Investing in Performance: Information and Merit-Based Incentives in K-12 Education

Vanitha Virudachalam, Sergei Savin
The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, savin@wharton.upenn.edu
vanitha@wharton.upenn.edu

Matthew P. Steinberg
Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104, steima@upenn.edu

The United States educational policy requires that K-12 students participate in annual standardized tests. As a result, school districts that have traditionally utilized ongoing "formative" assessments of student progress, are increasingly relying on additional, costly "interim" assessments. In addition, some districts are experimenting with merit-based incentives that tie teachers' bonuses to student performance on state tests. We examine the relationship between information on student performance and monetary incentives for teachers using a two-period principal-agent model. In our model, the school district (principal) chooses whether to invest in interim assessments, and, also, how much merit-based compensation to offer to teachers, while the teachers (agents) decide on the level of effort to exert in each period. We use two-state ("proficient" vs. "not proficient") Markovian dynamics to describe the evolution of student readiness for the tests, and assume the presence of information asymmetry between the teachers and the school district regarding the student readiness level.

Our analysis shows that, for schools that are not proficient at the beginning of the year, the return from merit-based incentives is always greater than the return from information derived from interim assessments. For schools that begin the year on track to achieve proficiency, there exist settings where investing in the interim assessment is optimal, such as when the district has a low budget and the formative assessment is reasonably accurate. However, we also establish that there are settings where the provision of additional information about the student mid-year performance has a demotivating effect on teachers.

1. Introduction

Performance-based contracts have long been identified as a way to incentivize workers when direct oversight is not possible. Yet, the effectiveness of such contracts depends on far more than simply the level of incentives offered – workers must also have access to the resources and information necessary to do their jobs, which often require additional monetary expenditures by the company.

In this paper, we study a system where employers often combine monetary incentives and investment in additional information: K-12 education in the United States. Student assessments as a source of information on the quality of ongoing educational process have been an integral part of the U.S. education system for decades (Linn 2000), but the form and extent of testing have varied considerably over time and across the individual states. With the passage of the No Child Left Behind Act of 2001 (NCLB), a greater emphasis was placed on frequent testing based on a well-defined set of standards. In particular, states were required to ensure student "proficiency" on state tests in reading and math by 2013-14, and the individual schools were required to show "adequate yearly progress" (AYP) for all student groups as well as key subgroups, such as minority students and students in special education (United States General Accounting Office 2003, Klein 2015, Dillon and Rotherham 2007). In 2009, the $4.35 billion Race to the Top Fund (RTTT) was launched with a focus on four core goals: updating educational standards; building data systems to measure student performance and inform educators; "recruit[ing], develop[ing], reward[ing], and retain[ing]" effective teachers and principals; and turning around low-performing schools (U.S. Department of Education 2009). Partly due to RTTT, school districts have been experimenting with pay-for-performance contracts for teachers designed to improve educational outcomes.

Perie et al. (2007) describe three types of assessments frequently used in the U.S. K-12 system: formative, summative, and interim. *Formative* assessments are used by educators to obtain ongoing information about student understanding of the material being taught. *Summative* assessments are given at the end of an instructional period, often a year or a semester, to check student knowledge against a broad set of content standards. These standards, set by an external entity rather than by an individual teacher, specify the body of knowledge and/or a set of skills a student should acquire at each grade level. Finally, *interim* assessments lie between these two types of assessments. Interim assessments are used to evaluate students against a specific set of achievement goals, and the results from these assessments guide teaching within the classroom and inform decisions more broadly at the school and district level. Interim assessments have become increasingly popular as districts seek to evaluate student performance across different schools on the standards expected to be met by the year's end. These assessments are prepared by external evaluation bodies and are considered to be much more reliable but also much more costly indicators of student performance than formative assessments. Thus, for a school district, investing in a mid-year interim assessment provides an accurate measure of student progress but can also divert funds from performance-based incentives for teachers. Although in practice school districts invest in both, currently much more money is allocated towards ongoing assessments than towards additional teacher compensation. For example, the Teacher Incentive Fund, established by Congress in 2006 to provide grants to support performance-based teacher and principal compensation in high-needs schools, allocated $225 million for such awards in 2016. At the same time, total U.S. spending on classroom assessments in 2013-14 was $1.3 billion per year, up from $434 million in 2001-02 (Cavanagh 2015). Yet, despite the vast sums of money being spent on these two approaches to improving educational outcomes, to the best of our knowledge, there are no studies of their relative effectiveness.

In this paper, we seek to address this gap in the extant literature. We analyze the problem faced by a school district that wants to maximize the probability of students being "proficient" on the end-of-the-year standardized test by allocating a fixed budget between an "interim" assessment and a merit-based incentive for teachers. In our model, the district plays the role of a principal that, on the one hand, can provide an agent (teachers) with financial incentives tied to the achieved educational outcome, and, on the other hand, can, at a cost, improve the quality of the information set under which the agent operates. Since the educational process unfolds over a protracted period of time (e.g., a year) and the additional information on the state of student proficiency is provided by the interim assessment in the middle of this process, our model uses a two-period dynamic principal-agent framework. If the district invests in an interim assessment midway through the school year, both the district and the teachers will know the state of student proficiency at that time; otherwise, they will receive less accurate information from a low-cost formative assessment. To reflect an information asymmetry between the school district and the teachers likely to be present in many educational settings, we assume that, while the district's knowledge of the mid-year state of student preparedness relies entirely on the assessment results, the teachers, through their daily interactions with students, may have an informational advantage over the district and may be able to gauge the exact state of student preparedness even under the "imperfect" formative assessment. To capture settings with varying degrees of this information asymmetry, we assume that the district can estimate the likelihood that teachers have perfect information on the mid-year state of student proficiency. In addition to the information provided by the interim assessment, the district may offer teachers a merit-based bonus, which they earn if a sufficient portion of their students show "proficiency" on the year-end standardized test. Teachers respond to both the merit-based incentive and the information they possess about mid-year student progress by choosing a dynamic policy that defines their effort levels. In our model, we use a scalar as a simplified representation of the multiple levers a teacher can use to influence educational outcomes, such as spending extra time working with students or creating detailed lesson plans.

Our analysis pursues two main goals. First, we aim to build a model that can assist school districts in resolving the trade-offs they face in allocating limited funds among popular mechanisms for achieving high student performance. While in practice each district knows the "price tag" associated with additional testing, there exist no clear guidance on estimating the benefit of a mid-year test in achieving the year-end "proficiency" targets. Thus, our main focus is on assessing the value of additional information brought in by the interim assessments for a particular school district using a small number of parameters that can be estimated in practical settings. Our second goal is to understand how the value of additional information is shaped by the effectiveness of teachers' efforts in improving and maintaining student performance, by the quality of information delivered by "noisy" but free formative assessments, and by the degree of information asymmetry between the teachers and the school district.

We quantify the degree of disadvantage suffered by schools with low-performing students. In particular, we show that for schools that begin the year already "behind" in terms of student performance, a school district is always better off investing their entire budget into the merit-based incentives. For schools that begin the year on track to achieve proficiency, we show that extra information on mid-year student performance may produce diametrically opposed results, depending on the accuracy of the formative assessment and the available budget. Counterintuitively, when teachers rely on a low-accuracy formative assessment, the provision of more accurate information on mid-year student performance via an interim assessment, irrespective of the cost, is demotivating and has negative consequences on the end-of-year student achievement outcome. In this case, in the absence of accurate mid-year information, teachers are more likely to believe that students are performing well, based on their knowledge of student proficiency at the beginning of the year. This, in turn, makes it easier to incentivize teachers to exert high levels of effort. At the same time, for a more accurate formative assessment, the extra information delivered by the interim assessment can be valuable for realistic budget values. This paradox arises because, as the formative assessment increases in accuracy, teachers place more weight on the assessment results and less on the fact that students began the year in a high-performance state. Then, the interim assessment is optimal when teachers are more likely to believe the true intermediate state is not-proficient under the formative assessment. Interestingly, we find that when teachers are able to perfectly detect high student performance, the district should "revert" to the formative assessment. In this case, the main effect of information is, again, demotivational, since it will reduce the likelihood that teachers exert effort when the school "slips" to a low-performance state in the middle of the year.

The rest of the paper is organized as follows. In Section 2, we discuss the relevant literature. Our model is presented in Section 3, followed by its analysis in Sections 4 and 5. Finally, in Section 6, we discuss our results and future avenues of research.

2. Literature Review

Our analysis draws on the principal-agent model literature in economics and operations management, as well as the literature on performance pay in K-12 education.

The role of information in principal-agent models has been studied extensively. In particular, there is much work that considers the optimal policy when the agent has, or can independently gain, private information about the production environment. For example, Baron and Myerson (1982) derive the optimal regulatory policy for a monopolistic firm with privately-known costs. Lewis and Sappington (1997) determine the optimal contract to incentivize the agent to acquire and reveal information. Crémer et al. (1998) study when it is optimal for the principal to induce the agent to gather additional information at a cost. On the other hand, multi-period dynamic models account only for a rather small, and a more recent, fraction of the vast principal-agent literature. Fudenberg et al. (1990) introduce stochastic elements to a dynamic principal-agent model and identify conditions under which a long-term contract can be implemented as a sequence of short-term contracts. Plambeck and Zenios (2000) provide an analysis of a previously intractable setting relying on assumptions about the “economic structure” of principal-agent interaction, and Fuloria and Zenios (2001) build upon this dynamic model in the context of healthcare contracting. Shumsky and Pinker (2003) study the compensation system a firm should offer a gatekeeper who has private knowledge about the complexity of a customer’s problem and their ability to treat it. Zhang and Zenios (2008) use a dynamic principal-agent model with hidden information, where the state is known to the agent but not to the principal. Chu and Sappington (2009) characterize the optimal contract when a principal and agent begin with symmetric information but the agent will ultimately acquire superior information.

Our approach follows the spirit of these models: in our model, the agent (teachers) solves a dynamic program to determine the optimal effort allocation policy. The dynamic nature of the agent’s response in our model is dictated by the setting we describe: the information brought in by additional costly testing is revealed “in the middle” of a protracted instruction period, potentially altering the agent’s decision-making process and, thus, requiring a “closed-loop” modeling approach. The principal’s ability to invest in the enhancement of the information set used by the agent is a distinguishing, novel feature of our analysis.

More broadly, our work is also related to the supply chain literature on asymmetric cost information. For example, Corbett and De Groote (2000) study a supplier’s optimal quantity discount policy when the buyer’s cost is unknown. Ha (2001) analyzes a supplier-buyer relationship with asymmetric cost information under stochastic, price-sensitive demand. Lutze and Özer (2008) characterize a promised-lead time contract, which includes an optimal promised lead time and corresponding payments, that a supplier should offer a retailer who has private information about shortage costs. Additionally, our research relates to work that characterizes the relationship between worker effort and labor costs, e.g. Tan and Netessine (2014).

In our model, the overall effort level of the agent is affected by the monetary incentive offered by the principal and contingent on achieving a pre-specified performance level. In this regard, our work adds to a rich stream of papers focused on contracting and performance pay, also commonly called “merit pay,” in K-12 education in the United States. On the theoretical side, Murnane and Cohen (1986) use the microeconomics contracting framework to argue that merit-pay contracts may be difficult to implement in education settings. They argue that the very nature of the teaching process makes it difficult for supervisors to articulate why some teachers may receive merit pay but others do not, which can lead to dissatisfaction for teachers that do not receive the reward. Similarly, Johnson (1984) points out potential negative effects of teacher-level merit compensation such as harmful competition among teachers and low morale. The lack of precise guidelines that teachers can follow to earn the reward is another complicating factor for the use of merit-based incentives. The concern about rewarding only some teachers within a school can be ameliorated through the use of school-level, rather than teacher-level, incentives (Clotfelter and Ladd 1996). Still, empirical work suggests that teacher-level incentives remain common, and in the presence of such incentives schools can mitigate the potential negative effects by making merit pay inconspicuous or awarding it to almost everyone (Murnane and Cohen 1986). In our work, we assume that teachers within a school form a homogeneous group that can earn a school-level reward. Additionally, although teaching remains as much an art as a science, providing teachers with timely information about their progress toward achieving performance targets may alleviate teachers' uncertainty about the path to earning a merit-based reward. The growing availability of interim assessments has made it easier for districts to do just that, and, in our model, we explore the new dynamics brought in by these assessments. In more recent work, Barlevy and Neal (2012) propose a pay-for-percentile incentive scheme that overcomes some of the unintended consequences of existing pay-for-performance schemes, such as coaching and scale manipulation. In our work, we take it as given that schools will use a generic incentive scheme and consider how a district should allocate resources to maximize the probability of achieving a goal. We do not specify the precise type of incentive scheme, only the transition probabilities that determine the likelihood that teachers will earn the reward.

The empirical evidence on the effects of merit pay remains mixed. Eberts et al. (2002) find that teacher-level merit pay improved student retention but negatively impacted student attendance and course passing rates, with student GPAs remaining unchanged. Figlio and Kenny (2007) use survey data from 390 schools to show that merit pay (defined as "at least one [teacher] ... reported having a merit pay bonus") is correlated with higher test scores. Springer et al. (2011) implemented a short-term, experimental teacher-level performance pay program in Metro Nashville Public Schools (MNPS) where teachers were eligible for up to $15,000 per year in bonuses based on student test-score gains. Although the program did not have a significant, lasting effect on student test scores, it did impact the way some teachers approached their jobs: while 80 percent of teachers believed that the program did not change their teaching practices, teachers in the "treatment group" were more likely to collaborate with other teachers and align their instruction with test preparation. Fryer (2013) studies an experimental school-level incentive program in over 200 high-needs New York City public schools, which was implemented as a randomized school-based trial from the 2007-08 school year through the 2009-10 school year. The author finds no evidence that financial incentives lead to improvements in student performance outcomes or in teacher or student behavior.

The results from the longer-term merit-pay experiments seem to suggest that the positive effects of such interventions, if any, may not show up immediately. Dee and Wyckoff (2015) analyze IMPACT, the teacher evaluation reform introduced during the 2009-10 school year in the District of Columbia Public Schools. IMPACT offers strong financial incentives for highly effective teachers, where effectiveness is determined based on multiple components, such as classroom observation scores and students' performance on standardized tests. The researchers find that for highly effective teachers, the base pay incentives for scoring "highly effective" for another year were associated with a seven-percentile increase in teacher effectiveness. Unlike many earlier studies that focused specifically on short-term, experimental performance pay programs, IMPACT is a multi-pronged, long-term program. Chiang et al. (2017) evaluate the Teacher Incentive Fund (TIF), a program which was established by the U.S. Congress in 2006 and "which provides grants to support performance-based compensation systems for teachers and principals in high-need schools." Specifically, they evaluate ten districts in which the pay-for-performance component of TIF was randomly assigned. The program was implemented over a four-year period, and, by the second year, it led to a slight increase in student achievement that held steady in the remaining years of the program. In this program, although most educators received a bonus, the actual bonus level was differentiated based on the performance of their students.

These empirical studies focus specifically on the impact of a performance-based incentive and do not consider the influence of mid-year assessments on teaching practices and student performance. In our analysis, we focus on identifying the school districts that, in the presence of merit-based teacher compensation, may benefit from additional information brought in by interim assessments as well as the the school districts that are better off using formative assessments.

3. Model: Combining Assessments and Merit-Based Pay to Achieve Proficiency

In this section, we present a dynamic principal-agent model that captures the interaction between the school district (principal) and the group of teachers at a school (agent). In our model, the district explores the option of investing in additional information on the state of student performance and providing incentives to maintain or achieve standards of performance, and teachers respond to information and incentives by selecting a dynamic policy defining their effort levels.

3.1. Time Horizon, System States and Actions

Consider a discrete-time, two-period model, with time indices $t = 0, 1$ corresponding to the beginning of periods 1 and 2, respectively, and the index $t = 2$ corresponding to the end of period 2 and indicating the time at which the proficiency of the student body at the school is measured via a state-administered standardized assessment. At time $t = 0, 1, 2$, the school proficiency is given by

$\beta_t \in \{P, N\}$, which indicates whether that school is "proficient" (P) or "not proficient" (N). We define "proficient" to mean that a sufficient fraction of the school's students are either on track to satisfy state-imposed learning standards at $t = 0, 1$ or satisfy these standards at $t = 2$. (See Table 1 for the description of our notation.)

In each period, teachers decide how much effort to allocate towards activities they believe will improve student performance. We use $e_t \geq 0$, $t = 0, 1$ to denote the teachers' effort level in period $t+1$. Two features of our approach to modeling teachers' effort are important to underscore. First, in our model, we assume that the school's teachers are homogeneous and act as a group, with e_t reflecting teachers' joint effort. This assumption approximates a more complex reality where the effort levels of individual teachers will vary, with e_t reflecting the "average" school-level effort. We believe such a modeling simplification is justified since, in practice, the proficient/not-proficient designation is often applied to the entire school, as are the performance-based incentives. Modeling the "average" effort level thus allows us to focus on the "first-order" effect of the incentives. Second, we approximate the multidimensional nature of efforts that teachers make in reality by a single "aggregate" measure represented by a scalar. Although no single measure can be a perfect representation of teachers' efforts, this scalar can be a proxy, for example, for the extra time that teachers may spend working with students.

We use vector $\mathbf{e} = (e_0, e_1)$ to represent the effort level decisions for the two periods. The evolution of the student proficiency state in each period is influenced by the state in the beginning of the period and by the teachers' effort decision in that period. Figure 1 illustrates the state transition diagram for the discrete-time Markov chain in each period, where $\alpha(e)$ represents the effort-dependent probability of transitioning from "N" in the beginning to "P" in the end of the time period, and $\delta(e)$ represents the respective probability of transitioning from "P" to "N."

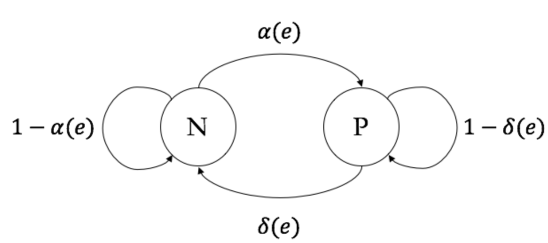


Figure 1 Transition probabilities between proficient ("P") and not proficient ("N") states during each time period as functions of effort level.

In modeling $\alpha(e)$ (the probability of moving to a proficient state) and $1 - \delta(e)$ (the probability of remaining in a proficient state), we use the simplest functional form that reflects the standard assumptions of monotonicity and non-increasing return-on-effort:

Notation	Description
$t = 0, 1, 2$	Time indices corresponding to beginning of period 1, beginning of period 2, and end of period 2, respectively
β_t	State of proficiency at time t, either proficient (P) or not proficient (N)
$\alpha(e)$	Probability of transitioning from N to P as a function of effort e at any period t
$\delta(e)$	Probability of transitioning from P to N as a function of effort e at any period t
A_N	Not-proficient response-to-effort parameter
A_P	Proficient response-to-effort parameter
λ	Effort level at which the marginal impact of effort on probability of transitioning to P is 0
q	Probability teachers know the true intermediate state regardless of assessment choice
z_I	District's choice of assessment, either interim (1) or formative (0)
X_1	Result of formative assessment at $t = 1$, either proficient (P) or not proficient (N)
$\phi_{P\|N}$, $\phi_{P\|P}$	Probability $X_1 = P$ given $\beta_1 = N, P$, respectively
π	Merit-based incentive
γ	Marginal cost of effort at time t
$\mathbf{S}_t$	State of the system at time t
B	School's available budget
F	Cost of interim assessment
$e_t(\mathbf{S}_t, \pi, z_I)$	Effort at time t as a function of state $\mathbf{S}_t$, merit-based incentive π, and assessment decision z_I
$(e_0^*(\mathbf{S}_0, \pi, z_I), e_1^*(\mathbf{S}_1, \pi, z_I))$	Teachers' optimal response policy
$Pr^*[\mathbf{S}_2 \| \mathbf{S}_0]$	District's estimate of the probability school is in proficient state at $t = 2$ under optimal teachers' response policy for fixed π and z_I
$Pr^*_{z_I}[\mathbf{S}_2 \| \mathbf{S}_0]$	Probability school is in proficient state at $t = 2$ under optimal teachers' response policy and optimal merit-based incentive for fixed z_I

Table 1 Description of model's notation.

Assumption 1.

$$\alpha(e) = \begin{cases} A_N\left(\frac{e}{\lambda}\right), & \text{if } 0 \le e \le \lambda, \\ A_N, & \text{if } \lambda \le e, \end{cases} \tag{1}$$

$$1 - \delta(e) = \begin{cases} A_P\left(\frac{e}{\lambda}\right), & \text{if } 0 \le e \le \lambda, \\ A_P, & \text{if } \lambda \le e, \end{cases} \tag{2}$$

where $\lambda > 0$ *and* $0 < A_N \le A_P \le 1$.

Under Assumption 1, probabilities $\alpha(e)$ and $1 - \delta(e)$ are monotonic and concave in e, with λ designating the maximum effort level producing an increase in the probability of being in the proficient state at the end of period. In reality, teachers always exhibit non-zero "base" effort that results in a non-zero probability of reaching proficiency, and our model treats e as "additional" effort that can be elicited through merit-based incentives, resulting in enhanced probability of reaching the proficient state. Note that in (1) and (2) we normalize both the base effort and the base proficiency probability to 0. The maximum probability values A_N and A_P reflect the *co-produced* nature of teaching, where the outcome depends both on the teachers' and students' efforts. Thus, teachers' efforts alone may not guarantee that the proficient state is reached if A_N and A_P are less than 1. The condition $A_N \leq A_P$ implies that it is more difficult to attain proficiency than to maintain it. The literature supports this assumption: Davison et al. (2004) suggest that groups of students often have difficulty overcoming even small achievement gaps, while Neal and Schanzenbach (2010) argue that the incentives for teachers in many school-accountability systems inevitably lead to students at the lowest end of the achievement distribution getting "left behind."

Finally, we assume that the transition dynamic described by (1)-(2) is stationary and does not depend on the time period. This stationarity assumption is reasonable given that the time periods we consider correspond to several months.

3.2. Interim and Formative Assessments: Cost and Information Structure

We assume that the teachers' choice of effort levels cannot be directly observed by the school district. Furthermore, the initial state β_0 is known to both the teachers and the district, and the final state β_2 will be made known to both parties after the final standardized assessment. However, both the teachers and the district may have imperfect knowledge of the intermediate state of the system β_1: students are assessed at the end of period 1 (at $t=1$) to determine whether the school is in the proficient or not-proficient state, but this assessment may be inaccurate. The degree of accuracy depends on the type of assessment used. In particular, the district chooses between two options: to administer an *interim* assessment or to rely exclusively on a *formative* assessment. An *interim* assessment has a fixed cost F that the district must incur and perfectly reveals β_1, the state of the system at $t=1$, whereas a *formative* assessment does not incur any additional cost for the district but is less accurate than the interim assessment. For either choice, both the teachers and the school district will know the results of the assessment at $t=1$. Additionally, while the district relies solely on assessments to gauge student progress, teachers are able to use multiple inputs based on their daily interactions with students. Therefore, we let teachers be one of two types: the type that knows the true intermediate state, β_1, regardless of the assessment choice ($T=1$) or the type that relies entirely on the assessment results ($T=0$). The district assumes that teachers are the first type with probability q.

We use $X_1 \in \{P, N\}$ to denote the result of the formative assessment. That is, $X_1 = P$ ($X_1 = N$) if the formative assessment indicates that the school is in the proficient (not-proficient) state at

$t=1$. The probability that the formative assessment result X_1 takes a particular value given the true intermediate state β_1 is captured by the parameters $\phi_{P|P}$ and $\phi_{P|N}$, where

$$Pr\left[X_1 = P | \beta_1 = P\right] = \phi_{P|P}, \tag{3}$$

$$Pr\left[X_1 = P | \beta_1 = N\right] = \phi_{P|N}. \tag{4}$$

We assume that the formative assessment can never perfectly assess the mid-year state, i.e. it is never the case that $\phi_{P|P} = 1$ and $\phi_{P|N} = 0$. Furthermore, we assume that the probability that the formative assessment returns the proficient result ($X_1 = P$) when the true intermediate state β_1 is proficient cannot be lower than the same probability when the true intermediate state β_1 is not proficient:

Assumption 2. $\phi_{P|P} \geq \phi_{P|N}$.

3.3. District Decisions and the Timeline of Events

For the analysis of the district's decision problem, we introduce the following notation. First, we use the binary variable z_I to indicate the district's choice of assessment type:

$$z_I = \begin{cases} 0, & \text{if district relies exclusively on formative assessment,} \\ 1, & \text{if district chooses interim assessment.} \end{cases} \tag{5}$$

Second, as noted above, we use the binary variable T to indicate whether teachers are the type that know the true state at $t=1$ regardless of assessment choice:

$$T = \begin{cases} 0, & \text{if teacher relies exclusively on assessment result to gauge student progress,} \\ 1, & \text{if teachers know the true intermediate state, } \beta_1\text{, regardless of } z_I, \end{cases} \tag{6}$$

where the district believes that $T=1$ with probability q.

At $t=0$, the district chooses the type of assessment to administer at the end of period 1 and offers teachers a compensation contract that includes a base pay component we normalize to zero and a merit (performance-based) pay component:

$$w = \pi \mathbf{1}_{\beta_2 = P}\left(\pi, z_I\right), \tag{7}$$

where $\pi > 0$ is the level of merit-based incentive and

$$\mathbf{1}_{\beta_2 = P}\left(\pi, z_I\right) = \begin{cases} 1, & \text{if } \beta_2 = P, \\ 0, & \text{if } \beta_2 = N. \end{cases} \tag{8}$$

Teachers know the type of assessment the district chose and the terms of the contract at $t=0$, and they receive compensation at $t=2$. The payment of the reward at a single point in time is consistent with current practice in the field of education. For example, Fryer (2013) describes an incentive scheme in New York City Public Schools in which teachers were given a reward based on annual performance targets. Chiang et al. (2017) study the implementation of the Teacher Incentive Fund (TIF) in ten school districts. They state that in seven out of ten of districts in the study, teachers received their one-time reward during the subsequent school year.

The timeline of events is illustrated in Figure 2. At the beginning of period 1 (at $t=0$), β_0 is known to both the teachers and the district. Based on this information, the district chooses the type of assessment (z_I) and the merit pay component (π). The teachers respond by determining the policy they will use in selecting their effort levels at $t=0$ and $t=1$. Given the initial state of the system and the effort level teachers select at $t=0$ (e_0), the system transitions to state β_1. Then, depending on the school district's choice of assessment, the school either conducts an interim assessment or relies on a formative assessment at the end of period 1. If the district invests in an interim assessment, both the teachers and the district will know the proficiency state β_1 (Figure 2a). If the district relies on a formative assessment, the teachers and the district will be given the result X_1; the district will use this to estimate the probability that the proficiency state β_1 is proficient or not proficient, whereas teachers will do the same only if they do not know the true intermediate state based on their personal knowledge (Figure 2b). Based on the assessment results and the teacher type, teachers select the effort level (e_1) to be applied in period 2 (at $t=1$). At the end of period 2 (at $t=2$), a standardized assessment is administered and the proficiency state β_2 is revealed to both the teachers and the school district. The teachers are then paid according to the compensation contract (7).

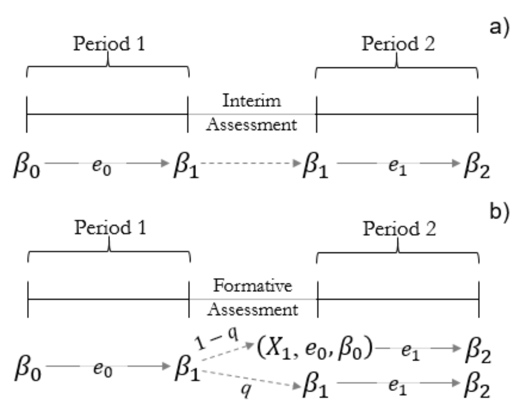


Figure 2 Timeline of events: proficiency states (β_0, β_1, and β_2), the outcome of the formative assessment (X_1), and teachers' actions (e_0 and e_1) when a) the district chooses interim assessment ($z_I=1$) and when b) the district relies exclusively on formative assessment ($z_I=0$).

We assume that both the district and the teachers are risk-neutral. Below we describe the problems faced by the teachers (agent) and the district (principal).

3.4. Teachers' Problem: Dynamic Response to District's Decisions

Given the type of assessment and merit-based contract that the district proposes, teachers choose the effort levels that maximize their expected merit-based compensation net of the cost of effort they incur. In modeling the teachers' cost, we use a simple linear functional form representing stationary and constant marginal cost-of-effort.

Assumption 3. *The teachers' cost-of-effort at time* $t=0,1$ *is given by* $c(e_t)=\gamma e_t$*, with* $\gamma>0$.

Note that the teachers' decision problem is represented by a two-period dynamic program, where their decision in period 2 depends on the information they receive at the end of period 1, and the decision in period 1 is influenced by the policy they adopt for period 2. In order to provide a formal description of the teachers' problem, we define

$$\mathbf{S}_0=\beta_0, \tag{9}$$

$$\mathbf{S}_1=\begin{cases}(X_1,e_0,\mathbf{S}_0), & \text{if } z_I=0 \text{ and } T=0,\\ \beta_1, & \text{if } z_I=1 \text{ or } T=1,\end{cases} \tag{10}$$

and

$$\mathbf{S}_2=\beta_2, \tag{11}$$

to describe the states of the system at $t=0,1$, and 2, respectively. The teachers will use the states at $t=0$ and $t=1$ to make their effort decisions at $t=0$ and $t=1$, respectively. Note that the state of the system at $t=1$ has different "content" depending on both the teacher type and whether the interim or formative assessment is used. In particular, if their information about the proficiency at $t=1$ is imprecise ($z_I=0$ and $T=0$), the teachers must use both the initial state $\mathbf{S}_0$ as well as their action taken at $t=0$ (e_0) to calculate the expected net earnings stemming from their action at $t=1$ (e_1), as we will show below. For each combination of the district's decisions (π,z_I), we can use the notation in (9)-(11) to express the dynamic program that teachers solve as

$$J_t(\mathbf{S}_t)=\max_{e_t\geq 0}\left[E\left[J_{t+1}\left(h_{t+1}\left(e_t,\mathbf{S}_t\right)\right)\right]-\gamma e_t\right],\quad t=0,1, \tag{12}$$

where the expectation is taken over the random state of the system $h_{t+1}(e_t,\mathbf{S}_t)$ at time $t+1$ and

$$J_2(\mathbf{S}_2)=\begin{cases}\pi, & \text{if } \mathbf{S}_2=P,\\ 0, & \text{if } \mathbf{S}_2=N.\end{cases} \tag{13}$$

For convenience, we summarize the description of the state $h_{t+1}(e_t,\mathbf{S}_t)$, for each state-action combination $(e_t,\mathbf{S}_t)$ in Lemma B1 in the Appendix.

To emphasize the connection between the district's decision and the teachers' response, we will use $(e_0^*(\mathbf{S}_0,\pi,z_I), e_1^*(\mathbf{S}_1,\pi,z_I))$ to denote the optimal effort policy, i.e., the policy that solves the dynamic program (12)-(13) for a given set of district decisions (π,z_I).

3.5. District's Problem: Choosing the Optimal Assessment-Incentive Combination

For each school, the district wants to select the assessment type and matching merit pay compensation to incentivize teachers to choose effort levels that will maximize the school's probability of being in the proficient state when the standardized test is administered. The total amount of investment in the information provided by the interim assessment and the incentive payments is limited by budget B. Because each district is likely to manage a number of schools, we assume that it is acceptable for payments to a particular school to exceed the allocated budget, as long as the budget constraint is satisfied in expectation. In order to formulate the district's decision problem, we use, at the slight abuse of notation, $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ to denote, for fixed π and z_I, the probability that the school is in the proficient state at $t = 2$ under the optimal-response teachers' policy $(e_0^*(\mathbf{S}_0, \pi, z_I), e_1^*(\mathbf{S}_1, \pi, z_I))$, given that the school starts in the state $\mathbf{S}_0$ and taking into account the district's uncertainty about teacher type. Then, for given initial performance state $\mathbf{S}_0$, the district's decision can be expressed as the following optimization problem:

$$\max_{\pi \geq 0, z_I \in \{0,1\}} Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] \tag{14}$$

$$\text{s.t. } \pi Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] + F z_I \leq B. \tag{15}$$

In summary, (14)-(15) and (12)-(13) describe a principal-agent problem where a principal selects the combination of information set and incentives for the agent and the agent's response is represented by a dynamic programming policy.

Below we provide an analysis of this principal-agent problem. First, in the next Section, we estimate "base-case" values for some of the parameters of our model using publicly available data. Next, for a given type of assessment, we describe the optimal merit pay selection and the optimal teachers' response policy (Section 4). Then, we analyze the problem of the optimal selection of the assessment type (Section 5).

3.6. Estimating the Base-Case Problem Parameters

To estimate "base-case" parameters for our model, we rely on the data from several sources. First, we use results from the DC Comprehensive Assessment System (DC CAS), the end-of-year standardized tests for District of Columbia Public Schools (DCPS) (District of Columbia Public Schools 2018b). This is a publicly-available dataset that includes eight years of school performance data (2006-07 through 2013-14) for both reading and math. For each school year, the number and percentage of test takers that fall into each proficiency category (below basic, basic, proficient, and advanced) are given by school and grade level for both the reading and math assessments.

We base school-level proficiency targets on the 2006-07 annual measurable objectives for DCPS, as stated in the *Assessment and Accountability Manual* (District of Columbia Office of the State

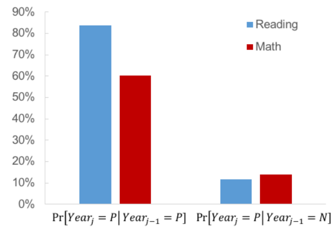


Figure 3 Probability of achieving proficiency (P) at the end of a given year based on the state of proficiency (P or N) in the previous year by subject for District of Columbia Public Schools from 2006-2014.

Superintendent of Education 2011). In particular, for each subject, we average the elementary and secondary targets for the 2006-07 school year and round to the nearest integer. Thus, we define a school as being in the proficient state if at least 45 percent of students are proficient or above in reading and at least 40 percent of students are proficient or above in math. We use these target values for all of the school years in the dataset. Using these values, we calculate that the overall probability that a proficient school remains in the proficient state the following year is 84 percent for the reading assessment and 60 percent for the math assessment, and the probability that a not-proficient school moves to the proficient state is 12 percent for reading and 14 percent for math. We average these and use 72 percent as an estimate of A_P and 13 percent as an estimation of A_N. Although we recognize that A_P and A_N are likely to depend on each school, we rely on the aggregate district-wide data in the absence of a sufficient number of data points for each individual school.

According to Topol et al. (2012), "school districts are spending an average of \$15–\$20 or more per student on interim assessments and data management systems to house their test data." DCPS enrollment was approximately $45,000$ during this period across 115 schools (District of Columbia Public Schools 2018a), which suggests a district-wide cost of at least \$675,000–\$900,000 if the district chooses to implement interim assessments, or approximately \$5,870–\$7,826 per school. Because Topol et al. (2012) suggest that this is a lower bound on interim assessment expenditures, for our base case, we let $F = \$7,500$.

The most difficult values to parameterize in our model are the maximum effort level producing an increase in the probability of transitioning to the proficient state (λ) and the marginal cost of effort (γ). The product $\lambda\gamma$ represents the maximum cost to teachers for exerting additional effort. Recall that IMPACT is one of the few programs where financial incentives produced measurable results; in that program, rewards ranged from \$5,000 to \$25,000 per teacher, although only a subset of teachers were eligible for rewards at the upper end of that range. Furthermore, the average number of full-time teachers at a DC public school is 35 (District of Columbia Public Schools

2018c). Based on these values, we assume that the average maximum cost of effort per teacher is \$10,000; for one school, $\lambda\gamma$ is \$350,000.

Finally, we recognize that the accuracy of formative assessments can vary greatly depending on the teacher and setting. Throughout the paper, we consider a case where the formative assessment has a high likelihood of returning a false positive result, $(\phi_{P|P}, \phi_{P|N}) = (0.9, 0.5)$. We also consider cases where the formative assessment detects proficiency reasonably well ($\phi_{P|P} = 0.9$) or perfectly ($\phi_{P|P} = 1$) and $\phi_{P|N}$ varies, and when the formative assessment detect low-performance reasonably well ($\phi_{P|N} = 0.1$) and $\phi_{P|P}$ varies.

The base case values of the parameters is shown in Table 2.

Parameter	**Value**
A_P	0.72
A_N	0.13
$\lambda\gamma$, in \$000's/school	350
F, in \$000's/school	7.5
Teachers per school	35

Table 2 The base-case parameter values.

In the following analysis we use the base-case parameter values to illustrate the properties of the optimal effort policies for teachers and the optimal merit-based incentives and assessments for the school district.

4. Impact of Assessment Choice

In this section, we begin by considering the case where both the district and teachers rely on a formative assessment (Figure 2b) of the state of student performance at $t = 1$. The district selects a merit-based incentive level π to maximize the probability that the school will be in the proficient state at $t = 2$. The formative assessment result X_1 does not perfectly reflect the state of proficiency at $t = 1$, β_1; the degree of imperfection is characterized by (3)-(4). For any given value of π, we solve the dynamic program that determines the optimal teachers' effort policy. Using this result, we then consider the teachers' effort levels in the case where teachers have perfect information about the intermediate state, either because the district invests in an interim assessment (Figure 2a) or based on the teachers' first-hand knowledge of student progress (Figure 2b). Finally, based on the teachers' response to incentives, we determine the optimal level of merit-based incentive under both formative and interim assessments.

4.1. Optimal Teachers' Effort Policy in the Presence of Merit-Based Incentive

The teachers' problem is a two-period dynamic program (12)-(13). We begin by characterizing, for given π, the teachers' effort decision at the beginning of period 2 (i.e., at $t=1$). The teachers' "profit-to-go" function at $t=1$ is

$$J_1(\mathbf{S}_1) = \max_{e_1 \geq 0} \left[\pi Pr\left[\mathbf{S}_2 = P | \mathbf{S}_1\right] - \gamma e_1\right], \tag{16}$$

where the state of the system at $t=1$ is given in (10), i.e.,

$$\mathbf{S}_1 = \begin{cases} (X_1, e_0, \mathbf{S}_0), & \text{if } z_I = 0 \text{ and } T = 0, \\ \beta_1, & \text{if } z_I = 1 \text{ or } T = 1. \end{cases} \tag{17}$$

Here and below we omit, for simplicity, the designation (π, z_I) when referring to the profit-to-go functions and the optimal effort levels. In particular, we use $e_1^*(\mathbf{S}_1)$ to denote the optimal effort level at $t=1$:

$$e_1^*(\mathbf{S}_1) = \arg\max_{e_1 \geq 0} \left[\pi Pr\left[\mathbf{S}_2 = P | \mathbf{S}_1\right] - \gamma e_1\right]. \tag{18}$$

For the analysis below, we use

$$\hat{\pi} = \frac{\pi}{\lambda\gamma} \quad \text{and} \quad \hat{e}_1^*(\mathbf{S}_1) = \frac{e_1^*(\mathbf{S}_1)}{\lambda} \tag{19}$$

to represent the *scaled merit-based incentive level* and *optimal scaled effort level*. Additionally, we use $\mathcal{P}_1(\mathbf{S}_1)$ to represent the *merit-based incentive threshold* at $t=1$, and

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \frac{\mathcal{P}_1(\mathbf{S}_1)}{\lambda\gamma}, \tag{20}$$

to represent the *scaled merit-based incentive threshold*. Finally, for convenience, we define

$$m(e_0, \mathbf{S}_0) = \begin{cases} 1 - \delta(e_0), & \text{if } \mathbf{S}_0 = P, \\ \alpha(e_0), & \text{if } \mathbf{S}_0 = N. \end{cases} \tag{21}$$

Proposition 1 describes the optimal effort level at $t=1$.

Proposition 1. *a) The optimal scaled effort level at $t=1$ is given by*

$$\hat{e}_1^*(\mathbf{S}_1) = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \hat{\mathcal{P}}_1(\mathbf{S}_1), \\ 1, & \text{if } \hat{\mathcal{P}}_1(\mathbf{S}_1) \leq \hat{\pi}, \end{cases} \tag{22}$$

where

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \frac{1}{Pr\left[\beta_1 = P | \mathbf{S}_1\right](A_P - A_N) + A_N} \tag{23}$$

and

$$Pr\left[\beta_1 = P | \mathbf{S}_1\right] = \begin{cases} \frac{\phi_{P|P} m(e_0,\mathbf{S}_0)}{\phi_{P|P} m(e_0,\mathbf{S}_0) + \phi_{P|N}(1-m(e_0,\mathbf{S}_0))}, & \text{if } \mathbf{S}_1 = (P, e_0, \mathbf{S}_0), \\ \frac{\left(1-\phi_{P|P}\right) m(e_0,\mathbf{S}_0)}{\left(1-\phi_{P|P}\right) m(e_0,\mathbf{S}_0) + \left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0))}, & \text{if } \mathbf{S}_1 = (N, e_0, \mathbf{S}_0). \end{cases} \tag{24}$$

b) For all values of $\hat{\pi}$, $\mathbf{S}_0$, and e_0, $\hat{e}_1^*(P, e_0, \mathbf{S}_0) \geq \hat{e}_1^*(N, e_0, \mathbf{S}_0)$.

When teachers do not know the true intermediate state ($T=0$ and $z_I=0$), they rely on both the result of the formative assessment X_1 and the effort level at $t=0$, e_0 to estimate the probability that the school is in the proficient state and select their effort level at $t=1$. This reflects the reality that teachers often use multiple inputs, e.g., ongoing daily observation, throughout the course of the school year to gauge student progress, weighting those inputs based on their experience. (22) reflects general features of the optimal effort levels that will apply under a broad range of concave response-to-effort functions that display smaller marginal response at the not-proficient state, and a broad range of convex cost-of-effort functions. First, as expected, the optimal effort level is a non-decreasing function of the merit-based incentive level selected by the school district. Second, the optimal effort level when the assessment result is proficient is at least as high as the optimal effort level when the result is not-proficient, holding all else constant. This is driven by the greater difficulty of achieving proficiency faced by teachers at a school in the not-proficient state. Thus, the schools that fall behind may require higher rewards for their teachers in order to reach the proficient state, even in the absence of any other differences between performing and non-performing schools. In practice, this need for extra incentives is further compounded by the fact that more effective teachers tend to be distributed towards more advantaged schools (Clotfelter et al. 2006).

The optimal effort level at $t=1$ under perfect information is the special case when $\phi_{P|P}=1$ and $\phi_{P|N}=0$. We describe this result in Section A1 in the Appendix.

Next, consider the teachers' effort level decision at $t=0$. The teachers' profit-to-go function at that time is

$$J_0(\mathbf{S}_0) = \max_{e_0 \geq 0} \left[E\left[J_1\left(h_1\left(e_0, \mathbf{S}_0 \right) \right) \right] - \gamma e_0 \right], \tag{25}$$

where for all z_I, from (9),

$$\mathbf{S}_0 = \beta_0 \tag{26}$$

describes the state of the system at $t=0$. The probability distribution of $h_1\left(e_0, \mathbf{S}_0\right)$ is given by (B2). We use $e_0^*\left(\mathbf{S}_0\right)$ to denote the effort level that optimizes this function:

$$e_0^*\left(\mathbf{S}_0\right) = \arg\max_{e_0 \geq 0} \left[E\left[J_1\left(h_1\left(e_0, \mathbf{S}_0 \right) \right) \right] - \gamma e_0 \right], \tag{27}$$

where we represent the *optimal scaled effort* by

$$\hat{e}_0^*\left(\mathbf{S}_0\right) = \frac{e_0^*\left(\mathbf{S}_0\right)}{\lambda}. \tag{28}$$

Furthermore, $\mathcal{P}_0\left(\mathbf{S}_0\right)$ is the *merit-based incentive threshold* at $t=0$, and

$$\hat{\mathcal{P}}_0\left(\mathbf{S}_0\right) = \frac{\mathcal{P}_0\left(\mathbf{S}_0\right)}{\lambda\gamma} \tag{29}$$

is the *scaled merit-based incentive threshold*.

For the analysis below, we define the following constants:

$$\mathcal{A}_1 = \frac{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_P\left(1-A_P\right)\left(1-\phi_{P|N}\right)}, \tag{30}$$

$$\mathcal{A}_2 = \frac{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}, \tag{31}$$

$$\mathcal{R}_1 = \frac{A_P\left(1-\phi_{P|P}\right)+\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}, \tag{32}$$

$$\mathcal{R}_2 = \frac{1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}, \tag{33}$$

and the following sets of (A_P, A_N) values:

$$\mathcal{S}_1 = \left\{(A_P, A_N)\,\middle|\, 0 \le A_P - A_N \le \frac{A_N}{A_P}\right\}, \tag{34}$$

$$\mathcal{S}_2 = \left\{(A_P, A_N)\,\middle|\, \frac{A_N}{A_P} < A_P - A_N \le \mathcal{A}_1\right\}, \tag{35}$$

$$\mathcal{S}_3 = \left\{(A_P, A_N)\,|\,\mathcal{A}_1 < A_P - A_N \le \mathcal{A}_2\right\}, \tag{36}$$

$$\mathcal{S}_4 = \left\{(A_P, A_N)\,|\,\mathcal{A}_2 < A_P - A_N\right\}. \tag{37}$$

Our subsequent results depend, on the one hand, upon the relationship between the individual values of the response-to-effort paramaters A_P and A_N, and, on the other hand, the set these values jointly belong to. The different sets of (A_P, A_N) pairs are illustrated in Figure 4. In general, there exist at most four distinct sets, and when teachers have perfect information, either through testing or because of the teacher type, the number of sets decreases to two (Figure 4b). Broadly, when A_P and A_N are close, the pair is in $\mathcal{S}_1$, and, as the distance between the parameters increases, the set to which the pair belongs changes to $\mathcal{S}_2$, then to $\mathcal{S}_3$, and finally, for some values of A_P, to $\mathcal{S}_4$. Observe that the shape of each region depends on the parameters $\phi_{P|P}$ and $\phi_{P|N}$, and $\mathcal{S}_4$ does not always span the entire range of A_P. As the difference between $\phi_{P|P}$ and $\phi_{P|N}$ increases, $\mathcal{S}_4$ spans a greater range of A_P values.

Using these regions, we characterize the optimal scaled effort level $\hat{e}_0^*(\mathbf{S}_0)$ and the scaled profit-to-go function $\hat{J}_0(\mathbf{S}_0)$ at $t=0$.

Proposition 2. *The optimal scaled effort at $t=0$ is*

$$\hat{e}_0^*(\mathbf{S}_0) = \begin{cases} 0, & \textit{if } 0 \le \hat{\pi} < \hat{\mathcal{P}}_0(\mathbf{S}_0), \\ 1, & \textit{if } \hat{\mathcal{P}}_0(\mathbf{S}_0) \le \hat{\pi}, \end{cases} \tag{38}$$

where

$$\hat{\mathcal{P}}_0(\mathbf{S}_0) = \begin{cases} \frac{1}{A_N(A_P-A_N)}, & \textit{if } \mathbf{S}_0 = N, \\ \frac{1}{A_P(A_P-A_N)}, & \textit{if } \mathbf{S}_0 = P \textit{ and } (A_P, A_N) \in \mathcal{S}_1 \cup \mathcal{S}_2, \\ \mathcal{R}_1, & \textit{if } \mathbf{S}_0 = P \textit{ and } (A_P, A_N) \in \mathcal{S}_3, \\ \mathcal{R}_2, & \textit{if } \mathbf{S}_0 = P \textit{ and } (A_P, A_N) \in \mathcal{S}_4. \end{cases} \tag{39}$$

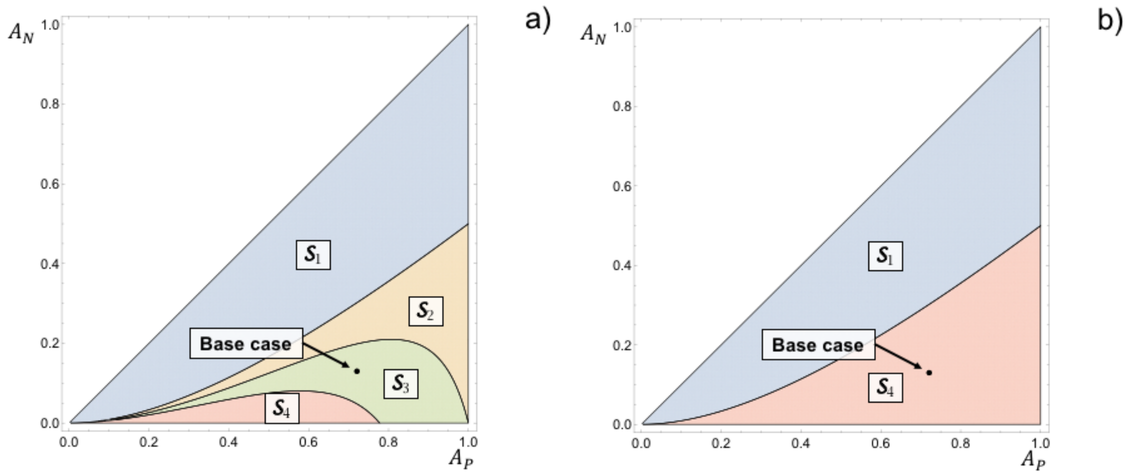


Figure 4 **The shaded areas indicate different regions of interest for pairs of the parameters A_P and A_N. In Region I, $(A_P, A_N) \in \mathcal{S}_1 \cup \mathcal{S}_2$; in Region II, $(A_P, A_N) \in \mathcal{S}_3$; and in Region III, $(A_P, A_N) \in \mathcal{S}_4$, where a) $\phi_{P|P} = 0.9$, $\phi_{P|N} = 0.5$, b) $\phi_{P|P} = 1$, $\phi_{P|N} = 0$.**

As with the effort level at $t = 1$, there exists a minimum positive level of the merit-based incentive necessary to induce a positive effort level. When the school begins in the not-proficient state, the optimal effort level and profit-to-go functions are independent of the accuracy parameters $\phi_{P|P}$ and $\phi_{P|N}$, due to the high level of reward necessary to induce positive effort at $t = 0$. Specifically, when the initial state is not proficient, if the reward is high enough to incentivize positive effort at $t = 0$, then it is always high enough to incentivize positive effort at $t = 1$, regardless of whether the intermediate state is proficient or not. Therefore, the uncertainty from the formative assessment is no longer a factor, and the optimal effort decision is driven entirely by the initial state and effort level at $t = 0$.

For schools that begin the year in the proficient state, the functional form of the optimal effort level may be influenced by the accuracy of the formative assessment. When the response-to-effort parameters A_P and A_N are sufficiently close, i.e., when $(A_P, A_N) \in \mathcal{S}_1 \cup \mathcal{S}_2$, the functional form of teacher effort does not depend on the accuracy parameters, although the accuracy parameters do impact the boundary for $\mathcal{S}_2$. This is to be expected: when the transition probability between the states does not significantly vary with the starting state, the accuracy of formative assessment results is less important. Presumably, this is the case for schools that have a long history of being in the proficient (not-proficient) state, where the probability of transitioning to the proficient state is likely to be high (low) regardless of the starting state in that particular instance. As A_P and A_N diverge, the accuracy parameters play a more important role in determining the functional form of effort and a positive value of the scaled reward is necessary to ensure a positive teacher effort.

In Section A2 in the Appendix, we describe additional properties of the merit-based incentive threshold at $t=0$, including the value of the threshold under perfect information; as when $t=1$, this is the special case when $\phi_{P|P}=1$ and $\phi_{P|N}=0$.

4.2. Optimal Merit-Based Incentive

Using the characterization of the optimal teachers' response policy in the presence of merit-based incentives supported by information about the state of the system at $t=1$, we analyze the district's decision on the optimal incentive level. The district's optimization problem, (14)-(15), is

$$\max_{\pi\geq 0} Pr^*[\mathbf{S}_2=P|\mathbf{S}_0] \tag{40}$$

$$\text{s.t. } \pi Pr^*[\mathbf{S}_2=P|\mathbf{S}_0]+Fz_I\leq B, \tag{41}$$

where we hold z_I fixed and $Pr^*\left[\mathbf{S}_2=P|\mathbf{S}_0\right]$ is the probability that the system will be in the proficient state at $t=2$ under the optimal teachers' response to the merit-based incentive π. We derive a complete characterization of this quantity in Lemma B3 in the Appendix.

We represent the *scaled budget* by $\hat{B}$ and the *scaled fixed cost of the interim assessment* by $\hat{F}$:

$$\hat{B}=\frac{B}{\lambda\gamma}, \quad \hat{F}=\frac{F}{\lambda\gamma}, \tag{42}$$

and define the following constants:

$$\mathcal{B}_1=\left(\frac{A_P\left(1-\phi_{P|P}\right)+\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}\right)\left(A_P^2+A_N\left(1-A_P\right)\right), \tag{43}$$

$$\mathcal{B}_2=\left(1+\frac{A_N\phi_{P|N}}{A_P\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}\right)\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right), \tag{44}$$

$$\mathcal{B}_3=\mathcal{R}_2\left(qA_P^2+(1-q)\left(A_P^2\phi_{P|P}+A_N\left(1-A_P\right)\phi_{P|N}\right)\right), \tag{45}$$

$$\mathcal{B}_4=\left(1+A_P\right)\left(q+(1-q)\left(\frac{A_P^2\phi_{P|P}+A_N\left(1-A_P\right)\phi_{P|N}}{A_P^2}\right)\right). \tag{46}$$

Recall that $\mathcal{R}_1$ and $\mathcal{R}_2$ are given in (32) and (33), and q characterizes the probability that teachers know the true state at $t=1$ regardless of the assessment choice.

It is straightforward to show that the district's estimate of the probability of achieving proficiency in the final state is a non-decreasing function of the merit-based incentive π (see Proposition B1 in the Appendix), a property that facilitates the search for the optimal level of the merit-based incentive. Proposition 3 describes the optimal choice of the scaled merit-based incentive $\hat{\pi}^*$ as a function of the district's scaled budget.

PROPOSITION 3. *If the district relies on the formative assessment ($z_I=0$) and there is a moderate level of information asymmetry ($0<q<1$), the optimal scaled merit-based incentive can be characterized as follows.*

a) For $\mathbf{S}_0 = N$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < 1, \\ \frac{1}{A_N}, & \text{if } 1 \le \hat{B} < 1 + \frac{1}{A_P - A_N}, \\ \frac{1}{A_N(A_P - A_N)}, & \text{if } 1 + \frac{1}{A_P - A_N} \le \hat{B}. \end{cases} \tag{47}$$

b) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_1$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < 1, \\ \frac{1}{A_N}, & \text{if } 1 \le \hat{B} < 1 + \frac{A_N}{A_P(A_P - A_N)}, \\ \frac{1}{A_P(A_P - A_N)}, & \text{if } 1 + \frac{A_N}{A_P(A_P - A_N)} \le \hat{B}. \end{cases} \tag{48}$$

c) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_2$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < (1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right), \\ \frac{1}{A_P(A_P - A_N)}, & \text{if } (1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right) \le \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N\left(1 - A_P^2\right)}{A_P^2}\right), \\ \frac{1+A_P}{A_P^2}, & \text{if } 1 + A_P + (1-q)\left(\frac{A_N\left(1 - A_P^2\right)}{A_P^2}\right) \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ \frac{1}{A_N}, & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}. \end{cases} \tag{49}$$

d) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, *and* $\left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right) \le 1 - \phi_{P|P}$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < (1-q)\mathcal{B}_1, \\ \mathcal{R}_1, & \text{if } (1-q)\mathcal{B}_1 \le \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N\left(1 - A_P^2\right)}{A_P^2}\right), \\ \frac{1+A_P}{A_P^2}, & \text{if } 1 + A_P + (1-q)\left(\frac{A_N\left(1 - A_P^2\right)}{A_P^2}\right) \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ \frac{1}{A_N}, & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}. \end{cases} \tag{50}$$

e) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, *and* $1 - \phi_{P|P} < \left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right)$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < q(1 + A_P), \\ \frac{1+A_P}{A_P^2}, & \text{if } q(1 + A_P) \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1, \\ \mathcal{R}_1, & \text{if } qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ \frac{1}{A_N}, & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}. \end{cases} \tag{51}$$

f) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and* $\left(\frac{A_P(A_P - A_N) - A_N}{A_P}\right)\phi_{P|N} \le 1 - \phi_{P|P}$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < q(1 + A_P), \\ \frac{1+A_P}{A_P^2}, & q(1 + A_P) \le \hat{B} < \mathcal{B}_3, \\ \mathcal{R}_2, & \text{if } \mathcal{B}_3 \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1, \\ \mathcal{R}_1, & \text{if } qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ \frac{1}{A_N}, & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}. \end{cases} \tag{52}$$

g) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and* $1 - \phi_{P|P} < \left(\frac{A_P(A_P - A_N) - A_N}{A_P}\right)\phi_{P|N}$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < (1-q)\,\mathcal{B}_2, \\ \mathcal{R}_2, & \text{if } (1-q)\,\mathcal{B}_2 \le \hat{B} < \mathcal{B}_4, \\ \frac{1+A_P}{A_P^2}, & \text{if } \mathcal{B}_4 \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1, \\ \mathcal{R}_1, & \text{if } qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ \frac{1}{A_N}, & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}. \end{cases} \tag{53}$$

As expected, the optimal reward is non-decreasing in the scaled budget $\hat{B}$. Furthermore, the formative assessment accuracy parameters $\phi_{P|P}$ and $\phi_{P|N}$ play a key role in determining the optimal reward only when the school begins the year in the proficient state, and when A_P and A_N are sufficiently different, i.e. $(A_P, A_N) \in \mathcal{S}_3 \cup \mathcal{S}_4$. As one would expect, the degree of the district's uncertainty about whether teachers have perfect knowledge of the intermediate state, q, matters in the cases where the accuracy of the mid-year assessment affects the teachers' effort decision. This includes the case where $(A_P, A_N) \in \mathcal{S}_2$, since the formative assessment accuracy affects the boundary of that region. We characterize the settings without information asymmetry ($q = 0$) and with extreme information asymmetry ($q = 1$) in Section A3 in the Appendix.

5. Choosing the Best Assessment

Using the analysis of Section 4, we now turn to the district's optimal choice of assessment for a particular school. Recall the district's optimization problem given in (14)-(15):

$$\max_{\pi \ge 0, z_I \in \{0,1\}} Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] \tag{54}$$

$$\text{s.t. } \pi Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] + Fz_I \le B. \tag{55}$$

For a given budget B and cost F, the optimal choice can be either the interim or formative assessment, or both. A complete characterization of the optimal assessment choice for any combination of problem parameters is presented in Proposition B2 in the Appendix.

As one might expect, there are two trivial settings where the choice of assessment does not change the probability that the school is in the proficient state at the end of the year. First, when the budget is sufficiently small ($B < B_L$), the school district can never afford to offer a reward that is high enough to induce teachers to exert positive effort. Therefore, there is always zero probability of achieving proficiency in the final period. Conversely, when the budget is large ($B \ge B_U$) and the cost of the interim assessment is sufficiently small, the school district can offer a reward that is high enough to incentivize maximum effort levels throughout the school year for any value of the state at $t = 1$, $\mathbf{S}_1$. In this second case, the probability of achieving proficiency reaches the maximum possible level regardless of the assessment choice. The values of B_L and B_U depend on the parameters.

Specifically, in Proposition B2, we show that B_L may depend on the accuracy parameters ($\phi_{P|N}$, $\phi_{P|P}$) and response-to-effort parameters (A_N, A_P) as well as the level of information asymmetry (q), whereas B_U depends only on the response-to-effort parameters.

In Proposition 4 below, we describe two settings where the district does not benefit from investing in additional information even for non-trivial budget levels.

Proposition 4. *Consider a setting where teachers have imperfect information about the state at* $t = 1$ $(q < 1)$.

a) When the school begins the year in the not-proficient state, $\mathbf{S}_0 = N$*, the formative assessment is optimal.*

b) Suppose the school begins the year in the proficient state, $\mathbf{S}_0 = P$*, and* $\phi_{P|P} < 1$, $\frac{A_N}{A_P^2} < A_P - A_N$, $q < \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}$. *Then, there exists* $\bar{\phi}_{P|N} < \phi_{P|P}$ *such that for all* $\phi_{P|N} \in \left(\bar{\phi}_{P|N}, \phi_{P|P}\right]$*, the formative assessment is optimal.*

Proposition 4a states that when the school begins the year in the not-proficient state ($\mathbf{S}_0 = N$), there is never any benefit from investing in the interim assessment. In this case, the low probability that the school will achieve proficiency at $t = 2$ means that a high level of reward is necessary to incentivize teachers to exert effort. Furthermore, the reward needed to incentivize teachers at $t = 0$ is always higher than the reward needed to incentivize teachers at $t = 1$, as shown in Corollary A4. Therefore, if the district offers a sufficiently high merit-based incentive, teachers will be incentivized to exert effort throughout the year, regardless of how students progress. On the other hand, if the reward offered is too low to incentivize effort at $t = 0$, teachers know with certainty that students will remain in the not-proficient state at $t = 1$. In either case, teachers' effort decisions are driven entirely by the merit-based incentive and do not change based on additional information. This finding is consistent with the way some urban districts incentivize effort. For example, DC Public Schools' IMPACT offers significantly higher rewards to teachers in high-poverty schools, recognizing the need to offer higher incentives for teachers facing greater challenges.

In Proposition 4b, we consider settings with low information asymmetry where the response-to-effort parameters are sufficiently different and when the initial state is proficient. When the formative assessment cannot perfectly recognize high-performance and the accuracy parameters are sufficiently close, it is always optimal for the district to rely on the formative assessment. In particular, we show that the formative assessment is the only optimal assessment choice for non-trivial budget values. Broadly, this suggests that when the formative assessment is not particularly informative ($\phi_{P|P}$ and $\phi_{P|N}$ are close), teachers will be inclined to rely on their knowledge of the initial state and the level of effort exerted at $t = 0$. Because of this, when the school begins the year in the proficient state, it is cheaper to incentivize effort both at $t = 0$ and at $t = 1$ under the formative assessment than to incentivize effort in any setting under the interim assessment. Then, any subsequent increases in the budget that allow the district to incentivize effort under the interim assessment can never result in a better chance of achieving proficiency in the final period under the interim assessment.

We illustrate the value of information in Figure 5, where value is determined by comparing the probability that the school is in the proficient state at the end of the year under the interim assessment (perfect information) and formative assessment (imperfect information). Specifically, we indicate whether information has a positive, negative, or zero value for regions based on the accuracy parameter $\phi_{P|P}$ and the district's scaled budget $\hat{B}$. Figure 5a shows the case where the interim assessment is free ($\hat{F}=0$) and there is no information asymmetry ($q=0$), and Figure 5b shows the case where information has a cost ($\hat{F}=0.5$) and there is some information asymmetry ($q=0.5$). In both figures, we see that information has zero value for small and large budgets. The boundary on the upper zero-value region corresponds to $\hat{B}_U$, the threshold indicating when the budget becomes trivially large, and the boundary on the lower zero-value region corresponds to $\hat{B}_L$, the threshold indicating the highest budget that is trivially small. Notice that as $\hat{F}$ increases, $\hat{B}_U$ increases while $\hat{B}_L$ decreases.

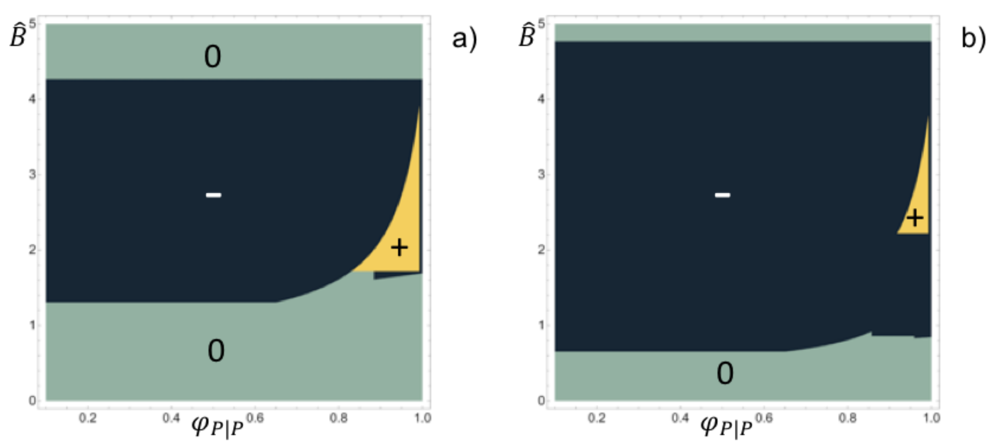


Figure 5 Regions showing where information has positive (+), negative (−), or zero value by accuracy parameter $\phi_{P|P}$ and scaled budget $\hat{B}$ when a) $\hat{F}=0$ and $q=0$ and b) $\hat{F}=0.5$ and $q=0.5$ ($A_P=0.72$, $A_N=0.13$, $\phi_{P|N}=0.1$).

Strikingly, there are large regions where information has a negative value. In particular, this is true for smaller values of $\phi_{P|P}$ and moderate, i.e. non-trivial, values of the budget, as described by Proposition 4b. For low values of $\phi_{P|P}$, the probability that $\beta_1=P$ given the formative assessment result X_1 is high for both $X_1=P$ and $X_1=N$, which results in low merit-based incentive thresholds.

Therefore, as the budget increases from zero, the district will first be able to afford incentivizing teachers under the formative assessment, for all possible scenarios (i.e., at $t=0$ and at $t=1$ for $X_1=P$ and $X_1=N$). Figure 5b suggests that this result holds even for large levels of information asymmetry.

For higher values of $\phi_{P|P}$, there is greater variation in whether or not information is valuable. The optimal assessment choice in this region is driven by two factors. First, when $\mathbf{S}_0=P$, the reward threshold at $t=1$ for $X_1=N$ under the formative assessment is increasing in $\phi_{P|P}$. This threshold is less than the reward threshold at $t=0$ for small values of $\phi_{P|P}$, but it exceeds the threshold at $t=0$ for large values of $\phi_{P|P}$. This is in contrast to the interim assessment, for which the reward threshold at $t=1$ when $\beta_1=N$ always exceeds the reward threshold at $t=0$ when $\mathbf{S}_0=P$. Second, recall that the district is constrained by its expected costs. Paradoxically, this means that a district may only be able to afford incentivizing effort under an assessment that results in a *lower* probability of the school achieving proficiency at $t=2$ even if the reward amount is *higher* than what it would have been under the alternative assessment choice.

In Figure 5a, as $\phi_{P|P}$ increases, information first becomes valuable (at $\phi_{P|P}\approx 0.85$) when the budget can no longer support a reward under the formative assessment; because the incentive threshold at $t=0$ under the formative assessment is increasing in $\phi_{P|P}$ in this region, so are the district's expected costs. However, at this point, the district can afford to offer a similar level of reward under the interim assessment, due to the lower probability of achieving proficiency at $t=2$. Hence, information has a positive value in this case.

Once $\phi_{P|P}$ is sufficiently high, the merit-based incentive threshold at $t=1$ when $X_1=N$ exceeds that under the formative assessment at $t=0$. For $\phi_{P|P}$ values in this region, as the budget increases, the district can first afford to incentivize positive effort at $t=0$ and at $t=1$ when the mid-year assessment result is proficient under the formative assessment; this results in a small region where information has a negative value. Once the budget is large enough to support incentivizing effort at $t=0$ and at $t=1$ for a proficient mid-year result under both the formative and interim assessments, the district is better off investing in the interim assessment. Finally, when the budget is large enough to support always incentivizing effort under the formative assessment, including when $X_1=N$, investing in the interim assessment becomes counterproductive.

Notice that as $\phi_{P|P}$ approaches 1, there is a small region of $\phi_{P|P}$ for which information always has negative value adjacent to the region where information has positive value. For these high values of $\phi_{P|P}$, teachers almost always exert effort when the true state is proficient, and they also exert effort when the true state is not proficient with probability $\phi_{P|N}$. If the district invested in information in this case, the additional benefit from always exerting effort when the true state is proficient is outweighed by the lost chance of recovering from falling into the not-proficient state.

Finally, as one might expect, as q increases, the set of values for which information has positive value shrinks and the set for which information has negative value increases. Interestingly, this occurs when the belief that teachers may have perfect information lowers the probability of achieving proficiency under the formative assessment, thereby decreasing the district's expected cost from offering rewards and allowing the districting to offer rewards at a lower budget.

6. Discussion

In this paper, we consider the tradeoffs faced by a school district that must allocate a fixed budget between merit-based incentives for teachers and an interim assessment that gives precise information about mid-year student proficiency. We study this setting using a dynamic, two-period principal-agent modeling framework with hidden information, where the school district is the principal and the teachers at a given school the agent. Using a stylized model, we characterize the teachers' optimal effort policy for a given level of merit-based incentive under each type of assessment. We also describe the district's optimal level of performance-based incentives and choice of assessment.

Notably, we establish that for low-performing schools, i.e., schools that begin the year being "behind," the school district never benefits from investing in additional information via the interim assessment. In this case, the challenge of achieving proficiency at the end of the year is great, and a commensurately high incentive must be offered to encourage teachers to exert effort throughout the school year. If such an incentive is offered, teachers will always exert maximal effort, regardless of mid-year student progress. Thus, districts are better off investing their entire budget in merit-based incentives.

For schools that begin the year on track to achieve proficiency, whether or not the additional information is a valuable investment depends on the accuracy of the formative assessment. Specifically, in the absence of clear information, teachers are inclined to believe that the state will remain proficient if they exert effort in the first half of the academic year. This, in turn, makes it easier to motivate teachers using incentives. In this case, the district is better off foregoing investment in the interim assessment; doing so enables them to incentivize maximal effort throughout the year.

On the other hand, when the formative assessment more accurately relays the true mid-year state of proficiency, the school district generally benefits from investing in information for low budget values. In this case, the relative accuracy of the formative assessment makes teachers less likely to rely on their knowledge of students' high performance at the beginning of the year, which consequently raises the level of merit-based incentive required to motivate them. An exception to this is when the formative assessment almost perfectly determines when the school is in the proficient state: then, investing in information has minimal impact on teachers' behavior when students peform well on the mid-year assessment, but reduces teachers' effort when students have fallen behind. In such settings, the school district should rely on the formative assessment. Finally, for larger budgets, the school district is able to afford to incentivize teachers to exert maximal effort throughout the year when they are relying on the formative assessment. In this case, investing in the interim assessment is no longer optimal.

We recognize that our model has several limitations. First, we characterize effort as a single-dimensional decision made solely by teachers. In practice, teaching and the effort put into it consist of many distinct components, such as lesson planning and professional development. Moreover, students also make an effort decision: they actively determine the type and level of their own effort to exert throughout the year. The information provided by an interim assessment may allow both students and teachers to target their effort more effectively, resulting in a higher probability of students succeeding at no additional "cost" to teachers. We expect that extending our work to include such features will broaden the range of settings for which mid-year perfromance information has a positive value. Furthermore, we focus on the level of proficiency attained by a school and do not take into account student "growth." Student growth targets are designed to standardize the level of effort necessary for teachers and districts to meet their performance goals, regardless of the initial state of student proficiency. In reality, both measures are important for school accountability. A simple way to adapt our model to this setting is by assuming that every school starts in the "proficient" state, and letting high-performing and low-performing schools vary in terms of how easily they can transition to the proficient state or recover from falling behind.

We believe the insights derived from our stylized model are noteworthy. In practice, school districts often focus on the potential benefits from providing additional information, but not on the potential drawbacks stemming from it. Our analysis identifies settings where extra information is beneficial, but also settings where it may have a demotivating effect. We view our model as the first step in exploring the rich and complex environment of education.

References

Barlevy G, Neal D (2012) Pay for percentile. *American Economic Review* 102(5):1805–31.

Baron DP, Myerson RB (1982) Regulating a monopolist with unknown costs. *Econometrica: Journal of the Econometric Society* 911–930.

Cavanagh S (2015) As McGraw-Hill Education leaves state testing, market thrives for classroom assessment. *Education Week* URL `https://www.edweek.org/ew/articles/2015/07/31/as-mcgraw-hill-education-leaves-state-testing-market.html`.

Chiang H, Speroni C, Herrmann M, Hallgren K, Burkander P, Wellington A (2017) Evaluation of the Teacher Incentive Fund: Final Report on Implementation and Impacts of Pay-for-Performance Across Four

Years (NCEE 2017-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Chu LY, Sappington DE (2009) Implementing high-powered contracts to motivate intertemporal effort supply. *The Rand Journal of Economics* 40(2):296–316.

Clotfelter CT, Ladd HF (1996) Recognizing and rewarding success in public schools. *Holding schools accountable: Performance-based reform in education* 23–64.

Clotfelter CT, Ladd HF, Vigdor JL (2006) Teacher-student matching and the assessment of teacher effectiveness. *Journal of human Resources* 41(4):778–820.

Corbett CJ, De Groote X (2000) A supplier's optimal quantity discount policy under asymmetric information. *Management science* 46(3):444–450.

Crémer J, Khalil F, Rochet JC (1998) Contracts and productive information gathering. *Games and Economic Behavior* 25(2):174–193.

Davison ML, Seo YS, Davenport Jr EC, Butterbaugh D, Davison LJ (2004) When do children fall behind? What can be done? *Phi Delta Kappan* 85(10):752.

Dee TS, Wyckoff J (2015) Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management* 34(2):267–297.

Dillon E, Rotherham A (2007) States evidence: What it means to make 'Adequate Yearly Progress' under NCLB. *Education sector explainers* .

District of Columbia Office of the State Superintendent of Education (2011) Assessment and Accountability Manual. URL `https://osse.dc.gov/publication/district-columbia-assessment-and-accountability-manual`.

District of Columbia Public Schools (2018a) DCPS at a Glance: Enrollment. URL `https://dcps.dc.gov/page/dcps-glance-enrollment`.

District of Columbia Public Schools (2018b) DCPS Data Set - DC CAS. URL `https://dcps.dc.gov/publication/dcps-data-set-dc-cas`.

District of Columbia Public Schools (2018c) Our Schools. URL `https://dcps.dc.gov/page/our-schools`.

Eberts R, Hollenbeck K, Stone J (2002) Teacher performance incentives and student outcomes. *Journal of Human Resources* 913–927.

Figlio DN, Kenny LW (2007) Individual teacher incentives and student performance. *Journal of Public Economics* 91(5):901–914.

Fryer RGJ (2013) Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics* 31(2):373–407.

Fudenberg D, Holmstrom B, Milgrom P (1990) Short-term contracts and long-term agency relationships. *Journal of economic theory* 51(1):1–31.

Fuloria PC, Zenios SA (2001) Outcomes-adjusted reimbursement in a health-care delivery system. *Management Science* 47(6):735–751.

Ha AY (2001) Supplier-buyer contracting: Asymmetric cost information and cutoff level policy for buyer participation. *Naval Research Logistics (NRL)* 48(1):41–64.

Johnson SM (1984) Merit pay for teachers: A poor prescription for reform. *Harvard Educational Review* 54(2):175–186.

Klein A (2015) No child left behind: An overview. *Education Week* URL `http://www.edweek.org/ew/section/multimedia/no-child-left-behind-overview-definition-summary.html`.

Lewis TR, Sappington DE (1997) Information management in incentive problems. *Journal of political Economy* 105(4):796–821.

Linn RL (2000) Assessments and accountability. *Educational researcher* 29(2):4–16.

Lutze H, Özer Ö (2008) Promised lead-time contracts under asymmetric information. *Operations Research* 56(4):898–915.

Murnane R, Cohen D (1986) Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review* 56(1):1–18.

Neal D, Schanzenbach DW (2010) Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics* 92(2):263–283.

Perie M, Marion S, Gong B, Wurtzel J (2007) The role of interim assessments in a comprehensive assessment system. *Washington, DC: The Aspen Institute* .

Plambeck EL, Zenios SA (2000) Performance-based incentives in a dynamic principal-agent model. *Manufacturing & Service Operations Management* 2(3):240–263.

Shumsky RA, Pinker EJ (2003) Gatekeepers and referrals in services. *Management Science* 49(7):839–856.

Springer MG, Ballou D, Hamilton L, Le VN, Lockwood JR, McCaffrey DF, Pepper M, Stecher BM (2011) Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching (POINT). *Society for Research on Educational Effectiveness* .

Tan TF, Netessine S (2014) When does the devil make work? an empirical study of the impact of workload on worker productivity. *Management Science* 60(6):1574–1593.

Topol B, Olson J, Roeber E, Hennon P (2012) Getting to higher-quality assessments: Evaluating costs, benefits and investment strategies.

United States General Accounting Office (2003) Title I: Characteristics of Tests Will Influence Expenses; Information Sharing May Help States Realize Efficiencies (GAO-03-389).

US Department of Education (2009) Race to the Top: Executive Summary.

Zhang H, Zenios S (2008) A dynamic principal-agent model with hidden information: Sequential optimality through truthful state revelation. *Operations Research* 56(3):681–696.

Appendix for "Investing in Performance: Information and Merit-Based Incentives in K-12 Education"

A1. Merit-Based Incentive Threshold at $t = 1$ Under Perfect Information

Corollary A1 describes the merit-based incentive threshold – that is, the minimum level of merit-based incentive for which teachers are incentivized to exert effort – in the case where either the district invests in a perfectly-informative interim assessment ($z_I = 1$) or the teachers are able to accurately gauge student knowledge regardless of the assessment type ($T = 1$).

COROLLARY A1. *When teachers have perfect information about the state at* $t = 1$*, the minimum level of scaled merit-based incentive necessary to induce positive effort at* $t = 1$ *under the optimal effort policy is given by*

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \begin{cases} \frac{1}{A_P}, & \text{if } \mathbf{S}_1 = P, \\ \frac{1}{A_N}, & \text{if } \mathbf{S}_1 = N. \end{cases} \tag{A1}$$

Proof of Corollary A1

The result follows from Proposition 1 using $\phi_{P|P} = 1$ and $\phi_{P|N} = 0$. □

The closed-form expressions for the teachers' optimal effort level at $t = 1$ provided in Corollary A1 are intuitive and stem from Assumptions 1 and 3, which we make on the functional form of the teachers' response-to-effort and cost-of-effort functions. Note that in the perfect information case, when $\mathbf{S}_1 = P$ ($\mathbf{S}_1 = N$), $\hat{e}_1^*(\mathbf{S}_1)$ depends only on A_P (A_N).

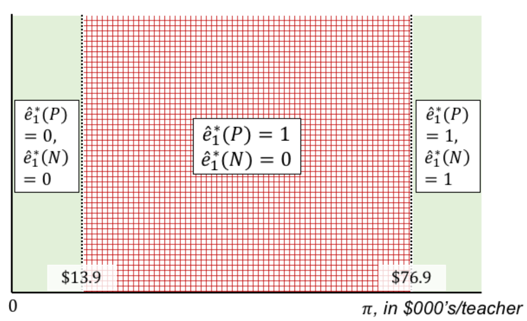


Figure A1 **Optimal scaled effort at $t = 1$, $\hat{e}_1^*$, as a function of the per-teacher merit-based incentive π for different values of the state S_1 when teachers have perfect information ("base-case" parameter set: $A_P = 0.72$, $A_N = 0.13$, $\lambda\gamma = \$350,000$, 35 teachers per school).**

The results of Corollary A1 are illustrated in Figure A1. The optimal scaled effort level at $t = 1$, $\hat{e}_1^*(\mathbf{S}_1)$, is shown as a function of the merit-based incentive level π for different values of the state $\mathbf{S}_1$. Figure A1 illustrates the dramatic difference, for realistic values of problem parameters, in the levels of incentives required to generate teachers' response in proficient and not-proficient states:

approximately, \$14,000 vs. \$77,000. While specific values of estimates we obtain rely heavily on our assumption set, the gap in required incentives is driven by the corresponding disparity in how far teachers' efforts go in the two states, and is likely to remain significant under a wide range of alternative assumptions.

A2. Properties of the Merit-Based Incentive Threshold at $t = 0$

In the following result, we describe the minimum reward necessary to incentivize positive effort at $t = 0$ for the special case where teachers know the true intermediate state ($z_I = 1$ or $T = 1$).

COROLLARY A2. *When teachers have perfect information about the intermediate state at $t = 1$, the minimum level of merit-based incentive necessary to induce positive effort at $t = 0$ under the optimal effort policy is*

$$\hat{\mathcal{P}}_0(\mathbf{S}_0) = \begin{cases} \frac{1}{A_N(A_P - A_N)}, & \textit{if } \mathbf{S}_0 = N, \\ \frac{1}{A_P(A_P - A_N)}, & \textit{if } \mathbf{S}_0 = P \textit{ and } (A_P, A_N) \in \mathcal{S}_1, \\ \frac{1 + A_P}{A_P^2}, & \textit{if } \mathbf{S}_0 = P \textit{ and } (A_P, A_N) \in \mathcal{S}_4. \end{cases} \quad \text{(A2)}$$

Proof of Corollary A2

The result is obtained from Proposition 2 using $\phi_{P|P} = 1$ and $\phi_{P|N} = 0$. □

When teachers have perfect information about the state at $t = 1$, the decision-making at $t = 0$ is different from that at $t = 1$ in an important way: at $t = 0$ the threshold value of the merit-based pay required to induce positive effort depends on the characteristics of both proficient and not-proficient states rather than only on the response-to-effort parameter of the state the school is in at $t = 0$. (In the last case in Corollary A2, the region determination depends on both A_P and A_N even though the reward threshold value does not.) This result is a direct consequence of the dynamic nature of the teachers' effort selection problem that must account for all the future states that the school may be in at $t = 1$.

Figure A2 compares the minimum level of scaled reward necessary to induce a positive effort level at $t = 0$, $\mathcal{P}_0(\mathbf{S}_0)$, as a function of the response-to-effort parameter A_N when teachers have imperfect and perfect information about the state at $t = 1$. When $\mathbf{S}_0 = N$, the value of the reward threshold is independent of the assessment accuracy and higher than that in the case where the school starts in the proficient state. When $\mathbf{S}_0 = P$, whether the reward threshold is higher or lower under uncertainty (in this case, $\phi_{P|P} = 0.9$ and $\phi_{P|N} = 0.5$) than under perfect information depends on the value of A_N. In particular, there are three regions of A_N in which a different effect dominates. For low values of A_N, teachers have a natural incentive to remain in the proficient state since A_N is significantly smaller than A_P. This effect is more pronounced under uncertainty than under perfect information ($z_I = 1$ or $T = 1$), which makes it less costly to induce positive effort under the

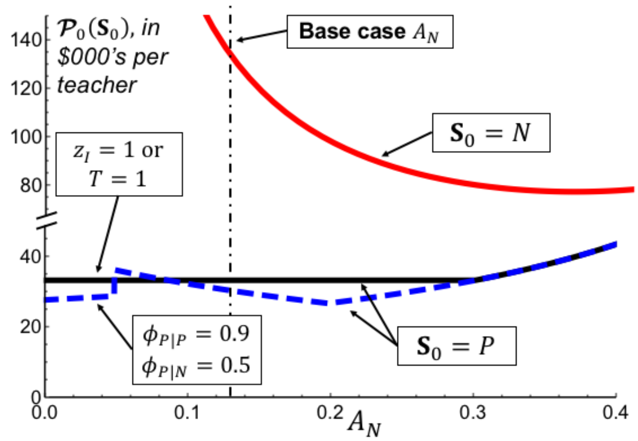


Figure A2 $\mathcal{P}_0(\mathbf{S}_0)$**, the minimum level of per-teacher merit-based incentive required to induce positive effort at $t = 0$ as a function of the response to effort parameter A_N for $\mathbf{S}_0 = N$ and $\mathbf{S}_0 = P$ under both perfect and imperfect ($\phi_{P|P} = 0.9$ and $\phi_{P|N} = 0.5$) information ($A_P = 0.72$).**

formative assessment. As A_N increases, there is a critical point at which the difference between A_N and A_P becomes secondary. Initially at this point, teachers overestimate the possibility of being in the not-proficient state relative to having perfect information, which results in a higher reward threshold under uncertainty. However, as A_N continues to increase, the greater responsiveness-to-effort under the not-proficient state eventually leads to a lower reward threshold under uncertainty. Finally, at another critical point of A_N, the proximity of A_N to A_P acts as a disincentive to exert effort at $t = 0$, so the reward threshold increases as A_N increases. Once A_N is sufficiently close to A_P, the threshold is independent of the accuracy of the formative assessment. In addition to these qualitative trends, it is important to point out two quantitative outcomes observed under realistic "base-case" parameter values. First, the monetary incentive required to induce teachers' response at $t = 0$ is several times higher in the not-proficient state (around \$135,000) than in the proficient state (around \$35,000). Second, in either state our estimates for this incentive level at $t = 0$ are around two times higher than those for the corresponding level at $t = 1$, reaching impractical values. Note that the accuracy of formative assessments only slightly modulates these outcomes.

To further illustrate the nuances of the relationship between accuracy and responsiveness-to-effort, Figure A3 shows the reward threshold under imperfect information for $\phi_{P|P} = 0.9$ as a function of $\phi_{P|N}$, relative to the perfect information case. The figure suggests that, for smaller values of $\phi_{P|N}$, a not-proficient assessment result is given greater weight in the imperfect information case than under perfect information, making it more costly to induce effort under uncertainty. However, as $\phi_{P|N}$ increases, becoming closer to $\phi_{P|P}$ and therefore less informative, teachers rely more on the initial state ($\mathbf{S}_0 = P$) and the effort at $t = 0$ than on the assessment result. This leads to under-weighting the likelihood of being in the not-proficient state at $t = 1$; consequently, a lower reward

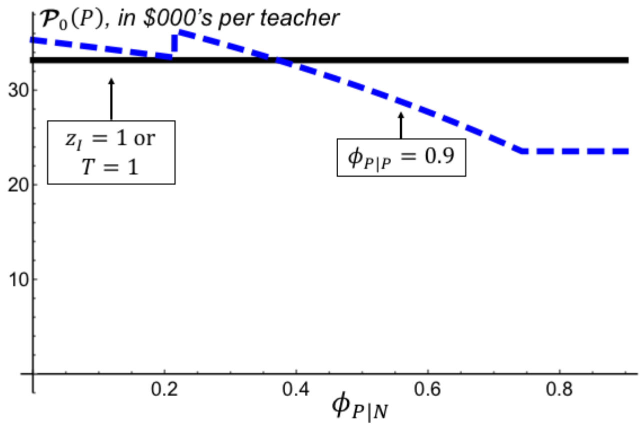


Figure A3 The minimum level of per-teacher merit-based incentive required to induce positive effort at $t = 0$ for $\mathbf{S}_0 = P$ as a function of the accuracy parameter $\phi_{P|N}$ and different values of $\phi_{P|P}$ relative to the perfect information case ($A_P = 0.72, A_N = 0.13$).

is necessary to induce positive effort. The discontinuity at $\phi_{P|N} = 0.21$ occurs when the (A_P, A_N) region changes from $\mathcal{S}_4$ to $\mathcal{S}_3$.

Figures A2 and A3 illustrate an interesting outcome: in some settings, it is less costly to incentivize teachers to exert effort when they have imperfect information about the intermediate state of the system, than under perfect information. In Corollary A3, we show that this outcome always holds for two special cases.

Corollary A3. *For $\mathbf{S}_0 = P$, suppose that either $\phi_{P|P} = 1$ and $\phi_{P|N} > 0$ or that $\phi_{P|P} = \phi_{P|N}$. Additionally, suppose A_N is sufficiently smaller than A_P, so that $(A_P, A_N) \in \mathcal{S}_4$ when $\phi_{P|P} = 1$ and $(A_P, A_N) \in \mathcal{S}_2$ when $\phi_{P|P} = \phi_{P|N}$. Then, the merit-based incentive threshold at $t = 0$, $\mathcal{P}_0(\mathbf{S}_0)$, is lower when teachers have imperfect information about the intermediate state than when they have perfect information.*

Proof of Corollary A3

From (31),

$$\mathcal{A}_2 = \frac{A_P^2 \left(1 - \phi_{P|P}\right) + A_N \left(1 - A_P\right) \left(1 - \phi_{P|N}\right)}{\left(1 - A_P\right) \left(1 - \phi_{P|N}\right) \left(A_P \phi_{P|P} - A_N \phi_{P|N}\right)}. \tag{A3}$$

When $\phi_{P|P} = 1$,

$$\mathcal{A}_2 = \frac{A_N}{A_P - A_N \phi_{P|N}}, \tag{A4}$$

and from (37),

$$(A_P, A_N) \in \mathcal{S}_4 \iff \mathcal{A}_2 < A_P - A_N \iff A_N < \left(A_P - A_N\right) \left(A_P - A_N \phi_{P|N}\right). \tag{A5}$$

For any value of A_P, this statement holds for sufficiently small A_N, since

$$\lim_{A_N \to 0^+} A_N = 0 < A_P^2 = \lim_{A_N \to 0^+} (A_P - A_N)\left(A_P - A_N \phi_{P|N}\right). \tag{A6}$$

Next, compare the merit-based incentive thresholds under perfect and imperfect information when $(A_P, A_N) \in \mathcal{S}_4$. If $\mathbf{S}_0 = P$ and $(A_P, A_N) \in \mathcal{S}_4$, from (39) and (A2),

$$\hat{\mathcal{P}}_0(\mathbf{S}_0) = \begin{cases} \mathcal{R}_2, & \text{if } z_I = 0 \text{ and } T = 0, \\ \frac{1+A_P}{A_P^2}, & \text{if } z_I = 1 \text{ or } T = 1, \end{cases} \tag{A7}$$

where from (33),

$$\mathcal{R}_2 = \frac{1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}. \tag{A8}$$

Then,

$$\mathcal{R}_2 \le \frac{1+A_P}{A_P^2} \iff \frac{1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} \le \frac{1+A_P}{A_P^2} \iff 1 - \phi_{P|P} \le \left((A_P - A_N) - \frac{A_N}{A_P}\right)\phi_{P|N}. \tag{A9}$$

Note that

$$(A_P, A_N) \in \mathcal{S}_4 \Rightarrow \frac{A_N}{A_P} \le A_P - A_N, \tag{A10}$$

so the right-hand side of (A9) is non-negative. Therefore, (A9) holds when $\phi_{P|P}$ is sufficiently large relative to $\phi_{P|N}$. In particular, this always holds when $\phi_{P|P} = 1$.

From (30),

$$\mathcal{A}_1 = \frac{A_P^2\left(1 - \phi_{P|P}\right) + A_N(1 - A_P)\left(1 - \phi_{P|N}\right)}{A_P^2\left(1 - \phi_{P|P}\right) + A_P(1 - A_P)\left(1 - \phi_{P|N}\right)}. \tag{A11}$$

When $\phi_{P|P} = \phi_{P|N}$,

$$\mathcal{A}_1 = \frac{A_P^2 + A_N(1 - A_P)}{A_P^2 + A_P(1 - A_P)}, \tag{A12}$$

and from (35),

$$(A_P, A_N) \in \mathcal{S}_2 \iff \frac{A_N}{A_P} < A_P - A_N \le \mathcal{A}_1 \iff \frac{A_N}{A_P} < A_P - A_N \le \frac{A_P^2 + A_N(1 - A_P)}{A_P^2 + A_P(1 - A_P)}. \tag{A13}$$

For any value of A_P, there exists A_N sufficiently small that this statement holds. Specifically, for all A_P, the above statement holds when $A_N = 0$.

Again, compare the merit-based incentive thresholds under perfect and imperfect information when A_N is sufficiently small relative to A_P, i.e. $(A_P, A_N) \in \mathcal{S}_2$ for imperfect information and $(A_P, A_N) \in \mathcal{S}_4$ for perfect information. In this case, if $\mathbf{S}_0 = P$, from (39) and (A2),

$$\hat{\mathcal{P}}_0(\mathbf{S}_0) = \begin{cases} \frac{1}{A_P(A_P - A_N)}, & \text{if } z_I = 0 \text{ and } T = 0, \\ \frac{1+A_P}{A_P^2}, & \text{if } z_I = 1 \text{ or } T = 1. \end{cases} \tag{A14}$$

Now,

$$\frac{1}{A_P\left(A_P-A_N\right)} \le \frac{1+A_P}{A_P^2} \iff A_P \le \left(1+A_P\right)\left(A_P-A_N\right) \iff \frac{A_N}{A_P} \le A_P - A_N, \tag{A15}$$

which always holds in this case. □

In Corollary A3, we consider two settings when the initial state is proficient: first, if the formative assessment correctly identifies when the true intermediate state is proficient but does not always identify when the true state is not-proficient, and second, if the formative assessment results are completely uninformative, i.e., where the probability of getting a proficient assessment result is independent of the true state of proficiency. In these settings, if the probability of transitioning to the proficient state is significantly higher when starting in the proficient state than in the not-proficient state, then it is cheaper to incentivize teachers to exert effort when they have imperfect information about the intermediate state. When $\phi_{P|P}=1$, the lower likelihood of a not-proficient result tempers the high "penalty" effect from a small A_N. In the case where $\phi_{P|P}=\phi_{P|N}$, the formative assessment provides no useful information, so teachers place greater weight on the initial state being proficient.

The following result formalizes sufficient conditions for our observation that the level of merit-based compensation necessary to incentivize positive effort at $t=0$ is higher than the corresponding threshold at $t=1$ for either state of the system.

Corollary A4. *Suppose the initial state is not proficient ($\mathbf{S}_0=N$) or suppose the initial state is proficient ($\mathbf{S}_0=P$) and $(A_P,A_N)\in\mathcal{S}_1$. Then, for all $\mathbf{S}_1$, the reward threshold to induce positive effort at $t=0$ is greater than the reward threshold to induce positive effort at $t=1$: $\mathcal{P}_0\left(\mathbf{S}_0\right)>\mathcal{P}_1\left(\mathbf{S}_1\right)$.*

Proof of Corollary A4

From (39),

$$\hat{\mathcal{P}}_0\left(\mathbf{S}_0\right)=\begin{cases}\frac{1}{A_N(A_P-A_N)}, & \text{if } \mathbf{S}_0=N,\\ \frac{1}{A_P(A_P-A_N)}, & \text{if } \mathbf{S}_0=P \text{ and } (A_P,A_N)\in\mathcal{S}_1\cup\mathcal{S}_2,\\ \mathcal{R}_1, & \text{if } \mathbf{S}_0=P \text{ and } (A_P,A_N)\in\mathcal{S}_3,\\ \mathcal{R}_2, & \text{if } \mathbf{S}_0=P \text{ and } (A_P,A_N)\in\mathcal{S}_4,\end{cases} \tag{A16}$$

and from (23) and (24),

$$\hat{\mathcal{P}}_1\left(\mathbf{S}_1\right)=\begin{cases}\frac{\phi_{P|P}m(e_0,\mathbf{S}_0)+\phi_{P|N}(1-m(e_0,\mathbf{S}_0))}{A_P\phi_{P|P}m(e_0,\mathbf{S}_0)+A_N\phi_{P|N}(1-m(e_0,\mathbf{S}_0))}, & \text{if } \mathbf{S}_1=\left(P,e_0,\mathbf{S}_0\right),\\ \frac{\left(1-\phi_{P|P}\right)m(e_0,\mathbf{S}_0)+\left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0))}{A_P\left(1-\phi_{P|P}\right)m(e_0,\mathbf{S}_0)+A_N\left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0))}, & \text{if } \mathbf{S}_1=\left(N,e_0,\mathbf{S}_0\right).\end{cases} \tag{A17}$$

Then, from (B22)-(B25),

$$\hat{\mathcal{P}}_1\left(X_1,e_0,\mathbf{S}_0\right) \le \frac{1}{A_N} \text{ for all } X_1 \text{ and } \mathbf{S}_0. \tag{A18}$$

Therefore, for all X_1, $\hat{\mathcal{P}}_1(X_1, e_0, N)$ is bounded above by $\hat{\mathcal{P}}_0(N)$.

Consider the case where $\mathbf{S}_0 = P$. For $(A_P, A_N) \in \mathcal{S}_1 \cup \mathcal{S}_2$, the threshold at $t = 0$ always exceeds the threshold at $t = 1$ if

$$\frac{1}{A_P(A_P - A_N)} \geq \frac{1}{A_N} \iff \frac{A_N}{A_P} \geq A_P - A_N. \tag{A19}$$

This always holds when $(A_P, A_N) \in \mathcal{S}_1$ and never holds when $(A_P, A_N) \in \mathcal{S}_2$.

When $(A_P, A_N) \in \mathcal{S}_3$, the threshold at $t = 0$ cannot exceed every reward threshold at $t = 1$, since

$$\hat{\mathcal{P}}_0(P) = \mathcal{R}_1 = \frac{A_P\left(1 - \phi_{P|P}\right) + (1 - A_P)\left(1 - \phi_{P|N}\right)}{A_P^2\left(1 - \phi_{P|P}\right) + A_N(1 - A_P)\left(1 - \phi_{P|N}\right)} \geq \frac{1}{A_N} \iff A_N \geq A_P, \tag{A20}$$

which is impossible under Assumption 1.

When $(A_P, A_N) \in \mathcal{S}_4$, it is possible for the threshold at $t = 0$ to exceed the threshold at $t = 1$, since

$$\hat{\mathcal{P}}_0(P) = \mathcal{R}_2 = \frac{1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} \geq \frac{1}{A_N} \iff \frac{A_N}{A_P\phi_{P|P}} \geq A_P - A_N. \tag{A21}$$

□

When the school begins the year in the not-proficient state or the probability of achieving proficiency does not significantly depend on the starting state, then inducing positive effort earlier in the year is more costly. In such settings, a district with a limited budget may benefit from allocating non-monetary resources or incentives during this period. For example, in practice classroom observations tend to occur towards the end of the school year; our results suggest there is additional value to be gained from scheduling a greater portion of classroom observations earlier in the year.

Finally, the following result describes the monotonicity properties of $\hat{e}_0^*$ for the perfect information case.

Corollary A5. *In the perfect information case, the optimal scaled effort $\hat{e}_0^*(\mathbf{S}_0)$ is non-decreasing in $\hat{\pi}$ and non-decreasing in A_P for $\mathbf{S}_0 = P$ and $\mathbf{S}_0 = N$. In addition, $\hat{e}_0^*(P)$ is non-increasing in A_N.*

Proof of Corollary A5

From (38), it is clear that $\hat{e}_0^*(\mathbf{S}_0)$ is non-decreasing in $\hat{\pi}$.

Next, consider $\hat{e}_0^*(P)$. When $\phi_{P|P} = 1$ and $\phi_{P|N} = 0$, $(A_P, A_N) \in \mathcal{S}_1$ and $(A_P, A_N) \in \mathcal{S}_4$ are equivalent to $0 \leq A_P \leq \frac{A_N + \sqrt{A_N^2 + 4A_N}}{2}$ and $\frac{A_N + \sqrt{A_N^2 + 4A_N}}{2} < A_P \leq 1$. Then, we have from (A2)

$$\hat{\mathcal{P}}_0(P) = \begin{cases} \frac{1}{A_P(A_P - A_N)}, & \text{if } 0 \leq A_P \leq \frac{A_N + \sqrt{A_N^2 + 4A_N}}{2}, \\ \frac{1 + A_P}{A_P^2}, & \text{if } \frac{A_N + \sqrt{A_N^2 + 4A_N}}{2} < A_P \leq 1. \end{cases} \tag{A22}$$

Suppose A_P is fixed. For sufficiently small A_N, the scaled reward threshold required to induce positive effort is $\frac{1+A_P}{A_P^2}$, which does not change with A_N. Once A_N is sufficiently large, so that $A_P \leq \frac{A_N+\sqrt{A_N^2+4A_N}}{2}$, the value of $\hat{\mathcal{P}}_0(P)$ is at least as large as the corresponding value when A_N is small:

$$\frac{1}{A_P(A_P-A_N)} \geq \frac{1+A_P}{A_P^2} \iff \frac{A_N}{A_P} \geq A_P - A_N, \tag{A23}$$

which holds for $A_P \leq \frac{A_N+\sqrt{A_N^2+4A_N}}{2}$. Furthermore, the threshold $\frac{1}{A_P(A_P-A_N)}$ is increasing in A_N. Therefore, when $\phi_{P|P}=1$ and $\phi_{P|N}=0$, the reward threshold is increasing in A_N, so $\hat{e}_0^*(P)$ is non-increasing in A_N.

Next, suppose that A_N is fixed. When A_P is small, the scaled reward threshold is $\frac{1}{A_P(A_P-A_N)}$, which is decreasing in A_P. As A_P increases, the reward threshold eventually becomes $\frac{1+A_P}{A_P^2}$. From (A23), this is smaller than the initial threshold value, and under Assumption 1, it is straightforward to show that this threshold is decreasing in A_P. Therefore, the reward threshold is decreasing in A_P, so $\hat{e}_0^*(P)$ is non-decreasing in A_P.

Now, consider $\hat{e}_0^*(N)$. The scaled reward threshold required for positive effort is $\frac{1}{A_N(A_P-A_N)}$, and this value is decreasing in A_P. Therefore, $\hat{e}_0^*(N)$ is non-decreasing in A_P. $\square$

Corollary A5 states that, as expected, a higher merit-based incentive results in higher effort levels. Furthermore, increasing the proficient response-to-effort parameter A_P results in a higher optimal effort. Intuitively, as A_P increases and the not-proficient response-to-effort parameter A_N remains fixed, it becomes comparatively easier to remain in the proficient state than to transition to the proficient state from the not-proficient state. Therefore, teachers are incentivized to exert a higher effort level in order to remain in the proficient state. The reverse logic holds when the initial state is proficient and A_N is increasing relative to a fixed A_P. These effects are evident in Figure A2.

However, the effect of A_N on optimal effort levels when starting in the not-proficient state is less straightforward. Figure A4 shows the optimal scaled effort level at $t=0$, $\hat{e}_0^*$, as a function of A_N when the system starts in the not-proficient state for fixed π. (Note that this result does not depend on the accuracy of teachers' information about mid-year student progress.) For low values of A_N, starting in the not-proficient state is a great disadvantage; teachers react to increases in A_N by scaling up their efforts. Once A_N becomes sufficiently high (and the not-proficient state becomes sufficiently close to the proficient one in terms of the effect of teachers' effort) the teachers react to further increases in A_N in a way that is similar to their reaction if they were in a proficient state. In practice, this means that teachers' effort levels earlier in the school year may slip if the consequences for students' future performance falling to the not-proficient state are either too severe or non-existent. Realistically, it is far more likely that consequences will be too severe, since students that fall behind are more likely to have a harder time transitioning to the proficient state. As π increases, the range of values of A_N which incentivize positive effort expands.

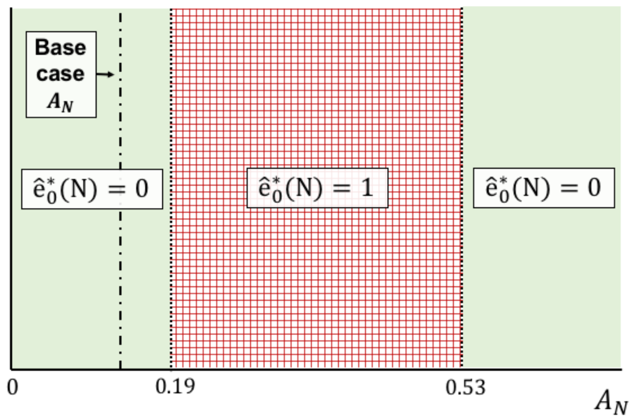


Figure A4 **Optimal scaled effort at $t = 0$ as a function of not-proficient response-to-effort parameter A_N when starting in the not-proficient state ($A_P = 0.72$, $\pi = \$100,000$/teacher).**

A3. Optimal Incentives for No-Information-Asymmetry and Maximum-Information-Asymmetry Cases

Note that, in the extreme settings with no information asymmetry between the district and the teachers ($q = 0$) and with maximum information asymmetry ($q = 1$), the optimal level of merit-based pay exhibits discontinuities and cannot always be obtained using the corresponding limits of the expressions in Proposition 3. The following result describes the optimal level of the scaled merit-based incentive $\hat{\pi}^*$ when the district chooses to rely on the formative assessment and believes that that the teachers have no additional knowledge of the intermediate state.

Proposition A1. *When the district relies on the formative assessment ($z_I = 0$) and believes that the teachers have no additional information about the intermediate state ($q = 0$), the optimal scaled merit-based incentive can be characterized as follows.*

a) For $\mathbf{S}_0 = N$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < 1, \\ \frac{1}{A_N}, & \text{if } 1 \le \hat{B} < 1 + \frac{1}{A_P - A_N}, \\ \frac{1}{A_N(A_P - A_N)}, & \text{if } 1 + \frac{1}{A_P - A_N} \le \hat{B}. \end{cases} \tag{A24}$$

b) For $\mathbf{S}_0 = P$ and $(A_P, A_N) \in \mathcal{S}_1$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < 1, \\ \frac{1}{A_N}, & \text{if } 1 \le \hat{B} < 1 + \frac{A_N}{A_P(A_P - A_N)}, \\ \frac{1}{A_P(A_P - A_N)}, & \text{if } 1 + \frac{A_N}{A_P(A_P - A_N)} \le \hat{B}. \end{cases} \tag{A25}$$

c) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_2$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < 1 + \frac{A_N}{A_P(A_P - A_N)}, \\ \frac{1}{A_P(A_P - A_N)}, & \text{if } 1 + \frac{A_N}{A_P(A_P - A_N)} \le \hat{B}. \end{cases} \tag{A26}$$

d) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_3$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < \mathcal{B}_1, \\ \mathcal{R}_1, & \text{if } \mathcal{B}_1 \le \hat{B}. \end{cases} \tag{A27}$$

e) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_4$,

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} < \mathcal{B}_2, \\ \mathcal{R}_2, & \text{if } \mathcal{B}_2 \le \hat{B} < \mathcal{B}_1, \\ \mathcal{R}_1, & \text{if } \mathcal{B}_1 \le \hat{B}. \end{cases} \tag{A28}$$

Proof of Proposition A1

When the district relies on the formative assessment ($z_I = 0$), the school district's maximization problem is as given in (40)-(41):

$$\max_{\pi \ge 0} Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] \tag{A29}$$

$$\text{s.t. } \pi Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] \le B, \tag{A30}$$

with $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ given in (B108)-(B114) but with $q = 0$.

In the proof of Proposition B1, we show that $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is a non-decreasing step function of $\hat{\pi}$. Therefore, the expression on the left-hand side of the district's constraint (41) is an increasing function of $\hat{\pi}$, and we must consider the value of the objective function (40) at each of the endpoints of each interval of $\hat{\pi}$ that corresponds to a "step." We assume that if $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is constant over a region of $\hat{\pi}$ and any value of $\hat{\pi}$ in that region is optimal, the district will choose the smallest value of $\hat{\pi}$ in that region.

To determine the optimal merit-based incentive $\hat{\pi}^*$ and the corresponding probability that the final state is proficient, we must consider the several cases stated in Proposition B1 that determine the characterization of $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$.

Begin with the case where $\mathbf{S}_0 = N$, where $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is given by (B108). Then, $\hat{\pi}^* = \frac{1}{A_N(A_P - A_N)}$ if

$$\frac{\lambda\gamma}{A_N(A_P - A_N)}(A_N(1 + A_P - A_N)) \le B \iff 1 + \frac{1}{A_P - A_N} \le \hat{B}. \tag{A31}$$

Similarly, $\hat{\pi}^* = \frac{1}{A_N}$ if

$$\frac{\lambda\gamma}{A_N}(A_N) \le B < \frac{\lambda\gamma}{A_N(A_P - A_N)}(A_N(1 + A_P - A_N)) \iff 1 \le \hat{B} < 1 + \frac{1}{A_P - A_N}. \tag{A32}$$

These results are stated in (A24). We follow similar steps for the cases where $\mathbf{S}_0 = P$, using the functional forms for the $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ given by (B109)-(B114). $\square$

As in the case for $0 < q < 1$, the optimal reward levels are non-decreasing in the budget, and the accuracy of the formative assessment matters most when the initial state is proficient and A_P and A_N are sufficiently different, i.e. $(A_P, A_N) \in \mathcal{S}_3 \cup \mathcal{S}_4$, since $\mathcal{R}_1$ and $\mathcal{R}_2$ both depend on $\phi_{P|P}$ and $\phi_{P|N}$. Interestingly, when $(A_P, A_N) \in \mathcal{S}_2 \cup \mathcal{S}_3$, depending on the size of the budget, it is optimal for the district to either offer no merit-based reward or a reward high enough to incentivize positive effort at both $t = 0$ and $t = 1$. Thus, for districts where there is a moderate difference between the marginal impact of effort in the proficient and not-proficient states and where there is no information asymmetry between the district and the teachers, the district is always better off choosing a reward level that will incentivize positive effort across the entire year, if the budget is sufficiently large.

Finally, Proposition A2 describes the optimal level of the scaled merit-based incentive $\hat{\pi}^*$ when the district chooses an interim assessment or there is maximum information asymmetry between the district and teachers.

PROPOSITION A2. *When the district invests in the interim assessment* $(z_I = 1)$ *or there is maximum information asymmetry between the school district and teachers* $(q = 1)$*, the optimal scaled merit-based incentive is as follows.*

a) For $\mathbf{S}_0 = N$*,*

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} - \hat{F}z_I < 1, \\ \frac{1}{A_N}, & \text{if } 1 \le \hat{B} - \hat{F}z_I < 1 + \frac{1}{A_P - A_N}, \\ \frac{1}{A_N(A_P - A_N)}, & \text{if } 1 + \frac{1}{A_P - A_N} \le \hat{B} - \hat{F}z_I. \end{cases} \tag{A33}$$

b) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_1$*,*

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} - \hat{F}z_I < 1, \\ \frac{1}{A_N}, & \text{if } 1 \le \hat{B} - \hat{F}z_I < 1 + \frac{A_N}{A_P(A_P - A_N)}, \\ \frac{1}{A_P(A_P - A_N)}, & \text{if } 1 + \frac{A_N}{A_P(A_P - A_N)} \le \hat{B} - \hat{F}z_I. \end{cases} \tag{A34}$$

c) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_4$*,*

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \le \hat{B} - \hat{F}z_I < 1 + A_P, \\ \frac{1 + A_P}{A_P^2}, & \text{if } 1 + A_P \le \hat{B} - \hat{F}z_I < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ \frac{1}{A_N}, & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B} - \hat{F}z_I. \end{cases} \tag{A35}$$

Proof of Proposition A2

In this case, the school district's maximization problem is as given in (40)-(41):

$$\max_{\pi \ge 0} Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] \tag{A36}$$

$$\text{s.t. } \pi Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] + Fz_I \le B, \tag{A37}$$

with $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ given in (B108)-(B114) but with $q = 1$.

In the proof of Proposition B1, we show that $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is a non-decreasing step function of $\hat{\pi}$. Therefore, the expression on the left-hand side of the district's constraint (41) is an increasing function of $\hat{\pi}$, and we must consider the value of the objective function (40) at each of the endpoints of each interval of $\hat{\pi}$ that corresponds to a "step." We assume that if $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is constant over a region of $\hat{\pi}$ and any value of $\hat{\pi}$ in that region is optimal, the district will choose the smallest value of $\hat{\pi}$ in that region.

To determine the optimal merit-based incentive $\hat{\pi}^*$ and the corresponding probability that the final state is proficient, we must consider the several cases stated in Proposition B1 that determine the characterization of $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$.

Begin with the case where $\mathbf{S}_0 = N$, where $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is given by (B108). Then, $\hat{\pi}^* = \frac{1}{A_N(A_P - A_N)}$ if

$$\frac{\lambda\gamma}{A_N(A_P - A_N)}\left(A_N\left(1 + A_P - A_N\right)\right) \leq B - Fz_I \iff 1 + \frac{1}{A_P - A_N} \leq \hat{B} - \hat{F}z_I. \tag{A38}$$

Similarly, $\hat{\pi}^* = \frac{1}{A_N}$ if

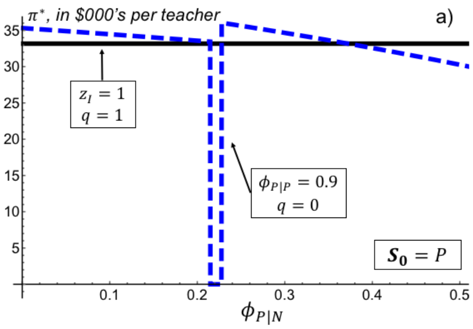
$$\frac{\lambda\gamma}{A_N}\left(A_N\right) \leq B - Fz_I < \frac{\lambda\gamma}{A_N(A_P - A_N)}\left(A_N\left(1 + A_P - A_N\right)\right) \iff 1 \leq \hat{B} - \hat{F}z_I < 1 + \frac{1}{A_P - A_N}. \tag{A39}$$

These results are stated in (A33). We follow similar steps for the cases where $\mathbf{S}_0 = P$, using the functional forms for the $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ given by (B109) and (B114); the other cases are not feasible in this setting.

□

In the setting where, under formative assessment, there is the highest degree information asymmetry between the district and the teachers, teachers are naturally inclined to exert high effort in order to maintain proficiency throughout the year, when the school starts the year in the proficient state. As a result, the school district can induce positive effort while offering teachers a smaller reward.

Figure A5 illustrates the non-monotonic effect of $\phi_{P|N}$ on the reward that the district should offer teachers. Figure A5a shows the case with no information asymmetry between the district and the teachers ($q = 0$). Similar to the value of the reward required to induce positive effort at $t = 0$ illustrated in Figure A3, there are two regions where the optimal reward is elevated: first, for near-zero $\phi_{P|N}$, and second, when $\phi_{P|N}$ is just large enough so that there is a high rate of false proficient results but the formative assessment results are still somewhat informative. In the first region, as $\phi_{P|N}$ approaches zero, the formative assessment almost always correctly assesses when the true state is not proficient, but underreports when the true state is proficient, leading teachers



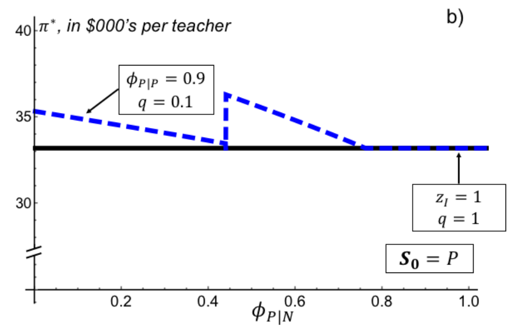


Figure A5 The optimal reward level π^*, in \$000's per teacher, as a function of the accuracy parameter $\phi_{P|N}$ for $S_0 = P$ and different values of $\phi_{P|P}$ when a) $q = 0$ and b) $q = 0.1$ compared to the reward under the interim assessment ($A_P = 0.72$; $A_N = 0.13$; $F = \$7,500$ per school; $B = \$700,000$ per school).

to overweight the possibility of being in the not-proficient state. Notice that for the second effect, there is a small region where the perceived probability of being in the not-proficient state is so high that the budget does not support the incentive level necessary to induce effort, and, therefore, the optimal reward drops to 0. Both effects decrease as $\phi_{P|N}$ increases. Ultimately, this leads to a lower optimal reward threshold than under the perfect information case, as the formative assessment carries little actionable information and teachers place greater weight on the initial state being proficient.

Figure A5b shows the case with small information asymmetry between the district and the teachers, i.e., the case where the district believes that there is small probability that teachers know the true intermediate state ($q = 0.1$). As in the case of $q = 0$, there remain regions of $\phi_{P|N}$ for which a higher incentive level is optimal. However, the non-zero probability that teachers may respond to the incentives as they would under perfect information results in the optimal reward level being bounded below by the reward under perfect information.

B1. Proofs of Analytical Results

LEMMA B1. *Define*

$$m(e_0,\mathbf{S}_0)=\begin{cases}1-\delta(e_0), & \text{if } \mathbf{S}_0=P,\\ \alpha(e_0), & \text{if } \mathbf{S}_0=N.\end{cases} \tag{B1}$$

Then,

$$Pr\left[h_1\left(e_0,\mathbf{S}_0\right)=\left(P,e_0,\mathbf{S}_0\right)\right]=m(e_0,\mathbf{S}_0)\phi_{P|P}+\left(1-m(e_0,\mathbf{S}_0)\right)\phi_{P|N}, \tag{B2}$$

and

$$Pr\left[h_2\left(e_1,\mathbf{S}_1\right)=P\right]=\left(1-\delta\left(e_1\right)\right)Pr\left[\beta_1=P|\mathbf{S}_1\right]+\alpha\left(e_1\right)\left(1-Pr\left[\beta_1=P|\mathbf{S}_1\right]\right), \tag{B3}$$

where

$$Pr\left[\beta_1=P|\mathbf{S}_1\right]=\begin{cases}\frac{\phi_{P|P}m(e_0,\mathbf{S}_0)}{\phi_{P|P}m(e_0,\mathbf{S}_0)+\phi_{P|N}(1-m(e_0,\mathbf{S}_0))}, & \text{if } \mathbf{S}_1=\left(P,e_0,\mathbf{S}_0\right),\\ \frac{\left(1-\phi_{P|P}\right)m(e_0,\mathbf{S}_0)}{\left(1-\phi_{P|P}\right)m(e_0,\mathbf{S}_0)+\left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0))}, & \text{if } \mathbf{S}_1=\left(N,e_0,\mathbf{S}_0\right).\end{cases} \tag{B4}$$

Proof of Lemma B1

When $z_I=0$ and $T=0$, the value of β_1 is not known with certainty, and at $t=1$, the result of the formative assessment, X_1, is revealed. Focusing on $X_1=P$, we have to consider two possible combinations for $\mathbf{S}_1$: (P,e_0,P) and (P,e_0,N). The probability of having the first combination is the sum of two probabilities: the one having $\beta_1=P$ and then generating $X_1=P$ $\left(\left(1-\delta\left(e_0\right)\right)\phi_{P|P}\right)$ and the one having $\beta_1=N$ and then generating $X_1=P$ $\left(\delta\left(e_0\right)\phi_{P|N}\right)$. Thus, we obtain the first line in (B2). The derivation of the second line in (B2) follows the same steps.

The analysis for $t=2$ involves two steps. First, (B3) connects the probability that $\beta_1=P$ to the probability that $\beta_2=P$ as well, accounting for transitions between $\beta_1=P$ and $\beta_1=N$ and $\beta_2=P$. Second, (B4) looks at four possible values of the state at $t=1$, $(X_1,e_0,\mathbf{S}_0)$: $\mathbf{S}_1=(P,e_0,P)$, $\mathbf{S}_1=(N,e_0,P)$, $\mathbf{S}_1=(P,e_0,N)$, and $\mathbf{S}_1=(N,e_0,N)$. For each of these states, (B1)-(B4) express the probability that $\beta_1=P$. Below we show the derivation of this probability for the case of $\mathbf{S}_1=(P,e_0,P)$; the derivations for the remaining cases follows the same steps. Using Bayes' rule, we have

$$\begin{aligned}&Pr\left[\beta_1=P|X_1=P,e_0,\mathbf{S}_0=P\right]\\ &=\frac{Pr\left[X_1=P|\beta_1=P,e_0,\mathbf{S}_0=P\right]Pr\left[\beta_1=P|e_0,\mathbf{S}_0=P\right]}{Pr\left[X_1=P|\beta_1=P,e_0,\mathbf{S}_0=P\right]Pr\left[\beta_1=P|e_0,\mathbf{S}_0=P\right]+Pr\left[X_1=P|\beta_1=N,e_0,\mathbf{S}_0=P\right]Pr\left[\beta_1=N|e_0,\mathbf{S}_0=P\right]}\\ &=\frac{\phi_{P|P}\left(1-\delta\left(e_0\right)\right)}{\phi_{P|P}\left(1-\delta\left(e_0\right)\right)+\phi_{P|N}\delta\left(e_0\right)}.\end{aligned} \tag{B5}$$

When $z_I=1$ or $T=1$, the probabilities are obtained by letting $\phi_{P|P}=1$ and $\phi_{P|N}=0$. In this case, β_1 is known exactly, and the response to the effort levels at both $t=0$ and $t=1$ is described by (1)-(2). □

Proof of Proposition 1

a) The maximization in the expression for the profit-to-go function for $t=1$ when $z_I=0$ and $T=0$ is carried over a concave function of e_1 for any value of $\mathbf{S}_1$. In particular, using (B3) and (16), this is represented by

$$J_1(\mathbf{S}_1) = \max_{e_1\geq 0}\left[\pi\left(\left(1-\delta\left(e_1\right)\right)Pr\left[\beta_1=P|\mathbf{S}_1\right]+\alpha\left(e_1\right)\left(1-Pr\left[\beta_1=P|\mathbf{S}_1\right]\right)\right)-\gamma e_1\right] \tag{B6}$$

$$= \max_{e_1\geq 0}\left[\pi\left(Pr\left[\beta_1=P|\mathbf{S}_1\right]\left(1-\delta\left(e_1\right)-\alpha\left(e_1\right)\right)+\alpha\left(e_1\right)\right)-\gamma e_1\right], \tag{B7}$$

with $\alpha\left(e_1\right)$ and $\delta\left(e_1\right)$ given by (1) and (2), respectively, and $Pr\left[\beta_1=P|\mathbf{S}_1\right]$ given by (B4). For clarity, we represent the expression under the maximization operator in (B7) by $f(e_1)$ for the remainder of this proof. Note that $f(e_1)$ is decreasing for $e_1>\lambda$.

Let

$$\hat{e}_1^*\left(\mathbf{S}_1\right)=\frac{e_1^*\left(\mathbf{S}_1\right)}{\lambda} \text{ and } \hat{J}_1\left(\mathbf{S}_1\right)=\frac{J_1\left(\mathbf{S}_1\right)}{\lambda\gamma}. \tag{B8}$$

Then, the scaled effort level maximizing (B7) is given by

$$\hat{e}_1^*\left(\mathbf{S}_1\right)=\begin{cases}0, & \text{if } 0\leq\hat{\pi}<\hat{\mathcal{P}}_1\left(\mathbf{S}_1\right),\\ 1, & \text{if } \hat{\mathcal{P}}_1\left(\mathbf{S}_1\right)\leq\hat{\pi},\end{cases} \tag{B9}$$

and the corresponding scaled profit-to-go function is

$$\hat{J}_1(\mathbf{S}_1)=\begin{cases}0, & \text{if } 0\leq\hat{\pi}<\hat{\mathcal{P}}_1\left(\mathbf{S}_1\right),\\ \hat{\pi}\left(Pr\left[\beta_1=P|\mathbf{S}_1\right]\left(A_P-A_N\right)+A_N\right)-1, & \text{if } \hat{\mathcal{P}}_1\left(\mathbf{S}_1\right)\leq\hat{\pi},\end{cases} \tag{B10}$$

where

$$\hat{\mathcal{P}}_1\left(\mathbf{S}_1\right)=\frac{1}{Pr\left[\beta_1=P|\mathbf{S}_1\right]\left(A_P-A_N\right)+A_N}. \tag{B11}$$

b) Under Assumptions 1 and 2, we have $Pr\left[\beta_1=P|\mathbf{S}_1=\left(P,e_0,\mathbf{S}_0\right)\right]\geq Pr\left[\beta_1=P|\mathbf{S}_1=\left(N,e_0,\mathbf{S}_0\right)\right]$, which can be shown using the expressions given in (B4) through simple algebraic transformations of this inequality. Therefore,

$$\hat{\mathcal{P}}_1\left(P,e_0,\mathbf{S}_0\right)\leq\hat{\mathcal{P}}_1\left(N,e_0,\mathbf{S}_0\right), \tag{B12}$$

so that, for all values of the scaled merit-based incentive,

$$\hat{e}_1^*\left(P,e_0,\mathbf{S}_0\right)\geq\hat{e}_1^*\left(N,e_0,\mathbf{S}_0\right). \tag{B13}$$

□

Proof of Proposition 2

In order to establish the statements of this Proposition, we will need the following result.

LEMMA B2.

$$\frac{A_N}{A_P}\leq\mathcal{A}_1\leq 1,\quad \mathcal{A}_1\leq\mathcal{A}_2. \tag{B14}$$

Proof of Lemma B2

From (30),

$$\mathcal{A}_1 = \frac{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_P\left(1-A_P\right)\left(1-\phi_{P|N}\right)}, \tag{B15}$$

and from (31),

$$\mathcal{A}_2 = \frac{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}. \tag{B16}$$

Under Assumptions 1 and 2, it is clear that $\mathcal{A}_1 \leq 1$, and that

$$\begin{aligned}\mathcal{A}_1 \geq \frac{A_N}{A_P} &\iff \frac{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_P\left(1-A_P\right)\left(1-\phi_{P|N}\right)} \geq \frac{A_N}{A_P}\\ &\iff A_P^2\left(A_P-A_N\right)\left(1-\phi_{P|P}\right) \geq 0.\end{aligned} \tag{B17}$$

Under the same assumptions, the following also holds:

$$\mathcal{A}_2 \geq \mathcal{A}_1 \iff A_P^2\left(1-\phi_{P|P}\right)+\left(1-A_P\right)\left(1-\phi_{P|N}\right)\left(A_P\left(1-\phi_{P|P}\right)+A_N\phi_{P|N}\right) \geq 0. \tag{B18}$$

The profit-to-go function for $t=0$ when $z_I=0$ and $T=0$ is given by (25). Expanding this gives

$$J_0(\mathbf{S}_0) = \max_{e_0\geq 0}\left[Pr\left[h_1\left(e_0,\mathbf{S}_0\right)=\left(N,e_0,\mathbf{S}_0\right)\right]J_1\left(N,e_0,\mathbf{S}_0\right)+Pr\left[h_1\left(e_0,\mathbf{S}_0\right)=\left(P,e_0,\mathbf{S}_0\right)\right]J_1\left(P,e_0,\mathbf{S}_0\right)-\gamma e_0\right], \tag{B19}$$

with $Pr\left[h_1\left(e_0,\mathbf{S}_0\right)\right]$ given in (B2) and $J_1\left(\mathbf{S}_1\right)$ given in (B10).

Now, the value of $J_1\left(\mathbf{S}_1\right)$ depends on $\hat{\mathcal{P}}_1\left(\mathbf{S}_1\right)$, where, from (23) and (24),

$$\hat{\mathcal{P}}_1\left(\mathbf{S}_1\right) = \begin{cases}\frac{\phi_{P|P}m(e_0,\mathbf{S}_0)+\phi_{P|N}(1-m(e_0,\mathbf{S}_0))}{\phi_{P|P}m(e_0,\mathbf{S}_0)A_P+\phi_{P|N}(1-m(e_0,\mathbf{S}_0))A_N}, & \text{if } \mathbf{S}_1=\left(P,e_0,\mathbf{S}_0\right),\\ \frac{\left(1-\phi_{P|P}\right)m(e_0,\mathbf{S}_0)+\left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0))}{\left(1-\phi_{P|P}\right)m(e_0,\mathbf{S}_0)A_P+\left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0))A_N}, & \text{if } \mathbf{S}_1=\left(N,e_0,\mathbf{S}_0\right).\end{cases} \tag{B20}$$

Using Assumptions 1 and 2, we can order the bounds on $\hat{\pi}$ from (B20):

$$\hat{\mathcal{P}}_1\left(P,e_0,\mathbf{S}_0\right) \leq \hat{\mathcal{P}}_1\left(N,e_0,\mathbf{S}_0\right). \tag{B21}$$

Furthermore, both terms in (B21) are decreasing functions of e_0 and are maximized when $e_0=0$ and minimized when $e_0=\lambda$. Therefore,

$$\frac{A_N\phi_{P|P}+\left(1-A_N\right)\phi_{P|N}}{A_N\left(A_P\phi_{P|P}+\left(1-A_N\right)\phi_{P|N}\right)} \leq \hat{\mathcal{P}}_1\left(P,e_0,N\right) \leq \frac{1}{A_N}, \tag{B22}$$

$$\frac{A_N\left(1-\phi_{P|P}\right)+\left(1-A_N\right)\left(1-\phi_{P|N}\right)}{A_N\left(A_P\left(1-\phi_{P|P}\right)+\left(1-A_N\right)\left(1-\phi_{P|N}\right)\right)} \leq \hat{\mathcal{P}}_1\left(N,e_0,N\right) \leq \frac{1}{A_N}, \tag{B23}$$

$$\frac{A_P\phi_{P|P}+\left(1-A_P\right)\phi_{P|N}}{A_P^2\phi_{P|P}+A_N\left(1-A_P\right)\phi_{P|N}} \leq \hat{\mathcal{P}}_1\left(P,e_0,P\right) \leq \frac{1}{A_N}, \tag{B24}$$

$$\frac{A_P\left(1-\phi_{P|P}\right)+\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)} \leq \hat{\mathcal{P}}_1\left(N,e_0,P\right) \leq \frac{1}{A_N}. \tag{B25}$$

Suppose that $\mathbf{S}_0 = N$. Then, we consider three regions of $\hat{\pi}$, where the bounds of each region are functions of e_0. In each region, we characterize the functional form of $J_0(N)$ and determine the optimal value of e_0.

For convenience, in the following analysis we represent the profit-to-go function in Region r by $J_0^r(\mathbf{S}_0)$, where $r = 1, 2, 3,$.

Region 1: $0 \le \hat{\pi} < \hat{\mathcal{P}}_1(P, e_0, N)$. In this case, (B19) becomes

$$J_0^1(N) = \max_{e_0 \ge 0} [-\gamma e_0]. \tag{B26}$$

Since this function is always decreasing in e_0, then $e_{0,1}^* = 0$ and $J_0^1(N) = 0$.

Region 2: $\hat{\mathcal{P}}_1(P, e_0, N) \le \hat{\pi} < \hat{\mathcal{P}}_1(N, e_0, N)$. In this case, (B19) becomes

$$J_0^2(N) = \max_{e_0 \ge 0} \left[\phi_{P|P}\alpha(e_0)(A_P\pi - \lambda\gamma) + \phi_{P|N}(1 - \alpha(e_0))(A_N\pi - \lambda\gamma) - \gamma e_0\right]. \tag{B27}$$

For clarity, represent the expression being maximized by $f(e_0)$. Then,

$$f'(e_0) = \begin{cases} \frac{A_N\phi_{P|P}}{\lambda}(A_P\pi - \lambda\gamma) - \frac{A_N\phi_{P|N}}{\lambda}(A_N\pi - \lambda\gamma) - \gamma, & \text{if } 0 \le e_0 \le \lambda, \\ -\gamma, & \text{if } \lambda \le e_0. \end{cases} \tag{B28}$$

This is clearly negative for $e_0 \ge \lambda$. Furthermore, $f'(e_0)$ is non-decreasing when $0 \le e_0 \le \lambda$ if and only if

$$\frac{A_N\phi_{P|P}}{\lambda}(A_P\pi - \lambda\gamma) - \frac{A_N\phi_{P|N}}{\lambda}(A_N\pi - \lambda\gamma) - \gamma \ge 0 \iff \hat{\pi} \ge \frac{1 + A_N\left(\phi_{P|P} - \phi_{P|N}\right)}{A_N\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}. \tag{B29}$$

Thus,

$$e_{0,2}^* = \begin{cases} 0, & \text{if } \hat{\pi} < \frac{1+A_N\left(\phi_{P|P}-\phi_{P|N}\right)}{A_N\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}, \\ \lambda, & \text{if } \frac{1+A_N\left(\phi_{P|P}-\phi_{P|N}\right)}{A_N\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)} \le \hat{\pi}, \end{cases} \tag{B30}$$

and

$$J_0^2(N) = \begin{cases} \phi_{P|N}(A_N\pi - \lambda\gamma), & \text{if } \hat{\pi} < \frac{1+A_N\left(\phi_{P|P}-\phi_{P|N}\right)}{A_N\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}, \\ \phi_{P|P}A_N(A_P\pi - \lambda\gamma) + \phi_{P|N}(1 - A_N)(A_N\pi - \lambda\gamma) - \lambda\gamma, & \text{if } \frac{1+A_N\left(\phi_{P|P}-\phi_{P|N}\right)}{A_N\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)} \le \hat{\pi}. \end{cases} \tag{B31}$$

From (B23), the maximum upper bound for $\hat{\pi}$ in this region is $\frac{1}{A_N}$. However,

$$\frac{1 + A_N\left(\phi_{P|P} - \phi_{P|N}\right)}{A_N\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} > \frac{1}{A_N}. \tag{B32}$$

Therefore, in this region, $e_0^* = 0$.

Region 3: $\hat{\mathcal{P}}_1(N, e_0, N) \le \hat{\pi}$. In this case, (B19) becomes

$$\begin{aligned} J_0^3(N) = & \max_{e_0 \ge 0} \Big[\big(1 - \alpha(e_0)\phi_{P|P} - (1 - \alpha(e_0))\phi_{P|N}\big) \\ & \left(\pi \left(\frac{\left(1 - \phi_{P|P}\right)\alpha(e_0)A_P + \left(1 - \phi_{P|N}\right)(1 - \alpha(e_0))A_N}{\left(1 - \phi_{P|P}\right)\alpha(e_0) + \left(1 - \phi_{P|N}\right)(1 - \alpha(e_0))} \right) - \lambda\gamma \right) \\ & + \left(\alpha(e_0)\phi_{P|P} + (1 - \alpha(e_0))\phi_{P|N}\right) \\ & \left(\pi \left(\frac{\phi_{P|P}\alpha(e_0)A_P + \phi_{P|N}(1 - \alpha(e_0))A_N}{\phi_{P|P}\alpha(e_0) + \phi_{P|N}(1 - \alpha(e_0))} \right) - \lambda\gamma \right) - \gamma e_0 \Big] \\ = & \max_{e_0 \ge 0} \left[\pi\left(\alpha(e_0)A_P + (1 - \alpha(e_0))A_N\right) - \lambda\gamma - \gamma e_0 \right]. \end{aligned} \tag{B33}$$

For clarity, represent the expression being maximized by $f(e_0)$. Then,

$$f'(e_0) = \begin{cases} A_N \frac{\pi}{\lambda}(A_P - A_N) - \gamma, & \text{if } 0 \le e_0 \le \lambda, \\ -\gamma, & \text{if } \lambda \le e_0. \end{cases} \tag{B34}$$

The function is strictly decreasing for $e_0 \ge \lambda$, and when $0 \le e_0 \le \lambda$, the function is non-decreasing if and only if

$$A_N \frac{\pi}{\lambda}(A_P - A_N) - \gamma \ge 0 \iff \hat{\pi} \ge \frac{1}{A_N(A_P - A_N)}. \tag{B35}$$

Thus,

$$e_{0,3}^* = \begin{cases} 0, & \text{if } \hat{\pi} < \frac{1}{A_N(A_P - A_N)}, \\ \lambda, & \text{if } \frac{1}{A_N(A_P - A_N)} \le \hat{\pi}, \end{cases} \tag{B36}$$

and the profit-to-go function is

$$J_0^3(N) = \begin{cases} A_N\pi - \lambda\gamma, & \text{if } \hat{\pi} < \frac{1}{A_N(A_P - A_N)}, \\ A_N\pi(1 + A_P - A_N) - 2\lambda\gamma, & \text{if } \frac{1}{A_N(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B37}$$

Combining the analysis of these three regions, we have that when $\mathbf{S}_0 = N$,

$$\hat{e}_0^* = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N(A_P - A_N)}, \\ 1, & \text{if } \frac{1}{A_N(A_P - A_N)} \le \hat{\pi}, \end{cases} \tag{B38}$$

and the corresponding profit-to-go function is

$$\hat{J}_0(N) = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N(A_P - A_N)}, \\ A_N(1 + A_P - A_N)\hat{\pi} - 2, & \text{if } \frac{1}{A_N(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B39}$$

Next, suppose that $\mathbf{S}_0 = P$. Again, we consider three regions for the value of $\hat{\pi}$. In each region, we characterize the functional form of $J_0(P)$ and determine the optimal value of e_0.

Region 1: $0 \le \hat{\pi} < \hat{\mathcal{P}}_1(P, e_0, P)$. In this case, (B19) becomes

$$J_0^1(P) = \max_{e_0 \ge 0} [-\gamma e_0]. \tag{B40}$$

Since this function is always decreasing in e_0, then

$$e_{0,1}^* = 0 \text{ and } J_0^1(P) = 0. \tag{B41}$$

Region 2: $\hat{\mathcal{P}}_1(P, e_0, P) \le \hat{\pi} < \hat{\mathcal{P}}_1(N, e_0, P)$. In this case, (B19) becomes

$$J_0^2(P) = \max_{e_0 \ge 0} \left[\phi_{P|P}(1-\delta(e_0))(A_P\pi - \lambda\gamma) + \phi_{P|N}\delta(e_0)(A_N\pi - \lambda\gamma) - \gamma e_0\right]. \tag{B42}$$

For clarity, represent the expression being maximized by $f(e_0)$. Then,

$$f'(e_0) = \begin{cases} \phi_{P|P}\frac{A_P}{\lambda}(A_P\pi - \lambda\gamma) - \phi_{P|N}\frac{A_P}{\lambda}(A_N\pi - \lambda\gamma) - \gamma, & \text{if } 0 \le e_0 \le \lambda, \\ -\gamma, & \text{if } \lambda \le e_0. \end{cases} \tag{B43}$$

This is negative for $e_0 \ge \lambda$ and non-decreasing when $0 \le e_0 \le \lambda$ if and only if

$$\begin{aligned} &\phi_{P|P}\frac{A_P}{\lambda}(A_P\pi - \lambda\gamma) - \phi_{P|N}\frac{A_P}{\lambda}(A_N\pi - \lambda\gamma) - \gamma \ge 0 \\ &\iff A_P\phi_{P|P}(A_P\hat{\pi} - 1) - A_P\phi_{P|N}(A_N\hat{\pi} - 1) \ge 1 \\ &\iff \hat{\pi}\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right) \ge \frac{1 + A_P\phi_{P|P} - A_P\phi_{P|N}}{A_P} \\ &\iff \hat{\pi} \ge \mathcal{R}_2, \end{aligned} \tag{B44}$$

where from (33),

$$\mathcal{R}_2 = \frac{1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}. \tag{B45}$$

Then, when $\hat{\pi}$ is in Region 2,

$$e_{0,2}^* = \begin{cases} 0, & \text{if } \hat{\pi} < \mathcal{R}_2, \\ \lambda, & \text{if } \mathcal{R}_2 \le \hat{\pi}, \end{cases} \tag{B46}$$

and the profit-to-go function is

$$J_0^2(P) = \begin{cases} \phi_{P|N}(A_N\pi - \lambda\gamma), & \text{if } \hat{\pi} < \mathcal{R}_2, \\ A_P\phi_{P|P}(A_P\pi - \lambda\gamma) + \phi_{P|N}(1 - A_P)(A_N\pi - \lambda\gamma) - \lambda\gamma, & \text{if } \mathcal{R}_2 \le \hat{\pi}. \end{cases} \tag{B47}$$

Consider the threshold value for $\hat{\pi}$ necessary to incentivize a positive effort level in (B46). We compare this to the maximum and minimum value of $\hat{\pi}$ in this region. First, the minimum value for the lower bound of this region of $\hat{\pi}$ is $\frac{A_P\phi_{P|P} + \phi_{P|N}(1-A_P)}{A_P\phi_{P|P}A_P + \phi_{P|N}(1-A_P)A_N}$. This is less than the threshold value for $\hat{\pi}$, since

$$\frac{A_P\phi_{P|P} + \phi_{P|N}(1 - A_P)}{A_P\phi_{P|P}A_P + \phi_{P|N}(1 - A_P)A_N} < \mathcal{R}_2 \iff A_P^2\phi_{P|P}\left(\phi_{P|N} - 1\right) < A_N\phi_{P|N}\left(1 - A_P\left(1 - \phi_{P|P}\right)\right), \tag{B48}$$

where the left-hand side of the inequality is negative and the right-hand side is positive under Assumptions 1 and 2.

From (B25), the maximum upper bound for $\hat{\pi}$ in this region is $\frac{1}{A_N}$. Comparing this maximum upper bound to the threshold value for $\hat{\pi}$ in (B47),

$$\mathcal{R}_2 \ge \frac{1}{A_N} \iff \frac{A_N}{A_P\phi_{P|P}} \ge A_P - A_N. \tag{B49}$$

Region 3: $\hat{\mathcal{P}}_1(N, e_0, P) \le \hat{\pi}$. In this case, (B19) becomes

$$\begin{aligned} J_0^3(P) = &\max_{e_0 \ge 0} \Big[\left(1 - (1-\delta(e_0))\,\phi_{P|P} - \delta(e_0)\,\phi_{P|N}\right) \\ &\left(\pi \left(\frac{\left(1-\phi_{P|P}\right)(1-\delta(e_0))\,A_P + \left(1-\phi_{P|N}\right)\delta(e_0) A_N}{\left(1-\phi_{P|P}\right)(1-\delta(e_0)) + \left(1-\phi_{P|N}\right)\delta(e_0)} \right) - \lambda\gamma \right) \\ &+ \left((1-\delta(e_0))\,\phi_{P|P} + \delta(e_0)\,\phi_{P|N}\right) \left(\pi \left(\frac{\phi_{P|P}(1-\delta(e_0))\,A_P + \phi_{P|N}\delta(e_0) A_N}{\phi_{P|P}(1-\delta(e_0)) + \phi_{P|N}\delta(e_0)} \right) - \lambda\gamma \right) - \gamma e_0 \Big] \\ = &\max_{e_0 \ge 0} \left[\left(\pi\left((1-\delta(e_0))\,A_P + \delta(e_0) A_N\right) - \lambda\gamma\right) - \gamma e_0 \right]. \end{aligned} \tag{B50}$$

For clarity, represent the expression being maximized by $f(e_0)$. Then,

$$f'(e_0) = \begin{cases} \frac{A_P \pi}{\lambda}(A_P - A_N) - \gamma, & \text{if } 0 \le e_0 \le \lambda, \\ -\gamma, & \text{if } \lambda \le e_0. \end{cases} \tag{B51}$$

The function is strictly decreasing for $e_0 \ge \lambda$, and when $0 \le e_0 \le \lambda$, the function is non-decreasing if and only if

$$\frac{A_P \pi}{\lambda}(A_P - A_N) - \gamma \ge 0 \iff \hat{\pi} \ge \frac{1}{A_P(A_P - A_N)}. \tag{B52}$$

Thus,

$$e_{0,3}^* = \begin{cases} 0, & \text{if } \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ \lambda, & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi}, \end{cases} \tag{B53}$$

and the corresponding profit-to-go function is

$$J_0^3(P) = \begin{cases} A_N \pi - \lambda\gamma, & \text{if } \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ \pi\left(A_N + A_P(A_P - A_N)\right) - 2\lambda\gamma, & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B54}$$

Whether the lower bound on $\hat{\pi}$ required for $e_0^* = \lambda$ is stronger than the lower bound on $\hat{\pi}$ for this region depends on the values of the parameters.

Using these results, we can consider four regions of $\hat{\pi}$, where the bounds are no longer a function of e_0:

- $0 \le \hat{\pi} < \frac{A_P \phi_{P|P} + (1-A_P)\phi_{P|N}}{A_P^2 \phi_{P|P} + A_N(1-A_P)\phi_{P|N}}$,
- $\frac{A_P \phi_{P|P} + (1-A_P)\phi_{P|N}}{A_P^2 \phi_{P|P} + A_N(1-A_P)\phi_{P|N}} \le \hat{\pi} < \mathcal{R}_1$,
- $\mathcal{R}_1 \le \hat{\pi} < \frac{1}{A_N}$,
- $\frac{1}{A_N} \le \hat{\pi} < \infty$.

Recall from (32) that

$$\mathcal{R}_1 = \frac{A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)}. \tag{B55}$$

Case 1: $0 \le \hat{\pi} < \frac{A_P\phi_{P|P}+(1-A_P)\phi_{P|N}}{A_P^2\phi_{P|P}+A_N(1-A_P)\phi_{P|N}}$. In this case, (B19) becomes

$$J_0(P) = J_0^1(P) = \max_{e_0 \ge 0} \left[-\gamma e_0\right], \tag{B56}$$

so that $e_0^* = 0$ and $J_0(P) = 0$, as given in (B41).

Case 2: $\frac{A_P\phi_{P|P}+(1-A_P)\phi_{P|N}}{A_P^2\phi_{P|P}+A_N(1-A_P)\phi_{P|N}} \le \hat{\pi} < \mathcal{R}_1$. In this case, (B19) becomes

$$J_0(P) = \begin{cases} J_0^1(P), & \text{if } 0 \le e_0^* < \frac{\lambda}{A_P}\left(1 - \frac{(A_P\hat{\pi}-1)\phi_{P|P}}{(A_P\hat{\pi}-1)\phi_{P|P}+(1-A_N\hat{\pi})\phi_{P|N}}\right), \\ J_0^2(P), & \text{if } \frac{\lambda}{A_P}\left(1 - \frac{(A_P\hat{\pi}-1)\phi_{P|P}}{(A_P\hat{\pi}-1)\phi_{P|P}+(1-A_N\hat{\pi})\phi_{P|N}}\right) \le e_0^*. \end{cases} \tag{B57}$$

From (B41),

$$e_{0,1}^* = 0 \text{ and } J_0^1(P) = 0, \tag{B58}$$

and, from (B46) and (B47),

$$e_{0,2}^* = \begin{cases} 0, & \text{if } \hat{\pi} < \mathcal{R}_2, \\ \lambda, & \text{if } \mathcal{R}_2 \le \hat{\pi}, \end{cases} \tag{B59}$$

and

$$J_0^2(P) = \begin{cases} \phi_{P|N}\left(A_N\pi - \lambda\gamma\right), & \text{if } \hat{\pi} < \mathcal{R}_2, \\ A_P\phi_{P|P}\left(A_P\pi - \lambda\gamma\right) + \phi_{P|N}\left(1-A_P\right)\left(A_N\pi - \lambda\gamma\right) - \lambda\gamma, & \text{if } \mathcal{R}_2 \le \hat{\pi}. \end{cases} \tag{B60}$$

Then, $e_0^* = \lambda$ if and only if $\mathcal{R}_2 \le \hat{\pi}$ and $\hat{\pi}$ satisfies the bounds for this case. Now, the threshold value for $\hat{\pi}$ necessary to ensure a positive effort level is greater than the lower bound on $\hat{\pi}$ in this case, since

$$\begin{aligned} \mathcal{R}_2 = \frac{1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} &\ge \frac{A_P\phi_{P|P} + \left(1-A_P\right)\phi_{P|N}}{A_P^2\phi_{P|P} + A_N\left(1-A_P\right)\phi_{P|N}} \\ \iff A_P\phi_{P|P}\left(A_P\left(1-\phi_{P|N}\right) + A_N\phi_{P|N}\right) &+ A_N\left(1-A_P\right)\phi_{P|N} \ge 0, \end{aligned} \tag{B61}$$

which clearly holds under Assumption 1 and 2.

Furthermore, the threshold value of $\hat{\pi}$ is less than the upper bound on $\hat{\pi}$ in this case if and only if

$$\begin{aligned} \mathcal{R}_2 < \mathcal{R}_1 &\iff \frac{1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} < \frac{A_P\left(1-\phi_{P|P}\right) + \left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right) + A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)} \\ &\iff \mathcal{A}_2 < A_P - A_N, \end{aligned} \tag{B62}$$

where from (31),

$$\mathcal{A}_2 = \frac{A_P^2\left(1-\phi_{P|P}\right) + A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}. \tag{B63}$$

Therefore, the optimal effort level in this case is determined by two subcases.

Subcase 1: If

$$\mathcal{A}_2 < A_P - A_N, \tag{B64}$$

then the optimal effort level is

$$e_0^* = \begin{cases} 0, & \text{if } \frac{A_P\phi_{P|P}+(1-A_P)\phi_{P|N}}{A_P^2\phi_{P|P}+A_N(1-A_P)\phi_{P|N}} \le \hat{\pi} < \mathcal{R}_2, \\ \lambda, & \text{if } \mathcal{R}_2 \le \hat{\pi} < \mathcal{R}_1, \end{cases} \tag{B65}$$

and the corresponding profit-to-go function is

$$J_0(P) = \begin{cases} 0, & \text{if } \frac{A_P\phi_{P|P}+(1-A_P)\phi_{P|N}}{A_P^2\phi_{P|P}+A_N(1-A_P)\phi_{P|N}} \le \hat{\pi} < \mathcal{R}_2, \\ A_P\phi_{P|P}\left(A_P\pi - \lambda\gamma\right) + \phi_{P|N}\left(1-A_P\right)\left(A_N\pi - \lambda\gamma\right) - \lambda\gamma, & \\ & \text{if } \mathcal{R}_2 \le \hat{\pi} < \frac{A_P\left(1-\phi_{P|P}\right)+(1-A_P)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_N(1-A_P)\left(1-\phi_{P|N}\right)}. \end{cases} \tag{B66}$$

Subcase 2: If

$$A_P - A_N \le \mathcal{A}_2, \tag{B67}$$

then the optimal effort level and the corresponding profit-to-go function are, respectively,

$$e_0^* = 0 \text{ and } J_0(P) = 0. \tag{B68}$$

Case 3: $\mathcal{R}_1 \le \hat{\pi} < \frac{1}{A_N}$. In this case, (B19) becomes

$$J_0(P) = \begin{cases} J_0^1(P), & \text{if } 0 \le e_0^* < \frac{\lambda}{A_P}\left(1 - \frac{(A_P\hat{\pi}-1)\phi_{P|P}}{(A_P\hat{\pi}-1)\phi_{P|P}+(1-A_N\hat{\pi})\phi_{P|N}}\right), \\ J_0^2(P), & \text{if } \frac{\lambda}{A_P}\left(1 - \frac{(A_P\hat{\pi}-1)\phi_{P|P}}{(A_P\hat{\pi}-1)\phi_{P|P}+(1-A_N\hat{\pi})\phi_{P|N}}\right) \le e_0^* < \frac{\lambda}{A_P}\left(1 - \frac{(A_P\hat{\pi}-1)\left(1-\phi_{P|P}\right)}{(A_P\hat{\pi}-1)\left(1-\phi_{P|P}\right)+(1-A_N\hat{\pi})\left(1-\phi_{P|N}\right)}\right), \\ J_0^3(P), & \text{if } \frac{\lambda}{A_P}\left(1 - \frac{(A_P\hat{\pi}-1)\left(1-\phi_{P|P}\right)}{(A_P\hat{\pi}-1)\left(1-\phi_{P|P}\right)+(1-A_N\hat{\pi})\left(1-\phi_{P|N}\right)}\right) \le e_0^*. \end{cases} \tag{B69}$$

The values of these profit-to-go functions and the associated optimal effort levels are as follows.

From (B41),

$$e_{0,1}^* = 0 \text{ and } J_0^1(P) = 0, \tag{B70}$$

and from (B46) and (B47),

$$e_{0,2}^* = \begin{cases} 0, & \text{if } \hat{\pi} < \mathcal{R}_2, \\ \lambda, & \text{if } \mathcal{R}_2 \le \hat{\pi}, \end{cases} \tag{B71}$$

and

$$J_0^2(P) = \begin{cases} \phi_{P|N}\left(A_N\pi - \lambda\gamma\right), & \text{if } \hat{\pi} < \mathcal{R}_2, \\ A_P\phi_{P|P}\left(A_P\pi - \lambda\gamma\right) + \phi_{P|N}\left(1-A_P\right)\left(A_N\pi - \lambda\gamma\right) - \lambda\gamma, & \text{if } \mathcal{R}_2 \le \hat{\pi}. \end{cases} \tag{B72}$$

Furthermore, from (B53) and (B54),

$$e_{0,3}^{*}=\begin{cases}0, & \text{if } \hat{\pi}<\frac{1}{A_P(A_P-A_N)}\\ \lambda, & \text{if } \frac{1}{A_P(A_P-A_N)}\leq\hat{\pi}\end{cases} \tag{B73}$$

and

$$J_0^3(P)=\begin{cases}A_N\pi-\lambda\gamma, & \text{if } \hat{\pi}<\frac{1}{A_P(A_P-A_N)}\\ \pi\left(A_N+A_P\left(A_P-A_N\right)\right)-2\lambda\gamma, & \text{if } \frac{1}{A_P(A_P-A_N)}\leq\hat{\pi}.\end{cases} \tag{B74}$$

Consider the threshold value for $\hat{\pi}$ necessary to ensure a positive effort level in (B53). This is greater than the lower bound on $\hat{\pi}$ in this case if and only if

$$\frac{1}{A_P\left(A_P-A_N\right)}\geq\mathcal{R}_1=\frac{A_P\left(1-\phi_{P|P}\right)+\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}\iff\mathcal{A}_1\geq A_P-A_N, \tag{B75}$$

where from (30),

$$\mathcal{A}_1=\frac{A_P^2\left(1-\phi_{P|P}\right)+A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_P\left(1-A_P\right)\left(1-\phi_{P|N}\right)}. \tag{B76}$$

Similarly, this is less than the upper bound on $\hat{\pi}$ in this case if and only if

$$\frac{1}{A_P\left(A_P-A_N\right)}<\frac{1}{A_N}\iff\frac{A_N}{A_P}<A_P-A_N. \tag{B77}$$

Combining this, we have three subcases.

Subcase 1: If

$$A_P-A_N\leq\frac{A_N}{A_P}, \tag{B78}$$

then the optimal effort level and the corresponding profit-to-go function are

$$e_0^{*}=0 \text{ and } J_0\left(P\right)=0. \tag{B79}$$

Subcase 2: If

$$\frac{A_N}{A_P}<A_P-A_N\leq\mathcal{A}_1, \tag{B80}$$

then the optimal effort level is

$$e_0^{*}=\begin{cases}0, & \text{if } \mathcal{R}_1\leq\hat{\pi}<\frac{1}{A_P(A_P-A_N)},\\ \lambda, & \text{if } \frac{1}{A_P(A_P-A_N)}\leq\hat{\pi}<\frac{1}{A_N},\end{cases} \tag{B81}$$

and the corresponding profit-to-go function is

$$J_0\left(P\right)=\begin{cases}0, & \text{if } \mathcal{R}_1\leq\hat{\pi}<\frac{1}{A_P(A_P-A_N)},\\ \pi\left(A_N+A_P\left(A_P-A_N\right)\right)-2\lambda\gamma, & \text{if } \frac{1}{A_P(A_P-A_N)}\leq\hat{\pi}<\frac{1}{A_N}.\end{cases} \tag{B82}$$

Subcase 3: If

$$\mathcal{A}_1 < A_P - A_N, \tag{B83}$$

then the optimal effort level and the corresponding profit-to-go function are

$$e_0^* = \lambda \text{ and } J_0(P) = \pi\left(A_N + A_P\left(A_P - A_N\right)\right) - 2\lambda\gamma. \tag{B84}$$

Case 4: $\frac{1}{A_N} \leq \hat{\pi} < \infty$. In this case, (B19) becomes

$$J_0(N) = J_0^3(P). \tag{B85}$$

Recall from (B53) and (B54) that

$$e_{0,3}^* = \begin{cases} 0, & \text{if } \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ \lambda, & \text{if } \frac{1}{A_P(A_P-A_N)} \leq \hat{\pi}, \end{cases} \tag{B86}$$

and the corresponding profit-to-go function is

$$J_0^3(P) = \begin{cases} A_N\pi - \lambda\gamma, & \text{if } \hat{\pi} < \frac{1}{A_P(A_P-A_N)}, \\ \pi\left(A_N + A_P\left(A_P - A_N\right)\right) - 2\lambda\gamma, & \text{if } \frac{1}{A_P(A_P-A_N)} \leq \hat{\pi}. \end{cases} \tag{B87}$$

Then, there are two subcases.

Subcase 1: If

$$\frac{1}{A_P\left(A_P - A_N\right)} \leq \frac{1}{A_N} \iff \frac{A_N}{A_P} \leq A_P - A_N, \tag{B88}$$

then

$$e_0^* = \lambda \text{ and } J_0(P) = \pi\left(A_N + A_P\left(A_P - A_N\right)\right) - 2\lambda\gamma. \tag{B89}$$

Subcase 2: If

$$\frac{1}{A_P\left(A_P - A_N\right)} > \frac{1}{A_N} \iff A_P - A_N < \frac{A_N}{A_P}, \tag{B90}$$

then

$$e_0^* = \begin{cases} 0, & \text{if } \frac{1}{A_N} \leq \hat{\pi} < \frac{1}{A_P(A_P-A_N)}, \\ \lambda, & \text{if } \frac{1}{A_P(A_P-A_N)} \leq \hat{\pi}, \end{cases} \tag{B91}$$

and the corresponding profit-to-go function is

$$J_0(P) = \begin{cases} A_N\pi - \lambda\gamma, & \text{if } \frac{1}{A_N} \leq \hat{\pi} < \frac{1}{A_P(A_P-A_N)}, \\ \pi\left(A_N + A_P\left(A_P - A_N\right)\right) - 2\lambda\gamma, & \text{if } \frac{1}{A_P(A_P-A_N)} \leq \hat{\pi}. \end{cases} \tag{B92}$$

Finally, we determine the optimal effort level and profit-to-go function for different regions of $A_P - A_N$. From Lemma B2, $\frac{A_N}{A_P} \leq \mathcal{A}_1 \leq \mathcal{A}_2$. Then, we consider four cases to determine the value of the optimal effort level and the optimal profit-to-go function.

Case 1: $(A_P, A_N) \in \mathcal{S}_1$ The optimal effort level is

$$\hat{e}_0^* = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ \lambda, & \text{if } \frac{1}{A_P(A_P - A_N)} \leq \hat{\pi}, \end{cases} \tag{B93}$$

and the corresponding profit-to-go function is

$$J_0(P) = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ \pi\left(A_N + A_P\left(A_P - A_N\right)\right) - 2\lambda\gamma, & \text{if } \frac{1}{A_P(A_P - A_N)} \leq \hat{\pi}. \end{cases} \tag{B94}$$

Case 2: $(A_P, A_N) \in \mathcal{S}_2$ The optimal effort level is

$$e_0^* = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ \lambda, & \text{if } \frac{1}{A_P(A_P - A_N)} \leq \hat{\pi}, \end{cases} \tag{B95}$$

and the corresponding profit-to-go function is

$$J_0\left(P\right) = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ \pi\left(A_N + A_P\left(A_P - A_N\right)\right) - 2\lambda\gamma, & \text{if } \frac{1}{A_P(A_P - A_N)} \leq \hat{\pi}. \end{cases} \tag{B96}$$

Case 3: $(A_P, A_N) \in \mathcal{S}_3$ The optimal effort level is

$$e_0^* = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_1, \\ \lambda, & \text{if } \mathcal{R}_1 \leq \hat{\pi}, \end{cases} \tag{B97}$$

and the corresponding profit-to-go function is

$$J_0(P) = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_1, \\ \pi\left(A_N + A_P\left(A_P - A_N\right)\right) - 2\lambda\gamma, & \text{if } \mathcal{R}_1 \leq \hat{\pi}. \end{cases} \tag{B98}$$

Case 4: $(A_P, A_N) \in \mathcal{S}_4$ The optimal effort level is

$$e_0^* = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_2, \\ \lambda, & \text{if } \mathcal{R}_2 \leq \hat{\pi}, \end{cases} \tag{B99}$$

and the corresponding profit-to-go function is

$$J_0(P) = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_2, \\ A_P\phi_{P|P}\left(A_P\pi - \lambda\gamma\right) + \phi_{P|N}\left(1 - A_P\right)\left(A_N\pi - \lambda\gamma\right) - \lambda\gamma, & \text{if } \mathcal{R}_2 \leq \hat{\pi} < \mathcal{R}_1, \\ \pi\left(A_N + A_P\left(A_P - A_N\right)\right) - 2\lambda\gamma, & \text{if } \mathcal{R}_1 \leq \hat{\pi}. \end{cases} \tag{B100}$$

□

LEMMA B3. *Recall that* $e_0^*(\mathbf{S}_0)$ *and* $e_1^*(\mathbf{S}_1)$ *are the optimal teachers' effort policies. Let* $\underline{e}_0^*(\mathbf{S}_0)$ *and* $\underline{e}_1^*(\mathbf{S}_1)$ *denote the policies in the special case where* $z = 1$ *or* $T = 1$.

Then, for the general case where the district chooses to rely on the formative assessment,

$$\begin{aligned} Pr^*[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = (1-q)\Big[&\left(1 - \delta\left(e_0^*\left(P\right)\right)\right)\left(1 - \delta\left(e_1^*\left(P, e_0^*\left(P\right), P\right)\right)\right)\phi_{P|P} \\ &+ \delta\left(e_0^*\left(P\right)\right)\alpha\left(e_1^*\left(P, e_0^*\left(P\right), P\right)\right)\phi_{P|N} \end{aligned}$$

$$
\begin{aligned}
&+\left(1-\delta\left(e_0^*(P)\right)\right)\left(1-\delta\left(e_1^*\left(N,e_0^*(P),P\right)\right)\right)\left(1-\phi_{P|P}\right)\\
&+\delta\left(e_0^*(P)\right)\alpha\left(e_1^*\left(N,e_0^*(P),P\right)\right)\left(1-\phi_{P|N}\right)\Big]\\
&+q\left[\left(1-\delta\left(\underline{e}_0^*(P)\right)\right)\left(1-\delta\left(\underline{e}_1^*\left(P,\underline{e}_0^*(P),P\right)\right)\right)\right.\\
&\left.+\delta\left(\underline{e}_0^*(P)\right)\alpha\left(\underline{e}_1^*\left(N,\underline{e}_0^*(P),P\right)\right)\right],
\end{aligned} \tag{B101}
$$

$$
\begin{aligned}
Pr^*[\mathbf{S}_2=P|\mathbf{S}_0=N]=(1-q)&\Big[\alpha\left(e_0^*(N)\right)\left(1-\delta\left(e_1^*\left(P,e_0^*(N),N\right)\right)\right)\phi_{P|P},\\
&+\left(1-\alpha\left(e_0^*(N)\right)\right)\alpha\left(e_1^*\left(P,e_0^*(N),N\right)\right)\phi_{P|N}\\
&+\alpha\left(e_0^*(N)\right)\left(1-\delta\left(e_1^*\left(N,e_0^*(N),N\right)\right)\right)\left(1-\phi_{P|P}\right)\\
&+\left(1-\alpha\left(e_0^*(N)\right)\right)\alpha\left(e_1^*\left(N,e_0^*(N),N\right)\right)\left(1-\phi_{P|N}\right)\Big]\\
&+q\left[\alpha\left(\underline{e}_0^*(N)\right)\left(1-\delta\left(\underline{e}_1^*\left(P,\underline{e}_0^*(N),N\right)\right)\right),\right.\\
&\left.+\left(1-\alpha\left(\underline{e}_0^*(N)\right)\right)\alpha\left(\underline{e}_1^*\left(N,\underline{e}_0^*(N),N\right)\right)\right].
\end{aligned} \tag{B102}
$$

The probability that the final state is proficient if the district relies on the interim assessment is obtained by letting $q=1$ in (B101) *and* (B102).

Proof of Lemma B3

We first determine the probability that the final state is proficient under the teachers' optimal effort decisions assuming no information asymmetry, $Pr^*[\mathbf{S}_2=P|\mathbf{S}_0,q=0]$. Using this result, we then calculate the probability that the final state is proficient for any q, $Pr^*[\mathbf{S}_2=P|\mathbf{S}_0]$.

When $\mathbf{S}_0=P$, we have

$$
Pr\left[\beta_1=P|\mathbf{S}_1\right]=\begin{cases}\frac{\phi_{P|P}\left(1-\delta\left(e_0^*(P)\right)\right)}{\phi_{P|P}\left(1-\delta\left(e_0^*(P)\right)\right)+\phi_{P|N}\left(\delta\left(e_0^*(P)\right)\right)}, & \text{if } \mathbf{S}_1=\left(P,e_0^*(P),\mathbf{S}_0\right),\\ \frac{\left(1-\phi_{P|P}\right)\left(1-\delta\left(e_0^*(P)\right)\right)}{\left(1-\phi_{P|P}\right)\left(1-\delta\left(e_0^*(P)\right)\right)+\left(1-\phi_{P|N}\right)\left(\delta\left(e_0^*(P)\right)\right)}, & \text{if } \mathbf{S}_1=\left(N,e_0^*(P),\mathbf{S}_0\right).\end{cases} \tag{B103}
$$

Then, with (B2) and (B3),

$$
\begin{aligned}
Pr^*[\mathbf{S}_2=P|\mathbf{S}_0=P,q=0]&=\left(1-\delta\left(e_1^*\left(X_1,e_0^*(P),P\right)\right)\right)Pr\left[\beta_1=P|\mathbf{S}_1\right]+\alpha\left(e_1^*\left(X_1,e_0^*(P),P\right)\right)\left(1-Pr\left[\beta_1=P|\mathbf{S}_1\right]\right)\\
&=Pr\left[\mathbf{S}_1=\left(P,e_0^*(P),P\right)\right]\left(\left(1-\delta\left(e_1^*\left(P,e_0^*(P),P\right)\right)\right)Pr\left[\beta_1=P|\mathbf{S}_1=\left(P,e_0^*(P),\mathbf{S}_0\right)\right]\right.\\
&\quad\left.+\alpha\left(e_1^*\left(P,e_0^*(P),P\right)\right)\left(1-Pr\left[\beta_1=P|\mathbf{S}_1=\left(P,e_0^*(P),\mathbf{S}_0\right)\right]\right)\right)\\
&\quad+Pr\left[\mathbf{S}_1=\left(N,e_0^*(P),P\right)\right]\left(\left(1-\delta\left(e_1^*\left(N,e_0^*(P),P\right)\right)\right)Pr\left[\beta_1=P|\mathbf{S}_1=\left(N,e_0^*(P),\mathbf{S}_0\right)\right]\right.\\
&\quad\left.+\alpha\left(e_1^*\left(N,e_0^*(P),P\right)\right)\left(1-Pr\left[\beta_1=P|\mathbf{S}_1=\left(N,e_0^*(P),\mathbf{S}_0\right)\right]\right)\right)\\
&=\left(1-\delta\left(e_0^*(P)\right)\right)\left(1-\delta\left(e_1^*\left(P,e_0^*(P),P\right)\right)\right)\phi_{P|P}\\
&\quad+\delta\left(e_0^*(P)\right)\alpha\left(e_1^*\left(P,e_0^*(P),P\right)\right)\phi_{P|N}\\
&\quad+\left(1-\delta\left(e_0^*(P)\right)\right)\left(1-\delta\left(e_1^*\left(N,e_0^*(P),P\right)\right)\right)\left(1-\phi_{P|P}\right)\\
&\quad+\delta\left(e_0^*(P)\right)\alpha\left(e_1^*\left(N,e_0^*(P),P\right)\right)\left(1-\phi_{P|N}\right).
\end{aligned} \tag{B104}
$$

To determine the probability that the final state is proficient, recall that the teachers rely on the assessment results with probability $1-q$ and know the true state with probability q. Then,

$$\begin{aligned}Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = &(1-q)\left[(1-\delta(e_0^*(P)))(1-\delta(e_1^*(P,e_0^*(P),P)))\phi_{P|P}\right.\\&+\delta(e_0^*(P))\alpha(e_1^*(P,e_0^*(P),P))\phi_{P|N}\\&+(1-\delta(e_0^*(P)))(1-\delta(e_1^*(N,e_0^*(P),P)))\left(1-\phi_{P|P}\right)\\&\left.+\delta(e_0^*(P))\alpha(e_1^*(N,e_0^*(P),P))\left(1-\phi_{P|N}\right)\right]\\&+q\left[(1-\delta(\underline{e}_0^*(P)))(1-\delta(\underline{e}_1^*(P,\underline{e}_0^*(P),P)))\right.\\&\left.+\delta(\underline{e}_0^*(P))\alpha(\underline{e}_1^*(N,\underline{e}_0^*(P),P))\right]\end{aligned}\tag{B105}$$

Following similar steps, when $\mathbf{S}_0 = N$, the probability that the final state is proficient assuming teachers rely only on the assessment result is

$$\begin{aligned}Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = N, q=0] = &\alpha(e_0^*(N))(1-\delta(e_1^*(P,e_0^*(N),N)))\phi_{P|P},\\&+(1-\alpha(e_0^*(N)))\alpha(e_1^*(P,e_0^*(N),N))\phi_{P|N}\\&+\alpha(e_0^*(N))(1-\delta(e_1^*(N,e_0^*(N),N)))\left(1-\phi_{P|P}\right)\\&+(1-\alpha(e_0^*(N)))\alpha(e_1^*(N,e_0^*(N),N))\left(1-\phi_{P|N}\right),\end{aligned}\tag{B106}$$

and the probability the final state is proficient for any q is

$$\begin{aligned}Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = N] = &(1-q)\left[\alpha(e_0^*(N))(1-\delta(e_1^*(P,e_0^*(N),N)))\phi_{P|P},\right.\\&+(1-\alpha(e_0^*(N)))\alpha(e_1^*(P,e_0^*(N),N))\phi_{P|N}\\&+\alpha(e_0^*(N))(1-\delta(e_1^*(N,e_0^*(N),N)))\left(1-\phi_{P|P}\right)\\&+(1-\alpha(e_0^*(N)))\alpha(e_1^*(N,e_0^*(N),N))\left(1-\phi_{P|N}\right)(\left.\right]\\&+q\left[\alpha(\underline{e}_0^*(N))(1-\delta(\underline{e}_1^*(P,\underline{e}_0^*(N),N))),\right.\\&\left.+(1-\alpha(\underline{e}_0^*(N)))\alpha(\underline{e}_1^*(N,\underline{e}_0^*(N),N))\right].\end{aligned}\tag{B107}$$

□

Proposition B1. *The probability that the state at $t=2$ is proficient $(\mathbf{S}_2 = P)$ can be described as follows.*

a) For $\mathbf{S}_0 = N$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = N] = \begin{cases}0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N},\\ A_N, & \text{if } \frac{1}{A_N} \le \hat{\pi} < \frac{1}{A_N(A_P - A_N)},\\ A_N(1+A_P-A_N), & \text{if } \frac{1}{A_N(A_P-A_N)} \le \hat{\pi}.\end{cases}\tag{B108}$$

b) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_1$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N}, \\ A_N, & \text{if } \frac{1}{A_N} \le \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ A_P^2 + A_N(1 - A_P), & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi}. \end{cases} \quad \text{(B109)}$$

c) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_2$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ (1-q)\left(A_P^2 + A_N(1 - A_P)\right), & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ A_P^2 + (1-q)\left(A_N(1 - A_P)\right), & \text{if } \frac{1+A_P}{A_P^2} \le \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N(1 - A_P), & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \quad \text{(B110)}$$

d) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, *and* $\left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right) \le 1 - \phi_{P|P}$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \mathcal{R}_1, \\ (1-q)\left(A_P^2 + A_N(1 - A_P)\right), & \text{if } \mathcal{R}_1 \le \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ A_P^2 + (1-q)\left(A_N(1 - A_P)\right), & \text{if } \frac{1+A_P}{A_P^2} \le \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N(1 - A_P), & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \quad \text{(B111)}$$

e) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, *and* $1 - \phi_{P|P} < \left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right)$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ qA_P^2, & \text{if } \frac{1+A_P}{A_P^2} \le \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + (1-q)\left(A_N(1 - A_P)\right), & \text{if } \mathcal{R}_1 \le \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N(1 - A_P), & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \quad \text{(B112)}$$

f) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and* $\left(\frac{A_P(A_P - A_N) - A_N}{A_P}\right)\phi_{P|N} \le 1 - \phi_{P|P}$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ qA_P^2, & \text{if } \frac{1+A_P}{A_P^2} \le \hat{\pi} < \mathcal{R}_2, \\ qA_P^2 + (1-q)\left(A_P^2\phi_{P|P} + A_N(1 - A_P)\phi_{P|N}\right), & \text{if } \mathcal{R}_2 \le \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + (1-q)A_N(1 - A_P), & \text{if } \mathcal{R}_1 \le \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N(1 - A_P), & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \quad \text{(B113)}$$

g) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and* $1 - \phi_{P|P} < \left(\frac{A_P(A_P - A_N) - A_N}{A_P}\right)\phi_{P|N}$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \mathcal{R}_2, \\ (1-q)\left(A_P^2\phi_{P|P} + A_N(1 - A_P)\phi_{P|N}\right), & \text{if } \mathcal{R}_2 \le \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ qA_P^2 + (1-q)\left(A_P^2\phi_{P|P} + A_N(1 - A_P)\phi_{P|N}\right), & \text{if } \frac{1+A_P}{A_P^2} \le \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + (1-q)A_N(1 - A_P), & \text{if } \mathcal{R}_1 \le \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N(1 - A_P), & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \quad \text{(B114)}$$

Proof of Proposition B1

First, consider the probability that the final state is proficient when the district relies on the formative assessment and there is no information asymmetry. This is (B101) and (B102) evaluated at $q = 0$.

Now, from (23) in Proposition 1, the minimum level of scaled merit-based incentive necessary to induce positive effort at $t = 1$ is

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \frac{1}{Pr[\beta_1 = P|\mathbf{S}_1](A_P - A_N) + A_N}, \tag{B115}$$

where, from (24),

$$Pr[\beta_1 = P|\mathbf{S}_1] = \begin{cases} \frac{\phi_{P|P} m(e_0,\mathbf{S}_0)}{\phi_{P|P} m(e_0,\mathbf{S}_0) + \phi_{P|N}(1-m(e_0,\mathbf{S}_0))}, & \text{if } \mathbf{S}_1 = (P, e_0, \mathbf{S}_0), \\ \frac{\left(1-\phi_{P|P}\right) m(e_0,\mathbf{S}_0)}{\left(1-\phi_{P|P}\right) m(e_0,\mathbf{S}_0) + \left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0))}, & \text{if } \mathbf{S}_1 = (N, e_0, \mathbf{S}_0). \end{cases} \tag{B116}$$

Therefore,

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \begin{cases} \frac{\phi_{P|P} m(e_0,\mathbf{S}_0) + \phi_{P|N}(1-m(e_0,\mathbf{S}_0))}{\phi_{P|P} m(e_0,\mathbf{S}_0) A_P + \phi_{P|N}(1-m(e_0,\mathbf{S}_0)) A_N}, & \text{if } \mathbf{S}_1 = (P, e_0, \mathbf{S}_0), \\ \frac{\left(1-\phi_{P|P}\right) m(e_0,\mathbf{S}_0) + \left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0))}{\left(1-\phi_{P|P}\right) m(e_0,\mathbf{S}_0) A_P + \left(1-\phi_{P|N}\right)(1-m(e_0,\mathbf{S}_0)) A_N}, & \text{if } \mathbf{S}_1 = (N, e_0, \mathbf{S}_0), \end{cases} \tag{B117}$$

and, under Assumptions 1 and 2,

$$0 \le \hat{\mathcal{P}}_1(P, e_0, \mathbf{S}_0) \le \hat{\mathcal{P}}_1(N, e_0, \mathbf{S}_0) \le \frac{1}{A_N}. \tag{B118}$$

Suppose $\mathbf{S}_0 = N$. From Proposition 2,

$$\hat{e}_0^*(N) = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N(A_P - A_N)}, \\ 1, & \text{if } \frac{1}{A_N(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B119}$$

Then, plugging (B119) into (B117),

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \begin{cases} \frac{1}{A_N}, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N(A_P - A_N)}, \\ \frac{1}{A_N}\left(\frac{A_N \phi_{P|P} + (1-A_N)\phi_{P|N}}{A_P \phi_{P|P} + (1-A_N)\phi_{P|N}}\right), & \text{if } \frac{1}{A_N(A_P - A_N)} \le \hat{\pi} \text{ and } \mathbf{S}_1 = (P, e_0, \mathbf{S}_0), \\ \frac{1}{A_N}\left(\frac{A_N\left(1-\phi_{P|P}\right) + (1-A_N)\left(1-\phi_{P|N}\right)}{A_P\left(1-\phi_{P|P}\right) + (1-A_N)\left(1-\phi_{P|N}\right)}\right), & \text{if } \frac{1}{A_N(A_P - A_N)} \le \hat{\pi} \text{ and } \mathbf{S}_1 = (N, e_0, \mathbf{S}_0). \end{cases} \tag{B120}$$

Therefore, from (22),

$$\hat{e}_1^*(P, e_0^*, N) = \hat{e}_1^*(N, e_0^*, N) = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N}, \\ 1, & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \tag{B121}$$

Plugging (B121) into (B102), the school will be in the proficient state at $t = 2$ when $q = 0$ with probability

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = N] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N}, \\ A_N, & \text{if } \frac{1}{A_N} \le \hat{\pi} < \frac{1}{A_N(A_P - A_N)}, \\ A_N(1 + A_P - A_N), & \text{if } \frac{1}{A_N(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B122}$$

Since this value does not depend on $\phi_{P|P}$ or $\phi_{P|N}$, the probability of achieving proficiency does not depend on the accuracy with which teachers know the intermediate state. Therefore, (B122) characterizes the probability of achieving proficiency for all $q \in [0,1]$. Furthermore, it is clear that $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = N]$ is non-decreasing in $\hat{\pi}$ and bounded below by 0 and above by 1 under Assumption 1.

Suppose $\mathbf{S}_0 = P$. Then, we consider two cases depending on the values of the parameters A_P, A_N, $\phi_{P|P}$, and $\phi_{P|N}$.

Case 1: If $(A_P, A_N) \in \mathcal{S}_1$, from Proposition 2, the optimal effort level at $t=0$ is

$$\hat{e}_0^* = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P-A_N)}, \\ 1, & \text{if } \frac{1}{A_P(A_P-A_N)} \le \hat{\pi}. \end{cases} \tag{B123}$$

Then, plugging this into (B117),

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \begin{cases} \frac{1}{A_N}, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P-A_N)}, \\ \frac{A_P\phi_{P|P}+(1-A_P)\phi_{P|N}}{A_P^2\phi_{P|P}+A_N(1-A_P)\phi_{P|N}}, & \text{if } \frac{1}{A_P(A_P-A_N)} \le \hat{\pi} \text{ and } \mathbf{S}_1 = (P, e_0, \mathbf{S}_0), \\ \frac{A_P\left(1-\phi_{P|P}\right)+(1-A_P)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right)+A_N(1-A_P)\left(1-\phi_{P|N}\right)}, & \text{if } \frac{1}{A_P(A_P-A_N)} \le \hat{\pi} \text{ and } \mathbf{S}_1 = (N, e_0, \mathbf{S}_0). \end{cases} \tag{B124}$$

In this case,

$$\frac{1}{A_N} \le \frac{1}{A_P(A_P-A_N)}. \tag{B125}$$

Then, the optimal effort level at $t=1$ is

$$\hat{e}_1^*(P, e_0^*, P) = e_1^*(N, e_0^*, P) = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N}, \\ 1, & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \tag{B126}$$

Plugging this into (B101) gives

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_N}, \\ A_N, & \text{if } \frac{1}{A_N} \le \hat{\pi} < \frac{1}{A_P(A_P-A_N)}, \\ A_P^2 + A_N(1-A_P), & \text{if } \frac{1}{A_P(A_P-A_N)} \le \hat{\pi}. \end{cases} \tag{B127}$$

Again, this value does not depend on $\phi_{P|P}$ or $\phi_{P|N}$, so the probability of achieving proficiency does not depend on the accuracy with which teachers know the intermediate state. Therefore, (B127) represents the probability of achieving proficiency for all $q \in [0,1]$. Furthermore, it is clear that $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P]$ is non-decreasing in $\hat{\pi}$ and bounded below by 0 and above by 1 under Assumption 1.

Case 2: For $\frac{A_N}{A_P} < A_P - A_N$, we again calculate the probability that the final state is proficient when $q=0$. Using these results, we then calculate the district's evaluation of the probability of achieving proficiency when $q=1$. We consider three subcases based on the parameter values.

Subcase 1: If $(A_P, A_N) \in \mathcal{S}_2$, from Proposition 2, the optimal effort level at $t=0$ is

$$\hat{e}_0^* = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ 1, & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B128}$$

Then, plugging this into (B117),

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \begin{cases} \frac{1}{A_N}, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ \frac{A_P \phi_{P|P} + (1-A_P)\phi_{P|N}}{A_P^2 \phi_{P|P} + A_N(1-A_P)\phi_{P|N}}, & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi} \text{ and } \mathbf{S}_1 = (P, e_0, \mathbf{S}_0), \\ \mathcal{R}_1, & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi} \text{ and } \mathbf{S}_1 = (N, e_0, \mathbf{S}_0), \end{cases} \tag{B129}$$

where, from (32),

$$\mathcal{R}_1 = \frac{A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)}. \tag{B130}$$

Now,

$$\frac{1}{A_P(A_P - A_N)} < \frac{1}{A_N}, \tag{B131}$$

and from (B75)

$$\frac{1}{A_P(A_P - A_N)} \ge \mathcal{R}_1 \iff \mathcal{A}_1 \ge A_P - A_N, \tag{B132}$$

which is a condition for this case. Recall from (30)

$$\mathcal{A}_1 = \frac{A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right) + A_P(1-A_P)\left(1-\phi_{P|N}\right)}. \tag{B133}$$

Furthermore, from Proposition 1,

$$\hat{\mathcal{P}}_1(P, e_0, \mathbf{S}_0) \le \hat{\mathcal{P}}_1(N, e_0, \mathbf{S}_0). \tag{B134}$$

Then, the optimal effort level at $t=1$ is

$$\hat{e}_1^*(P, e_0^*, P) = \hat{e}_1^*(N, e_0^*, P)^* = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ 1, & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B135}$$

Plugging this into (B101) when $q=0$ gives

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ A_P^2 + A_N(1-A_P), & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B136}$$

It is clear that $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P]$ is non-decreasing in $\hat{\pi}$ and bounded below by 0 and above by 1 under Assumption 1.

Subcase 2: If $(A_P, A_N) \in \mathcal{S}_3$, from Proposition 2, the optimal effort level at $t = 1$ is

$$\hat{e}_0^* = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_1, \\ 1, & \text{if } \mathcal{R}_1 \leq \hat{\pi}. \end{cases} \tag{B137}$$

Plugging this into (B117),

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \begin{cases} \frac{1}{A_N}, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_1, \\ \frac{A_P\phi_{P|P} + (1-A_P)\phi_{P|N}}{A_P^2\phi_{P|P} + A_N(1-A_P)\phi_{P|N}}, & \text{if } \mathcal{R}_1 \leq \hat{\pi} \text{ and } \mathbf{S}_1 = (P, e_0, \mathbf{S}_0), \\ \mathcal{R}_1, & \text{if } \mathcal{R}_1 \leq \hat{\pi} \text{ and } \mathbf{S}_1 = (N, e_0, \mathbf{S}_0). \end{cases} \tag{B138}$$

Now,

$$\mathcal{R}_1 \leq \frac{1}{A_N}, \tag{B139}$$

and, from Proposition 1,

$$\hat{\mathcal{P}}_1(P, e_0, \mathbf{S}_0) \leq \hat{\mathcal{P}}_1(N, e_0, \mathbf{S}_0). \tag{B140}$$

Then, the optimal effort level at $t = 2$ is

$$\hat{e}_1^*(P, e_0^*, P) = \hat{e}_1^*(N, e_0^*, P) = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_1, \\ 1, & \text{if } \mathcal{R}_1 \leq \hat{\pi}. \end{cases} \tag{B141}$$

Plugging this into (B101) when $q = 0$ gives

$$Pr^*[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + A_N(1 - A_P), & \text{if } \mathcal{R}_1 \leq \hat{\pi}. \end{cases} \tag{B142}$$

In the expression above, it is clear that $Pr^*[\mathbf{S}_2 = P | \mathbf{S}_0 = P]$ is non-decreasing in $\hat{\pi}$ and bounded below by 0 and above by 1 under Assumption 1.

Subcase 3: If $(A_P, A_N) \in \mathcal{S}_4$, from Proposition 2, the optimal effort level at $t = 0$ is

$$e_0^* = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_2, \\ \lambda, & \text{if } \mathcal{R}_2 \leq \hat{\pi}. \end{cases} \tag{B143}$$

Then, plugging this into (B117),

$$\hat{\mathcal{P}}_1(\mathbf{S}_1) = \begin{cases} \frac{1}{A_N}, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_2, \\ \frac{A_P\phi_{P|P} + (1-A_P)\phi_{P|N}}{A_P^2\phi_{P|P} + A_N(1-A_P)\phi_{P|N}}, & \text{if } \mathcal{R}_2 \leq \hat{\pi} \text{ and } \mathbf{S}_1 = (P, e_0, \mathbf{S}_0), \\ \mathcal{R}_1, & \text{if } \mathcal{R}_2 \leq \hat{\pi} \text{ and } \mathbf{S}_1 = (N, e_0, \mathbf{S}_0). \end{cases} \tag{B144}$$

From (B61),

$$\mathcal{R}_2 \geq \frac{A_P\phi_{P|P} + (1 - A_P)\phi_{P|N}}{A_P^2\phi_{P|P} + A_N(1 - A_P)\phi_{P|N}} \tag{B145}$$

always holds under Assumptions 1 and 2, and from (B62),

$$\mathcal{R}_2 < \mathcal{R}_1 \iff \mathcal{A}_2 < A_P - A_N, \tag{B146}$$

which is a condition for this case. Then, the optimal effort level at $t=1$ is

$$e_1^*(P, e_0^*, P) = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \mathcal{R}_2, \\ \lambda, & \text{if } \mathcal{R}_2 \le \hat{\pi}, \end{cases} \tag{B147}$$

and

$$e_1^*(N, e_0^*, P) = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \mathcal{R}_1, \\ \lambda, & \text{if } \mathcal{R}_1 \le \hat{\pi}, \end{cases} \tag{B148}$$

and, from (B101),

$$Pr^*[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \mathcal{R}_2, \\ A_P^2 \phi_{P|P} + A_N (1 - A_P) \phi_{P|N}, & \text{if } \mathcal{R}_2 \le \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + A_N (1 - A_P), & \text{if } \mathcal{R}_1 \le \hat{\pi}. \end{cases} \tag{B149}$$

In the expression above, it is clear that $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P]$ is non-decreasing in $\hat{\pi}$ and bounded below by 0 and above by 1 under Assumptions 1 and 2.

Using the results from the subcases, we calculate the probability that the system is in the proficient state when $q = 1$. Note that the teachers' response under perfect information is identical to the response under a perfectly accurate formative assessment ($\phi_{P|P} = 1, \phi_{P|N} = 0$). In that case, $\mathcal{A}_1 = \mathcal{A}_2 = \frac{A_N}{A_P}$, so Subcases 1 and 2 are never feasible. Furthermore, $\mathcal{R}_1 = \frac{1}{A_N}$ and $\mathcal{R}_2 = \frac{1+A_P}{A_P^2}$.

Therefore, in the case where there is maximum information asymmetry between the district and teachers ($q=1$) and $\frac{A_N}{A_P} < A_P - A_N$,

$$Pr^*[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ A_P^2, & \text{if } \frac{1+A_P}{A_P^2} \le \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N (1 - A_P), & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \tag{B150}$$

Finally, using the results from the subcases and (B150), we can characterize the the probability of achieving proficiency in the final state for any value of q for each subcase in Case 2.

Subcase 1: $(A_P, A_N) \in \mathcal{S}_2$. If $q = 0$, from (B136),

$$Pr^*[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ A_P^2 + A_N (1 - A_P), & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi}. \end{cases} \tag{B151}$$

Now,

$$\frac{1}{A_P (A_P - A_N)} < \frac{1 + A_P}{A_P^2} \iff \frac{A_N}{A_P} < A_P - A_N, \tag{B152}$$

which always holds in this case.

Therefore, for any q,

$$Pr^*[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{\pi} < \frac{1}{A_P(A_P - A_N)}, \\ (1-q)\left(A_P^2 + A_N (1 - A_P)\right), & \text{if } \frac{1}{A_P(A_P - A_N)} \le \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ A_P^2 + (1-q)\left(A_N (1 - A_P)\right), & \text{if } \frac{1+A_P}{A_P^2} \le \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N (1 - A_P), & \text{if } \frac{1}{A_N} \le \hat{\pi}. \end{cases} \tag{B153}$$

Subcase 2: $(A_P, A_N) \in \mathcal{S}_3$. If $q = 0$, from (B142),

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_1 \\ A_P^2 + A_N(1 - A_P), & \text{if } \mathcal{R}_1 \leq \hat{\pi}, \end{cases} \tag{B154}$$

where, from (32),

$$\mathcal{R}_1 = \frac{A_P\left(1 - \phi_{P|P}\right) + (1 - A_P)\left(1 - \phi_{P|N}\right)}{A_P^2\left(1 - \phi_{P|P}\right) + A_N(1 - A_P)\left(1 - \phi_{P|N}\right)}. \tag{B155}$$

First, note that

$$\mathcal{R}_1 \leq \frac{1}{A_N} \iff A_N \leq A_P, \tag{B156}$$

which holds under Assumption 1.

Furthermore,

$$\mathcal{R}_1 \leq \frac{1 + A_P}{A_P^2} \iff \left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right) \leq 1 - \phi_{P|P}. \tag{B157}$$

Therefore, we must consider two possibilities. If $\left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right) \leq 1 - \phi_{P|P}$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_1, \\ (1 - q)\left(A_P^2 + A_N(1 - A_P)\right), & \text{if } \mathcal{R}_1 \leq \hat{\pi} < \frac{1 + A_P}{A_P^2}, \\ A_P^2 + (1 - q)\left(A_N(1 - A_P)\right), & \text{if } \frac{1 + A_P}{A_P^2} \leq \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N(1 - A_P), & \text{if } \frac{1}{A_N} \leq \hat{\pi}. \end{cases} \tag{B158}$$

Finally, if $1 - \phi_{P|P} < \left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right)$,

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \frac{1 + A_P}{A_P^2}, \\ qA_P^2, & \text{if } \frac{1 + A_P}{A_P^2} \leq \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + (1 - q)\left(A_N(1 - A_P)\right), & \text{if } \mathcal{R}_1 \leq \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N(1 - A_P), & \text{if } \frac{1}{A_N} \leq \hat{\pi}. \end{cases} \tag{B159}$$

Subcase 3: $(A_P, A_N) \in \mathcal{S}_4$. If $q = 0$, from (B149),

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_2, \\ A_P^2\phi_{P|P} + A_N(1 - A_P)\phi_{P|N}, & \text{if } \mathcal{R}_2 \leq \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + A_N(1 - A_P), & \text{if } \mathcal{R}_1 \leq \hat{\pi}. \end{cases} \tag{B160}$$

First, recall from (B156) that $\mathcal{R}_1 \leq \frac{1}{A_N}$. Furthermore, when $(A_P, A_N) \in \mathcal{S}_4$, as defined in (37), then $\mathcal{R}_1 > \frac{1 + A_P}{A_P^2}$. To see this, first note that

$$\mathcal{R}_1 > \frac{1 + A_P}{A_P^2} \iff \frac{(1 - A_P)\left(A_P^2 - A_N(1 + A_P)\right)\left(1 - \phi_{P|N}\right)}{A_P^2} > 1 - \phi_{P|P} \tag{B161}$$

and

$$(A_P, A_N) \in \mathcal{S}_4 \iff \frac{A_P^2\left(1-\phi_{P|P}\right) + A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} < A_P - A_N$$
$$\iff 1-\phi_{P|P} < \frac{\left(1-A_P\right)\left(\left(A_P-A_N\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right) - A_N\right)\left(1-\phi_{P|N}\right)}{A_P^2}. \tag{B162}$$

Now,

$$\frac{\left(1-A_P\right)\left(A_P^2 - A_N\left(1+A_P\right)\right)\left(1-\phi_{P|N}\right)}{A_P^2} \geq \frac{\left(1-A_P\right)\left(\left(A_P-A_N\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right) - A_N\right)\left(1-\phi_{P|N}\right)}{A_P^2}$$
$$\iff A_P \geq A_P\phi_{P|P} - A_N\phi_{P|N}, \tag{B163}$$

which always holds under Assumptions 1 and 2. Therefore,

$$(A_P, A_N) \in \mathcal{S}_4 \Rightarrow \mathcal{R}_1 > \frac{1+A_P}{A_P^2}. \tag{B164}$$

Then, there are two possibilities to consider: $\frac{1+A_P}{A_P^2} \leq \mathcal{R}_2$ and $\mathcal{R}_2 < \frac{1+A_P}{A_P^2}$, where from (33)

$$\mathcal{R}_2 = \frac{1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}. \tag{B165}$$

Now,

$$\frac{1+A_P}{A_P^2} \leq \mathcal{R}_2 \iff \frac{\phi_{P|N}}{1-\phi_{P|P}} \leq \frac{A_P}{A_P^2 - A_NA_P - A_N}$$
$$\iff \left(\frac{A_P\left(A_P-A_N\right) - A_N}{A_P}\right)\phi_{P|N} \leq 1-\phi_{P|P}, \tag{B166}$$

where we use that $A_P - A_N > \frac{A_N}{A_P}$, since $(A_P, A_N) \in \mathcal{S}_4$.

Therefore, if

$$\left(\frac{A_P\left(A_P-A_N\right) - A_N}{A_P}\right)\phi_{P|N} \leq 1-\phi_{P|P}, \tag{B167}$$

then

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ qA_P^2, & \frac{1+A_P}{A_P^2} \leq \hat{\pi} < \mathcal{R}_2, \\ qA_P^2 + (1-q)\left(A_P^2\phi_{P|P} + A_N\left(1-A_P\right)\phi_{P|N}\right), & \text{if } \mathcal{R}_2 \leq \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + (1-q)A_N\left(1-A_P\right), & \text{if } \mathcal{R}_1 \leq \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N\left(1-A_P\right), & \text{if } \frac{1}{A_N} \leq \hat{\pi}. \end{cases} \tag{B168}$$

Otherwise, if

$$1-\phi_{P|P} < \left(\frac{A_P\left(A_P-A_N\right) - A_N}{A_P}\right)\phi_{P|N}, \tag{B169}$$

then

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \mathcal{R}_2, \\ (1-q)\left(A_P^2 \phi_{P|P} + A_N (1-A_P) \phi_{P|N}\right), & \text{if } \mathcal{R}_2 \leq \hat{\pi} < \frac{1+A_P}{A_P^2}, \\ qA_P^2 + (1-q)\left(A_P^2 \phi_{P|P} + A_N (1-A_P) \phi_{P|N}\right), & \text{if } \frac{1+A_P}{A_P^2} \leq \hat{\pi} < \mathcal{R}_1, \\ A_P^2 + (1-q) A_N (1-A_P), & \text{if } \mathcal{R}_1 \leq \hat{\pi} < \frac{1}{A_N}, \\ A_P^2 + A_N (1-A_P), & \text{if } \frac{1}{A_N} \leq \hat{\pi}. \end{cases} \tag{B170}$$

It is straightforward to show that, in all of these cases, $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P]$ and $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = P]$ are non-decreasing in $\hat{\pi}$ and are always between 0 and 1.

□

Proof of Proposition 3

When the district relies on the formative assessment ($z_I = 0$), the school district's maximization problem is as given in (40)-(41):

$$\max_{\pi \geq 0} Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] \tag{B171}$$

$$\text{s.t. } \pi Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0] \leq B, \tag{B172}$$

with $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ given in (B108)-(B114).

In the proof of Proposition B1, we show that $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is a non-decreasing step function of $\hat{\pi}$. Therefore, the expression on the left-hand side of the district's constraint (41) is an increasing function of $\hat{\pi}$, and we must consider the value of the objective function (40) at each of the endpoints of each interval of $\hat{\pi}$ that corresponds to a "step." We assume that if $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is constant over a region of $\hat{\pi}$ and any value of $\hat{\pi}$ in that region is optimal, the district will choose the smallest value of $\hat{\pi}$ in that region.

To determine the optimal merit-based incentive $\hat{\pi}^*$ and the corresponding probability that the final state is proficient, we must consider the several cases stated in Proposition B1 that determine the characterization of $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$.

We begin with the case where $\mathbf{S}_0 = N$ and $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ is given by (B108). Then, $\hat{\pi}^* = \frac{1}{A_N(A_P - A_N)}$ if

$$\frac{\lambda\gamma}{A_N(A_P - A_N)}(A_N(1 + A_P - A_N)) \leq B \iff 1 + \frac{1}{A_P - A_N} \leq \hat{B}. \tag{B173}$$

Similarly, $\hat{\pi}^* = \frac{1}{A_N}$ if

$$\frac{\lambda\gamma}{A_N}(A_N) \leq B < \frac{\lambda\gamma}{A_N(A_P - A_N)}(A_N(1 + A_P - A_N)) \iff 1 \leq \hat{B} < 1 + \frac{1}{A_P - A_N}. \tag{B174}$$

These results are stated in (47). We follow similar steps for the cases where $\mathbf{S}_0 = P$, using the functional forms for $Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0]$ given by (B109)-(B114).

□

Lemma B4. *The probability that the final state is proficient under the optimal incentive levels (47)-(53) when $q<1$ and $z_I=0$ can be expressed as follows.*

a) For $\mathbf{S}_0=N$,

$$Pr^*_{z_I}[\mathbf{S}_2=P|\mathbf{S}_0=N]=\begin{cases}0, & \text{if } 0\le\hat{B}<1,\\ A_N, & \text{if } 1\le\hat{B}<1+\frac{1}{A_P-A_N},\\ A_N\left(1+A_P-A_N\right), & \text{if } 1+\frac{1}{A_P-A_N}\le\hat{B}.\end{cases}\tag{B175}$$

b) For $\mathbf{S}_0=P$ *and* $(A_P,A_N)\in\mathcal{S}_1$,

$$Pr^*_{z_I}[\mathbf{S}_2=P|\mathbf{S}_0=P]=\begin{cases}0, & \text{if } 0\le\hat{B}<1,\\ A_N, & \text{if } 1\le\hat{B}<1+\frac{A_N}{A_P(A_P-A_N)},\\ A_P^2+A_N\left(1-A_P\right), & \text{if } 1+\frac{A_N}{A_P(A_P-A_N)}\le\hat{B}.\end{cases}\tag{B176}$$

c) For $\mathbf{S}_0=P$ *and* $(A_P,A_N)\in\mathcal{S}_2$,

$$Pr^*_{z_I}[\mathbf{S}_2=P|\mathbf{S}_0=P]=\begin{cases}0, & \text{if } 0\le\hat{B}<(1-q)\left(1+\frac{A_N}{A_P(A_P-A_N)}\right),\\ (1-q)\left(A_P^2+A_N\left(1-A_P\right)\right), & \\ \quad \text{if } (1-q)\left(1+\frac{A_N}{A_P(A_P-A_N)}\right)\le\hat{B}<1+A_P+(1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right), & \\ A_P^2+(1-q)\left(A_N\left(1-A_P\right)\right), & \\ \quad \text{if } 1+A_P+(1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right)\le\hat{B}<1+A_P\left(\frac{A_P}{A_N}-1\right), & \\ A_P^2+A_N\left(1-A_P\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right)\le\hat{B}.\end{cases}\tag{B177}$$

d) For $\mathbf{S}_0=P$, $(A_P,A_N)\in\mathcal{S}_3$, *and* $\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right)\le 1-\phi_{P|P}$,

$$Pr^*_{z_I}[\mathbf{S}_2=P|\mathbf{S}_0=P]=\begin{cases}0, & \text{if } 0\le\hat{B}<(1-q)\mathcal{B}_1,\\ (1-q)\left(A_P^2+A_N\left(1-A_P\right)\right), & \\ \quad \text{if } (1-q)\mathcal{B}_1\le\hat{B}<1+A_P+(1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right), & \\ A_P^2+(1-q)A_N\left(1-A_P\right), & \\ \quad \text{if } 1+A_P+(1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right)\le\hat{B}<1+A_P\left(\frac{A_P}{A_N}-1\right), & \\ A_P^2+A_N\left(1-A_P\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right)\le\hat{B}.\end{cases}\tag{B178}$$

e) For $\mathbf{S}_0=P$, $(A_P,A_N)\in\mathcal{S}_3$, *and* $1-\phi_{P|P}<\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right)$,

$$Pr^*_{z_I}[\mathbf{S}_2=P|\mathbf{S}_0=P]=\begin{cases}0, & \text{if } 0\le\hat{B}<q(1+A_P),\\ qA_P^2, & \text{if } q(1+A_P)\le\hat{B}<qA_P^2\mathcal{R}_1+(1-q)\mathcal{B}_1,\\ A_P^2+(1-q)A_N\left(1-A_P\right), & \text{if } qA_P^2\mathcal{R}_1+(1-q)\mathcal{B}_1\le\hat{B}<1+A_P\left(\frac{A_P}{A_N}-1\right),\\ A_P^2+A_N\left(1-A_P\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right)\le\hat{B}.\end{cases}\tag{B179}$$

f) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and* $\left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N} \leq 1-\phi_{P|P}$,

$$Pr^*_{z_I}[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \textit{if } 0 \leq \hat{B} < q(1+A_P), \\ qA_P^2, & \textit{if } q(1+A_P) \leq \hat{B} < \mathcal{B}_3, \\ qA_P^2 + (1-q)\left(A_P^2\phi_{P|P} + A_N(1-A_P)\phi_{P|N}\right), & \\ \qquad \textit{if } \mathcal{B}_3 \leq \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1, & \\ A_P^2 + (1-q)A_N(1-A_P), & \\ \qquad \textit{if } qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N}-1\right), & \\ A_P^2 + A_N(1-A_P), & \textit{if } 1 + A_P\left(\frac{A_P}{A_N}-1\right) \leq \hat{B}. \end{cases} \tag{B180}$$

g) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and* $1-\phi_{P|P} < \left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N}$,

$$Pr^*_{z_I}[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \textit{if } 0 \leq \hat{B} < (1-q)\mathcal{B}_2, \\ (1-q)\left(A_P^2\phi_{P|P} + A_N(1-A_P)\phi_{P|N}\right), & \\ \qquad \textit{if } (1-q)\mathcal{B}_2 \leq \hat{B} < \mathcal{B}_4, & \\ qA_P^2 + (1-q)\left(A_P^2\phi_{P|P} + A_N(1-A_P)\phi_{P|N}\right), & \\ \qquad \textit{if } \mathcal{B}_4 \leq \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1, & \\ A_P^2 + (1-q)A_N(1-A_P), & \\ \qquad \textit{if } qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N}-1\right), & \\ A_P^2 + A_N(1-A_P), & \textit{if } 1 + A_P\left(\frac{A_P}{A_N}-1\right) \leq \hat{B}. \end{cases} \tag{B181}$$

Proof of Lemma B4

The expessions in the Lemma are straightforward to calculate using the probability functions from Proposition B1 and the optimal merit-based incentive from Proposition 3.

When $\mathbf{S}_0 = N$, from (B108),

$$Pr^*[\mathbf{S}_2 = P|\mathbf{S}_0 = N] = \begin{cases} 0, & \text{if } 0 \leq \hat{\pi} < \frac{1}{A_N}, \\ A_N, & \text{if } \frac{1}{A_N} \leq \hat{\pi} < \frac{1}{A_N(A_P-A_N)}, \\ A_N(1+A_P-A_N), & \text{if } \frac{1}{A_N(A_P-A_N)} \leq \hat{\pi}, \end{cases} \tag{B182}$$

and from (47),

$$\hat{\pi}^* = \begin{cases} 0, & \text{if } 0 \leq \hat{B} < 1, \\ \frac{1}{A_N}, & \text{if } 1 \leq \hat{B} < 1 + \frac{1}{A_P-A_N}, \\ \frac{1}{A_N(A_P-A_N)}, & \text{if } 1 + \frac{1}{A_P-A_N} \leq \hat{B}. \end{cases} \tag{B183}$$

Then,

$$Pr^*_{z_I}[\mathbf{S}_2 = P|\mathbf{S}_0 = N] = \begin{cases} 0, & \text{if } 0 \leq \hat{B} < 1, \\ A_N, & \text{if } 1 \leq \hat{B} < 1 + \frac{1}{A_P-A_N}, \\ A_N(1+A_P-A_N), & \text{if } 1 + \frac{1}{A_P-A_N} \leq \hat{B}. \end{cases} \tag{B184}$$

We evalaute the remaining cases using the same approach.

Note that the setting where $q=0$, for which the optimal merit-based incentive is given in Proposition A1, is characterized by this case as well. To see this, notice that when $q=0$, $\mathcal{B}_2=\mathcal{B}_3$, where $\mathcal{B}_2$ and $\mathcal{B}_3$ are given in (44) and (45), respectively.

□

LEMMA B5. *The probability that the final state is proficient under the optimal incentive levels* (47)-(53) *when $q=1$ can be expressed as follows.*

a) For $\mathbf{S}_0=N$,

$$Pr^*_{z_I}[\mathbf{S}_2=P|\mathbf{S}_0=N]=\begin{cases}0, & \text{if } 0\le \hat{B}-\hat{F}z_I<1,\\ A_N, & \text{if } 1\le \hat{B}-\hat{F}z_I<1+\frac{1}{A_P-A_N},\\ A_N\left(1+A_P-A_N\right), & \text{if } 1+\frac{1}{A_P-A_N}\le \hat{B}-\hat{F}z_I.\end{cases} \tag{B185}$$

b) For $\mathbf{S}_0=P$ and $(A_P,A_N)\in\mathcal{S}_1$,

$$Pr^*_{z_I}[\mathbf{S}_2=P|\mathbf{S}_0=P]=\begin{cases}0, & \text{if } 0\le \hat{B}-\hat{F}z_I<1,\\ A_N, & \text{if } 1\le \hat{B}-\hat{F}z_I<1+\frac{A_N}{A_P(A_P-A_N)},\\ A_P^2+A_N\left(1-A_P\right), & \text{if } 1+\frac{A_N}{A_P(A_P-A_N)}\le \hat{B}-\hat{F}z_I.\end{cases} \tag{B186}$$

c) For $\mathbf{S}_0=P$, $(A_P,A_N)\in\mathcal{S}_4$,

$$Pr^*_{z_I}[\mathbf{S}_2=P|\mathbf{S}_0=P]=\begin{cases}0, & \text{if } 0\le \hat{B}-\hat{F}z_I<1+A_P,\\ A_P^2, & \text{if } 1+A_P\le \hat{B}-\hat{F}z_I<1+A_P\left(\frac{A_P}{A_N}-1\right),\\ A_P^2+A_N\left(1-A_P\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right)\le \hat{B}-\hat{F}z_I.\end{cases} \tag{B187}$$

Proof of Lemma B5

This is a special case of Lemma B4 with $q=1$, $\phi_{P|P}=1$, $\phi_{P|N}=0$, and the available budget $\hat{B}$ is replaced by $\hat{B}-\hat{F}z_I$.

□

PROPOSITION B2. *Let Z^* denote the set of optimal decisions, z_I^*, for the district.*

1. *For any budget value B, either there exists $F^*\le B$ such that*

$$Z^*=\begin{cases}\{1\}\ \text{or}\ \{0,1\}, & \text{if } F\le F^*,\\ \{0\}, & \text{if } F>F^*,\end{cases} \tag{B188}$$

   or $Z^*=\{0\}$ for all F.

2. *There exists $B_L\ge 0$ such that, for $B<B_L$, $Z^*=\{0,1\}$ for all $F\le B$.*
3. *There exists $B_U\ge B_L$ such that, for $B\ge B_U$,*

$$Z^*=\begin{cases}\{0,1\}, & \text{if } F\le F^*<B,\\ \{0\}, & \text{if } F>F^*.\end{cases} \tag{B189}$$

4. *Define*

$$\hat{F}^* = \frac{F^*}{\lambda\gamma}, \hat{B}_L = \frac{B_L}{\lambda\gamma}, \hat{B}_U = \frac{B_U}{\lambda\gamma}. \tag{B190}$$

Then, the values of $\hat{F}^*$, $\hat{B}_L$, *and* $\hat{B}_U$ *are given as follows.*

a) For $\mathbf{S}_0 = N$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < 1, \\ \hat{B} - 1, & \text{if } 1 \le \hat{B} < 1 + \frac{1}{A_P - A_N}, \\ \hat{B} - 1 - \frac{1}{A_P - A_N}, & \text{if } 1 + \frac{1}{A_P - A_N} \le \hat{B}, \end{cases} \tag{B191}$$

and $\hat{B}_L = \hat{B}_U = 1$.

b) For $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < 1, \\ \hat{B} - 1, & \text{if } 1 \le \hat{B} < 1 + \frac{A_N}{A_P(A_P - A_N)}, \\ \hat{B} - 1 - \frac{A_N}{A_P(A_P - A_N)}, & \text{if } 1 + \frac{A_N}{A_P(A_P - A_N)} \le \hat{B}, \end{cases} \tag{B192}$$

and $\hat{B}_L = \hat{B}_U = 1$.

c) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and either* $q = 1$, *or* $\phi_{P|P} = 1$, $\phi_{P|N} = 0$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < 1 + A_P, \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B193}$$

and $\hat{B}_L = \hat{B}_U = 1 + A_P$.

d) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_2$, *and* $q < \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < (1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right), \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B194}$$

and the interim assessment is never optimal, for any value of $\hat{F}$, *if*

$$(1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right) \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B195}$$

Additionally,

$$\hat{B}_L = \begin{cases} 1 + A_P, & \text{if } q \le \frac{A_N - A_P^2(A_P - A_N)}{A_P(A_P - A_N) + A_N}, \\ (1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right), & \text{if } \frac{A_N - A_P^2(A_P - A_N)}{A_P(A_P - A_N) + A_N} < q < \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}, \end{cases} \tag{B196}$$

and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

e) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_2$, *and* $\frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)} \le q < 1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < (1-q)\left(1+\frac{A_N}{A_P(A_P-A_N)}\right), \\ \hat{B}-1-A_P, & \text{if } 1+A_P \le \hat{B} < 1+A_P+(1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right), \\ \hat{B}-1-A_P\left(\frac{A_P}{A_N}-1\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right) \le \hat{B}, \end{cases} \tag{B197}$$

and the interim assessment is never optimal, for any value of $\hat{F}$, *if*

$$\begin{aligned} &(1-q)\left(1+\frac{A_N}{A_P\,(A_P-A_N)}\right) \le \hat{B} < 1+A_P \text{ or} \\ &1+A_P+(1-q)\left(\frac{A_N\,(1-A_P^2)}{A_P^2}\right) \le \hat{B} < 1+A_P\left(\frac{A_P}{A_N}-1\right). \end{aligned} \tag{B198}$$

Additionally, $\hat{B}_L = (1-q)\left(1+\frac{A_N}{A_P(A_P-A_N)}\right)$ *and* $\hat{B}_U = 1+A_P\left(\frac{A_P}{A_N}-1\right)$.

f) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, *and either* $\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right) \le 1-\phi_{P|P}$, $q < \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}$, *or* $(A_P, A_N) \in \mathcal{S}_3$, $1-\phi_{P|P} < \left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right)$, $q = 0$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < (1-q)\,\mathcal{B}_1, \\ \hat{B}-1-A_P\left(\frac{A_P}{A_N}-1\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right) \le \hat{B}, \end{cases} \tag{B199}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_1 \le \hat{B} < 1+A_P\left(\frac{A_P}{A_N}-1\right). \tag{B200}$$

Additionally, when $\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right) \le 1-\phi_{P|P}$,

$$\hat{B}_L = \begin{cases} 1+A_P, & \text{if } q < \mathcal{Q}_1, \\ (1-q)\,\mathcal{B}_1, & \text{if } \mathcal{Q}_1 \le q < \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}, \end{cases} \tag{B201}$$

and when $1-\phi_{P|P} < \left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right)$ *and* $q=0$, $\hat{B}_L = 1+A_P$. *In all cases,* $\hat{B}_U = 1+A_P\left(\frac{A_P}{A_N}-1\right)$.

g) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, $\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right) \le 1-\phi_{P|P}$, *and* $\frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)} \le q < 1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < (1-q)\,\mathcal{B}_1, \\ \hat{B}-1-A_P, & \text{if } 1+A_P \le \hat{B} < 1+A_P+(1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right), \\ \hat{B}-1-A_P\left(\frac{A_P}{A_N}-1\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right) \le \hat{B}, \end{cases} \tag{B202}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_1 \le \hat{B} < 1+A_P \text{ or } 1+A_P+(1-q)\left(\frac{A_N\,(1-A_P^2)}{A_P^2}\right) \le \hat{B} < 1+A_P\left(\frac{A_P}{A_N}-1\right). \tag{B203}$$

Additionally, $\hat{B}_L = (1-q)\,\mathcal{B}_1$ *and* $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

h) For $\mathbf{S}_0 = P$ *and either* $(A_P, A_N) \in \mathcal{S}_3$, $1 - \phi_{P|P} < \left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right)$, $0 < q < 1$, *or* $(A_P, A_N) \in \mathcal{S}_4$, $\max\left(\frac{A_P(A_P - A_N) - A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N} \le 1 - \phi_{P|P}$, *and* $0 < q < 1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < q(1+A_P), \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B204}$$

and the interim assessment is never the optimal choice if

$$q(1+A_P) \le \hat{B} < 1 + A_P \text{ or } qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B205}$$

Additionally, $\hat{B}_L = q\,(1+A_P)$ *and* $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

i) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $\left(\frac{A_P(A_P - A_N) - A_N}{A_P}\right)\phi_{P|N} \le 1 - \phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, *and* $0 < q < 1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < q(1+A_P), \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \le \hat{B} < \mathcal{B}_3, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B206}$$

and the interim assessment is never the optimal choice if

$$q(1+A_P) \le \hat{B} < 1 + A_P \text{ or } \mathcal{B}_3 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B207}$$

Additionally, $\hat{B}_L = q\,(1+A_P)$ *and* $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

j) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and either* $\left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N} \le 1 - \phi_{P|P} < \left(\frac{A_P(A_P - A_N) - A_N}{A_P}\right)\phi_{P|N}$, $q < 1$, *or* $\max\left(\frac{A_P(A_P - A_N) - A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N} \le 1 - \phi_{P|P}$, $A_N\phi_{P|N}\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right) < A_P^2\left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)$, *and* $q = 0$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < (1-q)\,\mathcal{B}_2, \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B208}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_2 \le \hat{B} < 1 + A_P \text{ or } qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B209}$$

Additionally, $\hat{B}_L = (1-q)\,\mathcal{B}_2$ *and* $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

k) For $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and either* $1 - \phi_{P|P} < \min\left(\frac{A_P(A_P - A_N) - A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, $0 \le q < \mathcal{Q}_6$, *or* $\left(\frac{A_P(A_P - A_N) - A_N}{A_P}\right)\phi_{P|N} \le 1 - \phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, $q = 0$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < (1-q)\,\mathcal{B}_2, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B210}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_2 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B211}$$

Additionally, if $1-\phi_{P|P} < \min\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$*, then*

$$\hat{B}_L = \begin{cases} 1 + A_P, & \text{if } 0 \leq q < \mathcal{Q}_5, \\ (1-q)\,\mathcal{B}_2, & \text{if } \mathcal{Q}_5 \leq q < \mathcal{Q}_6, \end{cases} \tag{B212}$$

and if $\left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N} \leq 1-\phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$ *and* $q=0$*, then*

$$\hat{B}_L = \begin{cases} 1 + A_P, & \text{if } A_N\phi_{P|N}\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right) \geq A_P^2\left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right), \\ \mathcal{B}_2, & \text{if } A_N\phi_{P|N}\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right) < A_P^2\left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right). \end{cases} \tag{B213}$$

For all cases, $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

l) For $\mathbf{S}_0 = P$*,* $(A_P, A_N) \in \mathcal{S}_4$*,* $1-\phi_{P|P} < \min\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$*, and* $\mathcal{Q}_6 \leq q < 1$*,*

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\,\mathcal{B}_2, \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \leq \hat{B} < \mathcal{B}_4, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B214}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_2 \leq \hat{B} < 1 + A_P \text{ or } \mathcal{B}_4 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B215}$$

Additionally, $\hat{B}_L = (1-q)\,\mathcal{B}_2$ *and* $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

m) For $\mathbf{S}_0 = P$*,* $(A_P, A_N) \in \mathcal{S}_4$*,* $\max\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N} \leq 1 - \phi_{P|P}$*,* $A_N\phi_{P|N}\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right) \geq A_P^2\left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)$*, and* $q=0$*,*

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < \mathcal{B}_2, \\ \hat{B} - 1 - A_P, & \text{if } \mathcal{B}_2 \leq \hat{B} < \mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B216}$$

and the interim assessment is never the optimal choice if

$$\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B217}$$

Additionally, $\hat{B}_L = 1 + A_P$ *and* $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

Proof of Proposition B2

In order to establish the statements of this Proposition, we will need the following result.

LEMMA B6. *We define the following constants:*

$$\mathcal{Q}_1 = \frac{A_P}{A_P^2 + A_N(1-A_P)} \left(\frac{(A_N(1-A_P) - A_P)\left(1-\phi_{P|P}\right) + (1-A_P)(A_P - 2A_N)\left(1-\phi_{P|N}\right)}{A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)} \right), \tag{B218}$$

$$\mathcal{Q}_2 = \frac{A_P^2}{A_N(1-A_P)} + 1 - \frac{(1+A_P)\left(A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)\right)}{A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)}, \tag{B219}$$

$$\mathcal{Q}_3 = \frac{\frac{1+A_P}{\mathcal{R}_2} - A_P^2\phi_{P|P} - A_N(1-A_P)\phi_{P|N}}{A_P^2 - A_P^2\phi_{P|P} - A_N(1-A_P)\phi_{P|N}}, \tag{B220}$$

$$\mathcal{Q}_4 = \frac{\mathcal{B}_1 - 1 - A_P}{\mathcal{B}_1 - A_P^2\mathcal{R}_1}, \tag{B221}$$

$$\mathcal{Q}_5 = 1 - \frac{A_P(1+A_P)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}{\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right)\left(A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right) + A_N\phi_{P|N}\right)}, \tag{B222}$$

$$\mathcal{Q}_6 = \frac{A_N(1-A_P)\phi_{P|N} - A_P^2\left(1-\phi_{P|P}\right)}{A_P^2\phi_{P|P} + A_N(1-A_P)\phi_{P|N}}, \tag{B223}$$

where, from (43),

$$\mathcal{B}_1 = \left(\frac{A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)} \right)\left(A_P^2 + A_N(1-A_P)\right). \tag{B224}$$

These constants have the following properties.

a) $\mathcal{Q}_2 \geq 1$.

b) If $\left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N} \leq 1 - \phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, *then* $\mathcal{Q}_3 \geq 1$.

c) If $(A_P, A_N) \in \mathcal{S}_4$, *then* $\mathcal{Q}_4 > 1$.

d) $\mathcal{Q}_5 < 1$.

e) $\mathcal{Q}_6 < 1$.

f) If $1 - \phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, *then* $\mathcal{Q}_6 > 0$.

Proof of Lemma B6

a) Using (B219),

$$\begin{aligned}
\mathcal{Q}_2 \geq 1 \iff & \frac{A_P^2}{A_N(1-A_P)} + 1 - \frac{(1+A_P)\left(A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\right)}{A_P\left(1-\phi_{P|P}\right) + (1-A_P)} \geq 1 \\
\iff & \frac{A_P^2}{A_N(1-A_P)} \geq \frac{(1+A_P)\left(A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\right)}{A_P\left(1-\phi_{P|P}\right) + (1-A_P)} \\
\iff & A_P^2\left(A_P - A_N\left(1-A_P^2\right)\right)\left(1-\phi_{P|P}\right) \geq \left(A_N^2\left(1-A_P^2\right) - A_P^2\right)(1-A_P),
\end{aligned} \tag{B225}$$

where the left-hand side of the inequality is always nonnegative and the right-hand side of the inequality is always nonpositive under Assumptions 1 and 2. Therefore, the statement always holds.

b) Suppose $\left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N} \leq 1-\phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$. From (B220),

$$\mathcal{Q}_3 = \frac{\frac{1+A_P}{\mathcal{R}_2} - A_P^2\phi_{P|P} - A_N(1-A_P)\phi_{P|N}}{A_P^2 - A_P^2\phi_{P|P} - A_N(1-A_P)\phi_{P|N}}. \tag{B226}$$

The denominator in this expression is negative, since

$$A_P^2 - A_P^2\phi_{P|P} - A_N(1-A_P)\phi_{P|N} < 0 \iff 1-\phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}, \tag{B227}$$

which holds by assumption in this case. Then, using (33),

$$\begin{aligned}
\mathcal{Q}_3 \geq 1 &\iff \frac{\frac{1+A_P}{\mathcal{R}_2} - A_P^2\phi_{P|P} - A_N(1-A_P)\phi_{P|N}}{A_P^2 - A_P^2\phi_{P|P} - A_N(1-A_P)\phi_{P|N}} \geq 1 \\
&\iff \frac{(1+A_P)A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}{1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)} \leq A_P^2 \\
&\iff A_P^2(1+A_P)\phi_{P|P} - A_N A_P(1+A_P)\phi_{P|N} \leq A_P^2 + A_P^3\left(\phi_{P|P}-\phi_{P|N}\right) \\
&\iff 0 \leq A_P\left(1-\phi_{P|P}\right) + \left(A_N(1+A_P) - A_P^2\right)\phi_{P|N} \\
&\iff \left(\frac{(A_P(A_P-A_N)-A_N)}{A_P}\right)\phi_{P|N} \leq 1-\phi_{P|P},
\end{aligned} \tag{B228}$$

where, again, the last line holds by assumption in this case.

c) Recall that

$$\begin{aligned}
(A_P, A_N) \in \mathcal{S}_4 &\iff \frac{A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)}{(1-A_P)\left(1-\phi_{P|N}\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} < A_P - A_N \\
&\iff 1-\phi_{P|P} < \frac{1-A_P}{A_P^2}\left(1-\phi_{P|N}\right)\left((A_P-A_N)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right) - A_N\right).
\end{aligned} \tag{B229}$$

From (B221),

$$\mathcal{Q}_4 = \frac{\mathcal{B}_1 - 1 - A_P}{\mathcal{B}_1 - A_P^2\mathcal{R}_1}, \tag{B230}$$

where under Assumption 1 and using the definitions of $\mathcal{R}_1$ and $\mathcal{B}_1$ from (32) and (43),

$$\mathcal{B}_1 - A_P^2\mathcal{R}_1 = A_N(1-A_P)\left(\frac{A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)}\right) > 0. \tag{B231}$$

Then,

$$\begin{aligned}
\mathcal{Q}_4 > 1 &\iff \frac{\mathcal{B}_1 - 1 - A_P}{\mathcal{B}_1 - A_P^2\mathcal{R}_1} > 1 \\
&\iff A_P^2\mathcal{R}_1 > 1 + A_P \\
&\iff (1-A_P)\left(A_P^2 - A_N(1+A_P)\right)\left(1-\phi_{P|N}\right) > A_P^2\left(1-\phi_{P|P}\right) \\
&\iff \left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right) > 1-\phi_{P|P}.
\end{aligned} \tag{B232}$$

This upper bound on $1-\phi_{P|P}$ is weaker than existing upper bound in this case, since

$$\begin{aligned}
&\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right) > \frac{1-A_P}{A_P^2}\left(1-\phi_{P|N}\right)\left((A_P-A_N)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)-A_N\right)\\
\iff& A_P^2-A_P^2\left(A_P-A_N\right)-A_N > \left(A_P-A_N\right)\left(1-A_P\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)-A_N\left(1-A_P\right)\\
\iff& A_P\left(1-A_P\right)\left(A_P-A_N\right) > \left(A_P-A_N\right)\left(1-A_P\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)\\
\iff& A_P > A_P\phi_{P|P}-A_N\phi_{P|N}, \qquad \text{(B233)}
\end{aligned}$$

which always holds under Assumptions 1 and 2.

d) Using (B222),

$$\begin{aligned}
\mathcal{Q}_5<1 \iff& 1-\frac{A_P\left(1+A_P\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}{\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)\left(A_P\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)+A_N\phi_{P|N}\right)}<1\\
\iff& 0<\frac{A_P\left(1+A_P\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}{\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)\left(A_P\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)+A_N\phi_{P|N}\right)}, \qquad \text{(B234)}
\end{aligned}$$

which always holds under Assumptions 1 and 2.

e) Using (B223),

$$\mathcal{Q}_6<1 \iff \frac{A_N\left(1-A_P\right)\phi_{P|N}-A_P^2\left(1-\phi_{P|P}\right)}{A_P^2\phi_{P|P}+A_N\left(1-A_P\right)\phi_{P|N}}<1 \iff A_P^2>0, \qquad \text{(B235)}$$

which always holds under Assumption 1.

f) Suppose $1-\phi_{P|P}<\frac{A_N(1-A_P)}{A_P^2}\phi_{P|N}$. Then,

$$\mathcal{Q}_6>0 \iff \frac{A_N\left(1-A_P\right)\phi_{P|N}-A_P^2\left(1-\phi_{P|P}\right)}{A_P^2\phi_{P|P}+A_N\left(1-A_P\right)\phi_{P|N}}>0 \iff \left(\frac{A_N\left(1-A_P\right)}{A_P^2}\right)\phi_{P|N}>1-\phi_{P|P}, \qquad \text{(B236)}$$

which holds by the initial assumption. □

Let Z^* denote the set of optimal decisions, z_I^*, for the district, so $Z^*\in\{\{0\},\{1\},\{0,1\}\}$. Define $F^*\le B$ such that

$$Z^*=\begin{cases}\{1\}\ \text{ or }\ \{0,1\}, & \text{if } F\le F^*,\\ \{0\}, & \text{if } F>F^*.\end{cases} \qquad \text{(B237)}$$

That is, for a given budget value, if F^* exists, it is the maximum value of the cost of the interim assessment F for which the interim assessment is an optimal choice of assessment.

Furthermore, define $B_L\ge 0$ such that for $B<B_L$, both the interim and formative assessment are optimal choices ($Z^*=\{0,1\}$) for all $F\le B$. Additionally, define $B_U\ge B_L$ such that for $B\ge B_U$, both the interim and formative assessment are optimal choices ($Z^*=\{0,1\}$) when $F\le F^*<B$, and only the formative assessment is optimal ($Z^*=\{0\}$) when $F>F^*$.

For different combinations of the parameters and initial state $\mathbf{S}_0$, we consider whether F^* exists, and, if so, determine its value. Using this, we can determine the value of B_L and B_U in each case.

For $\mathbf{S}_0 = N$, recall from (B175) that

$$Pr^*_{z_I}[\mathbf{S}_2 = P|\mathbf{S}_0 = N] = \begin{cases} 0, & \text{if } 0 \le \hat{B} - \hat{F}z_I < 1, \\ A_N, & \text{if } 1 \le \hat{B} - \hat{F}z_I < 1 + \frac{1}{A_P - A_N}, \\ A_N\left(1 + A_P - A_N\right), & \text{if } 1 + \frac{1}{A_P - A_N} \le \hat{B} - \hat{F}z_I. \end{cases} \tag{B238}$$

Then, for $\mathbf{S}_0 = N$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < 1, \\ \hat{B} - 1, & \text{if } 1 \le \hat{B} < 1 + \frac{1}{A_P - A_N}, \\ \hat{B} - 1 - \frac{1}{A_P - A_N}, & \text{if } 1 + \frac{1}{A_P - A_N} \le \hat{B}, \end{cases} \tag{B239}$$

and $\hat{B}_L = \hat{B}_U = 1$.

For $\mathbf{S}_0 = P$ and $(A_P, A_N) \in \mathcal{S}_1$, recall from (B176) that

$$Pr^*_{z_I}[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{B} - \hat{F}z_I < 1, \\ A_N, & \text{if } 1 \le \hat{B} - \hat{F}z_I < 1 + \frac{A_N}{A_P(A_P - A_N)}, \\ A_P^2 + A_N\left(1 - A_P\right), & \text{if } 1 + \frac{A_N}{A_P(A_P - A_N)} \le \hat{B} - \hat{F}z_I. \end{cases} \tag{B240}$$

Then, for $\mathbf{S}_0 = P$ and $(A_P, A_N) \in \mathcal{S}_1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < 1, \\ \hat{B} - 1, & \text{if } 1 \le \hat{B} < 1 + \frac{A_N}{A_P(A_P - A_N)}, \\ \hat{B} - 1 - \frac{A_N}{A_P(A_P - A_N)}, & \text{if } 1 + \frac{A_N}{A_P(A_P - A_N)} \le \hat{B}, \end{cases} \tag{B241}$$

and $\hat{B}_L = \hat{B}_U = 1$.

We now consider the case where $\mathbf{S}_0 = P$ and $(A_P, A_N) \in \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4$. When the district believes teachers have perfect information (which occurs if $\phi_{P|P} = 1$ and $\phi_{P|N} = 0$; $q = 1$; or $z_I = 1$), then the probability that the final state is proficient is determined by evaluating (B180)-(B181) for $q = 1$, $\phi_{P|P} = 1$, and $\phi_{P|N} = 0$. (Note that we consider only these equations since $\mathcal{S}_2$ and $\mathcal{S}_3$ are empty sets for these parameter values.) In this case,

$$Pr^*_{z_I}\left[\mathbf{S}_2 = P|\mathbf{S}_0 = P\right] = \begin{cases} 0, & \text{if } 0 \le \hat{B} - \hat{F}z_I < 1 + A_P, \\ A_P^2, & \text{if } 1 + A_P \le \hat{B} - \hat{F}z_I < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ A_P^2 + A_N\left(1 - A_P\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B} - \hat{F}z_I. \end{cases} \tag{B242}$$

Then, when $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, and $q = 1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < 1 + A_P, \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B243}$$

and $\hat{B}_L = \hat{B}_U = 1 + A_P$.

When $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4$, and $q < 1$, we must compare (B242) when $z_I = 1$ to each of the settings described by (B177)-(B181). Therefore, we consider five cases.

Case 1: $\mathbf{S}_0 = P$ *and* $(A_P, A_N) \in \mathcal{S}_2$. If $z_I = 0$, from (B177),

$$Pr^*_{z_I}[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{B} < (1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right), \\ (1-q)\left(A_P^2 + A_N(1-A_P)\right), & \\ \quad \text{if } (1-q)\left(1 + \frac{A_N}{A_P(A_P-A_N)}\right) \leq \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right), & \\ A_P^2 + (1-q)\left(A_N(1-A_P)\right), & \\ \quad \text{if } 1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right) \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), & \\ A_P^2 + A_N(1-A_P), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}. \end{cases} \tag{B244}$$

We consider two subcases.

Subcase 1: Suppose

$$(1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right) \geq 1 + A_P \iff q \leq \frac{A_N - A_P^2(A_P - A_N)}{A_P(A_P - A_N) + A_N}. \tag{B245}$$

If $0 \leq \hat{B} < (1-q)\left(1 + \frac{A_N}{A_P(A_P-A_N)}\right)$,

$$\hat{F}^* = \hat{B}. \tag{B246}$$

If $(1-q)\left(1 + \frac{A_N}{A_P(A_P-A_N)}\right) \leq \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right)$, then the probability when $z_I = 1$ exceeds that when $z_I = 0$ if and only if

$$A_P^2 \geq (1-q)\left(A_P^2 + A_N(1-A_P)\right) \iff q \geq \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}. \tag{B247}$$

But, this bound on q is infeasible given (B245), since

$$\frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)} > \frac{A_N - A_P^2(A_P - A_N)}{A_P(A_P - A_N) + A_N} \iff A_P - A_N > \frac{A_N}{A_P}, \tag{B248}$$

which always holds in this case. Therefore, the interim assessment is never optimal.

If $1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right) \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$, the interim assessment is never optimal for $q < 1$.

Finally, if $1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}$,

$$\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B249}$$

In this case, $\hat{B}_L = 1 + A_P$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

Subcase 2: Suppose instead that

$$\frac{A_N - A_P^2(A_P - A_N)}{A_P(A_P - A_N) + A_N} < q. \tag{B250}$$

If $0 \leq \hat{B} < (1-q)\left(1+\frac{A_N}{A_P(A_P-A_N)}\right)$, then

$$\hat{F}^* = \hat{B}. \tag{B251}$$

If $(1-q)\left(1+\frac{A_N}{A_P(A_P-A_N)}\right) \leq \hat{B} < 1+A_P$, the interim assessment is never optimal, even for $\hat{F}=0$, as long as $q<1$.

If $1+A_P \leq \hat{B} < 1+A_P+(1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right)$, then the probability the final state is proficient when $z_I=1$ exceeds that when $z_I=0$ if and only if

$$A_P^2 \geq (1-q)\left(A_P^2+A_N\left(1-A_P\right)\right) \iff q \geq \frac{A_N\left(1-A_P\right)}{A_P^2+A_N\left(1-A_P\right)}. \tag{B252}$$

This is a stronger lower bound on q than that in (B250), since

$$\frac{A_N\left(1-A_P\right)}{A_P^2+A_N\left(1-A_P\right)} > \frac{A_N-A_P^2\left(A_P-A_N\right)}{A_P\left(A_P-A_N\right)+A_N} \iff A_P-A_N > \frac{A_N}{A_P}, \tag{B253}$$

which always holds in this case. Then,

$$\hat{F}^* = \hat{B}-1-A_P. \tag{B254}$$

If instead

$$q < \frac{A_N\left(1-A_P\right)}{A_P^2+A_N\left(1-A_P\right)}, \tag{B255}$$

then the interim assessment is never optimal, even for $\hat{F}=0$.

If $1+A_P+(1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right) \leq \hat{B} < 1+A_P\left(\frac{A_P}{A_N}-1\right)$, the interim assessment is never optimal for $q<1$.

Finally, if $1+A_P\left(\frac{A_P}{A_N}-1\right) \leq \hat{B}$,

$$\hat{F}^* = \hat{B}-1-A_P\left(\frac{A_P}{A_N}-1\right). \tag{B256}$$

In this case, $\hat{B}_L=(1-q)\left(1+\frac{A_N}{A_P(A_P-A_N)}\right)$ and $\hat{B}_U=1+A_P\left(\frac{A_P}{A_N}-1\right)$.

Combining this, we have that if $\mathbf{S}_0=P$, $(A_P,A_N)\in\mathcal{S}_2$, and $q<\frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}$, then

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\left(1+\frac{A_N}{A_P(A_P-A_N)}\right), \\ \hat{B}-1-A_P\left(\frac{A_P}{A_N}-1\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right) \leq \hat{B}, \end{cases} \tag{B257}$$

and the interim assessment is never optimal, for any value of $\hat{F}$, if

$$(1-q)\left(1+\frac{A_N}{A_P\left(A_P-A_N\right)}\right) \leq \hat{B} < 1+A_P\left(\frac{A_P}{A_N}-1\right). \tag{B258}$$

Additionally,

$$\hat{B}_L = \begin{cases} 1 + A_P, & \text{if } q \le \frac{A_N - A_P^2 (A_P - A_N)}{A_P (A_P - A_N) + A_N}, \\ (1-q)\left(1 + \frac{A_N}{A_P (A_P - A_N)}\right), & \text{if } \frac{A_N - A_P^2 (A_P - A_N)}{A_P (A_P - A_N) + A_N} < q < \frac{A_N (1 - A_P)}{A_P^2 + A_N (1 - A_P)}, \end{cases} \tag{B259}$$

and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_2$, and $q \ge \frac{A_N (1 - A_P)}{A_P^2 + A_N (1 - A_P)}$, then

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < (1-q)\left(1 + \frac{A_N}{A_P (A_P - A_N)}\right), \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \le \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N \left(1 - A_P^2\right)}{A_P^2}\right), \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B260}$$

and the interim assessment is never optimal, for any value of $\hat{F}$, if

$$\begin{aligned} &(1-q)\left(1 + \frac{A_N}{A_P (A_P - A_N)}\right) \le \hat{B} < 1 + A_P \text{ or} \\ &1 + A_P + (1-q)\left(\frac{A_N (1 - A_P^2)}{A_P^2}\right) \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \end{aligned} \tag{B261}$$

Additionally, $\hat{B}_L = (1-q)\left(1 + \frac{A_N}{A_P (A_P - A_N)}\right)$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

Case 2: $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, *and* $\left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right) \le 1 - \phi_{P|P}$. Then, if $z_I = 0$, from (B178),

$$Pr^*_{z_I}[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{B} < (1-q)\,\mathcal{B}_1, \\ (1-q)\left(A_P^2 + A_N (1 - A_P)\right), & \text{if } (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N \left(1 - A_P^2\right)}{A_P^2}\right), \\ A_P^2 + (1-q)\left(A_N (1 - A_P)\right), & \text{if } 1 + A_P + (1-q)\left(\frac{A_N \left(1 - A_P^2\right)}{A_P^2}\right) \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right), \\ A_P^2 + A_N (1 - A_P), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B262}$$

where from (43),

$$\mathcal{B}_1 = \left(\frac{A_P\left(1 - \phi_{P|P}\right) + (1 - A_P)\left(1 - \phi_{P|N}\right)}{A_P^2\left(1 - \phi_{P|P}\right) + A_N (1 - A_P)\left(1 - \phi_{P|N}\right)}\right)\left(A_P^2 + A_N (1 - A_P)\right). \tag{B263}$$

Again, we compare the bounds on the budget terms in (B242) and (B262). Now,

$$\mathcal{B}_1 \ge 1 + A_P \iff (A_N (1 - A_P) - A_P)\left(1 - \phi_{P|P}\right) \ge (1 - A_P)(2A_N - A_P)\left(1 - \phi_{P|N}\right). \tag{B264}$$

We consider two cases.

Subcase 1: Suppose

$$(1-q)\,\mathcal{B}_1 > 1 + A_P \iff \mathcal{Q}_1 > q, \tag{B265}$$

where, as stated in (B218),

$$\mathcal{Q}_1 = \frac{A_P}{A_P^2 + A_N(1-A_P)}\left(\frac{(A_N(1-A_P)-A_P)\left(1-\phi_{P|P}\right) + (A_P - A_P^2 - 2A_N(1-A_P))\left(1-\phi_{P|N}\right)}{A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)}\right). \tag{B266}$$

If $0 \le \hat{B} < (1-q)\,\mathcal{B}_1$, then $\hat{F}^* = \hat{B}$. If $(1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right)$, then, again, the probability of final proficiency under the interim assessment exceeds that under the formative if and only if

$$A_P^2 \ge (1-q)\left(A_P^2 + A_N(1-A_P)\right) \iff q \ge \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}. \tag{B267}$$

However, this is never feasible: the lower bound on q is greater than the upper bound for this case, given by (B265), since

$$\begin{aligned}\frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)} \ge \mathcal{Q}_1 &\iff A_P^2\left(1-\phi_{P|P}\right) \ge (1-A_P)\left(A_P^2 - A_N(1+A_P)\right)\left(1-\phi_{P|N}\right)\\ &\iff 1-\phi_{P|P} \ge \left(1-(A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right),\end{aligned} \tag{B268}$$

and the last line is a condition for this case. Therefore, the interim assessment is never optimal.

If $1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right) \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$, the interim assessment is never optimal. Finally, if $1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}$,

$$\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B269}$$

In this case, $\hat{B}_L = 1 + A_P$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

Subcase 2: Suppose

$$\mathcal{Q}_1 \le q. \tag{B270}$$

If $0 \le \hat{B} < (1-q)\,\mathcal{B}_1$, $\hat{F}^* = \hat{B}$. If $(1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P$, the interim assessment is never the optimal choice.

If $1 + A_P \le \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right)$, then, again, the probability of final proficiency under the interim assessment exceeds that under the formative if and only if

$$A_P^2 \ge (1-q)\left(A_P^2 + A_N(1-A_P)\right) \iff q \ge \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}. \tag{B271}$$

As shown in (B268), this bound on q is stronger than the bound in (B270). Then, if $q \geq \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}$,

$$\hat{F}^* = \hat{B} - 1 - A_P, \tag{B272}$$

and if $q < \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}$, the interim assessment is never optimal.

If $1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right) \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$ and $q < 1$, the interim assessment is never optimal.

Finally, if $1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}$,

$$\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B273}$$

In this case, $\hat{B}_L = (1-q)\,\mathcal{B}_1$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

When combining the results from Subcases 1 and 2, recall from (B268) that

$$\mathcal{Q}_1 \leq \frac{A_N\,(1-A_P)}{A_P^2 + A_N\,(1-A_P)}. \tag{B274}$$

Then, if $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, $\left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right) \leq 1 - \phi_{P|P}$, and $q < \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\,\mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B275}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B276}$$

Additionally,

$$\hat{B}_L = \begin{cases} 1 + A_P, & \text{if } q < \mathcal{Q}_1, \\ (1-q)\,\mathcal{B}_1, & \text{if } \mathcal{Q}_1 \leq q < \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}, \end{cases} \tag{B277}$$

and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, $\left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right) \leq 1 - \phi_{P|P}$, and $q \geq \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\,\mathcal{B}_1, \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \leq \hat{B} < 1 + A_P + (1-q)\left(\frac{A_N\left(1-A_P^2\right)}{A_P^2}\right), \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B278}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_1 \leq \hat{B} < 1 + A_P \text{ or } 1 + A_P + (1-q)\left(\frac{A_N\,(1-A_P^2)}{A_P^2}\right) \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B279}$$

Additionally, $\hat{B}_L = (1-q)\,\mathcal{B}_1$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

Case 3: $\mathbf{S}_0 = P$*,* $(A_P, A_N) \in \mathcal{S}_3$*, and* $1-\phi_{P|P} < \left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right)$. If $z_I = 0$, from (B179),

$$Pr^*_{z_I}[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{B} < q(1+A_P), \\ qA_P^2, & \text{if } q(1+A_P) \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1, \\ A_P^2 + (1-q)\left(A_N\left(1-A_P\right)\right), & \text{if } qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N}-1\right), \\ A_P^2 + A_N\left(1-A_P\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N}-1\right) \le \hat{B}. \end{cases} \tag{B280}$$

First, consider the case where $q = 0$. Then,

$$Pr^*_{z_I}[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \le \hat{B} < \mathcal{B}_1, \\ A_P^2 + A_N\left(1-A_P\right), & \text{if } \mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N}-1\right). \end{cases} \tag{B281}$$

In this case, it always holds that

$$\mathcal{B}_1 \ge 1 + A_P. \tag{B282}$$

To see this, recall from (B264) that

$$\begin{aligned} \mathcal{B}_1 \ge 1 + A_P &\iff \left(A_N\left(1-A_P\right)-A_P\right)\left(1-\phi_{P|P}\right) \ge \left(1-A_P\right)\left(2A_N-A_P\right)\left(1-\phi_{P|N}\right) \\ &\iff 1-\phi_{P|P} \le \left(\frac{\left(1-A_P\right)\left(2A_N-A_P\right)}{A_N\left(1-A_P\right)-A_P}\right)\left(1-\phi_{P|N}\right). \end{aligned} \tag{B283}$$

(B283) always holds, since, comparing this upper bound on $1-\phi_{P|P}$ to the upper bound in this case,

$$\begin{aligned} &\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right) < \left(\frac{\left(1-A_P\right)\left(2A_N-A_P\right)}{A_N\left(1-A_P\right)-A_P}\right)\left(1-\phi_{P|N}\right) \\ \iff &\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(A_N\left(1-A_P\right)-A_P\right) > \left(1-A_P\right)\left(2A_N-A_P\right) \\ \iff &\left(A_P^2 - A_P^2\left(A_P-A_N\right)-A_N\right)\left(A_P-A_N\left(1-A_P\right)\right) + A_P^3\left(A_P-2A_N\right) < A_P^2\left(A_P-2A_N\right) \\ \iff &A_P^2\left(A_P-A_N\right)\left(1+A_N\left(1-A_P\right)\right) - A_N\left(A_P-A_N\left(1-A_P\right)\right) < A_P^2\left(A_P-2A_N\right) \\ \iff &A_N\left(1-A_P^2\right) - A_NA_P\left(1-A_P^2\right) < A_P\left(1-A_P\right)\left(1-A_P^2\right) \\ \iff &A_N < A_P, \end{aligned} \tag{B284}$$

which holds under Assumption 1.

Then, comparing (B281) to (B242), when $0 \le \hat{B} < \mathcal{B}_1$, $\hat{F}^* = \hat{B}$. When $\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N}-1\right)$, the interim assessment is never optimal. Finally, when $1 + A_P\left(\frac{A_P}{A_N}-1\right) \le \hat{B}$, $\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N}-1\right)$.

Then,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < \mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N}-1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N}-1\right) \le \hat{B}, \end{cases} \tag{B285}$$

and the interim assessment is never the optimal choice if

$$\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B286}$$

Additionally, $\hat{B}_L = 1 + A_P$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

Next, consider the case where $0 < q < 1$. If $0 \le \hat{B} < q(1 + A_P)$, then $\hat{F}^* = \hat{B}$.

Consider the bounds on the budget terms in (B242) and (B280). Now,

$$\begin{aligned}
&1 + A_P \ge qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \\
&\iff 1 + A_P \ge qA_P^2\left(\frac{A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)}{A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)}\right) \\
&+ (1-q)\left(\frac{\left(A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)\right)\left(A_P^2 + A_N(1-A_P)\right)}{A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)}\right) \\
&\iff q \ge \mathcal{Q}_2,
\end{aligned} \tag{B287}$$

where, as stated in (B219),

$$\mathcal{Q}_2 = \frac{A_P^2}{A_N(1-A_P)} + 1 - \frac{(1+A_P)\left(A_P^2\left(1-\phi_{P|P}\right) + A_N(1-A_P)\left(1-\phi_{P|N}\right)\right)}{\left(A_P\left(1-\phi_{P|P}\right) + (1-A_P)\left(1-\phi_{P|N}\right)\right)}. \tag{B288}$$

From Lemma B6, $\mathcal{Q}_2 \ge 1$. Therefore, since $q < 1 \le \mathcal{Q}_2$, (B287) is never feasible. If $q(1+A_P) \le \hat{B} < 1 + A_P$, the interim assessment is never optimal. If $1 + A_P \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1$ and since $q < 1$,

$$\hat{F}^* = \hat{B} - 1 - A_P. \tag{B289}$$

If $qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$ and since $q < 1$, the interim assessment is never optimal.

Finally, for all $0 < q < 1$, if $1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}$,

$$\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B290}$$

Combining this, if $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, $1 - \phi_{P|P} < \left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right)$, and $0 < q < 1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < q(1+A_P), \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B291}$$

and the interim assessment is never the optimal choice if

$$q(1+A_P) \le \hat{B} < 1 + A_P \text{ or } qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B292}$$

Additionally, $\hat{B}_L = q(1 + A_P)$ and $\hat{B}_U = 1 + A_P \left(\frac{A_P}{A_N} - 1\right)$.

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, $1 - \phi_{P|P} < \left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right)$, and $q = 0$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < \mathcal{B}_1, \\ \hat{B} - 1 - A_P \left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P \left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B293}$$

and the interim assessment is never the optimal choice if

$$\mathcal{B}_1 \leq \hat{B} < 1 + A_P \left(\frac{A_P}{A_N} - 1\right). \tag{B294}$$

Additionally, $\hat{B}_L = 1 + A_P$ and $\hat{B}_U = 1 + A_P \left(\frac{A_P}{A_N} - 1\right)$.

Case 4: $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and* $\left(\frac{A_P(A_P - A_N) - A_N}{A_P}\right)\phi_{P|N} \leq 1 - \phi_{P|P}$. If $z_I = 0$, from (B180),

$$Pr^*_{z_I}[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{B} < q(1 + A_P), \\ qA_P^2, & \text{if } q(1 + A_P) \leq \hat{B} < \mathcal{B}_3 \\ qA_P^2 + (1 - q)\left(A_P^2 \phi_{P|P} + A_N(1 - A_P)\phi_{P|N}\right), & \\ & \text{if } \mathcal{B}_3 \leq \hat{B} < qA_P^2 \mathcal{R}_1 + (1 - q)\mathcal{B}_1, \\ A_P^2 + (1 - q) A_N (1 - A_P), & \text{if } qA_P^2 \mathcal{R}_1 + (1 - q)\mathcal{B}_1 \leq \hat{B} < 1 + A_P \left(\frac{A_P}{A_N} - 1\right), \\ A_P^2 + A_N(1 - A_P), & \text{if } 1 + A_P \left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B295}$$

where from (45),

$$\mathcal{B}_3 = \mathcal{R}_2 \left(qA_P^2 + (1 - q)\left(A_P^2 \phi_{P|P} + A_N (1 - A_P)\phi_{P|N}\right)\right), \tag{B296}$$

and from (32) and (33),

$$\mathcal{R}_1 = \frac{A_P\left(1 - \phi_{P|P}\right) + (1 - A_P)\left(1 - \phi_{P|N}\right)}{A_P^2\left(1 - \phi_{P|P}\right) + A_N(1 - A_P)\left(1 - \phi_{P|N}\right)}, \tag{B297}$$

$$\mathcal{R}_2 = \frac{1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}. \tag{B298}$$

We first consider the special case where $q = 0$. Recall that when $q = 0$, $\mathcal{B}_2 = \mathcal{B}_3$. Then,

$$Pr^*_{z_I}[\mathbf{S}_2 = P | \mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{B} < \mathcal{B}_2 \\ A_P^2 \phi_{P|P} + A_N(1 - A_P)\phi_{P|N}, & \text{if } \mathcal{B}_2 \leq \hat{B} < \mathcal{B}_1, \\ A_P^2 + A_N(1 - A_P), & \text{if } \mathcal{B}_1 \leq \hat{B}. \end{cases} \tag{B299}$$

In this case, it always holds that $\mathcal{B}_1 \geq 1 + A_P$. To see this, recall from (B264) that

$$\mathcal{B}_1 < 1 + A_P \iff \left(1 - \phi_{P|P}\right) > \left(\frac{(1 - A_P)(2A_N - A_P)}{A_N(1 - A_P) - A_P}\right)\left(1 - \phi_{P|N}\right). \tag{B300}$$

Furthermore, recall from (B229) that

$$(A_P, A_N) \in \mathcal{S}_4 \Rightarrow 1 - \phi_{P|P} < \frac{1 - A_P}{A_P^2}\left(1 - \phi_{P|N}\right)\left((A_P - A_N)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right) - A_N\right). \tag{B301}$$

Then, (B300) is never feasible, since, comparing bounds on $1-\phi_{P|P}$,

$$\begin{aligned}
&\frac{1-A_P}{A_P^2}\left(1-\phi_{P|N}\right)\left(\left(A_P-A_N\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)-A_N\right)>\left(\frac{\left(1-A_P\right)\left(2A_N-A_P\right)}{A_N\left(1-A_P\right)-A_P}\right)\left(1-\phi_{P|N}\right)\\
&\iff \left(\left(A_P-A_N\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)-A_N\right)\left(A_N\left(1-A_P\right)-A_P\right)<A_P^2\left(2A_N-A_P\right)\\
&\iff \left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)\left(A_N\left(1-A_P\right)-A_P\right)<-\left(A_P^2+A_N\left(1-A_P\right)\right)\\
&\iff \left(A_P\phi_{P|P}-A_N\phi_{P|N}-A_P\right)\left(A_P-A_N\right)>A_N\left(1-A_P\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)\right), \qquad \text{(B302)}
\end{aligned}$$

which never holds under Assumption 1, since the left-hand side of the inequality is negative and the right-hand side of the inequality is positive.

Furthermore, note that

$$\begin{aligned}
&\mathcal{B}_2\geq 1+A_P\\
&\iff \left(1+\frac{A_N\phi_{P|N}}{A_P\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)}\right)\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)\geq 1+A_P\\
&\iff \left(A_P^3\phi_{P|P}+\left(1-A_P\right)A_NA_P\phi_{P|N}\right)\left(\phi_{P|P}-\phi_{P|N}\right)\geq A_P^3\phi_{P|P}-\left(1+A_P^2\right)A_N\phi_{P|N}\\
&\iff A_N\phi_{P|N}\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)\geq A_P^2\left(1-\left(\phi_{P|P}-\phi_{P|N}\right)\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right), \qquad \text{(B303)}
\end{aligned}$$

and

$$A_P^2\geq A_P^2\phi_{P|P}+A_N\left(1-A_P\right)\phi_{P|N}\iff 1-\phi_{P|P}\geq\left(\frac{A_N\left(1-A_P\right)}{A_P^2}\right)\phi_{P|N}. \qquad \text{(B304)}$$

Then, suppose $1-\phi_{P|P}\geq\left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$ and

$$A_N\phi_{P|N}\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)\geq A_P^2\left(1-\left(\phi_{P|P}-\phi_{P|N}\right)\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right). \qquad \text{(B305)}$$

If $0\leq\hat{B}<1+A_P$, then $\hat{F}^*=\hat{B}$. If $1+A_P\leq\hat{B}<\mathcal{B}_2$, then $\hat{F}^*=\hat{B}$. If $\mathcal{B}_2\leq\hat{B}<\mathcal{B}_1$, then $\hat{F}^*=\hat{B}-1-A_P$. If $\mathcal{B}_1\leq\hat{B}<1+A_P\left(\frac{A_P}{A_N}-1\right)$, then the formative assessment is always optimal. If $1+A_P\left(\frac{A_P}{A_N}-1\right)\leq\hat{B}$, then $\hat{F}^*=\hat{B}-1-A_P\left(\frac{A_P}{A_N}-1\right)$.

Next, suppose $1-\phi_{P|P}<\left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$ and

$$A_N\phi_{P|N}\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)\geq A_P^2\left(1-\left(\phi_{P|P}-\phi_{P|N}\right)\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right). \qquad \text{(B306)}$$

If $0\leq\hat{B}<1+A_P$, then $\hat{F}^*=\hat{B}$. If $1+A_P\leq\hat{B}<\mathcal{B}_2$, then $\hat{F}^*=\hat{B}$. If $\mathcal{B}_2\leq\hat{B}<\mathcal{B}_1$, then the formative assessment is always optimal. If $\mathcal{B}_1\leq\hat{B}<1+A_P\left(\frac{A_P}{A_N}-1\right)$, then the formative assessment is always optimal. If $1+A_P\left(\frac{A_P}{A_N}-1\right)\leq\hat{B}$, then $\hat{F}^*=\hat{B}-1-A_P\left(\frac{A_P}{A_N}-1\right)$.

Third, suppose $1-\phi_{P|P}\geq\left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$ and

$$A_N\phi_{P|N}\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)<A_P^2\left(1-\left(\phi_{P|P}-\phi_{P|N}\right)\right)\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right). \qquad \text{(B307)}$$

If $0 \le \hat{B} < \mathcal{B}_2$, then $\hat{F}^* = \hat{B}$. If $\mathcal{B}_2 \le \hat{B} < 1 + A_P$, then the formative assessment is always optimal. If $1 + A_P \le \hat{B} < \mathcal{B}_1$, then $\hat{F}^* = \hat{B} - 1 - A_P$. If $\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$, then the formative assessment is always optimal. If $1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}$, then $\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right)$.

Finally, suppose

$$A_N\phi_{P|N}\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right) < A_P^2\left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right) \tag{B308}$$

and $1 - \phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$. If $0 \le \hat{B} < \mathcal{B}_2$, then $\hat{F}^* = \hat{B}$. If $\mathcal{B}_2 \le \hat{B} < 1 + A_P$, then the formative assessment is always optimal. If $1 + A_P \le \hat{B} < \mathcal{B}_1$, then the formative assessment is always optimal. If $\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$, then the formative assessment is always optimal. If $1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}$, then $\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right)$.

We can combine these results as follows.

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $\max\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N} \le 1 - \phi_{P|P}$, $A_N\phi_{P|N}\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right) \ge A_P^2\left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)$, and $q = 0$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < \mathcal{B}_2, \\ \hat{B} - 1 - A_P, & \text{if } \mathcal{B}_2 \le \hat{B} < \mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B309}$$

and the interim assessment is never the optimal choice if

$$\mathcal{B}_1 \le \hat{B} \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B310}$$

Additionally, $\hat{B}_L = 1 + A_P$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $\max\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N} \le 1 - \phi_{P|P}$, $A_N\phi_{P|N}\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right) < A_P^2\left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)$, and $q = 0$, then

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < \mathcal{B}_2, \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \le \hat{B} < \mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B311}$$

and the interim assessment is never the optimal choice if

$$\mathcal{B}_2 \le \hat{B} \le \hat{B} < 1 + A_P \text{ and } \mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B312}$$

Additionally, $\hat{B}_L = \mathcal{B}_2$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

Finally, if $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $\left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N} \le 1 - \phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, and $q = 0$, then

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \le \hat{B} < \mathcal{B}_2, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \le \hat{B}, \end{cases} \tag{B313}$$

and the interim assessment is never the optimal choice if

$$\mathcal{B}_2 \le \hat{B} \le \hat{B} < 1 + A_P \left(\frac{A_P}{A_N} - 1 \right). \tag{B314}$$

Additionally,

$$\hat{B}_L = \begin{cases} 1 + A_P, & \text{if } A_N \phi_{P|N} \left(1 + A_P \left(\phi_{P|P} - \phi_{P|N}\right)\right) \ge A_P^2 \left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right) \left(A_P \phi_{P|P} - A_N \phi_{P|N}\right), \\ \mathcal{B}_2, & \text{if } A_N \phi_{P|N} \left(1 + A_P \left(\phi_{P|P} - \phi_{P|N}\right)\right) < A_P^2 \left(1 - \left(\phi_{P|P} - \phi_{P|N}\right)\right) \left(A_P \phi_{P|P} - A_N \phi_{P|N}\right), \end{cases} \tag{B315}$$

and $\hat{B}_U = 1 + A_P \left(\frac{A_P}{A_N} - 1 \right)$.

For $0 < q < 1$, we consider three possible subcases:

- $1 + A_P < \mathcal{B}_3$
- $\mathcal{B}_3 \le 1 + A_P < q A_P^2 \mathcal{R}_1 + (1 - q) \mathcal{B}_1$
- $q A_P^2 \mathcal{R}_1 + (1 - q) \mathcal{B}_1 \le 1 + A_P$

First,

$$\begin{aligned} 1 + A_P < \mathcal{B}_3 &\iff \frac{1 + A_P}{\mathcal{R}_2} - A_P^2 \phi_{P|P} - A_N (1 - A_P) \phi_{P|N} < q \left(A_P^2 - A_P^2 \phi_{P|P} - A_N (1 - A_P) \phi_{P|N} \right) \\ &\iff \begin{cases} \mathcal{Q}_3 < q, & \text{if } 1 - \phi_{P|P} > \left(\frac{A_N (1 - A_P)}{A_P^2} \right) \phi_{P|N}, \\ \mathcal{Q}_3 > q, & \text{if } 1 - \phi_{P|P} < \left(\frac{A_N (1 - A_P)}{A_P^2} \right) \phi_{P|N}, \\ A_P - A_N < \frac{A_N}{A_P^2}, & \text{if } 1 - \phi_{P|P} = \left(\frac{A_N (1 - A_P)}{A_P^2} \right) \phi_{P|N}, \end{cases} \\ &\iff \begin{cases} \mathcal{Q}_3 < q, & \text{if } 1 - \phi_{P|P} > \left(\frac{A_N (1 - A_P)}{A_P^2} \right) \phi_{P|N}, \\ 1 > q, & \text{if } 1 - \phi_{P|P} < \left(\frac{A_N (1 - A_P)}{A_P^2} \right) \phi_{P|N}, \\ A_P - A_N < \frac{A_N}{A_P^2}, & \text{if } 1 - \phi_{P|P} = \left(\frac{A_N (1 - A_P)}{A_P^2} \right) \phi_{P|N}, \end{cases} \end{aligned} \tag{B316}$$

where, as stated in (B220),

$$\mathcal{Q}_3 = \frac{\frac{1 + A_P}{\mathcal{R}_2} - A_P^2 \phi_{P|P} - A_N (1 - A_P) \phi_{P|N}}{A_P^2 - A_P^2 \phi_{P|P} - A_N (1 - A_P) \phi_{P|N}}, \tag{B317}$$

and from Lemma B6, we apply that if $\left(\frac{A_P (A_P - A_N) - A_N}{A_P} \right) \phi_{P|N} \le 1 - \phi_{P|P} < \left(\frac{A_N (1 - A_P)}{A_P^2} \right) \phi_{P|N}$, then $\mathcal{Q}_3 \ge 1$. Furthermore, recall that we assume $q < 1$ in this case.

Second,

$$\begin{aligned} q A_P^2 \mathcal{R}_1 + (1 - q) \mathcal{B}_1 \le 1 + A_P &\iff q \left(A_P^2 \mathcal{R}_1 - \mathcal{B}_1 \right) \le 1 + A_P - \mathcal{B}_1 \\ &\iff q \ge \mathcal{Q}_4, \end{aligned} \tag{B318}$$

where, as stated in (B221),

$$\mathcal{Q}_4 = \frac{\mathcal{B}_1 - 1 - A_P}{\mathcal{B}_1 - A_P^2 \mathcal{R}_1}, \tag{B319}$$

and we use that

$$A_P^2 \mathcal{R}_1 < \mathcal{B}_1 \iff 0 < A_P \left(1 - \phi_{P|P}\right) + (1 - A_P) \left(1 - \phi_{P|N}\right), \tag{B320}$$

which holds under Assumptions 1 and 2. However, from Lemma B6, $\mathcal{Q}_4 > 1$ when $(A_P, A_N) \in \mathcal{S}_4$. Therefore, this case is never feasible.

Third,

$$\begin{aligned}
&\mathcal{B}_3 \leq 1 + A_P < q A_P^2 \mathcal{R}_1 + (1-q)\mathcal{B}_1 \\
&\iff \begin{cases} q \leq \mathcal{Q}_3 \text{ and } q < \mathcal{Q}_4, & \text{if } 1 - \phi_{P|P} > \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}, \\ \mathcal{Q}_3 \leq q < \mathcal{Q}_4, & \text{if } 1 - \phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}, \\ A_P - A_N \geq \frac{A_N}{A_P^2} \text{ and } q < \mathcal{Q}_4, & \text{if } 1 - \phi_{P|P} = \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}, \end{cases} \\
&\iff \begin{cases} q \leq \mathcal{Q}_3, & \text{if } 1 - \phi_{P|P} > \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}, \\ A_P - A_N \geq \frac{A_N}{A_P^2}, & \text{if } 1 - \phi_{P|P} = \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}, \end{cases}
\end{aligned} \tag{B321}$$

where we apply Lemma B6.

In all cases, if $0 \leq \hat{B} < q(1 + A_P)$,

$$\hat{F}^* = \hat{B}. \tag{B322}$$

For larger budget levels, we must separately consider each of the cases described above.

Subcase 1: Suppose one of the followings sets of conditions holds:

$$\begin{aligned}
&\mathcal{Q}_3 < q \text{ and } 1 - \phi_{P|P} > \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}, \\
&\text{or } 1 - \phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}, \\
&\text{or } A_P - A_N < \frac{A_N}{A_P^2} \text{ and } 1 - \phi_{P|P} = \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}.
\end{aligned} \tag{B323}$$

Then, $1 + A_P < \mathcal{B}_3$.

If $q(1 + A_P) \leq \hat{B} < 1 + A_P$, the interim assessment is never optimal. If $1 + A_P \leq \hat{B} < \mathcal{B}_3$ and $q < 1$,

$$\hat{F}^* = \hat{B} - 1 - A_P. \tag{B324}$$

If $\mathcal{B}_3 \leq \hat{B} < q A_P^2 \mathcal{R}_1 + (1-q)\mathcal{B}_1$, the probability of final proficiency under the interim assessment exceeds that under the formative if and only if

$$\begin{aligned}
&A_P^2 \geq q A_P^2 + (1-q)\left(A_P^2 \phi_{P|P} + A_N (1-A_P)\phi_{P|N}\right) \\
&\iff A_P^2 \left(1 - \phi_{P|P}\right) - A_N (1-A_P)\phi_{P|N} \geq q\left(A_P^2\left(1-\phi_{P|P}\right) - A_N(1-A_P)\phi_{P|N}\right) \\
&\iff 1 - \phi_{P|P} \geq \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}.
\end{aligned} \tag{B325}$$

Then, if $1-\phi_{P|P} \geq \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$,

$$\hat{F}^* = \hat{B} - 1 - A_P. \tag{B326}$$

If instead $1-\phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, the interim assessment is not optimal.

If $qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$ and $q<1$, the interim assessment is never optimal.

Subcase 2: Suppose one of the following sets of conditions holds:

$$\begin{aligned} q \leq \mathcal{Q}_3 \text{ and } 1-\phi_{P|P} &> \left(\frac{A_N\,(1-A_P)}{A_P^2}\right)\phi_{P|N}, \\ \text{or } A_P - A_N \geq \frac{A_N}{A_P^2} \text{ and } 1-\phi_{P|P} &= \left(\frac{A_N\,(1-A_P)}{A_P^2}\right)\phi_{P|N}. \end{aligned} \tag{B327}$$

Then, $\mathcal{B}_3 \leq 1 + A_P < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1$.

If $q(1+A_P) \leq \hat{B} < 1+A_P$, the interim assessment is never optimal. If $1+A_P \leq \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1$, the probability under the interim assessment exceeds that under the formative if and only if

$$\begin{aligned} A_P^2 &\geq qA_P^2 + (1-q)\left(A_P^2\phi_{P|P} + A_N\,(1-A_P)\,\phi_{P|N}\right) \\ \iff 1-\phi_{P|P} &\geq \left(\frac{A_N\,(1-A_P)}{A_P^2}\right)\phi_{P|N}, \end{aligned} \tag{B328}$$

which is a necessary condition in this case, as stated in (B327). Then,

$$\hat{F}^* = \hat{B} - 1 - A_P. \tag{B329}$$

If $qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$, the interim assessment is never optimal.

Finally, for all q, if $1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}$,

$$\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B330}$$

Combining this, if $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $\max\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N} \leq 1-\phi_{P|P}$, and $0<q<1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < q(1+A_P), \\ \hat{B} - 1 - A_P, & \text{if } 1+A_P \leq \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B331}$$

and the interim assessment is never the optimal choice if

$$q\,(1+A_P) \leq \hat{B} < 1 + A_P \text{ or } qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B332}$$

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $\left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N} \leq 1-\phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, and $0 < q < 1$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < q(1+A_P), \\ \hat{B} - 1 - A_P, & \text{if } 1+A_P \leq \hat{B} < \mathcal{B}_3, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N}-1\right) \leq \hat{B}, \end{cases} \tag{B333}$$

and the interim assessment is never the optimal choice if

$$q(1+A_P) \leq \hat{B} < 1 + A_P \text{ or } \mathcal{B}_3 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B334}$$

Additionally, in both cases, $\hat{B}_L = q\left(1+A_P\right)$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N}-1\right)$.

Case 5: $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, *and* $1-\phi_{P|P} < \left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N}$. If $z_I = 0$,

$$Pr^*_{z_I}[\mathbf{S}_2 = P|\mathbf{S}_0 = P] = \begin{cases} 0, & \text{if } 0 \leq \hat{B} < (1-q)\,\mathcal{B}_2, \\ (1-q)\left(A_P^2\phi_{P|P} + A_N\left(1-A_P\right)\phi_{P|N}\right), & \\ & \text{if } (1-q)\,\mathcal{B}_2 \leq \hat{B} < \mathcal{B}_4, \\ qA_P^2 + (1-q)\left(A_P^2\phi_{P|P} + A_N\left(1-A_P\right)\phi_{P|N}\right), & \\ & \text{if } \mathcal{B}_4 \leq \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1, \\ A_P^2 + (1-q)\,A_N\left(1-A_P\right), & \\ & \text{if } qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N}-1\right), \\ A_P^2 + A_N\left(1-A_P\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N}-1\right) \leq \hat{B} - \hat{F}z_I. \end{cases} \tag{B335}$$

Recall that

$$(A_P, A_N) \in \mathcal{S}_4 \iff \frac{A_P^2\left(1-\phi_{P|P}\right) + A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} < A_P - A_N. \tag{B336}$$

We consider four possible cases:

- $1 + A_P < (1-q)\,\mathcal{B}_2$
- $(1-q)\,\mathcal{B}_2 \leq 1 + A_P < \mathcal{B}_4$
- $\mathcal{B}_4 \leq 1 + A_P < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1$
- $qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \leq 1 + A_P$

First,

$$1 + A_P < (1-q)\,\mathcal{B}_2 \iff q < \mathcal{Q}_5, \tag{B337}$$

where from (44),

$$\mathcal{B}_2 = \left(1 + \frac{A_N\phi_{P|N}}{A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}\right)\left(1 + A_P\left(\phi_{P|P} - \phi_{P|N}\right)\right), \tag{B338}$$

and from (B222) and Lemma B6,

$$\mathcal{Q}_5 = 1 - \frac{A_P\left(1+A_P\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}{\left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)\left(A_P\left(A_P\phi_{P|P}-A_N\phi_{P|N}\right)+A_N\phi_{P|N}\right)} < 1. \tag{B339}$$

Additionally, from (46),

$$\mathcal{B}_4 = \left(1+A_P\right)\left(q+\left(1-q\right)\left(\frac{A_P^2\phi_{P|P}+A_N\left(1-A_P\right)\phi_{P|N}}{A_P^2}\right)\right). \tag{B340}$$

Since $\left(1-q\right)\mathcal{B}_2 < \mathcal{B}_4$, then

$$\begin{aligned} 1+A_P < \left(1-q\right)\mathcal{B}_2 \Rightarrow 1+A_P < \mathcal{B}_4 & \\ \iff A_P^2\left(1-\phi_{P|P}\right) - A_N\left(1-A_P\right)\phi_{P|N} &< q\left(A_P^2\left(1-\phi_{P|P}\right)-A_N\left(1-A_P\right)\phi_{P|N}\right) \\ \iff 1-\phi_{P|P} &< \left(\frac{A_N\left(1-A_P\right)}{A_P^2}\right)\phi_{P|N}. \end{aligned} \tag{B341}$$

Second,

$$\left(1-q\right)\mathcal{B}_2 \le 1+A_P < \mathcal{B}_4 \iff q \ge \mathcal{Q}_5 \text{ and } 1-\phi_{P|P} < \left(\frac{A_N\left(1-A_P\right)}{A_P^2}\right)\phi_{P|N}. \tag{B342}$$

Third,

$$1+A_P < qA_P^2\mathcal{R}_1 + \left(1-q\right)\mathcal{B}_1 \iff q < \mathcal{Q}_2, \tag{B343}$$

so

$$\begin{aligned} \mathcal{B}_4 \le 1+A_P < qA_P^2\mathcal{R}_1+\left(1-q\right)\mathcal{B}_1 &\iff q<\mathcal{Q}_2 \text{ and } 1-\phi_{P|P} \ge \left(\frac{A_N\left(1-A_P\right)}{A_P^2}\right)\phi_{P|N} \\ &\iff 1-\phi_{P|P} \ge \left(\frac{A_N\left(1-A_P\right)}{A_P^2}\right)\phi_{P|N}, \end{aligned} \tag{B344}$$

where we use $q<1$ and, from Lemma B6, $\mathcal{Q}_2 \ge 1$.

Fourth,

$$qA_P^2\mathcal{R}_1+\left(1-q\right)\mathcal{B}_1 \le 1+A_P \iff q \ge \mathcal{Q}_2, \tag{B345}$$

but under Lemma B6, this is never feasible.

Using these results, we must consider three cases. In all cases, if $0 \le \hat{B} < \left(1-q\right)\mathcal{B}_2$,

$$\hat{F}^* = \hat{B}. \tag{B346}$$

For larger budget levels, we must separately consider each case.

Subcase 1: Suppose $q < \mathcal{Q}_5$ and $1-\phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$. If $(1-q)\,\mathcal{B}_2 \le \hat{B} < \mathcal{B}_4$, the probability under the interim assessment exceeds that under the formative if and only if

$$A_P^2 \ge (1-q)\left(A_P^2\phi_{P|P} + A_N\left(1-A_P\right)\phi_{P|N}\right) \iff q \ge \mathcal{Q}_6, \tag{B347}$$

where from (B223),

$$\mathcal{Q}_6 = \frac{A_N\left(1-A_P\right)\phi_{P|N} - A_P^2\left(1-\phi_{P|P}\right)}{A_P^2\phi_{P|P} + A_N\left(1-A_P\right)\phi_{P|N}}. \tag{B348}$$

However, this is never feasible in this case, since

$$\begin{aligned}\mathcal{Q}_6 \le \mathcal{Q}_5 \iff & A_P\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)\left(1-\left(\phi_{P|P}-\phi_{P|N}\right)\right) \le \\ & \left(1+A_P\left(\phi_{P|P}-\phi_{P|N}\right)\right)\left(A_N\phi_{P|N} + A_P\left(1-\phi_{P|P}\right)\right) \\ \iff & \left(\frac{A_P\left(A_P-A_N\right)-A_N}{A_P}\right)\phi_{P|N} \le 1-\phi_{P|P},\end{aligned} \tag{B349}$$

which contradicts a condition for this case. Then, the interim assessment is never optimal.

If $\mathcal{B}_4 \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1$, the probability under the interim assessment exceeds that under the formative if and only if

$$A_P^2 \ge qA_P^2 + (1-q)\left(A_P^2\phi_{P|P} + A_N\left(1-A_P\right)\phi_{P|N}\right) \iff 1-\phi_{P|P} \ge \left(\frac{A_N\left(1-A_P\right)}{A_P^2}\right)\phi_{P|N}, \tag{B350}$$

which violates a condition in this case. Then, the interim assessment is never optimal.

If $qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N}-1\right)$ and $q<1$, the interim assessment is never optimal.

Subcase 2: Suppose $q \ge \mathcal{Q}_5$ and $1-\phi_{P|P} < \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$. If $(1-q)\,\mathcal{B}_2 \le \hat{B} < 1 + A_P$, the interim assessment is never optimal. If $1 + A_P \le \hat{B} < \mathcal{B}_4$, the probability under the interim assessment exceeds that under the formative if and only if (B347) holds. As shown in (B349), $\mathcal{Q}_6 > \mathcal{Q}_5$, in this case. Therefore, if $q \ge \mathcal{Q}_6$,

$$\hat{F}^* = \hat{B} - 1 - A_P, \tag{B351}$$

and if $\mathcal{Q}_5 \le q < \mathcal{Q}_6$, the interim assessment is never optimal.

If $\mathcal{B}_4 \le \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1$, the probability under the interim assessment exceeds that under the formative if and only if (B350) holds, but this violates a condition of this case. Therefore, the interim assessment is never optimal.

If $qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \le \hat{B} < 1 + A_P\left(\frac{A_P}{A_N}-1\right)$ and $q<1$, the interim assessment is never optimal.

Subcase 3: Suppose $1-\phi_{P|P} \geq \left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$. If $(1-q)\mathcal{B}_2 \leq \hat{B} < 1+A_P$, the interim assessment is never optimal.

If $1 + A_P \leq \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1$, the probability under the interim assessment exceeds that under the formative if and only if (B350) holds, which is a necessary condition in this case. Then,

$$\hat{F}^* = \hat{B} - 1 - A_P. \tag{B352}$$

If $qA_P^2\mathcal{R}_1 + (1-q)\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$ and $q < 1$, the interim assessment is never optimal.

Finally, for all q, if $1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}$,

$$\hat{F}^* = \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B353}$$

Combining this, if $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $q < \mathcal{Q}_5$, and $1 - \phi_{P|P} < \min\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\mathcal{B}_2, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B354}$$

and the interim assessment is never the optimal choice if

$$(1-q)\mathcal{B}_2 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B355}$$

Additionally, in this case, $\hat{B}_L = 1 + A_P$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $1 - \phi_{P|P} < \min\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, and $q \geq \mathcal{Q}_6$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\mathcal{B}_2, \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \leq \hat{B} < \mathcal{B}_4, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B356}$$

and the interim assessment is never the optimal choice if

$$(1-q)\mathcal{B}_2 \leq \hat{B} < 1 + A_P \text{ or } \mathcal{B}_4 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B357}$$

Additionally, in this case, $\hat{B}_L = (1-q)\mathcal{B}_2$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, $1 - \phi_{P|P} < \min\left(\frac{A_P(A_P-A_N)-A_N}{A_P}, \frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N}$, and $\mathcal{Q}_5 \leq q < \mathcal{Q}_6$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\mathcal{B}_2, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B358}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_2 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B359}$$

Additionally, in this case, $\hat{B}_L = (1-q)\,\mathcal{B}_2$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

If $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_4$, and $\left(\frac{A_N(1-A_P)}{A_P^2}\right)\phi_{P|N} \leq 1 - \phi_{P|P} < \left(\frac{A_P(A_P-A_N)-A_N}{A_P}\right)\phi_{P|N}$,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\,\mathcal{B}_2, \\ \hat{B} - 1 - A_P, & \text{if } 1 + A_P \leq \hat{B} < qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1, \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B360}$$

and the interim assessment is never the optimal choice if

$$(1-q)\,\mathcal{B}_2 \leq \hat{B} < 1 + A_P \text{ or } qA_P^2\mathcal{R}_1 + (1-q)\,\mathcal{B}_1 \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B361}$$

Additionally, in this case, $\hat{B}_L = (1-q)\,\mathcal{B}_2$ and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

□

Proof of Proposition 4

a) This follows from the statement of Proposition B2a, since $B_L = B_U$ in this case.

b) We first show that for $\phi_{P|N}$ sufficiently close to $\phi_{P|P}$, $\mathcal{S}_4$ is the empty set for all values of A_P and A_N. We then consider the optimal choice of assessment for $(A_P, A_N) \in \mathcal{S}_2 \cup \mathcal{S}_3$.

Recall from (37) that

$$\mathcal{S}_4 = \{(A_P, A_N)\,|\,\mathcal{A}_2 < A_P - A_N\}, \tag{B362}$$

where from (31),

$$\mathcal{A}_2 = \frac{A_P^2\left(1-\phi_{P|P}\right) + A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)}. \tag{B363}$$

Now,

$$\begin{aligned} &\frac{A_P^2\left(1-\phi_{P|P}\right) + A_N\left(1-A_P\right)\left(1-\phi_{P|N}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right)} < A_P - A_N \\ &\iff \frac{A_P^2\left(1-\phi_{P|P}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)} + A_N < \left(A_P - A_N\right)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right). \end{aligned} \tag{B364}$$

The left-hand side of the inequality is increasing in $\phi_{P|N}$ and the right-hand side of the inequality is decreasing in $\phi_{P|N}$. Taking the limit of each side as $\phi_{P|N}$ approaches $\phi_{P|P}$ gives the following results:

$$\lim_{\phi_{P|N}\to\phi_{P|P}} \frac{A_P^2\left(1-\phi_{P|P}\right)}{\left(1-A_P\right)\left(1-\phi_{P|N}\right)} + A_N = \frac{A_P^2}{\left(1-A_P\right)} + A_N \tag{B365}$$

and

$$\lim_{\phi_{P|N}\to\phi_{P|P}} (A_P - A_N)\left(A_P\phi_{P|P} - A_N\phi_{P|N}\right) = \phi_{P|P}(A_P - A_N)^2. \tag{B366}$$

Then, for $\phi_{P|N}$ sufficiently large, (B364) cannot hold, since

$$\frac{A_P^2}{(1-A_P)} + A_N < \phi_{P|P}(A_P - A_N)^2 \iff \frac{A_P^2}{(1-A_P)(A_P-A_N)^2} + A_N < \phi_{P|P}, \tag{B367}$$

but the left-hand side of the inequality is greater than 1 and the right-hand side is less than or equal to 1. Then, there exists $\bar{\phi}_{P|N,4} < \phi_{P|P}$ such that for all $\phi_{P|N} \geq \bar{\phi}_{P|N,4}$, $\mathcal{S}_4$ is null.

Next, we consider the district's optimal policy when q is small $\left(q < \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}\right)$, $\phi_{P|N}$ sufficiently large ($\phi_{P|N} \geq \bar{\phi}_{P|N,4}$), and A_N and A_P sufficiently different ($A_P - A_N > \frac{A_N}{A_P^2}$, so $(A_P, A_N) \in \mathcal{S}_2 \cup \mathcal{S}_3$). Notice that $\frac{A_N}{A_P^2} \geq \frac{A_N}{A_P}$, so when $A_P < 1$, the feasible set of (A_P, A_N) is a subset of $\mathcal{S}_2 \cup \mathcal{S}_3$.

When $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_2$, and $q < \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}$, then F^*, B_L, and B_U are given by Case d in Proposition B2 ((B194) and (B195)). Recall that in this case,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\left(1 + \frac{A_N}{A_P(A_P-A_N)}\right), \\ \hat{B} - 1 - A_P\left(\frac{A_P}{A_N} - 1\right), & \text{if } 1 + A_P\left(\frac{A_P}{A_N} - 1\right) \leq \hat{B}, \end{cases} \tag{B368}$$

and the interim assessment is never optimal, for any value of $\hat{F}$, if

$$(1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right) \leq \hat{B} < 1 + A_P\left(\frac{A_P}{A_N} - 1\right). \tag{B369}$$

Additionally,

$$\hat{B}_L = \begin{cases} 1 + A_P, & \text{if } q \leq \frac{A_N - A_P^2(A_P-A_N)}{A_P(A_P-A_N)+A_N}, \\ (1-q)\left(1 + \frac{A_N}{A_P(A_P-A_N)}\right), & \text{if } \frac{A_N - A_P^2(A_P-A_N)}{A_P(A_P-A_N)+A_N} < q < \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}, \end{cases} \tag{B370}$$

and $\hat{B}_U = 1 + A_P\left(\frac{A_P}{A_N} - 1\right)$.

Now, q is non-negative by assumption, but

$$\frac{A_N}{A_P^2} < A_P - A_N \Rightarrow \frac{A_N - A_P^2(A_P - A_N)}{A_P(A_P - A_N) + A_N} < 0. \tag{B371}$$

Therefore, for $0 \leq q < \frac{A_N(1-A_P)}{A_P^2 + A_N(1-A_P)}$,

$$\hat{B}_L = (1-q)\left(1 + \frac{A_N}{A_P(A_P - A_N)}\right). \tag{B372}$$

Then, for all $\hat{B} \in \left[\hat{B}_L, \hat{B}_U\right)$, only the formative assessment is optimal.

Next, we determine the optimal assessment when $(A_P, A_N) \in \mathcal{S}_3$. When $\mathbf{S}_0 = P$, $(A_P, A_N) \in \mathcal{S}_3$, and $q = 0$, then F^*, B_L, and B_U are given by Case f in Proposition B2 ((B199) and (B200)). To determine these values for non-zero q, consider the following inequality:

$$\left(1 - (A_P - A_N) - \frac{A_N}{A_P^2}\right)\left(1 - \phi_{P|N}\right) \leq 1 - \phi_{P|P}. \tag{B373}$$

The left-hand side of the above inequality is decreasing in $\phi_{P|N}$ and the right-hand side is fixed. Then,

$$\lim_{\phi_{P|N}\to\phi_{P|P}} \left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right) = \left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|P}\right), \tag{B374}$$

and

$$\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|P}\right) \leq 1-\phi_{P|P} \iff -(A_P-A_N)-\frac{A_N}{A_P^2} \leq 0, \tag{B375}$$

which always holds under Assumption 1. In fact, the left-hand side of the inequality is always strictly less than 0. Therefore, there exists $\bar{\phi}_{P|N,3} < \phi_{P|P}$ such that for all $\phi_{P|N} \geq \bar{\phi}_{P|N,3}$, (B373) holds. Then, for $\phi_{P|N} \geq \bar{\phi}_{P|N,3}$, this Case f in Proposition B2 holds for all $q < \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}$. Recall that in this case,

$$\hat{F}^* = \begin{cases} \hat{B}, & \text{if } 0 \leq \hat{B} < (1-q)\mathcal{B}_1, \\ \hat{B}-1-A_P\left(\frac{A_P}{A_N}-1\right), & \text{if } 1+A_P\left(\frac{A_P}{A_N}-1\right) \leq \hat{B}, \end{cases} \tag{B376}$$

and the interim assessment is never the optimal choice if

$$(1-q)\mathcal{B}_1 \leq \hat{B} < 1+A_P\left(\frac{A_P}{A_N}-1\right). \tag{B377}$$

Additionally, when $\left(1-(A_P-A_N)-\frac{A_N}{A_P^2}\right)\left(1-\phi_{P|N}\right) \leq 1-\phi_{P|P}$,

$$\hat{B}_L = \begin{cases} 1+A_P, & \text{if } q < \mathcal{Q}_1, \\ (1-q)\mathcal{B}_1, & \text{if } \mathcal{Q}_1 \leq q < \frac{A_N(1-A_P)}{A_P^2+A_N(1-A_P)}, \end{cases} \tag{B378}$$

where from (B218),

$$\mathcal{Q}_1 = \frac{A_P}{A_P^2+A_N(1-A_P)}\left(\frac{(A_N(1-A_P)-A_P)\left(1-\phi_{P|P}\right)+(1-A_P)(A_P-2A_N)\left(1-\phi_{P|N}\right)}{A_P\left(1-\phi_{P|P}\right)+(1-A_P)\left(1-\phi_{P|N}\right)}\right). \tag{B379}$$

In all cases, $\hat{B}_U = 1+A_P\left(\frac{A_P}{A_N}-1\right)$.

Now, $\mathcal{Q}_1$ is a non-increasing function of $\phi_{P|N}$, since

$$\frac{\partial \mathcal{Q}_1}{\partial \phi_{P|N}} = -\frac{A_P(1-A_P)(A_P(A_P-A_N+1-A_N)+A_N(1-A_P))\left(1-\phi_{P|P}\right)}{\left(A_P^2+A_N(1-A_P)\right)\left(A_P\left(1-\phi_{P|P}\right)+(1-A_P)\left(1-\phi_{P|N}\right)\right)^2} \leq 0. \tag{B380}$$

Note that when $\phi_{P|P}=1$, Q_1 is positive and no longer depends on $\phi_{P|N}$. For $\phi_{P|P}<1$,

$$\begin{aligned} &\mathcal{Q}_1 = 0 \\ &\iff \frac{A_P}{A_P^2+A_N(1-A_P)}\left(\frac{(A_N(1-A_P)-A_P)\left(1-\phi_{P|P}\right)+(1-A_P)(A_P-2A_N)\left(1-\phi_{P|N}\right)}{A_P\left(1-\phi_{P|P}\right)+(1-A_P)\left(1-\phi_{P|N}\right)}\right) = 0 \\ &\iff (1-A_P)(A_P-2A_N)\left(1-\phi_{P|N}\right) = (A_P-A_N(1-A_P))\left(1-\phi_{P|P}\right) \\ &\iff \phi_{P|N} = 1-\frac{(A_P-A_N(1-A_P))\left(1-\phi_{P|P}\right)}{(1-A_P)(A_P-2A_N)}. \end{aligned} \tag{B381}$$

This value is bounded above by $\phi_{P|P}$, since

$$\begin{aligned} 1-\frac{\left(A_P-A_N\left(1-A_P\right)\right)\left(1-\phi_{P|P}\right)}{\left(1-A_P\right)\left(A_P-2A_N\right)}<\phi_{P|P} \\ \iff 1-\phi_{P|P}<\frac{\left(A_P-A_N\left(1-A_P\right)\right)\left(1-\phi_{P|P}\right)}{\left(1-A_P\right)\left(A_P-2A_N\right)} \\ \iff -\frac{A_P^2}{\left(1-A_P\right)}<A_N, \end{aligned} \tag{B382}$$

which always holds under Assumption 1. Therefore, for $\phi_{P|P}<1$ and $\phi_{P|N}>1-\frac{(A_P-A_N(1-A_P))\left(1-\phi_{P|P}\right)}{(1-A_P)(A_P-2A_N)}$, the formative assessment is the optimal choice for all $\hat{B}\in\left[\hat{B}_L,\hat{B}_U\right)$.

Finally, let

$$\bar{\phi}_{P|N}=\max\left(\bar{\phi}_{P|N,3},\bar{\phi}_{P|N,4},1-\frac{\left(A_P-A_N\left(1-A_P\right)\right)\left(1-\phi_{P|P}\right)}{\left(1-A_P\right)\left(A_P-2A_N\right)}\right). \tag{B383}$$

Then, for $\phi_{P|P}<1$, $\phi_{P|N}\in\left(\bar{\phi}_{P|N},\phi_{P|P}\right]$, and $A_P-A_N>\frac{A_N}{A_P^2}$, the formative assessment is always optimal.

□