Department of Economics and Statistics "Cognetti de Martiis"
Campus Luigi Einaudi, Lungo Dora Siena 100/A, 10153 Torino (Italy)
www.est.unito.it

Dipartimento
Economia e Statistica
Cognetti de Martiis

Working Paper Series

08/24

PERFORMANCE IN PASS-FAIL ASSESSMENTS

GIUSEPPE BERTOLA

UNIVERSITÀ DEGLI STUDI DI TORINO

The Department of Economics and Statistics "Cognetti de Martiis" publishes research papers authored by members and guests of the Department and of its research centers. ISSN: 2039-4004

Performance in Pass-Fail Assessments

Giuseppe Bertola

(Università di Torino, CEPR, CESifo)

This draft April 8, 2024

Abstract

Trading the cost of better performance off the probability that an imprecise test's performance estimate falls short of the pass threshold, an assessee may perform above the threshold (and fear failure because of negative errors) or below it (and hope to pass because of positive errors). This paper characterizes in general and with a specific model how that choice depends on the test's precision and threshold, which the assessor may choose to elicit high performance, and on other features of the assessee's problem, which the assessor must take as given and may not know exactly. When it is costly for the assessor to increase precision and assessees are heterogeneous it is optimal for some to fear failing, for others to hope to pass. This is empirically true in the results of exams that conform to the model's assumptions. The model is also applicable to editorial and research funding processes.

JEL: I23, D80

Keywords: First-order approach, Principal-Agent, Exam results.

For useful suggestions thanks without implication to Winfried Koeniger, Stefano Dughera, David Levine, and seminar participants.

1 Introduction

Economists and other academics devote much time and effort to writing and grading exams, processing admission and grant applications, preparing and refereeing research proposals and paper submissions. Yet, not much economic research studies these and many other familiar situations, such as driver's license written tests, where an assessee devotes costly effort to enhancing performance and earns a reward if an imprecise measure of performance exceeds a threshold. The vast literature on the economics of education typically supposes that the output of education depends continuously on educational inputs, and so does the smaller literature on the economic choices of teachers and researchers (e.g. Becker, 1979). Journal rejections and acceptances as discrete outcomes have been analyzed by sociological research (e.g. Hargens, 1988) and by econometric studies focused on personal connections between authors and editors (Colussi, 2018 and its references), but editorial and research funding decisions are absent from Stephan's (1996) survey of the economics of science.

Academic interactions are certainly tribal enough to be amenable to sociological or anthropological explanations, but rational economic behavior also plays a role in imprecise assessments that grant discrete rewards to assessees who privately choose costly performance. Humans who choose how to prepare exams they would like to pass or papers they would like to be accepted are aware that tests can return false positive and negative results, and their performance depends on the likelihood of failing despite strenuous preparation, or passing despite weak preparation. Hence, the implications of test parameters are more complex than in tests that aim to detect chemicals or viruses.

This paper studies how a rational assessee should choose costly performance when facing a test with given threshold and precision, and how rational assessors might choose those test parameters so as to elicit performance. The topic and results are novel, but the formal framework is related to various strands of literature. The incentive effects of a reward conditional on a noisy estimate of performance exceeding a threshold are similar to those of prizes and penalties in human resources management. Some of that literature characterizes the performance implications of precision. O'Keeffe, Viscusi, and Zeckhauser (1984) study symmetric contests where the pass threshold is at the mode of the measurement error's density, which becomes a steeper function of performance as measurement becomes more precise. Higher precision however need not have that implication, because flattens the density outside the region between the inflection points of a normal density. Kono and Yagi (2008) show that lower precision can increase effort in asymmetric contests with heterogeneous ability (or exams that assess relative performance), a result that also appears in Bertola's (2021) characterization of student effort in exam preparation, and the performance effects of the pass threshold in exams are characterized by Kuhen and Landeras (2014). These contributions treat the agent's local optimality condition as a constraint for the principal's choice set, a "first-order approach" that is valid only under conditions that may not be satisfied in those and other applications (Mirrlees, 1999; Rogerson, 1985; Ábrahám, Koehne, and Pavoni 2011; Bertola and Koeniger 2015).

What follows sets up and characterizes the simplest among possibly very complicated models of tests that aim to elicit the performance they assess and cannot rely on the assessee's first-order condition to guide the assessor's choice of test parameters. The problem is similar to that of a principal who knows that more than one performance choice may satisfy the agent's optimality conditions and maximizes it subject to "no-jump" constraints (Ke and Ryan, 2018 and their references). In a not very restrictive and broadly realistic class of models where the ratio of an increasing marginal cost of performance to the discrete pass reward crosses at most three times a single-peaked error density function, the no-jump constraint simply requires the assessor to ensure that the assessee chooses the larger solution of the first-order condition.

At a more substantive level the paper adapts and applies insights from contracting models to the problems of students and researchers who prepare exams or papers for submissions, and of examiners and editors who assess rationally chosen performance rather than a chemical or biological phenomenon. The principal of an employment contract aims at maximizing the revenue from selling the worker's output, net of wage and monitoring costs. The assessee's performance may generate revenue for the assessor, for example when a journal's subscription price depends on the quality of published papers. Even when it does not, the assessment generally does aim to better performance which, for example, increases the social surplus net of preparation costs generated by competent driving and public education. However, an assessor's instruments are not the same as those of principals who optimize the implications of agents' performance by setting and paying contractual rewards. In driver license tests, university exams, and research assessments the reward for passing the test and the cost of improving performance depend on market values and on the assessees' specific circumstances, which an assessor must take as given and may not know exactly. Conversely, the intensity of the noise that obfuscates effort from the principals' point of view may be assumed to be exogenously low enough to validate the first-order approach (Lazear and Rosen, 1981, footnote 2), but precision is an obvious costly choice when exams can be more or less complex, and editorial processes can use more or less competent and numerous referees.

Section 2 studies the problem of a rational assessee who finds it costly to improve performance, may not know exactly how preparation effort improves performance, and is rewarded randomly when the assessor's imprecise estimate of the assessee's performance exceeds a threshold. The assessee's first-order condition equates the marginal cost of preparation effort to its marginal effect on the pass probability and is sufficient for a local maximum when the second-order condition is satisfied. Under mild regularity conditions these conditions have at most two solutions, and the assessee's optimal choice is the one that yields the higher expected value.

Section 3 studies the problem of an assessor who prefers better performance, estimates it with a degree of precision that may be increased at a cost, and can choose the threshold. The test's precision and threshold parameters may locally increase or decrease the assessee's chosen performance. The higher fear-failing performance is optimal when at given precision the threshold is sufficiently low, or at given threshold precision is sufficiently high. Both the threshold and precision can increase while respecting that "no-jump" constraint to elicit increasingly high performance. In the infinite precision limit the threshold would set to zero the expected value of performing just above it, and passing with probability one. When high precision is costly or impossible to achieve, however, the assessor's finite precision choice equates its marginal cost to its marginal benefit in terms of higher performance, and may be low enough to imply slackness of the no-jump constraint.

Section 4 inspects numerically and characterizes analytically the model's solution when the test's imprecision is described by a normally distributed error. Normality is approximately realistic when passing the test requires satisfying sufficiently many criteria drawn from of an unbounded set and each is satisfied with a probability that increases in performance. This stylized test structure arguably fits higher education exams and research assessment better than one where the assessee chooses whether or not to meet with probability one some elements of a finite set of which the assessor tests only a subset, as in Eeckhout, Persico, and Todd's (2010) model of tests of rote learning and rule enforcement. Because the density is symmetric, performance increases in the threshold when the assessee optimally performs below it. Because higher precision increases the normal density between its inflection points, the direction of precision's effect depends on whether endogenous performance differs from the pass threshold by more or less than one standard deviation. At given threshold, performance is locally maximal in precision when the pass probability is about 84% for an assessee who fears to be randomly failed (as in Bertola, 2021), about 16% for an assessee who hopes to pass. The assessor prefers the higher fear-failing performance, but can elicit it only if its expected value is higher than that of the lower hope-to-pass performance. The assessor's choice of precision and threshold must satisfy that no-jump constraint, and when precision is costly equates its marginal cost to its marginal productivity.

If parameters are not known with certainty, to obtain high expected or average performance it is optimal to elicit both hope-to-pass and fear-failing behavior, and in a population of heterogeneous assessees the distribution of chosen performance dips around the threshold. Section 5 finds that the empirical distribution of a standard exam's results with about 50% pass frequency do conform to this distinctive implication and discusses what the paper's modeling framework can reveal about other assessments.

2 The assessee's problem

An assessee can improve performance x at cost $c(x)$ and obtains a reward $\upsilon > 0$ if an assessor's estimate of performance is higher than a threshold ζ. Preparation effort does not exactly determine performance (uncertainty about the pass threshold would have similar implications), and the assessor estimates performance imprecisely. Formally, when the assessee aims to perform at level x the realized performance is $y = x + \eta$, where the random variable η with variance σ_η^2 indexes uncertainty about effort's productivity. The assessor estimates y with an independently distributed error ε with variance σ_ε^2. Both the η and ε errors are zero in expectation: the assessee is rationally aware of being poorly informed, and the assessor's estimate is unbiased.

The test is passed when $y+\varepsilon = x+\eta+\varepsilon > \zeta$. Denoting the overall precision of the assessment with

$$\vartheta \equiv 1/\sigma \in (0, \infty) \text{ for } \sigma \equiv \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2} \tag{1}$$

and writing $\text{Prob}(\eta + \varepsilon < z) = F\left(z/\sigma\right)$ for $F\left(\cdot\right)$ the standardized distribution function of $\eta + \varepsilon$, the probability of passing the threshold is

$$\begin{aligned} \text{Prob}\left(x + \eta + \varepsilon > \zeta\right) &= 1 - \text{Prob}\left(\eta + \varepsilon < \zeta - x\right) \\ &= 1 - F\left(\left(\zeta - x\right)\vartheta\right) \equiv P(x; \zeta, \vartheta) \end{aligned}$$

and the expected value of the test for the assessee is

$$V(x; \zeta, \vartheta) = P(x; \zeta, \vartheta)\upsilon - c(x). \tag{2}$$

2.1 Optimality

The assessee maximizes (2). Because $\upsilon > 0$, the optimal performance choice

$$x^* = \arg\max_x \left[P(x; \zeta, \vartheta) - c(x)/\upsilon\right] \tag{3}$$

depends on the pass probability function $P(x;\zeta,\vartheta)$ and on the unitless cost/reward function $c(x)/\upsilon$. Suppose

Assumptions 1 ***(a)*** *The cost $c(\cdot)$ is convex, and zero at or below a finite performance level $\alpha < \zeta$ achievable at no cost; the marginal cost $c'(\cdot)$ is continuous at a α and increasing at all costly performance levels:*

$$c(x) = 0\ \forall x \leq \alpha;\ c'(\alpha) = 0,\ c'(x) > 0,\ c''(x) > 0\ \forall\ x > \alpha. \tag{4}$$

(b) *The distribution function of the error is strictly less than unity at α, twice continuously differentiable, and strictly increasing at all costly performance levels:*

$$F\left(\left(\zeta - \alpha\right)\vartheta\right) < 1,\ F'\left(\left(\zeta - x\right)\vartheta\right) > 0\ \forall\ x > \alpha \tag{5}$$

This ensures that it is possible (and potentially optimal) for the assessee to pass the test when choosing a low-cost positive performance. The Appendix shows

Fact 1 *Assumptions 1 imply that x^* is an interior solution of problem (3) for all $\vartheta < \infty$, and satisfies*

$$\vartheta F'\left(\left(\zeta - x^*\right)\vartheta\right) - c'(x^*)/\upsilon \quad = \quad 0, \tag{6}$$

$$-\vartheta^2 F''\left(\left(\zeta - x^*\right)\vartheta\right) - c''(x^*)/\upsilon \quad < \quad 0. \tag{7}$$

Assumptions 1 rule out a corner solution at costless effort and ensure that participating in the assessment has positive value at an optimal interior choice, where (6) equates to first order the positive density of errors that trigger failure to the marginal cost/reward ratio of higher performance, and (7) ensures local maximization requiring that ratio to increase in performance more strongly than the pass probability.

2.2 Multiple local maxima

The solution (3) of the assessee's problem satisfies (6) and (7), but those conditions need not uniquely identify it. To see why, note that (3) has the same form as a firm's profit-maximization problem with cost function $c(x)/\upsilon$ and revenue $P\left(x;\zeta,\vartheta\right) = 1 - F\left(\left(\zeta - x\right)\vartheta\right)$. The cost is convex by (4), but the marginal revenue is globally concave and first-order conditions uniquely identify the assessee's optimal choice only if $-\vartheta^2 F''\left(\left(\zeta - x\right)\vartheta\right) < 0$ is negative for all x, i.e. the error's density decreases everywhere. If estimation errors are unimodal instead, performance has locally increasing marginal implications for the pass probability. Hence, the conditions for

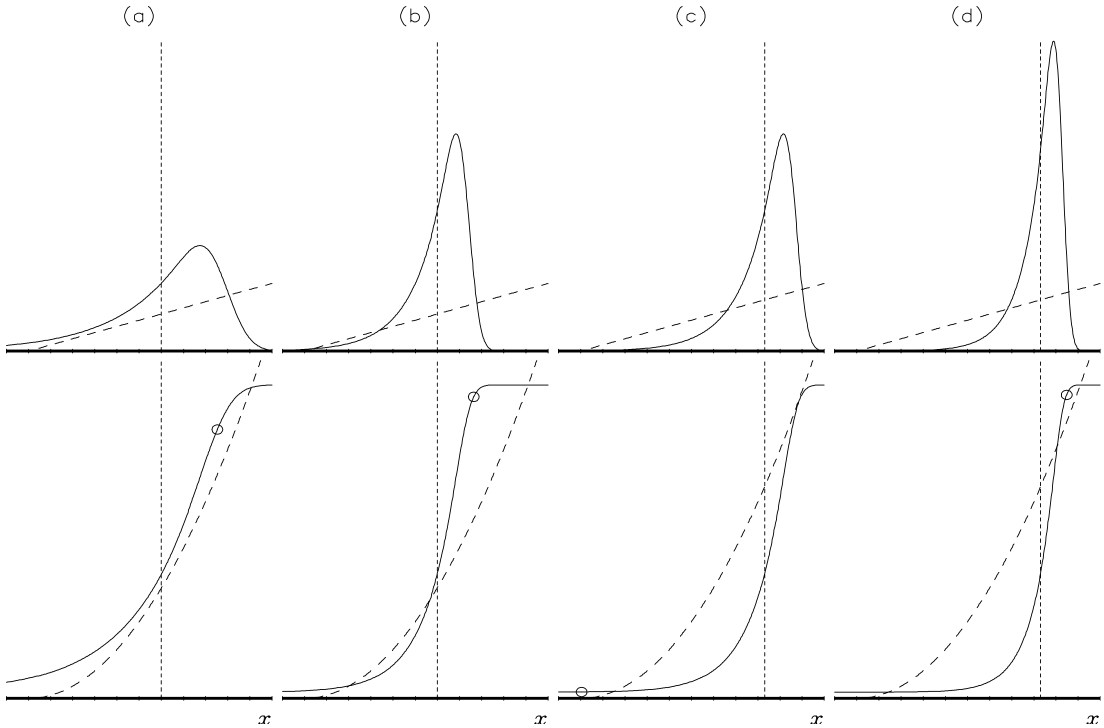

Figure 1: Top panels plot error densities and marginal cost/reward, bottom panels plot pass probabilities and total cost/reward, as functions of performance on the horizontal axis. Vertical line at the pass threshold, circles at the performance and pass probability that maximize the assessee's expected value.

a local maximum can be satisfied by more than one performance choice.

It greatly simplifies characterization to suppose

Assumptions 2 *In addition to Assumptions 1, (a)* $F''((\zeta - x)\vartheta) = 0$ *for only one* x*, (b)* $\vartheta F'((\zeta - x)\vartheta) = c'(x)/\upsilon$ *for at most three values of* x*.*

The first new Assumption restricts the error's density to be unimodal, so $\vartheta F'((\zeta - x)\vartheta)$ has a single peak as a function of x and the increasing marginal cost/reward function can cross it at most once along its declining portion. The second new Assumption excludes the peculiar possibility that the marginal cost/reward function may cross it more than twice along its upward-sloping portion, or even coincide with it locally. If that were to happen, the first-order condition would have many or even a continuum of solutions, but

Fact 2 *Assumptions 2 imply that (6) is solved by at most three values of* x *and (7) is satisfied at the smallest and (if there are three) largest solution, which can be globally optimal only if it implies a strictly positive value for (2).*

To see this consider the unimodal pass densities $\vartheta F'((\zeta - x)\vartheta)$ and marginal cost/reward functions in Figure 1. As functions of x the densities are skewed to the left: this is what happens if $F'(z)$ is right-skewed, i.e. large positive errors are more likely than large negative errors, as may be realistic in some settings. All satisfy Assumptions 2 so $c'(x)/v$, increasing monotonically from zero, must cross $F'((\zeta - x)\vartheta) > 0$ from below at least once in the top diagrams. In Figure 1(a) that crossing is the only local maximum and globally maximizes the assessee's objective function, i.e. the distance between the pass probability and the cost/reward ratio in the bottom diagram. This validates a first-order approach to the assessor's problem, but the Assumptions made so far do not rule out other realistic possibilities. In Figures 1(b), (c), (d) the marginal cost/reward function crosses $F'((\zeta - x)\vartheta)$ at three points, which by (7) identify a local maximum if the density grows slower than the marginal cost/reward function (as at the first crossing) or decreases (as at the last crossing). The middle crossing identifies a minimum, the other two are candidates for global optimality. For a given $c(x)/v$, which of the local maxima is global depends on the shape and position of $F'((\zeta - x)\vartheta)$. In Figure 1(b) the higher performance is globally optimal. For the same cost/reward function, the rightward shift of the distribution implied by a higher threshold makes the lower local maximum global in Figure 1(c), where the higher locally optimal performance has negative value, and the smaller and less costly solution of the first-order condition is globally optimal. Higher precision has the opposite implication in Figure 1(d), where a similarly shifted but more concentrated distribution increases the pass probability at the higher local optimum and makes it globally optimal.

3 The assessor's problem

An assessor who knows the other parameters of the assessee's problem and can solve it knows exactly the performance that can only imprecisely be estimated when observing the test's result. What follows characterizes the assessor's choice of the test's threshold ζ and precision ϑ supposing that

Assumptions 3 *The assessor maximizes an objective that increases linearly in the assessee's performance; may choose the threshold freely; and can increase precision at a cost.*

At the cost of some notational complication the assessor's objective function could increase in performance non-linearly, for example because it accounts for the assessee's cost of improving it, with very similar implications. Other objectives and constraints may be realistic and their implications are discussed briefly at the end of the paper.

3.1 Performance implications of test parameters

Whether there are multiple local maxima and which one is global depends on the assessee's cost/reward function, which the assessor takes as given, and on the position and shape of the pass probability density, which the assessor may control choosing the threshold and precision of the test. A higher threshold shifts to the right the density and the distribution, moving from Figure 1(b) to Figure 1(c). Higher precision concentrates the density and steepens the distribution, moving from Figure 1(a) to Figure 1(b), or from Figure 1(c) to Figure 1(d).

The test's precision and threshold influence the $\vartheta F'\left(\left(\zeta - x^*\right)\vartheta\right)$ marginal pass probability benefit of performance. If variation of the threshold or of precision implies a higher pass density at the assessee's optimum then performance increases along the marginal cost/revenue schedule. This is what happens if $x^* > \zeta$ and parameter changes heighten fear of failure, or $x^* < \zeta$ and parameter changes dash hopes of passing. Differentiating (6) and using (7) the Appendix shows

Fact 3 *Assumptions 1 imply that*

$$x^* \text{ increases in } \zeta \text{ at given } \vartheta \text{ if } \quad F''\left(\left(\zeta - x^*\right)\vartheta\right)\vartheta^2 \quad > \quad 0, \tag{8}$$

$$x^* \text{ increases in } \vartheta \text{ at given } \zeta \text{ if } \quad \left(\zeta - x^*\right)\vartheta\frac{F''\left(\left(\zeta - x^*\right)\vartheta\right)}{F'\left(\left(\zeta - x^*\right)\vartheta\right)} \quad > \quad -1. \tag{9}$$

By (8) a more demanding threshold increases performance if the density has negative slope, which by Assumption 2(a) requires $\zeta - x^*$ to be below the error density's mode. By (9) the direction of performance's response to precision depends on whether the elasticity of the estimation error's density exceeds negative unity.

Fact 3 follows from the conditions that establish Fact 1 when performance varies continuously, but variation of the parameters can trigger a jump of the assessee's optimal choice of performance. To see how the $V(x_H; \zeta, \vartheta) \leq V(x_F; \zeta, \vartheta)$ condition that selects $x^* = x_F$ as the global solution of the assessee's problem depends on the test's threshold and precision, an additional regularity condition is useful:

Assumptions 4 *In addition to Assumptions 1, when x_H and x_F both solve (6) and (7) then $x_H < \zeta < x_F$.*

Assuming that when there are two local maxima one is above and one below the pass threshold only mildly restricts the model's functional forms, ruling out zero-mean error densities that are so asymmetric and heavy-tailed as to be crossed by the marginal cost/revenue function twice on the same side of the threshold. The smaller solution of (6) has subscript H for "hope", because an assessee who chooses a performance below the pass threshold ζ hopes to pass when error realizations such that $\eta+\varepsilon>\zeta-x^*$ bring realized performance above the threshold. The larger solution of (6) has subscript F for "fear" because an unlucky realization $\eta+\varepsilon<\zeta-x^*$ may trigger failure for an assessee who, erring on the side of caution, chooses a performance above the pass threshold, and fears failing. Then,

Fact 4 *Assumptions 4 imply*
(a) *A higher threshold decreases $V(x_H;\zeta,\vartheta)$ more strongly than $V(x_F;\zeta,\vartheta)$,*

$$0>\frac{\partial}{\partial\zeta}V(x_H;\zeta,\vartheta)>\frac{\partial}{\partial\zeta}V(x_F;\zeta,\vartheta) \tag{10}$$

hence for each ϑ there is an upper bound on the values of ζ for which $x^*=x_F$.
(b) *Higher precision decreases $V(x_H;\zeta,\vartheta)$ and increases $V(x_F;\zeta,\vartheta)$,*

$$\frac{\partial}{\partial\vartheta}V(x_H;\zeta,\vartheta)<0,\ \frac{\partial}{\partial\vartheta}V(x_F;\zeta,\vartheta)>0 \tag{11}$$

hence for each ζ there is a lower bound on the values of ϑ for which $x^*=x_F$.

A higher pass threshold decreases the assessment's value because it implies a lower pass probability at the optimal performance (which varies but, by the envelope theorem, has no first-order value implications when the first-order condition is satisfied). The Appendix derives (10) showing that the value decreases more strongly at the larger solution of the first-order condition if the marginal cost/reward function is increasing, hence higher there than at the smaller solution. Because a more demanding threshold tends to favor hope-to-pass behavior, the threshold needs to be low enough for fear-failing behavior to be preferable at given precision. To derive (11) the Appendix shows that if, as assumed, two local maxima bracket the pass threshold then higher precision decreases the expected value of the assessment at the smaller solution, where it reduces the pass probability at the locally optimal performance (which again varies with no first-order value effects), and increases it at the higher solution.

When the assessee chooses x_F but is almost indifferent to choosing x_H, total differentiation of $V(x_H;\zeta,\vartheta)=V(x_F;\zeta,\vartheta)$ yields

$$\frac{d\vartheta}{d\zeta}=\frac{\frac{\partial}{\partial\zeta}V(x_H;\zeta,\vartheta)-\frac{\partial}{\partial\zeta}V(x_F;\zeta,\vartheta)}{\frac{\partial}{\partial\vartheta}V(x_F;\zeta,\vartheta)-\frac{\partial}{\partial\vartheta}V(x_H;\zeta,\vartheta)},$$

which is positive by (10) and (11). To prevent a downward jump, a higher threshold requires higher precision. Hence, it is possible to elicit higher performance by raising both the threshold and precision so as to keep the assessee's expected value larger at fear-failing than at hope-to-pass performance. As precision approaches infinity the failure probability approaches zero and performance tends to a level just above $\bar{\zeta}$, the highest threshold compatible with non-negative expected value.

3.2 Costly precision

A perfectly precise test could in the limit remove the information asymmetry that makes it necessary to fail unlucky assessees in order to preserve incentives to perform. This is obviously not what happens in a reality where precision cannot costlessly increase to infinity. The assessor may reduce the estimation error's variance by asking more numerous and detailed questions in exams, or using more numerous and competent referees in editorial process, but will find it optimal to do so only as long as the marginal cost of higher precision is lower than its marginal performance benefit. At the optimum,

Fact 5 *The test parameters that per Assumptions 3 maximize performance $x^*(\vartheta, \zeta)$, net of a differentiable increasing cost $C(\vartheta)$ of precision, necessarily satisfy*

$$V(x_H; \zeta, \vartheta) - V(x_F; \zeta, \vartheta) \leq 0 \quad (12)$$

$$\partial x^*(\vartheta, \zeta)/\partial \zeta \geq 0, \quad (13)$$

$$\partial x^*(\vartheta, \zeta)/\partial \vartheta - C'(\vartheta) \leq 0, \quad (14)$$

with complementary slackness: the inequalities in (13) and (14) can be strict only if (12) holds with equality.

When (12) holds with strict inequality and performance is a continuous function of ζ and ϑ, marginal variation of ζ or ϑ cannot violate the no-jump constraint, hence (13) and (14) must hold with equality. The optimal choice of the ζ threshold neither increases nor decreases performance, and by (8) the optimal performance's distance from the threshold is at the mode of the error's density. Precision's performance effect equals its positive marginal cost $C'(\vartheta)$ at a finite optimal ϑ which is smaller if, in terms of the value of performance for the assessor, $C'(\vartheta)$ is larger.

When instead (12) holds with equality, strict inequality is generally optimal in (13) and (14). A higher threshold would improve performance if it responded continuously, but by increasing the left-hand side of (12) with equality triggers a downward jump of performance. Reducing precision would increase $x^*(\vartheta, \zeta) - C(\vartheta)$ if it were a continuous concave function of ϑ, but similarly triggers a downward jump of performance when it makes it optimal for the assessee to choose the smaller hope-to-pass solution of the first-order condition.

3.3 Hidden information

In reality information is asymmetric not only because the assessee's action is obscured by imprecise measurement, but also because the assessor does not know exactly how performance depends on the threshold and precision. Improving performance may require more or less intense effort and passing may be more or less desirable depending on individual circumstances that are known to the assessee, but not to the assessor and third parties. Moreover, communicating precision and threshold exactly to assessees is difficult.

If an assessee's performance depends on parameters that have a continuous distribution known by an assessor who intends to maximize expected performance, a higher threshold improves performance with the probability of parameters that satisfy (8), worsens it with that of parameters that do not satisfy (8) or lead to a violation of (12) and a downward jump of performance. The implications of precision similarly depend on whether (9) is satisfied and whether lower precision triggers a jump from the higher to the lower solution of the assessee's first-order condition.

As long as discrete jumps have infinitesimal measure on the set of performance reactions indexed by i, the effects of threshold and precision variation are continuous in expectation. Hence,

Fact 6 *When the assessor knows the assessee's reaction function only up to a continuous probability distribution, the test parameters that per Assumptions 3 maximize expected performance* $E\left[x_i^*(\vartheta,\zeta)\right]$ *over the continuous index* i *of performance reaction functions, net of an increasing cost* $C(\vartheta)$ *of precision, necessarily satisfy*

$$\partial E\left[x_i^*(\vartheta,\zeta)\right]/\partial\zeta = 0, \tag{15}$$

$$\partial E\left[x_i^*(\vartheta,\zeta)\right]/\partial\vartheta = C'(\vartheta). \tag{16}$$

Condition (15) weighs the assessee's possible reactions by their probability density and, because variation of the threshold is costless, imposes that they locally integrate to zero. Hence, the optimal choice of threshold necessarily implies that performance may be higher or lower than the threshold with positive probability, and that hope and fear coexist in expectation. Condition (16) equates the expected performance effect of the threshold to the marginal cost of precision. When the no-jump constraint would be binding under certainty, the assessor's choice of parameters implies that it is expected to be loose, so as to balance the probabilities that moving performance closer to the constraint would marginally improve performance or trigger a discrete downward jump.

4 Normally distributed errors

Standardization of the zero-mean error by its precision ϑ allowed the derivations above to analyze that parameter's and the threshold's effects on the assessee's performance. Precision is not a primitive structural parameter in general, because structural features of a test determine not only the second but also the higher moments of the error's mean-zero distribution.[1] It is however straightforward to characterize and interpret the solution of the assessee's and assessor's problems when errors are normally distributed, a tractable assumption that is not unusual and, as shown below, is approximately valid for realistic assessment structures.

If both ε and η are normally distributed, so is their sum, and its standardized distribution $F(\cdot)$ is the unit-variance normal probability function $\Phi(\cdot)$ with density

$$\Phi'(z) = e^{-z^2/2}/\sqrt{2\pi} \equiv \phi(z) \tag{17}$$

Assuming a simple quadratic form for the cost function

$$c(x) = \begin{cases} 0 & \text{for } x \leq \alpha, \\ (x-\alpha)^2\gamma/2 & \text{for } \alpha < x \end{cases} \tag{18}$$

results in an explicit model with few interpretable parameters that conforms to all the Assumptions above. In particular, symmetry of the density makes it impossible for two solutions of (6) to be on the same side of the threshold, and the marginal cost/reward schedule implied by (18) cannot cross the bell-shaped normal density more than three times.[2]

It is not possible to obtain explicit expressions for the solutions of the first-order condition and of the no-jump condition, but some analytic results ease interpretation of numerical illustrations. The Appendix shows that (17) and (18) imply

$$\lim_{|\zeta|\to\infty} x^* = \lim_{\vartheta\downarrow 0} x^* = \alpha : \tag{19}$$

[1] Bertola (2023, Appendix) outlines in some detail how the variation of precision at given mean may be isolated from that of other moments in a structural model.

[2] A marginal cost function with constant slope $\gamma > 0$ eases derivation of some results, mentioned in the following footnotes, that would be similar but more complicated for other increasing and convex explicit cost functions.

the optimal performance tends to be near the costless level when the threshold is very remote or the test is very imprecise, both of which make it nearly pointless to exert effort in order to increase the pass probability. Setting $x^* > \alpha$ is worth the cost only if $\vartheta > 0$ and possibly very negative normal errors may trigger a failure. At the other extreme, a perfectly precise test should be passed with probability one if the cost $c(\zeta)$ of preparation at the threshold is less than the reward υ, failed with probability one at the costless level α otherwise. The Appendix shows that (18) implies

$$\lim_{\vartheta\to\infty} x^* = \begin{cases} \zeta & \text{if } \zeta < \alpha + \sqrt{2\upsilon/\gamma} \approx \alpha + 1.41\sqrt{\upsilon/\gamma} \equiv \bar{\zeta}, \\ \alpha & \text{otherwise.} \end{cases} \tag{20}$$

When imprecision makes it optimal to choose a performance between α and ζ, symmetry of the normal density simplifies the interpretation of diagrams that as in Figure 1 identify optimal solutions of the first-order condition, and functional forms (17) and (18) make it possible to see clearly how variation of the threshold and of precision determines performance. In the following numerical illustrations the parameters of the cost/reward function (18) are $\upsilon = 1$, $\alpha = 0$, $\gamma = 2$. Because performance is costless at $x = 0$ and prohibitively costly at $x > \bar{\zeta} = 1$ by (20) its potentially optimal levels range from zero to unity, like pass probabilities, and the diagrams fit neatly in a unit square. Other parameters would shift and rescale the diagrams without altering their qualitative features.

4.1 The threshold's performance implications

Figure 2 shows how the threshold ζ determines the assessee's optimal choice x^* for two values of precision. A higher threshold ζ shifts the density and pass probability functions horizontally to the right, as between Figures 1(b) and Figure 1(c). In Figure 2(a) higher thresholds increase the assessee's optimal performance until $x^* = \zeta$, at failure probability $\Phi(0) = 50\%$. Further increases of the threshold reduce performance, and the pass probability decrease smoothly while remaining above the cost/reward function. For this to happen precision must be sufficiently low.[3] In Figure 2(b) a higher value of precision implies that as ζ grows performance decreases gradually only until it approaches the cost/reward function, where it drops discontinuously because the expected value becomes lower at fear-fail than at hope-to-pass.

Figure 3 plots the performance and pass probability implications of the threshold ζ at various

[3] The Appendix shows that $\vartheta \leq \left(\alpha\gamma/\upsilon + 1/\sqrt{\upsilon/\gamma}\right)/\phi(0)$ ensures that the assessee's value is positive at the fear-failing local maximum. The bound is about 3.54 for the $\alpha = 0$, $\gamma = 2$, $\upsilon = 1$ parameters of the numerical solution.

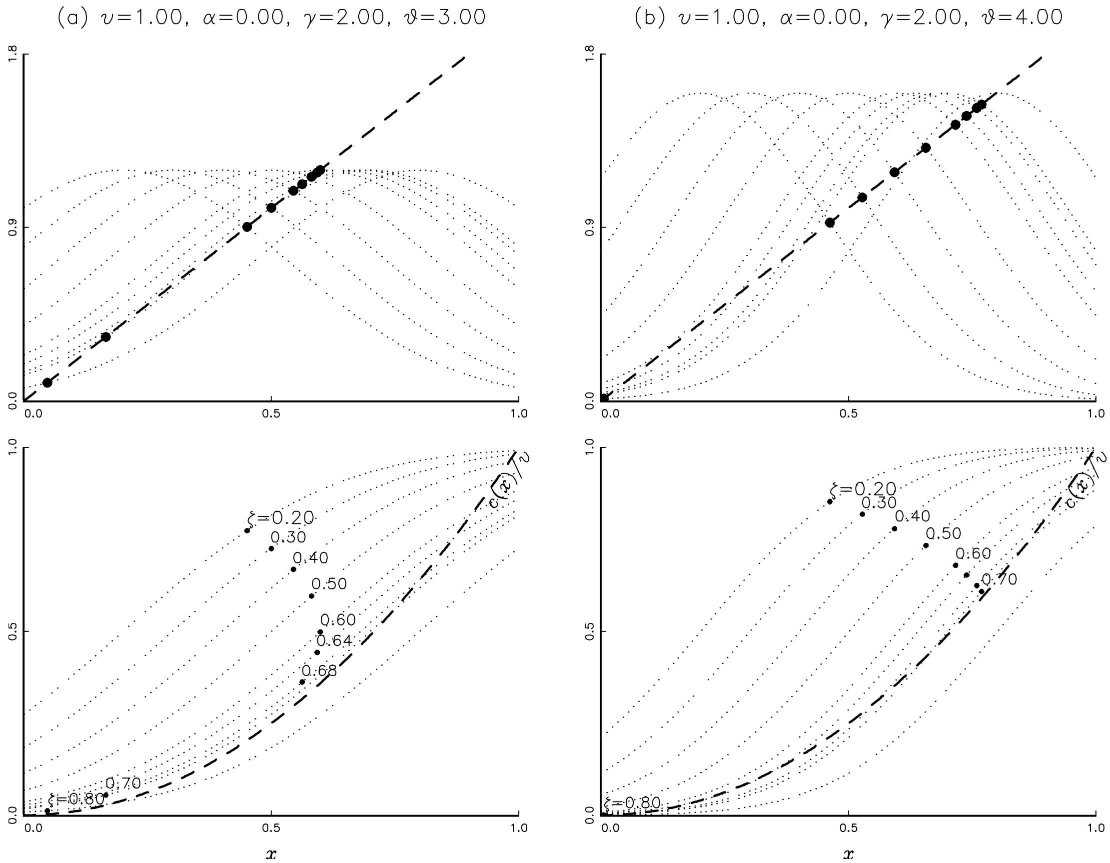

Figure 2: Densities and probability functions for various thresholds and two precisions. Thick dashed lines plot the marginal and total cost/reward functions, dots mark the assessee's choice of performance.

values of precision ϑ. Because $F''(z)$ is positive when $z < 0$, it immediately follows from (8) that

$$\frac{\partial x^*}{\partial \zeta} > 0 \text{ iff } x^* > \zeta \text{ and } P(x^*; \zeta, \vartheta) > 0.5 : \tag{21}$$

when performance is above the threshold, it increases in ζ at given ϑ. Under normality, and for any symmetric zero-mean error distribution, the assessee is more likely to pass than to fail when fearing failure, so the threshold increases performance when the pass probability is larger than 50%. In Figure 3, higher thresholds ζ increase the assessee's performance x^* when $x^* > \zeta$ and the curves are above the dashed 45^0 line in the top diagram, and above 50% in the bottom diagram. When precision is low the sign of the effect changes smoothly through zero. At higher precision, when the threshold becomes so high as to make hope-to-pass behavior optimal then performance jumps down, around the $x^* = \zeta$ locus at 50% pass probability.

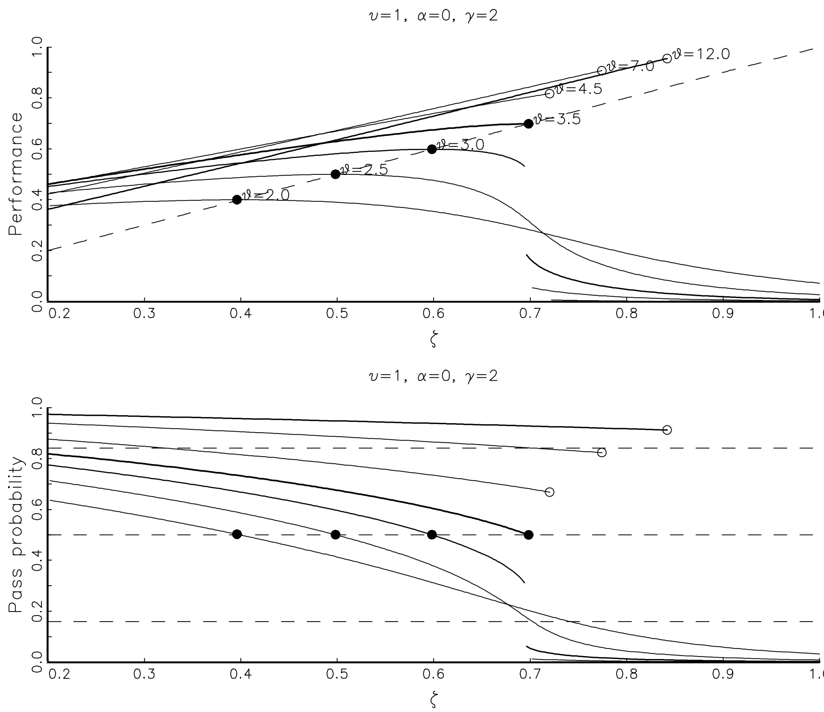

Figure 3: Implications of continuous threshold variation at various precisions.

4.2 Precision's performance implications

Figure 4 shows how precision determines performance and the pass probability for two values of the threshold. In Figure 4(a), performance increases monotonically in precision along the linear marginal cost/reward schedule through the mode of the density and a 50% pass probability, moving smoothly from hope-to-pass to fear-failing local optima of the assessee's problem which, as in Figure 1(a) and Figure 1(b), identify the global optimum. Fear-failing performance remains optimal at higher ϑ, and begins to decrease when performance reaches $x_F^* = \zeta + 1/\vartheta$. Indeed, the Appendix shows that Fact 3 and normality imply

$$\frac{\partial x^*}{\partial \vartheta} > 0 \text{ if } |(\zeta - x^*)\,\vartheta| < 1 \iff \Phi(-1) < P(x^*; \zeta, \vartheta) < \Phi(1), \tag{22}$$

Higher precision improves performance when the distance between x^* and ζ is in absolute value smaller than the inverse of precision, i.e. performance is within one standard deviation of the pass threshold. To see why, note that a more concentrated normal distribution shifts probability away from the tails and towards the region between the standardized normal density's inflection points at -1 and 1. In that region, higher precision strengthens the marginal impact of performance on the pass probability, making it optimal to choose a higher and more costly performance, and the pass probability $P(x; \zeta, \vartheta)$ varies between the $\Phi(-1) \approx 0.16$ and $\Phi(1) \approx 0.84$ values it takes at the boundaries.

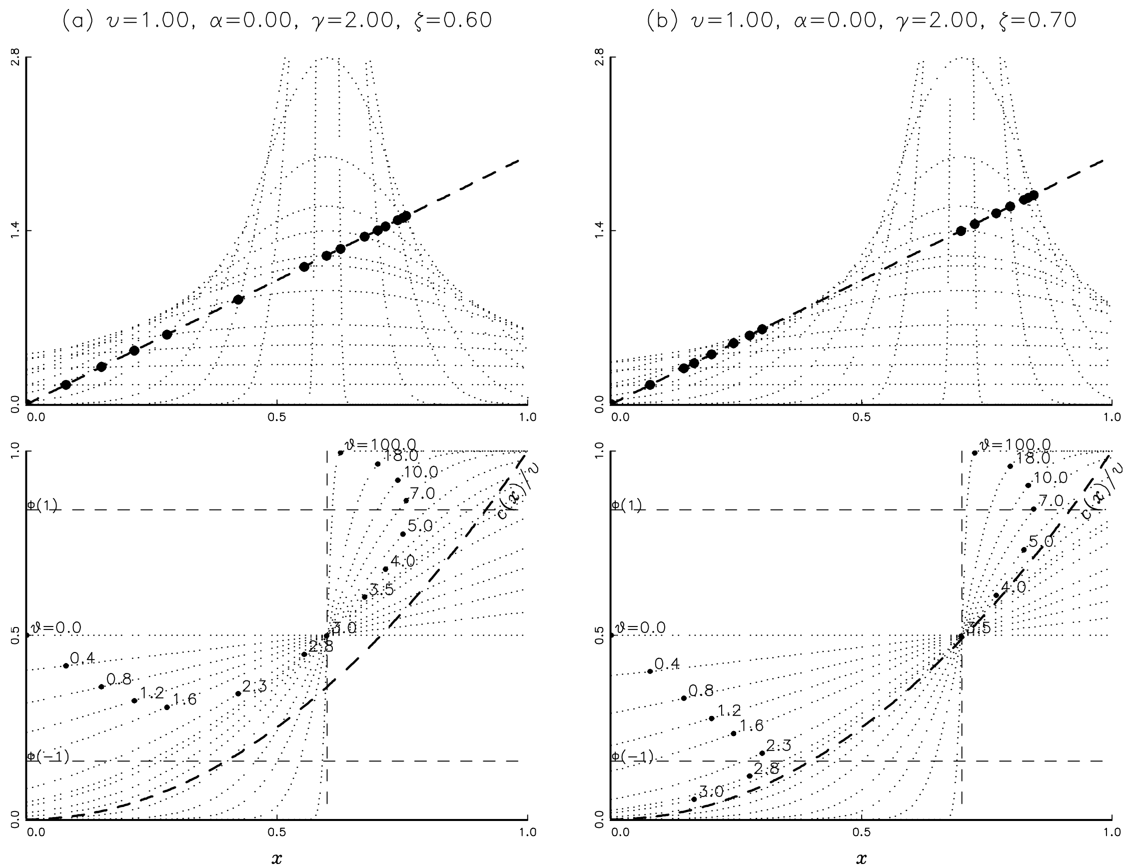

Figure 4: Densities and probability functions for various precisions and two thresholds. Thick dashed lines plot the marginal and total cost/reward functions, dots mark the assessee's choice of performance.

Performance is continuous in precision and there is a unique local performance maximum if the threshold is low enough.[4] This is what happens in Figure 4(a) but in Figure 4(b), where the threshold is higher, performance increases as precision grows from zero, begins to decrease when it reaches $\zeta - 1/\vartheta$, and continues to decrease until precision is so high as to make it optimal for the assessee to trade the cost of discretely higher performance off a discretely higher pass probability above 50%, and performance jumps around ζ from a hope-to-pass to a fear-failing local maximum. As precision increases further, performance grows to $x_F^* = \zeta + 1/\vartheta$, where it begins to decrease.

Figure 5 plots performance and pass probabilities as functions of precision for various thresholds. By (22), precision ϑ locally increases performance x^* when $\zeta - 1/\vartheta < x^* < \zeta + 1/\vartheta$, i.e.

[4] The Appendix shows that performance cannot equal $\zeta - 1/\vartheta$ at any precision if

$$\zeta \leq \alpha + \sqrt{4\phi(1)\upsilon/\gamma} \approx \alpha + 0.98\sqrt{\upsilon/\gamma} \equiv \hat{\zeta}.$$

For the parameters used in the figures $\hat{\zeta} \approx 0.69$.

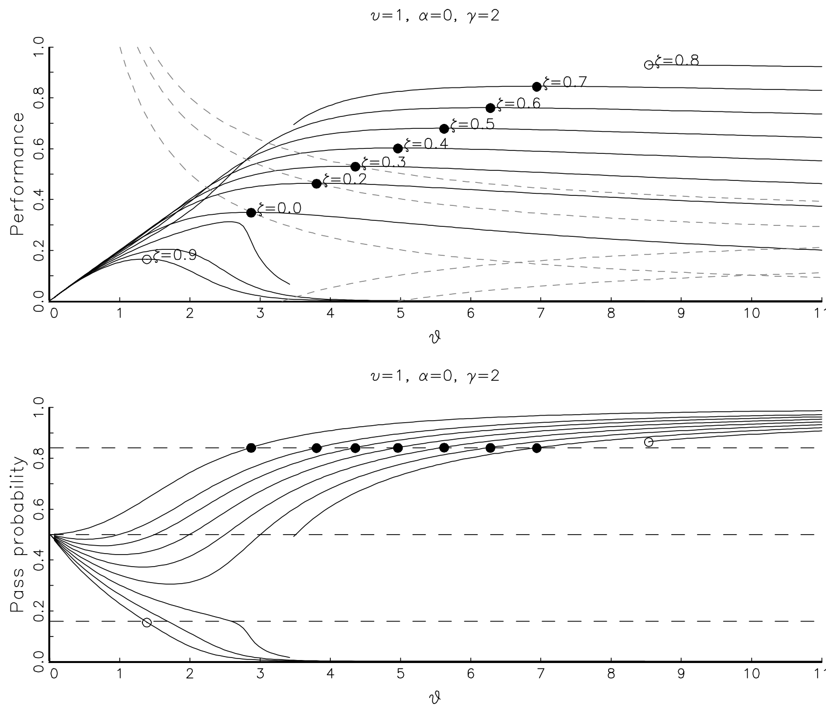

Figure 5: Implications of continuous variation of precision at various thresholds.

when the curves are between the dashed hyperbola drawn for each value of the threshold ζ in the top diagram, and between the dashed lines drawn at $\Phi(1)$ and $\Phi(-1)$ in the bottom diagram. This is what happens for all values of ζ as ϑ increases from zero. The curves drawn for sufficiently low thresholds grow up to $\zeta + 1/\vartheta$, where performance implies a $\Phi(1)$ probability of passing, then decrease towards the threshold that as $\vartheta \to \infty$ implies a unitary pass probability. At higher thresholds the curves decrease upon reaching $\zeta - 1/\vartheta$, where the assessee hopes to pass with probability $\Phi(-1)$, then jump upward, grow further towards $\zeta + 1/\vartheta$, and decrease towards that limit as $\vartheta \to \infty$.[5] For higher thresholds performance jumps to a level above $\zeta + 1/\vartheta$, at a pass probability higher than $\Phi(1)$, and immediately begin to decrease. This is what the line drawn for $\zeta = 0.8$ does in Figure 5, where the line drawn for $\zeta = 0.9$ has yet to jump up at the highest $\vartheta = 11$ precision considered in the figure.

[5] The Appendix shows that this happens if $\hat{\zeta} < \zeta < \tilde{\zeta}$ for

$$\tilde{\zeta} \equiv \alpha + \frac{\Phi(1) - \phi(1)/2}{\sqrt{\Phi(1)/2}} \sqrt{\upsilon/\gamma} \approx \alpha + 1.11\sqrt{\upsilon/\gamma},$$

which prevents performance from equalling $\zeta + 1/\vartheta$ at any precision. For the parameters in the figures $\tilde{\zeta} \approx 0.785$.

4.3 The assessor's performance menu

Choosing the threshold ζ and precision ϑ the assessor determines the assessee's performance on the nonlinear menu offered by the non-monotonic function $x^*(\zeta,\vartheta)$ illustrated in Figures 3 and 5, which is concave when continuous. At given precision, the threshold maximizes performance at the points marked by full circles in Figure 3 which, because the normal density is symmetric, lie along a 45^o line in the top diagram and imply 50% pass probabilities in the bottom diagram. At given threshold, precision maximizes performance at the points marked by full circles in Figure 5. Because the assessee's marginal cost is constant, those points lie on a straight line in the top diagram. Because the error's distribution is normal, they all imply passing probability $\Phi(1) \approx 0.84$ in the bottom diagram.

Such interior maxima do not always satisfy the no-jump condition. When they do not, corner solutions elicit maximal performance, and empty circles identify them in Figures 3 and 5. At given precision, corner optima increase in the threshold with declining slope in the top diagram of Figure 3, and imply increasingly high pass probabilities in the bottom diagram. In Figure 5, precision needs to be high enough to prevent a downward jump when the threshold is ζ =0.8. An empty circle also appears at the lower local performance maximum $\zeta - 1/\vartheta$ when $\zeta = 0.9$ implies that an upward jump to a global constrained maximum lies beyond the end of the precision axis, as it certainly does when $\zeta > \bar{\zeta}$ prevents optimality of fear-failing performance at any precision.

If the assessor could choose the threshold and precision freely maximal performance would be elicited at the infinite precision limit of the no-jump-constrained empty dots in Figure 3, which as ϑ grows get closer to the 45^0 line and $\bar{\zeta} = 1$, and imply pass probabilities that tend to unity. The explicit functional forms help see what can prevent this. As precision grows towards infinity its performance implications become smaller in Figure 5, and if its marginal cost is positive the assessor will find it optimal choose a finite value which may be low enough to loosen the no-jump constraint, and validate a first-order-condition characterization of the assessor's problem.

4.4 Cost of precision

A normal error distribution approximates realistic tests where both the cost of performance and the cost of precision have natural interpretations. Consider an assessment based on N criteria that is passed when at least Z are satisfied. If an assessee who chooses performance x satisfies each criterion with probability $p(x)$, the realized proportion of satisfied criteria is approximately normal with mean $p(x) + 1/(2N)$, where the continuity correction $1/(2N)$ accounts for the fact that the normal cumulative distribution approximates a discrete binomial probability, and variance

$$\sigma_\varepsilon^2 = p(x)(1 - p(x))/N. \tag{23}$$

As in the stylized model above, the pass criterion requires the intended performance that the assessor infers from the realized proportion of satisfied criteria to exceed a threshold $\zeta = Z/N$. The normal approximation to the error's distribution is accurate when N is large and $p(x)$ is not far from 0.5. The derivations above further approximate this test structure by not allowing precision to depend on the student's performance choice x. Normality of the composite error can also be realistic if η averages many types of errors. Denoting $p(x)(1-p(x)) \equiv \xi$, if ξ is close to 1/4 and N is large the test's precision (1) is approximately

$$\vartheta = 1/\sqrt{\sigma_\eta^2 + \xi/N} \tag{24}$$

at given x. The assessor must take σ_η^2 as given but can increase N, for example asking more numerous questions in exams or referees in editorial processes. Precision ϑ requires $N = \xi\vartheta^2/\left(1 - \sigma_\eta^2\vartheta^2\right)$ assessment criteria, so if the unit cost of additional assessment criteria is constant at χ the cost of precision $C(\vartheta) = \xi\chi\vartheta^2/\left(1 - \sigma_\eta^2\vartheta^2\right)$ is increasing and convex in ϑ as long as $\vartheta^2 < 1/\sigma_\eta^2$. As ϑ tends to $1/\sigma_\eta$ its slope $C'(\vartheta) = 2\xi\chi\vartheta/\left(\vartheta^2\sigma_\eta^2 - 1\right)^2$ tends to infinity because overall precision is bounded by the precision of the assessee's perception of performance's implications.

4.5 Optimal test parameters

When the structure of the assessee's problem is known with certainty the solution of the assessor's problem satisfies the conditions in Fact 5. The diagrams on the left in Figure 6 depict its objective function for parameters such that the solution of the nonlinear system of equations in ϑ and ζ formed by (13) and (14) with equality violates (12): because the no-jump constraint binds, the threshold and precision optimally to place the assessee's performance on the edge of a precipice. In the diagrams on the right, increasing precision is more costly and the irreducible imprecision indexed by σ_η^2 is larger. This makes it optimal for precision to be low and for performance to be distant from the no-jump constraint, and the assessor's first-order conditions hold. Equality in (13) requires equality in (21), so the assessee chooses to perform exactly at the threshold and has 50% pass probability. By (14) with equality, the optimal choice of precision ensures that its performance effect equals its marginal cost.

If the assessor knows the assessee's problem only up to a probability distribution the optimal choices satisfy the conditions on expectations of the assessee's reactions in Fact 6. Figure

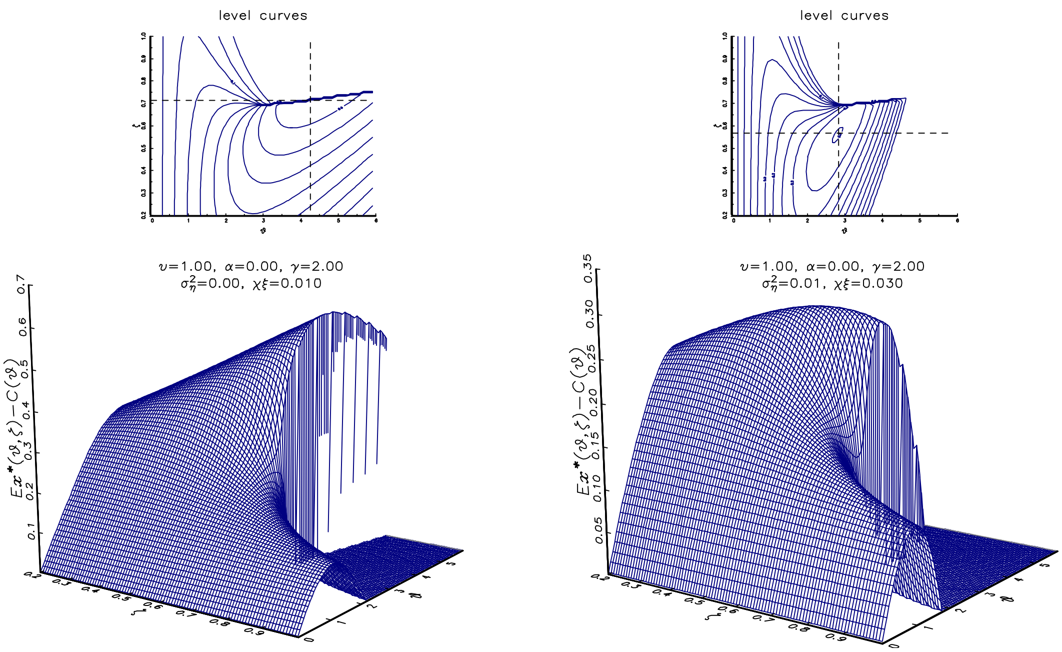

Figure 6: Level curves and three-dimensional view of the assessor's objective as a function of costly precision ϑ and threshold ζ. The assessor knows all the assessee's parameter with certainty.

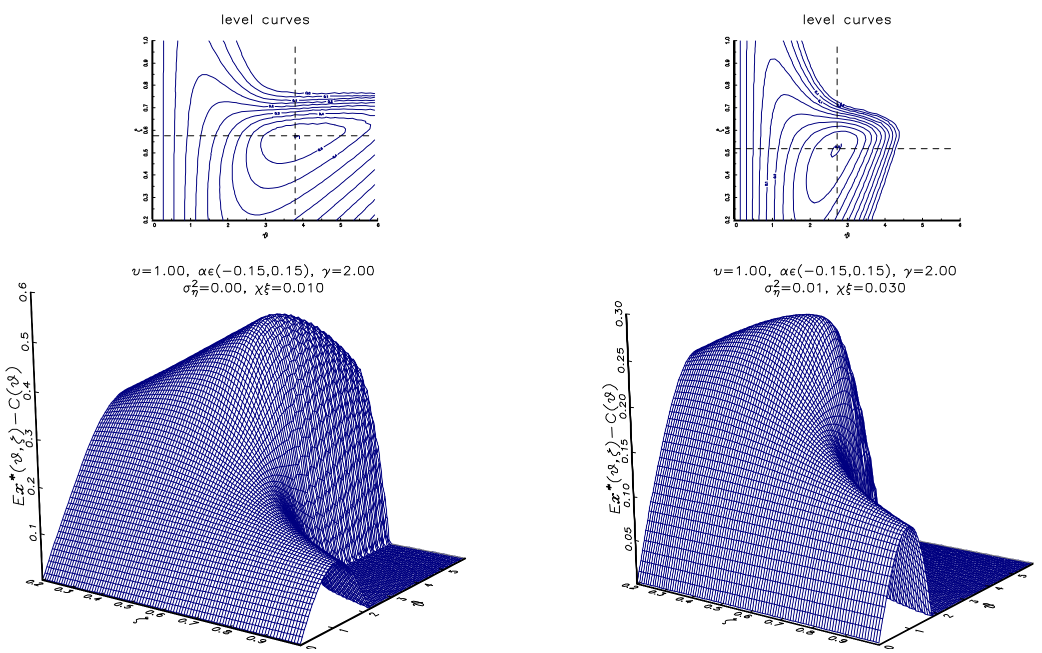

Figure 7: Level curves and three-dimensional view of the assessor's objective as a function of costly precision ϑ and threshold ζ. The assessor knows α only up to a triangular distribution centered at the same value known with certainty in previous figures.

7 illustrates supposing that the horizontal intercept α of the cost function has a triangular distribution centered at zero, while all other parameters are known and the same as in Figure 6. On the left, uncertainty about the precise location of the no-jump constraint leads the assessor to cautiously place expected performance quite some distance away from the edge of the sheer cliff visible in Figure 6. Uncertainty has similar but much milder implications in the diagrams on the right, where the parameters are such that the assessor would choose an interior solution under certainty.

The assessor's problem under uncertainty and its solution are easier to inspect from a two-dimensional perspective that highlights heterogeneity of assessee performance. Figure 8 fixes precision at the assessor's optimal choice and, like Figure 3, show how various thresholds would determine performance and pass probabilities that, for concreteness and to derive observable implications, can be viewed as fractions of a large population of assessees. The width of gray lines is proportional to the frequencies implied for each α value by the symmetric triangular distribution used to compute the averages plotted by solid lines. At given threshold, performance is higher for larger values of α. As the threshold becomes more demanding, performances initially increase and diverge across individuals, then jump down or decline continuously, and eventually converge.

Figure 9 is similar to Figure 5, but displays performance reaction functions for a common ζ and various values of α rather than for a single α and various values of ζ. For $\vartheta \approx 0$ all performances start from their respective costless levels, and their distribution is that of α. As precision increases above zero, performance of strong assessees grows faster than that of weaker assessees, which at some point begins to decrease, and might later jump back up.

In Figure 8's left-hand diagrams the no-jump constraint would be binding under certainty and is likely to bind for many assessees when α is uncertain. To prevent downward jumps that would reduce average performance, the assessor chooses high precision and the threshold marked by a vertical line, where few assessees choose hope-to-pass performance and the average pass probability is about 75% for the figure's parameterized example. In the diagrams on the right precision cannot easily be increased, so under certainty the no-jump constraint would not be binding and it would be optimal to ensure that a single assessee with the average parameters would neither hope nor fear, perform at the threshold, and pass with 50% probability. When assessees are unobservably heterogeneous the assessor's optimal threshold, again marked by a vertical line, satisfies the first-order condition in Fact 6: average performance is maximal in the top diagram, there are many assessees on both sides of the pass threshold, and in the bottom diagram the average pass probability differs from 50% only a little because of Jensen's

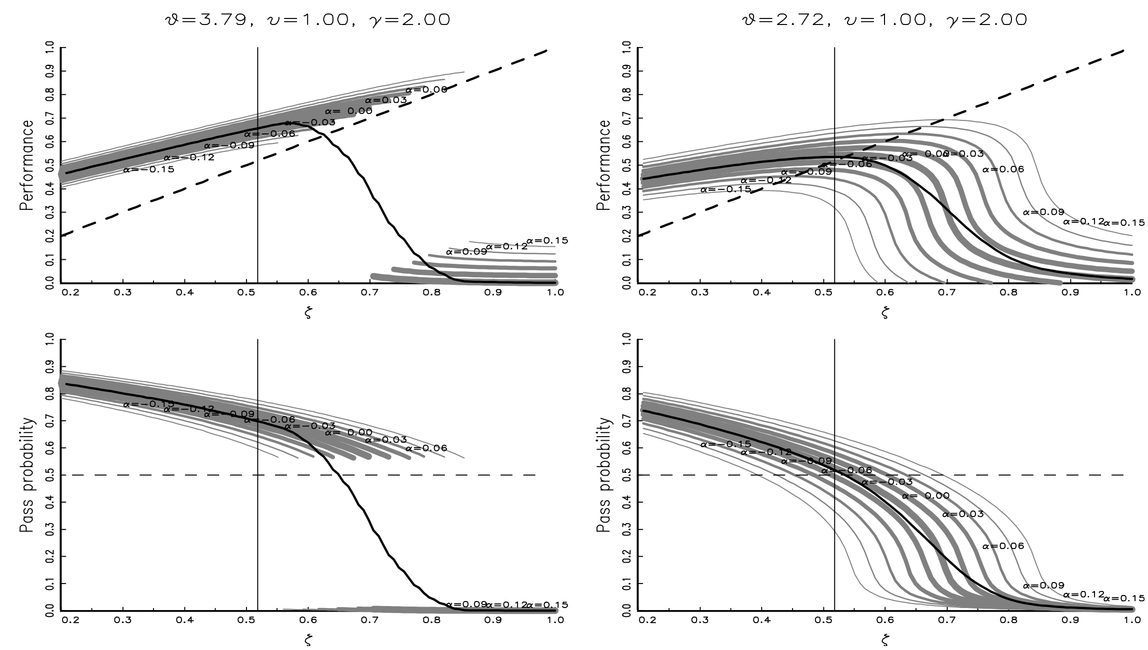

Figure 8: Performance and pass probability as functions of the pass threshold for various values of the costless performance α and on average. Dashed lines: 45^0 in the top panels, 50% in the bottom panels.

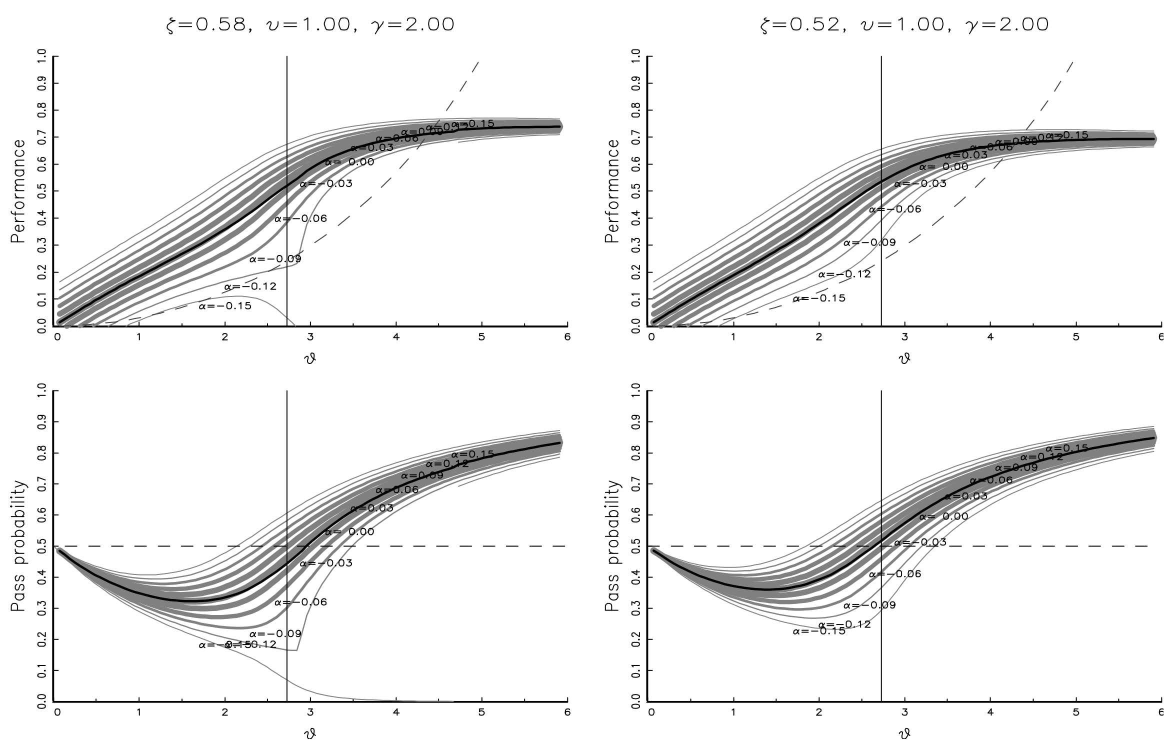

Figure 9: Performance and pass probability as functions of precision for various values of the costless performance α and on average. Dashes lines: assessor's cost of precision in the top panela, 50% pass probability in the bottom panels.

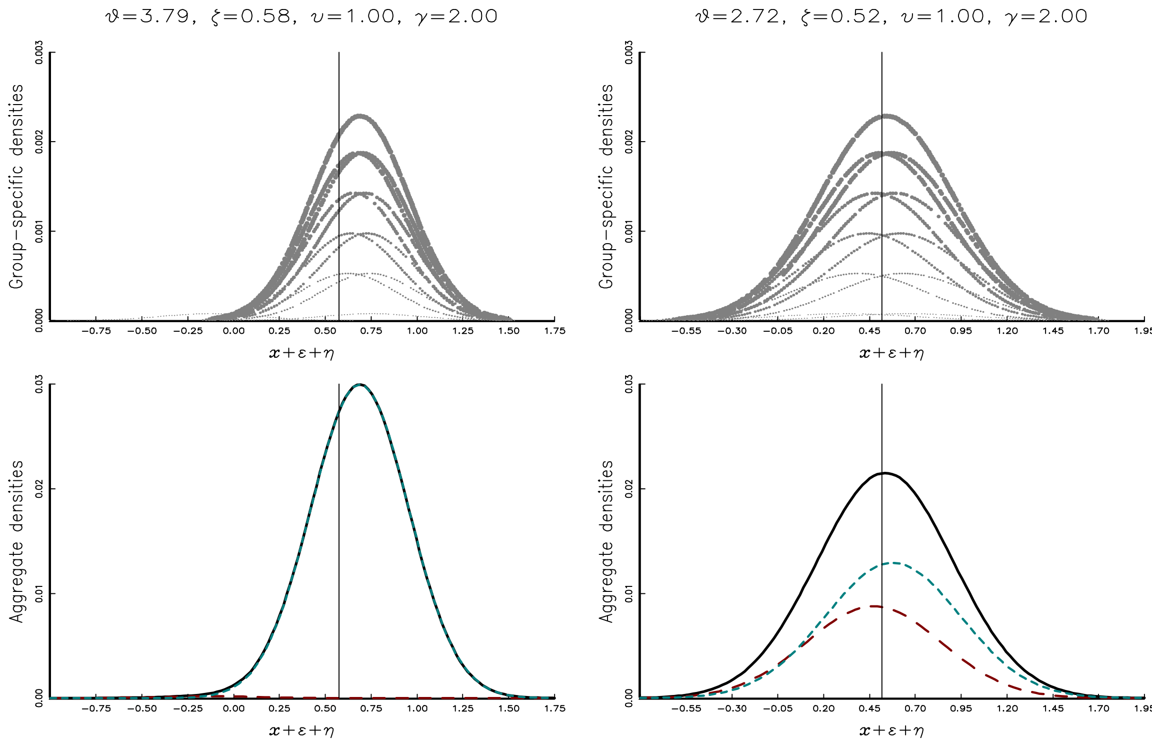

Figure 10: Densities of observed performance for each α in the top panels, and for the total population and hope-to-pass and fear-failing subpopulations in the bottom panels.

inequality. The assessor's optimal choice of precision in Figure 9 also satisfies its first-order condition: precision has a marginal average performance effects that equals its marginal cost in the top diagram, and again distributes assessees on both sides of the 50% pass probability that distinguishes those who fear failing from those who hope to pass.

As test parameters increase from zero the performance of assessees who choose $x^* > \zeta$ and fear failing first diverges from that of assesses who choose $x^* < \zeta$ and hope to pass, then converges. Figure 10 displays the cross-sectional densities of observed performances implied by the optimal threshold and precision. As estimation errors are added to the heterogeneous assessees' chosen performance, discretely different optimal choices of stronger assessees (who choose fear-failing performances) and weaker ones (who choose hope-to-pass performances) imply that the overall distribution of results mixes two approximately normal distributions. In the diagrams on the left, where the optimal precision is high and very few assessees hope to pass, the weight in the mix of the lower-mean distribution is very small. In the diagrams on the right, where the parameters imply lower precision at the optimum, assessees who hope to pass are almost as numerous as those who fear failing, and their weight in the mix would be higher still if even lower precision made the hope-to-pass and fear-failing choices more different.

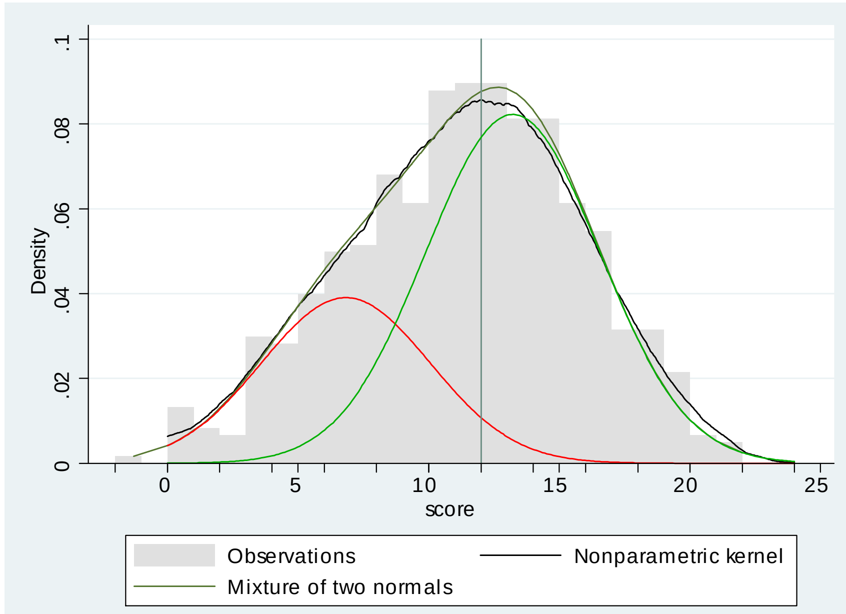

Figure 11: Raw data histogram of a set of exam scores, and mixture of two normals estimated by the Stata command `fmm:regress` (or, almost identically, by the Stata external program `denormix`). A chi-square test (p-value 0.5887) and a one-sample cumulative Kolmogorov-Smirnov test (p-value 0.490) comfortably fail to reject the hypothesis that the grades are drawn from the estimated mixture. The fit of a single normal is much worse (p-value 0.082). Estimation of a mixture of 3 normals is numerically difficult and at convergence cannot be rejected with less confidence (p-value 0.382).

5 The model and reality

Some assessments do conform to the model's assumptions, and generate data that conform to its implications. Consider for example an exam that the author and colleagues, aiming on behalf of society to induce students to learn Economics, administer to very heterogeneous students. All take the same exam, which cannot be tailored to any of their observable characteristics, and can retake it they fail, so the stakes at each attempt are low. Some aim at good grades but, because Economics is one of the most difficult exams from their perspective, most just aim to pass, which is what happens if they score at least 12 points on $N = 22$ questions with 4 possible answers that award a point if correct, and subtract a third of a point if wrong. The score ranges from 22 down to zero (the expected value of random answers) and below if there are many incorrect choices.

The model above approximates this situation supposing that each question is answered correctly with a probability that students can increase at a cost and the examiner estimates as the realized proportion of correct answers. Students pass when that proportion is at least $12/22 = 0.545$, and its distribution is approximately normal when the probability of correct answers is not very different from 50%. In these data the overall pass frequency is about 47% and the distribution of the empirical scores in Figure 11 is very closely approximated, as in Figure 10, by a mixture of two normal random variables: one, with weight 31.5%, is centered at 6.83, the other is centered at 13.22, and their standard deviations are very similar at 3.22 and 3.32. About a third of the students hope to pass, performing some 5 points below the pass threshold on average, and about two thirds fear failure, performing a little more than a point above the pass threshold on average.

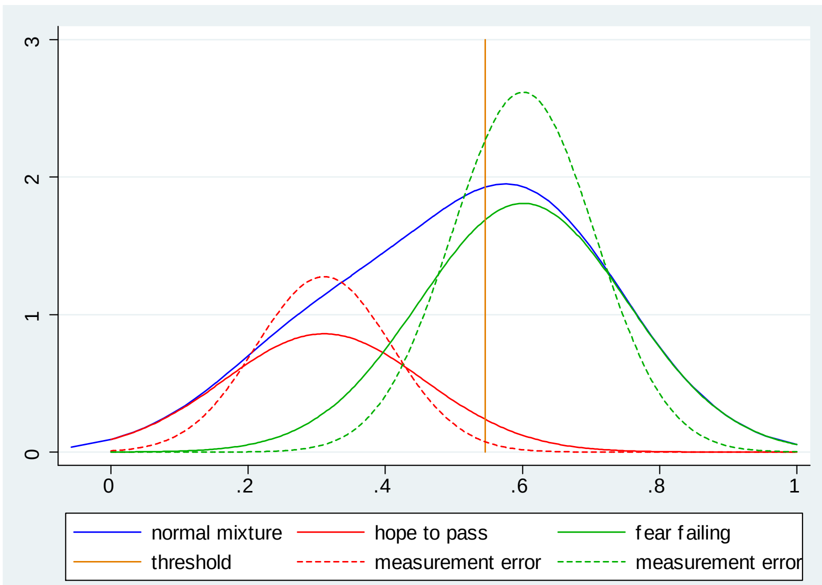

Figure 12: Solid lines plot the estimated normal mixture from Figure 11 scaled to the 0,1 interval. Dashed lines plot normal approximations of measurement errors with 22 questions.

The formal model approximates this exam's structure and the resulting student performances very well even though it assumes that the distribution's functional form remains constant as precision varies, performance influences its mean but not its variance, and all questions are equally likely to be answered correctly at given preparation.[6] The good fit makes it possible to gauge the role of test precision and other sources of uncertainty. Figure 12 plots with continuous lines the fitted mixture of the proportion of correct answers (normal densities centered at 0.310 and 0.601 around the 0.545 pass threshold, with standard deviations about 0.15 and proportions 0.31 and 0.69) and with dashes the distribution of normal measurement errors for students who perform at the centers of the two normals. By (23) their standard deviations $\sqrt{.310(1-.310)/22} = 0.099$ and $\sqrt{.601(1-.601)/22} = 0.104$ are very similar to each other and to the $\sqrt{0.545\,45(1-0.545\,45)/22} = 0.106$ standard deviation of measurement errors when per-

[6]It also neglects penalties for wrong answers, which imply a somewhat different distribution and induce risk-taking behavior by students who hope to pass.

formance is at the threshold. Within each of the mixed distributions the measurement error that the model attributes to the limited precision of an exam with a finite number of questions accounts for only about $0.1^2/0.15^2 = 44\%$ of the empirical variance. The rest of the observed score dispersion is due to heterogeneity of costs and/or rewards among students and to their uncertain perception of their own competence relative to the threshold, both of which the examiners must take as given.

More numerous questions would not appreciably increase the exam's precision: devising, asking, and grading 33 instead of 22 questions (at considerable additional cost for both the examiner and the students) would reduce the standard deviation of errors from about 0.10 only to about 0.08. The pass frequency is close to the 50% implied by a performance-maximizing assessor's choice of threshold under limited precision, so these data do not deny that the examiner aims at maximizing performance but do indicate that, as in many other assessments, precision is quite low.

5.1 Discussion

Teachers and examiners are certainly aware that exam results depend on the students' talent and preparation as well as on random factors, which are more or less relevant in differently precise exams administered to more or less heterogeneous and poorly-informed student populations. Editors know that referees' assessments randomly depend on their mood and competence, and typically mention that acceptance would be "too unlikely" when issuing a desk rejection. The analytical characterization in Section 2 of the assessee's choice of performance can help them see how they should choose the precision and threshold of their exams and editorial processes if, like the assessor considered in Section 3, they prefer better performance.

The explicit functional forms of Section 4 flesh out this general insight with sharp characterization results that are approximately valid for nearly normal distributions and nearly linear marginal cost functions, highlight more general features of assessments where chosen performance may be above or below the pass threshold, and appear realistic for some, such as such as that inspected in this section. They are however restrictive enough to make falsifiable predictions not only about the distribution of performance among passing and failing assessees, which for editorial and admission processes may not be easy to observe across admitted students or accepted papers and their rejected counterparts, but also about pass frequencies, which can be very high at high precision but cannot be less than 50% as long as the assessor can choose the threshold, maximizes performance, and the distribution of rational errors is normal or more generally has symmetric density around its zero mean.

This begs the question of what may rationalize the much lower acceptance rates at most journals and admission rates at selective schools. It may be the case that the pass probability does not depend on the preparation effort of assessees, who take the test to obtain an assessment of their own exogenous ability (as in Bertola, 2023) and are sufficiently uncertain about it to fail with very high probability. If performance is more plausibly at least to some extent endogenous, the general framework offers insights into how different assumptions may explain very high failure rates.

As long as assessors do prefer better performance, their choice of threshold can imply very low pass rates at the mode of the error density if that density is not symmetric. When precision and skewness are both influenced by the test's structure derivations are considerably more complicated than under normality, but further research could study situations where only a few exam questions or few referees are drawn from a set with a highly skewed distribution of difficulty or friendliness, or where the assessees' uncertainty about their own exogenous ability implies that it is under- or over-estimated with very different probabilities.

As long as the error density is not strongly asymmetric, rejection rates much higher than 50% must result from a combination of demanding pass thresholds and highly imprecise assessment. To prevent optimality of hope-to-pass behavior more demanding thresholds should be associated with higher precision, yet need not be if for assessors it is costly not only to increase precision but also to decrease the threshold, even when doing so would improve performance. This can happen when high rejection rates are viewed as signals of quality for journals and universities and, because of capacity constraints, among very numerous assessees most must be rejected (Bertola, 2024).

Appendix

Derivation of (6) and (7)

Because $0 < P(x;\zeta,\vartheta) < 1$ the objective function $V(x;\zeta,\vartheta) = P(x;\zeta,\vartheta)\upsilon - c(x)$ is bounded above for any finite υ. Assumptions 1 imply that $V(x;\zeta,\vartheta)$ is increasing at $x = \alpha$, and its maximum value is interior. To see this, consider

$$V(\alpha+\Delta;\zeta,\vartheta) - V(\alpha;\zeta,\vartheta) = \int_{\alpha}^{\alpha+\Delta} \left(F'\left(\left(\zeta - z\right)\vartheta\right)\vartheta\upsilon - c'(z)\right)dz$$

and note that $V(\alpha+\Delta;\zeta,\vartheta) > V(\alpha;\zeta,\vartheta)$ for some $\Delta > 0$ because the integrand is positive when $F'\left(\left(\zeta - z\right)\vartheta\right)\vartheta\upsilon$ is positive at and above α and $c'(z)$ growing at finite slope from zero at α. Writing $V(\alpha;\zeta,\vartheta) = \left(1 - F\left(\left(\zeta-\alpha\right)\vartheta\right)\right)\upsilon - c(\alpha)$ and noting that $F\left(\left(\zeta-\alpha\right)\vartheta\right) < 1$ and $c(\alpha) = 0$ establishes that $V(\alpha;\zeta,\vartheta) > 0$ for at least some $\Delta > 0$.

Differentiability of $F\left(\left(\zeta - z\right)\vartheta\right)$ and $c(x)$ in Assumptions 1 implies that the objective function is twice continuously differentiable, hence its interior maximum satisfies the first- and second-order conditions

$$\frac{\partial}{\partial x}P(x;\zeta,\vartheta)\upsilon - c'(x) \ = 0,\ \frac{\partial^2}{\partial x^2}P(x;\zeta,\vartheta)\upsilon - c''(x) < 0$$

which with $\frac{\partial}{\partial x}P(x;\zeta,\vartheta) = \vartheta F'\left(\left(\zeta - x\right)\vartheta\right)$, $\frac{\partial^2}{\partial x^2}P(x;\zeta,\vartheta) = -\vartheta^2 F''\left(\left(\zeta - x\right)\vartheta\right)$, and $\upsilon > 0$ are the same as (6) and (7).

Derivation of (8) and (9)

Total differentiation of (6) with respect to ζ yields

$$\frac{dx^*(\vartheta,\zeta)}{d\zeta} = \frac{\vartheta^2 F''\left(\left(\zeta - x^*\right)\vartheta\right)}{\vartheta^2 F''\left(\left(\zeta - x^*\right)\vartheta\right) + c''(x^*)\gamma/\upsilon}.$$

The denominator is positive by (7), and (8) follows.

Total differentiation of (6) with respect to ϑ yields

$$\frac{dx^*(\vartheta,\zeta)}{d\vartheta} = \frac{F'\left(\left(\zeta - x^*\right)\vartheta\right) + \left(\zeta - x^*\right)\vartheta F''\left(\left(\zeta - x^*\right)\vartheta\right)}{\vartheta^2 F''\left(\left(\zeta - x^*\right)\vartheta\right) + c''(x^*)\gamma/\upsilon},$$

which by (7) is positive if

$$F'\left(\left(\zeta - x^*\right)\vartheta\right) + \left(\zeta - x^*\right)\vartheta F''\left(\left(\zeta - x^*\right)\vartheta\right) > 0$$

or, dividing by $F'\left(\left(\zeta - x^*\right)\vartheta\right) > 0$ by Assumptions 1,

$$1 + \left(\zeta - x^*\right)\vartheta\frac{F''\left(\left(\zeta - x^*\right)\vartheta\right)}{F'\left(\left(\zeta - x^*\right)\vartheta\right)} > 0, \tag{A.1}$$

which is the same as (9).

Derivation of (10) and (11)

As long as x_H and x_F respond continuously to ζ and ϑ their variation has no first-order effect on $V(x_H;\zeta,\vartheta)$ and $V(x_F;\zeta,\vartheta)$ by condition (6). Whether $V(x_H;\zeta,\vartheta) \leq V(x_F;\zeta,\vartheta)$ or, by (2),

$$P(x_H;\zeta,\vartheta)\upsilon - c(x_H) \leq P(x_F;\zeta,\vartheta)\upsilon - c(x_F) \tag{A.2}$$

therefore depends on ζ and ϑ only through the pass probability $P(x;\zeta,\vartheta) = 1 - F\left((\zeta - x)\vartheta\right)$.

Differentiating with respect to ζ,

$$\begin{aligned} \frac{\partial}{\partial\zeta}V(x_H;\zeta,\vartheta) &= \frac{\partial}{\partial\zeta}\left[-F\left((\zeta - x_H)\vartheta\right)\upsilon\right] = -\upsilon F'\left((\zeta - x_H)\vartheta\right)\vartheta, \\ \frac{\partial}{\partial\zeta}V(x_F;\zeta,\vartheta) &= \frac{\partial}{\partial\zeta}\left[-F\left((\zeta - x_F)\vartheta\right)\upsilon\right] = -\upsilon F'\left((\zeta - x_F)\vartheta\right)\vartheta \end{aligned}$$

are both negative by Assumption 1. By the first-order condition (6),

$$-\upsilon F'\left((\zeta - x_H)\vartheta\right)\vartheta = -c'(x_H), \quad -\upsilon F'\left((\zeta - x_F)\vartheta\right)\vartheta = -c'(x_F),$$

and (10) follows because $c'(x_H) < c'(x_F)$ for $x_H < x_F$ and $c''(x) > 0$ by Assumption (1).

Differentiating with respect to ϑ

$$\begin{aligned} \frac{\partial}{\partial\vartheta}V(x_H;\zeta,\vartheta) &= \frac{\partial}{\partial\vartheta}\left[-F\left((\zeta - x_H)\vartheta\right)\upsilon\right] = -\upsilon F'\left((\zeta - x_H)\vartheta\right)(\zeta - x_H) < 0, \\ \frac{\partial}{\partial\vartheta}V(x_F;\zeta,\vartheta) &= \frac{\partial}{\partial\vartheta}\left[-F\left((\zeta - x_F)\vartheta\right)\upsilon\right] = -\upsilon F'\left((\zeta - x_F)\vartheta\right)(\zeta - x_F) > 0 : \end{aligned}$$

the inequalities follow from $x_H < \zeta < x_F$ by Assumption 4, and establish (11).

Derivation of (22)

For the normal distribution $F'(z) = \phi(z) = e^{-z^2/2}/\sqrt{2\pi}$ and $F''(z) = \phi'(z) = -ze^{-\frac{1}{2}z^2}/\sqrt{2\pi}$. In (A.1) the elasticity simplifies to

$$(\zeta - x^*)\vartheta \frac{-(\zeta - x^*)\vartheta e^{-\frac{1}{2}((\zeta - x^*)\vartheta)^2}/\sqrt{2\pi}}{e^{-\frac{1}{2}((\zeta - x^*)\vartheta)^2}/\sqrt{2\pi}} = -\left((\zeta - x^*)\vartheta\right)^2,$$

and the inequality is satisfied if $-\left((\zeta - x^*)\vartheta\right)^2 > -1$ or, equivalently, the first implication in (22). In the region where precision increases performance the pass probability $P(x;\zeta,\vartheta) = 1 - \Phi\left((\zeta - x)\vartheta\right) = \Phi((x - \zeta)\vartheta)$ is larger than $\Phi(-1)$ at $x = \zeta - 1/\vartheta$, and smaller than $\Phi(-1)$ at $x = \zeta + 1/\vartheta$.

Derivation of (19) and (20)

The assessee's first-order condition always holds under Assumptions 1, and by (17) and (18) reads

$$\vartheta\phi\left((\zeta - x)\vartheta\right) = (x - \alpha)\gamma/\upsilon. \tag{A.3}$$

Rearranging,

$$x^* = \alpha + \vartheta\phi\left((\zeta - x^*)\vartheta\right)\upsilon/\gamma \tag{A.4}$$

is larger than α $\forall\zeta \in (-\infty, \infty)$, $\forall\vartheta \in (0, \infty)$ because $\vartheta\phi(z\vartheta) > 0$ for all z and $\vartheta > 0$. In the limit,

$$\lim x^* = \alpha + \lim \vartheta\phi\left(\left(\zeta - x^*\right)\vartheta\right)\upsilon/\gamma$$

implies (19) because $\vartheta \lim_{|\zeta|\to\infty} \phi\left(\left(\zeta - x^*\right)\vartheta\right) = 0$ for $\vartheta > 0$ and $\lim_{|\zeta|\to\infty}\left(\zeta - x^*\right) < \infty$, as x^* would otherwise tend to infinity with ζ and contradict (A.4), and $\lim_{\vartheta\to 0} \vartheta\phi\left(\left(\zeta - x^*\right)\vartheta\right) = 0$ because $\phi(z)$ is finite for all z and all $\vartheta < \infty$.

To evaluate $\lim_{\vartheta\to\infty} \vartheta\phi\left(\left(\zeta - x^*\right)\vartheta\right)$ it is necessary to consider the choice criterion (A.2). As precision tends to infinity the $x_F > \zeta$ solution of (A.3) tends to ζ from above and in the limit sets the pass probability to unity. The $x_H < \zeta$ solution tends to α because

$$\lim_{\vartheta\to\infty} \vartheta\phi\left(\left(\zeta - x\right)\vartheta\right) \propto \lim_{\vartheta\to\infty} \vartheta \exp\left(-\left(\zeta - x\right)^2\vartheta^2/2\right) = 0 \text{ for } x \neq \zeta,$$

and must be the optimal performance x^* if the value of the test is not positive at x_F in the limit. The condition

$$\lim_{\vartheta\to\infty}\left(1 - \Phi\left(\left(\zeta - x_F\right)\vartheta\right) - (x_F - \alpha)^2\gamma/\left(2\upsilon\right)\right) \leq 0$$

with $1 - \Phi\left(\lim_{\vartheta\to\infty}\left(\zeta - x_F\right)\vartheta\right) = 1$, $\lim_{\vartheta\to\infty} x_F = \zeta$ requires $(\zeta - \alpha)^2\gamma/\left(2\upsilon\right) \geq 1$, or $(\zeta - \alpha)\sqrt{\gamma/\upsilon} \geq \sqrt{2} \approx 1.41$, and yields the expression (20) for $\bar{\zeta}$, the threshold that rules out fear-failing optimal behavior at near-infinite precision.

Derivation of the expression in footnote 3

The globally optimal performance x^* can coincide with he threshold ζ only if $x = \zeta$ locally maximizes the assessee's objective function, which by (A.3) requires $\gamma\zeta = \vartheta\phi(0)\upsilon$, and yields a strictly positive expected value, which by (A.2) with $\Phi(\zeta - \zeta) = 0.5$ requires $0.5\upsilon - (\zeta - \alpha)^2\gamma/2 > 0$. Combining these requirements yields a quadratic inequality in ϑ

$$\frac{1}{2} > \frac{1}{2}(\vartheta\phi(0)\upsilon/\gamma - \alpha)^2\gamma/\upsilon,$$

for which the positive ϑ root is the expression in footnote 3. If the inequality is not satisfied then x^* cannot equal ζ, and jumps discontinuously around it.

Derivation of the expression in footnote 4

Increasing precision ϑ can let x^* reach $\zeta - 1/\vartheta$ and begin to decrease only if $x = \zeta - 1/\vartheta$ for some ϑ solves (A.3), which with $(\zeta - x)\vartheta = 1$ and $\vartheta = 1/\left(\zeta - x\right)$ reads

$$\phi(1)/\left(\zeta - x\right) = (x - \alpha)\gamma/\upsilon.$$

This quadratic equation has no real solution, hence the assessee's optimal performance increases monotonically on $(0, \zeta + 1/\vartheta]$, if

$$(\zeta - \alpha)^2 - 4\phi(1)\upsilon/\gamma < 0,$$

which can be solved for $\hat{\zeta}$ as a function of the other parameters.

Derivation of the expression in footnote 5

If variation of precision brings x^* to reach $\zeta + 1/\vartheta$ and begin to decrease in precision, $x = \zeta + 1/\vartheta$ must for some ϑ solve condition (A.3), which with $\phi((\zeta - x)\,\vartheta) = \phi(-1) = \phi(1)$ and $\vartheta = 1/\,(x - \zeta)$ reads

$$\frac{1}{x-\zeta}\phi(1) = (x-\alpha)\frac{\gamma}{\upsilon}, \tag{A.5}$$

and be the assessee's optimal performance, which requires its expected value to be strictly positive: with $(\zeta - x)\vartheta = -1$ and $(1 - \Phi(-1)) = \Phi(1)$,

$$\Phi(1)\upsilon - \gamma\,(x-\alpha)^2\,/2 > 0. \tag{A.6}$$

Inserting in (A.6) the solution of quadratic equation (A.5) that is larger than α,

$$x = \alpha + \frac{1}{2}\left(\zeta - \alpha + \sqrt{(\zeta-\alpha)^2 + 4\phi(1)\upsilon/\gamma}\right),$$

yields a quadratic inequality in $(\zeta - \alpha)\sqrt{\gamma/\upsilon}$,

$$(\zeta-\alpha)\sqrt{\gamma/\upsilon} + \sqrt{\left((\zeta-\alpha)\sqrt{\gamma/\upsilon}\right)^2 + 4\phi(1)} < 2\sqrt{2\Phi(1)},$$

solved by

$$\left(\tilde{\zeta} - \alpha\right)\sqrt{\upsilon/\gamma} < \frac{\Phi(1) - \phi(1)/2}{\sqrt{\Phi(1)/2}}$$

which can be rearranged to the expression for $\tilde{\zeta}$ in the footnote.

References

Ábrahám, Árpad, Sebastian Koehne, and Nicola Pavoni (2011) “On the First-Order Approach in Principal-Agent Models with Hidden Borrowing and Lending” Journal of Economic Theory 146, 1331–1361.

Becker, William (1979) "Professorial Behavior Given A Stochastic Reward Structure" American Economic Review 69, 1010-17

Bertola, Giuseppe (2021) "Exam Precision and Learning Effort" Economic Letters 207, 110020

Bertola, Giuseppe (2023) “University Dropout Problems and Solutions” Journal of Economics 138:3, 221-248

Bertola, Giuseppe (2024) “Equilibrium Failure Rates in Tests of Endogenous Performance” mimeo

Bertola, Giuseppe and Winfried Koeniger (2015) “Hidden Insurance in a Moral-Hazard Economy” RAND Journal of Economics 46:4, 777–790

Colussi, Tommaso (2018) "Social Ties in Academia: A Friend Is a Treasure" The Review of Economics and Statistics 100, 45–50.

Eeckhout, Jan, Nicola Persico, and Petra E. Todd (2010) "A Theory of Optimal Random Crackdowns" American Economic Review 100, 1104-1135

Hargens, Lowell L. (1988) "Scholarly Consensus and Journal Rejection Rates" American Sociological Review 53:1, 139-151

Ke, Rongzhu and Christopher Thomas Ryan (2018) "A general solution method for moral hazard problems" Theoretical Economics 13, 1425–1481

Kono, Hisaki and Nobuyuki Yagi (2008) “Heterogeneous Contests and Less Informative Signals” Japanese Economic Review 59:1, 113– 126

Kuehn, Zoë, and Pedro Landeras (2014) "The Effect of Family Background on Student Effort" The B.E. Journal of Economic Analysis and Policy 14:4, 1337-403

Lazear, Edward P., and Sherwin Rosen (1981) "Rank-Order Tournaments as Optimum Labor Contracts" Journal of Political Economy 89:5, 841-864

Mirrlees, James (1999) "The Theory of Moral Hazard and Unobservable Behaviour: Part I" The Review of Economic Studies 66:1, 3–21

O'Keeffe, Mary, W. Kip Viscusi and Richard J. Zeckhauser (1984) "Economic Contests: Comparative Reward Schemes" Journal of Labor Economics 2:1 27-56

Rogerson, William P. (1985) "The First-Order Approach to Principal-Agent Problems" Econometric 53:6, 1357–1367

Stephan, Paula E. (1996) "The Economics of Science" Journal of Economic Literature 34:3, 1199-1235