policy does not stifle additional learning because those students would not learn that additional material even if the questions were not announced.

Proposition 2 speaks directly to this issue. The optimal q is increasing in ability. As the ability of the students in question rises, it is better to increase the number of items that are subject to testing. High ability students are more easily motivated, so spreading incentives thin, which provides enough incentive to learn, increases the range of items that will be learned. The trade-off between weakening incentives on any given item and increasing the number of items that a student may learn moves in the direction of more items as ability rises.

"No Child Left Behind" emphasizes high-stakes testing only for low performing schools. Although all schools are required to take the test, high performing schools are in little danger of failing to meet the standards. As such, the test provides only weak incentive to those schools. If there is any monitoring incentive at all for upper quality schools, it is provided through more indirect stochastic methods. But failing schools are in the range where the high-stakes test matters. As a result, the NCLB system acts as a bifurcated program, producing high-stakes testing for those who attend problem schools and stochastic monitoring (at best) for those who go to schools that are doing well. The ability variable a in Proposition 2 can be interpreted as reflecting the cost of schooling rather than the ability of the underlying students. Because a is a shifter, which simply displaces the $j(V)$ distribution, a school with higher costs of teaching is one where a is lower. Whether the costs are higher because the students have lower ability or because the schools have fewer resources, poorer teachers, or poor family support is of no consequence to the proposition. The model provides a rationale for the No-child-left-behind approach since the regime appropriate for high ability students is stochastic monitoring where q is closer to 1, whereas the regime appropriate for children in poorly performing schools is likely to be high-stakes testing.

III.B. *Age, Background, Difficulty, and Test Form*

Proposition 2 provides intuition on why testing and monitoring methods vary across grades and schools. Consider the learning ability of young children relative to college age students. It is more costly for young children to learn academic subjects than for older ones, but probably cheaper for them to learn language.[15] As a result, the distribution of learning costs associated with academic subjects is higher for younger children than for older ones. This is parameterized by assigning a lower value of a to younger children than to older ones. Proposition 2 dictates that the optimal q should be lower for younger children. At the extreme, $q = m/n$, so that they are told exactly what is on the test. Spelling tests given to elementary school children generally specify exactly which ten words must be known for Friday's test. By the time students reach graduate school, only the papers and books and sometimes only the general subject area from which the test will be drawn are announced.

Analogously, children who are in honors classes are likely to have lower costs of learning than those in remedial classes. Indeed, tests in honors classes are less predictable, pose questions that are extensions of material learned, and are drawn from a

15. There is a large body of literature on learning different skills at different ages. Perhaps best known for these ideas is Piaget and Inhelder [1969].

larger body of knowledge than those in remedial classes. In short, q is higher in honors classes than in remedial classes.

III.C. Separating Teacher and Student Incentives

The discussion has been put in terms of motivating students, but most of the thought behind specific programs like “No Child Left Behind” is that it is the teacher, not the student who needs motivating. The model as set up can be interpreted to refer to teachers instead of students.

Suppose that teachers have full control over what is learned by the student. Interpret V as the teacher’s cost of teaching the student an item of knowledge. Let K be the penalty associated with her student failing to answer a question correctly in the high-stakes environment or as the penalty that the teacher faces if the student is detected to be ignorant of an item of knowledge. Then all of the above analysis holds exactly as written, and nothing is changed.

The problem of interest, though, is how are teachers motivated. Many would argue that the current system of random monitoring does not motivate teachers at all. Teachers are motivated by intrinsic considerations only, and intrinsic motivation is insufficient to induce some teachers to do the right thing. Again, the issue is one of heterogeneity as well as motivation, but let us consider the incentive issue in a world of homogeneous teachers first.

Intrinsic motivation might be thought to serve as the main motivator for tenured teachers whose salaries are fixed and jobs are secure, being virtually independent of performance. Intrinsic motivation is best modeled by assuming that $j(V) > 0$ for $V < 0$. That is, for some values of V, the cost of imparting knowledge is negative. Even if teachers received no compensation for the amount of knowledge their students acquired, they would still choose to provide some knowledge to each student, namely $J(0)$ items would be taught. Teachers with high amounts of intrinsic motivation teach even in the absence of any extrinsic penalty.

Proposition 2 is relevant here also. The more costly or difficult is teaching, or the less able are the teachers, the lower is the optimal q. Good teachers should not be told explicitly what to teach and on which items their students will be tested. Vague descriptions of curriculum requirements (with q close to 1) are more appropriate for intrinsically motivated or able teachers.[16]

The assumption that K is exogenous can be altered. Under the current system, it is unlikely that information about a teacher's ability to raise students' test scores would become public information or part of what determines their compensation. But it is useful to consider what could be done if the school opted to implement an optimal compensation structure. Given that the social value of learning a particular item is γ, the school would simply set $q = 1$ so that all items in the knowledge set are potentially tested and choose K such that the expected penalty equals γ. If the number of questions is given as m, then the teacher would be fined K such that

$$Km/n = \gamma,$$

or

$$K = \gamma n/m,$$

for each question that each of her students misses on the exam. The teacher would teach the item whenever

$$V < \text{expected fine},$$

or

$$V < Km/n,$$

or

$$V < \gamma,$$

which is the efficiency condition.

III.D. Monitoring Input or Output?

Formally, the model has been structured in terms of monitoring output not input. Much of the discussion of high-stakes testing views stochastic monitoring as being based on input. For example, in the absence of high-stakes tests, teachers could be monitored by having the principal visit the classroom on either a predicted or random basis. As is shown here, input monitoring is accommodated by the model already presented.

16. Other issues with teachers and students involve team problems. Because both have an incentive to free ride on the other's effort, the standard result that effort of each party falls short of the optimum holds. But there is little about the student-teacher team that distinguishes it from other partnership problems, which have been analyzed. See, for example, Holmstrom [1981] and Kandel and Lazear [1992].

Think of teachers as being in the classroom for n minutes, and let one item of knowledge be conveyed if the teacher bears cost $V \sim j(V)$ as before. The principal announces that he will monitor classes for m minutes (per teacher) and that q of the n minutes of total teaching time are subject to monitoring. If he finds that the teacher has not conveyed the information in the minute during which he is in the room, the teacher will be fined K for that minute in lower salary. (Of course, K may be zero or close to it.) Setting $q = m/n$ is tantamount to telling the teacher exactly when the principal will visit the room. Setting $q = 1$ tells the teacher that all minutes are equally likely to be monitored. Then the expected penalty is Km/qn, just as before and the teacher's decision is to teach if

$$V < Km/qn.$$

Everything in the prior setup applies to monitoring on the basis of input. Both interpretations, monitoring on input or monitoring on output, are consistent with any given level of q. Whether monitoring is done on the basis of input or output relates to the costs of measuring by each method and is not special to teaching. That issue has been analyzed elsewhere.[17]

A high-stakes test creates incentives for teachers and students to find out what will be tested. As such, it is closest to the case of setting $q = m/n$. The current alternative, which is to monitor input and sometimes output in a stochastic fashion, is formally treated at having a $q > m/n$ and in the limit, equal to 1. Because the current situation tends to be coupled with low stakes, i.e., low values of K associated with "infractions," teachers have little incentive to attempt to discover when and how the monitoring will be done. It is for this reason that the typical situation in schools corresponds more closely to high values of q and high-stakes testing to low values of q. But this argument suggests that the choice of K and of q are not independent. When the stakes are raised, there is a natural tendency by those being monitored to learn the specifics of what will be monitored, which induces a positive, endogenous relation of q to K.

17. See, for example, Lazear [1986, 2003].

IV. Extensions and Further Discussion

In this section some details, additional implications, and directions for future research are considered.

IV.A. Interdependence of Learning

As in the speeding structure, independence of V has been assumed. Having learned one item does not affect in a direct way the cost of learning another item. This is unrealistic in three respects. First, a student may have a capacity for learning so that as the amount of studying increases, he is unable to absorb new material at the same cost. Only a limited number of items can be remembered in one sitting. Second, and working in the opposite direction is that learning begets learning. It is easier to learn calculus after algebra has been mastered. Third, because the distribution of V also includes variation across people, interactions between students is ignored. But peer pressure and identity might be important. Coleman [1961] was the first to argue this and found that the group in which an individual classified himself (e.g., bookworm, athlete . . .) was a predictor of academic performance. Akerlof and Kranton [2002] summarize and build on this notion to explain why academic performance may vary greatly from school to school, even when resources do not. Building in some form of dependence is possible, but complex and is ignored in this formulation.

IV.B. Test Design and Learning Incentives

It is possible to ask how test design, and in particular scoring, affects incentives. For example, one very large, high-stakes test could be given or many smaller, low-stakes tests could be required. The incentives to study or teach are very different under the two approaches.

Additionally, exams could be graded pass/fail or in a continuous fashion. The pass/fail structure is more like a tournament against a standard, where the standard is calibrated on the basis of previous classes' performances. A continuous grading structure is like paying a piece rate. It is already known that the incentive effects of the two different structures vary, depending on the nature of the payoff scheme and the heterogeneity of the underlying population.[18]

On a different note, good exams are neither too easy nor too

18. See Lazear [2000] and Lazear and Rosen [1981].

difficult, and this is primarily for statistical reasons, but also because of the effect on incentives. On a very easy exam, a careless mistake can cause a student to fall well below the rest of the class. Such exams have low signal-to-noise ratios. On a very difficult exam, average scores are very low, and it is difficult to distinguish among people because all do so poorly. Again, the signal-to-noise ratio is low. As for incentives, it is a general principle in incentive theory that when noise is high relative to the signal, incentives are diminished. The optimal test difficulty should take this incentive effect as well as statistical issues into account.

Investigation of these issues is left to subsequent work.

IV.C. A Theoretical Observation on Optimal Enforcement

The model used above is a generalization of the usual optimal enforcement literature. Rather than having the probability of detection being the same over all actions, this structure allows for different probabilities of detection. In the case of education, it is that qn items have a probability of detection equal to m/qn and $(1 - q)n$ items have a probability of detection of zero. This generalization can be extended. At the extreme, it is possible to allow for each item (or each mile) to be given a different probability of detection. Eeckhout, Persico, and Todd [2004] show that when potential criminals are homogeneous, it is never optimal to use more than two different enforcement probabilities. In general, enforcement resources could be spread over the entire set of possible crimes so as to minimize social loss.[19]

Different characteristics lead to different strategies. It has already been shown that when detection is expensive or costs of learning or teaching are high, it is best to go to a corner, having the questions announced in advance. Since it is only the difference between social value and social cost, $\gamma - V$, that is relevant, it is possible to turn this around and think of situations where the difference is high because γ is high, not because V is low. For example, it might be optimal to put the police on more congested roads where the social cost of an accident is especially high. Similarly, it might be better to test certain skills like basic literacy over others because they have higher social value. The model can be interpreted to

19. I thank Thomas Dohmen for pointing out this generalization.

address these issues, but more structure would be needed before specific implications about items tested could be provided.[20]

V. Conclusion

Speeding, terrorism, tax fraud, inspections in the military, and teaching to the test are all symptoms of the same kind of incentive problem. Individuals become aware of the rules, obey them within a narrow range, and disregard them everywhere else.

The analysis has shown that providing well-defined requirements dominates stochastic incentives for individuals for whom compliance costs are high. In the context of education, this means that predictable tests are best used for high cost learners or low ability types and stochastic monitoring, where students are not informed in exact terms what will be required of them, provides better incentives for low cost learners or high ability types.

Put differently, a "good test" is a well-defined concept once incentives are considered. Good tests are not necessarily those that draw evenly from the knowledge base, or those for which scores generalize to predict other aspects of knowledge mastery. Sometimes, especially for high cost learners or for failing teachers, tests that are predictable are best at providing incentives to learn or to teach. For high ability students or successful teachers, somewhat less predictable, more amorphous tests are best.

Additional results are provided.

1. If teachers have low degrees of intrinsic motivation, then well-defined high-stakes tests are best, but for teachers with high intrinsic motivation, a more randomized accountability system is efficient.
2. Number of questions and randomness are complements. When testing is cheap, and more questions can be on the exam, the optimal proportion of items which are subject to testing rises. When testing is expensive and only very few questions can be asked, it is better to announce those questions, or the student will learn almost nothing.
3. Exam specifics are made known to younger children and to students who are assigned difficult material because revealing exam questions provides better incentives.

20. Persico [2002] discusses briefly the idea that using differential rates of enforcement may be optimal.

Appendix 1

As before, let there be Z targets where targets 1 through Z_1 have value V_1 to the terrorists and γ_1 to the citizens. For targets between Z_1 and Z, $V = V_2$, and $\gamma = \gamma_2$ with $V_2 > V_1$ and $\gamma_2 > \gamma_1$.

The highest q_1 consistent with deterring terrorism on V_1 targets, is

$$q_1 = KG/V_1Z$$

since expected penalty is then

$$\frac{KG}{q_1Z} = \frac{KG}{(KG/V_1Z)Z} = V_1, \tag{A1}$$

so expected penalty exactly equals the value of terrorism to the terrorist. Similarly, the highest q_2 consistent with deterring terrorism on V_2 targets, is

$$q_2 = KG/V_2Z. \tag{A2}$$

As before, there are G enforcers. If all enforcers are used on the γ_1 value targets, then it is possible to deter

$$q_1Z = KG/V_1$$

terrorist acts. If all are used on the γ_2 value targets, it is possible to deter

$$q_2Z = KG/V_2$$

terrorist acts. It is clear that

$$q_2Z < q_1Z$$

because $V_2 > V_1$. To make things simple, assume that $q_1Z < Z_1$ and $q_2Z < Z - Z_1$ (see Figure II).

Because it is possible to specialize in either high value or low value targets, no mixing ever occurs. If it is better to place enforcement resources on one high value target than on more than one low valued target, it is better to do the same for the second high valued target and so forth since all high valued targets are identical and all low valued targets are identical.

The assumption that there are enough of both high and low valued targets to exhaust all enforcement resources is realistic. There are many more targets of high value (government buildings, hospitals, stadiums, auditoriums, and amusement parks where many people congregate) than there are enforcers. There

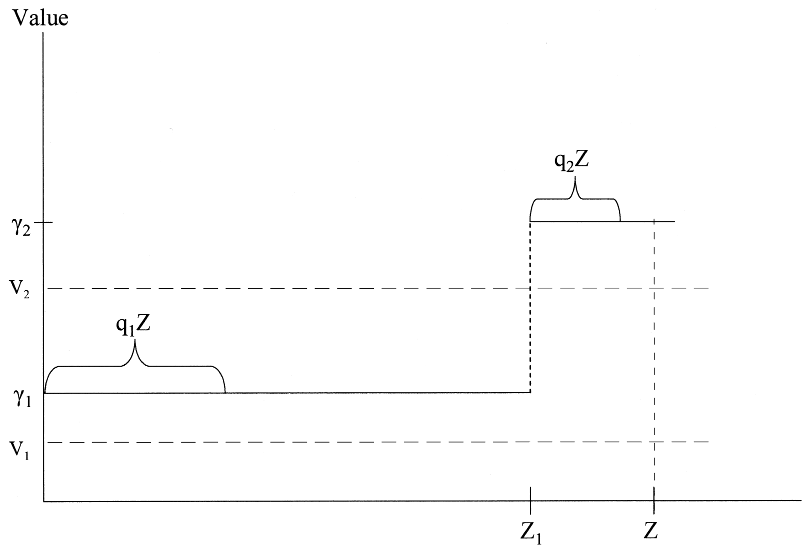

FIGURE II

are innumerable targets of low value consisting of individual homes, small retail stores, automobiles, etc.

It is possible to specialize in one type of target or another and since the value of all targets within a class is the same, it is best to focus on one type of target only. The trade-off is that it is easier to deter acts against low valued targets (because $V_1 < V_2$) and so more can be protected. Fewer high valued targets can be protected, but each target spared has higher value.

Ignoring the utility value of terrorism to the terrorist (a natural approach) in social value, enforcement should be concentrated on high valued targets if and only if

$$\gamma_2 q_2 Z > \gamma_1 q_1 Z$$

or using (A1) and (A2), if

$$\gamma_2 \frac{KG}{V_2} Z > \gamma_1 \frac{KG}{V_1} Z,$$

which is

$$\gamma_2/\gamma_1 > V_2/V_1. \tag{A3}$$

It is better to focus on high valued targets over low valued ones if and only if citizens value those targets relative to low valued ones more than terrorists do.

Appendix 2

Proof of Proposition 1

$$FSD(q) = n[qS(X) + (1 - q)S(0)].$$

The expected penalty on the eligible items is

$$X = K\frac{m}{qn}$$

so (6) can be written as

$$\text{(A4)} \qquad FSD(q) = n\left[qS\left(K\frac{m}{qn}\right) + (1 - q)S(0)\right].$$

Differentiate with respect to q to obtain

$$\text{(A5)} \qquad \frac{\partial}{\partial q} = n\left[S\left(K\frac{m}{qn}\right) - S(0) - K\frac{m}{qn}S'\left(K\frac{m}{qn}\right)\right].$$

In order for it to be optimal to state exactly which questions are on the exam, q must be equal to m/n. This happens if $\partial/\partial q$ in (A5) is positive at $q = m/n$. Then, increasing q will only increase social damage so the corner solution is best. The requirement is that

$$S(0) - S(K) < -KS'(K).$$

A sufficient condition for this to hold is that $S(X)$ is concave over the range of X from 0 to K (see Figure III). In order for the $S(X)$ function to be concave between 0 and K, it is necessary that

$$j'(X)(X - \gamma) + j(X) < 0$$

from (5). Since it is likely, especially in the education structure, that the expected penalty will be well below the social damage, γ, necessary is that $j'(X)$ is positive, which means that the density function is increasing over the range 0 to K.

It is also possible that the other corner solution is optimal, where students are told that all items are subject to testing. For $q = 1$ to be optimal, $\partial/\partial q$ in (A4) must be negative when $q = 1$ or

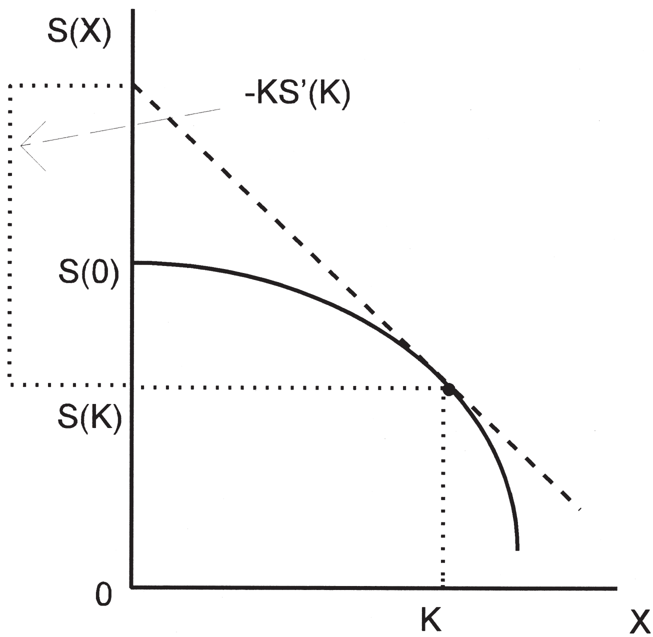

FIGURE III

$$\frac{\partial}{\partial q} = n\left[S\left(K\frac{m}{n}\right) - S(0) - K\frac{m}{n}S'\left(K\frac{m}{n}\right)\right] < 0,$$

which requires that

$$S(0) - S\left(K\frac{G}{Z}\right) > -K\frac{G}{Z}S'\left(K\frac{G}{Z}\right).$$

This holds if S is convex throughout the relevant range.

Interior solutions, where $m/n < q < 1$, are also possible. Using (A5), they occur when

$$S\left(K\frac{m}{qn}\right) - S(0) - K\frac{m}{qn}S'\left(K\frac{m}{qn}\right) = 0$$

for $m/n < q < 1$.

Proof of Proposition 2

The student learns an item if $V - a < x$ or if $V < x + a$. Define $c(x,a) = x + a$. Desired is to show, given an interior solution for q, $(m/n < q < 1)$, that

$$\frac{\partial q^*}{\partial a} = -\left.\frac{\partial FOC/\partial a}{\partial FOC/\partial q}\right|_{q=q^*} > 0.$$

Since q^* is at the minimum, the second-order condition implies that

$$\left.\frac{\partial FOC}{\partial q}\right|_{q=q^*} > 0.$$

By the chain rule

(A6)
$$\frac{\partial FOC}{\partial q} = \left.\frac{\partial FOC}{\partial x}\frac{\partial x}{\partial q}\right|_{x=Km/q^*n} = \left.\frac{\partial FOC}{\partial x}\left(-\frac{Km}{q^{*2}n}\right)\right|_{x=Km/q^*n} > 0.$$

Given $c(x,a) = x + a$,

$$\frac{\partial S}{\partial x} \equiv (c(x,a) - \gamma)j(c(x,a)) \equiv \frac{\partial S}{\partial a}.$$

Because the FOC is made up of $S(\,)$ and $S'(\,)$ only,

$$\frac{\partial FOC}{\partial a} \equiv \frac{\partial FOC}{\partial x}. \tag{A7}$$

Combining (A6) and (A7),

$$\left.\frac{\partial FOC}{\partial a}\left(-\frac{Km}{q^{*2}n}\right)\right|_{q=q^*} > 0,$$

which implies that

$$\frac{\partial FOC}{\partial a} < 0\Big|_{q=q^*},$$

so

$$\frac{\partial q^*}{\partial a} = -\left.\frac{\partial FOC/\partial a < 0}{\partial FOC/\partial q > 0}\right|_{q=q^*} > 0.$$

Proof of Proposition 3

The first-order condition for the optimal q is

$$\frac{\partial}{\partial q} = n\left[S\left(K\frac{m}{qn}\right) - S(0) - K\frac{m}{qn}S'\left(K\frac{m}{qn}\right)\right] = 0.$$

Using the implicit function theorem on,

$$\frac{\partial q}{\partial m} = -\frac{S''(Km/qn)(-K^2m/q^2n)}{S''(Km/qn)(K^2m^2/q^3n)} = \frac{q}{m} > 0.$$

Hoover Institution and Graduate School of Business, Stanford University

References

Akerlof, George A., and Rachel E. Kranton, "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature,* XL (2002), 1167–2001.

Becker, Gary, "Crime and Punishment: An Economic Approach," *Journal of Political Economy,* LXXVI (1968), 169–217.

Coleman, James, *The Adolescent Society: The Social Life of the Teenager and Its Impact on Education* (New York, NY: The Free Press, 1961).

Eeckhout, Jan, Nicola Persico, and Petra Todd, "Estimating Deterrence Effects Using Crackdowns: Theory and Evidence," University of Pennsylvania, 2004.

Hoffman, James V., Lori Czop Assaf, and Scott G. Paris, "High-Stakes Testing in Reading: Today in Texas, Tomorrow?" *The Reading Teacher,* LIV (2001), 482–492.

Hölmstrom, Bengt, "Contractual Models of the Labor Market," *American Economic Review Papers and Proceedings,* LXXI (1981), 308–313.

Jones, M. Gail, Brett D. Jones, Belinda Hardin, Lisa Chapman, Tracie Yarbrough, and Marcia Davis, "The Impact of High-Stakes Testing on Teachers and Students in North Carolina," *Phi Delta Kappan,* LXXXI (1999), 199–203.

Kandel, Eugene, and Edward P. Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy,* C (1992), 801–817.

Koretz, Daniel M., Robert L. Linn, Stephen B. Dunbar, and Lorrie A. Shepard, "The Effects of High-Stakes Testing on Achievement: Preliminary Findings about Generalization across Tests." Originally presented in R. L. Linn (Chair), *Effects of High-Stakes Testing on Instruction and Achievement,* symposium presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago, April 5, 1991.

Lazear, Edward P., "Salaries and Piece Rates," *Journal of Business,* LIX (1986), 405–431.

——, "The Power of Incentives," *American Economic Review,* XC (2000), 410–414.

——, "Teacher Incentives," *Swedish Economic Policy Review,* X (2003), 179–214.

Lazear, Edward P., and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy,* LXXXIX (1981), 841–864.

McBee, Maridyth M., and Laura L. B. Barnes, "The Generalizability of a Performance Assessment Measuring Achievement in Eighth-Grade Mathematics," *Applied Measurement in Education,* XI (1998), 179–194.

Meisels, Samuel J., "High-Stakes Testing in Kindergarten," *Educational Leadership,* XLVI (1989), 16–22.

Persico, Nicola, "Racial Profiling, Fairness and Effective Policing," *American Economic Review,* XCII (2002), 1472–1497.

Piaget, Jean, and Bärbel Inhelder, *The Psychology of the Child,* Translated by Helen Weaver (New York, NY: Basic Books, 1969).

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement." National Bureau of Economic Research Working Paper No. 6691, revised 2001.