Teacher professional development as a signal*

Inefficient sorting and its implications for optimal training provision

Olga Meshcheriakova,** Stan Vermeulen,***

Abstract

Providing on-the-job training opportunities for teachers is a key policy tool for improving the quality of education. However, without an understanding of teacher incentives, it is unclear whether training will be taken by those who would improve their teaching the most from attending the program. In this paper we develop a model on the decision to enter on-the-job training for teachers. We argue that because performance is imperfectly observable, teachers will incorporate the signaling value of signing up for training into their decision. When participation is voluntary, sorting into training is inefficient: programs aimed at low ability teachers will be underutilized while programs aimed at the top will be overused. We show that offering training has spillover effects: introducing advanced courses increases participation in basic courses. We discuss the implications of these results for program evaluation and implementation, and make several testable predictions for assessing the validity of our model.

Keywords: Teachers, Training, Human Capital, Signaling

JEL: I2, J24, J45, M53

---

* The authors thank Timothy Bond, Lex Borghans, Tamás Dávid-Barrett, Mirko Draca, Bart Golsteyn, Susanna Loeb, Trudie Schils, and seminar participants at Maastricht University, Universität Tübingen, Kiel Institute for the World Economy, as well as participants at the European Association of Labour Economists meeting 2016 for their valuable comments.
** Department of Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands. Phone nr: +31 43 388 3841. o.meshcheriakova@maastrichtuniversity.nl
*** Department of Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands. c.vermeulen@maastrichtuniversity.nl

1. Introduction

The quality of teachers is an essential determinant of educational outcomes in society. According to Hanushek (2011), no other school attribute even comes close to its influence on student achievement. Therefore providing on-the-job training opportunities for teachers is a key instrument for improving the quality of education (European Commission, 2012; OECD, 2014), and policy makers devote a lot of resources to stimulate teachers to enter professional development programs. A recent study by Jacob and McGovern (2015) estimate that around 6 to 9% of school districts' operating budget is being used for teacher professional development in the United States. In the Netherlands, around 10% of the total budget allocated to schools is reserved for this purpose (Ministerie van Onderwijs, Cultuur en Wetenschap, 2014). Teachers also appear to make use of the professional development opportunities they are given. Results from the OECD's TALIS study show that, across the developed countries, 88% of teachers have participated in at least one professional development program in the past year (OECD, 2014). For the United States, Wei, Darling-Hammond and Adamson (2010) report a very similar participation rate (87.5% for subject content training) based on the 2008 Schools and Staffing Teacher Survey. However, less is known about how teachers select into available professional development programs. Without an understanding of teacher incentives, it is unclear whether training will be taken by those who would benefit the most from attending the program in terms of ability. This leads to difficulties in evaluating the true added value of the training program, and in designing the most efficient implementation strategy.

In this paper, we model the decision of teachers to enter on-the-job training. We argue that because performance is imperfectly observable, teachers will incorporate the signaling value of signing up for specific types of training into their decision making process. We show that when training participation is voluntary, sorting into training is inefficient. Programs aimed at improving teachers at the low end of the ability distribution will be underutilized, while programs that aim to improve the top end of the distribution will be overused. Furthermore, we show that in a situation where teachers can choose between a basic and an advanced training program increasing the attractiveness of the advanced program increases the participation rate for the basic program. The intuition behind this result is that improvements in an advanced training program stimulate more teachers to sign up for advanced training. This, in turn, decreases the negative signal associated with signing up for basic training, as the average ability of those who do not take any training decreases. Following the same logic, increasing the attractiveness of the basic program reduces participation in advanced programs. Thus, the availability of different types of training attenuates the signal of each particular training program.

These results have several implications. First, by investing in programs targeting the top of the ability distribution, policy makers can affect the entire pool of teachers and improve average ability overall. In other words, simply providing the option of an advanced training course can be an effective instrument to induce low ability teachers to self-select into training more efficiently. Secondly, evaluations of the effectiveness of single programs should take the potential for spillover effects into account. Even if an advanced course is ineffective on average, its mere existence increases participation in basic courses. Assuming the basic course is effective, this would lead to an increase in average teacher quality. Finally, when designing training programs, policy makers should take the signal the course will give into consideration. Our model predicts that as a course becomes more basic, fewer teachers will sign up for this course unless it is unrealistically effective.

Our model leads to several empirically testable predictions. First, teachers that gain more from sending a positive signal will sort into training based less on gains in objective ability than those for who the signal is of little value. This implies, for example, that teachers that are on a fixed term contract are predicted to sort into advanced training more than their colleagues on permanent contracts. Second, in schools which offer a higher variety of training programs, participation rates among teachers will be higher. Finally, schools that offer and stimulate more advanced training opportunities will have a higher average level of teacher quality, independent of the effectiveness of the advanced training programs.

Our paper contributes to the literature on teacher professional development. Studies on teacher professional development programs tend to focus on evaluating the effectiveness of particular programs (see e.g. Yoon et al. (2007) for an overview). Not much attention has been given to analysing how teachers choose among the different training options available to them. We show that even when a program is effective in raising teacher quality, it may fail to attract any participants because of the potential negative information it signals about teachers' initial ability. Conversely, we show that programs aimed at high ability teachers improve the efficiency of teacher sorting even if these programs themselves are ineffective on average. Therefore, we argue that evaluations of the effectiveness of single programs should take the potential for these types of spillover effects into account.

More generally, our paper also contributes to the on-the-job training literature. While there have been many papers on on-the-job training in general (e.g. Becker, 1962; Hashimoto, 1981; Acemoglu, 1997; Acemoglu & Pischke, 1998, 1999a, 1999b; Autor, 2001; Leuven, 2005), less consideration has been given to the decision to participate in training specifically for the public sector. As public sector employees, teachers face weak performance incentives and high levels of wage compression. Additionally, monitoring of performance is difficult in the teaching profession. Therefore, models that predict training participation for private sector employees do not completely capture the considerations of teachers.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of the relevant literature and motivation for our model. Section 3 introduces the teacher decision model. Section 4 discusses the empirical predictions and Section 5 concludes.

2. Motivation and Background Literature

In this section we will discuss some explanations for why employees sign up for on-the-job training, and argue why they do not fully capture teachers' considerations. First, we look at the returns to on-the-job training in terms of wages. Secondly, we consider intrinsic motivation as a reason for training participation. Finally, we introduce career concerns and make the case that these, combined with imperfectly observable performance, influence teachers' decision to participate in training because of the signal it confers to their employer.

In general, workers' decision to enter training programs can be thought of from a human capital point of view, in which they decide to enter when the immediate costs of training are outweighed by the net present value of future benefits. These benefits can be either internal (e.g. intrinsic motivation) or external (e.g. increased wages). Regarding external benefits, most empirical studies find that training in the workplace has a positive effect on employee wages (e.g. Frazis & Loewenstein, 2005; Bassanini, Booth, Brunello, De Paola, & Leuven, 2007; Hansson, 2008; Haelermans & Borghans, 2012; Fouarge, Schils, & de Grip, 2013;), although some studies do find nonsignificant returns as well (e.g. Leuven & Oosterbeek, 2004; Leuven & Oosterbeek, 2008). For teachers the future external benefits of training are not that obvious, as in most countries they are paid a fixed salary that is independent of their performance (Lazear, 2003). Teachers themselves also recognize a lack of external incentives to participate in training as a problem (OECD, 2014). One reason that there is no direct monetary benefit of being an excellent teacher, compared to being a poor one, is likely because teacher performance is difficult to quantify. Many factors interact in the educational production function to generate student outcomes, making it hard for schools to reward their teachers based on their objective output.

Training participation in absence of external incentives can occur when performance is internally valued by employees. The increase in performance is a future benefit on its own. Studies on job performance of public sector employees consider intrinsic motivation a main factor in explaining why they choose to exert effort in their job (Wilson, 1989; Dixit, 2002; Prendergast, 2007; Delfgaauw & Dur, 2008). When we think of training participation as a type of job effort that increases future performance, this argument could be an explanation for why teachers participate in on-the-job training. When teachers derive internal value from improving their students' outcomes, and they believe entering professional development programs has a positive influence on their teaching quality, they will enter training even without external incentives.

While we agree that intrinsic motivation plays a role in the training participation decision, we argue that it cannot fully account for the empirical findings in the professional development program effectiveness literature. If internal valuation of one's teaching ability is the main factor in explaining why teachers go into professional development programs, evaluations of these programs are biased towards finding a positive effect if self-selection is not controlled for. Especially considering that there is already a selection bias in which programs are likely to be evaluated. Pritchett (2002) shows that advocates of ineffective programs have an incentive to not subject their program to independent evaluations. Ineffective programs are less likely to be evaluated as a result. In the empirical literature, we therefore expect to find an overestimation of the true effectiveness of the average program.

However, most studies on the effect of specific professional development programs on teacher quality find very little impact on student performance (Yoon et al., 2007; Blank & De Las Alas, 2009; Jacob & McGovern, 2015). Evidence from recent US studies also point to zero effects of professional development programs on student achievement (Jacob & Lefgren, 2004; Garet et al. 2008; Garet et al. 2010; Garet et al. 2011). Conclusions drawn from these null findings could be that many professional development programs are poorly designed, or that the studies evaluating their effectiveness are flawed. Yoon et al. (2007), for example, argue that the majority of professional development evaluation studies have some methodological issues. There is also a large literature discussing the optimal design of professional development programs that conclude that many programs fall short of meeting best practice standards (see Wei, Darling-Hammond, Andree, Richardson, & Orphanos (2009) for an overview).

We argue that the disappointing returns to most professional development programs are partly caused by inefficient selection into training. This inefficient selection into training results from teachers' rational response to their working environment. While teachers' salaries may not depend on their performance, in most countries their salary does increase with tenure (OECD, 2014). Teachers want to stay employed and have career concerns (Gibbons & Murphy, 1992). Because performance is imperfectly observed, teachers have an incentive to strategically divulge information from which their employer will infer they are of high quality. Signing up for particularly advanced forms of professional development programs could be such a signal. Conversely, signing up for a course that improves very elementary skills sends a signal that the teacher does not even master the basics. Advanced courses will be oversubscribed, while courses that improve basic skills will be unpopular. Because program evaluations only look at the objective gains in teaching quality, those who sign up for the signal alone while gaining nothing from the program will bias the observed effect size downwards.

From the preceding we can conclude that, while there are no direct monetary benefits from *being* a more productive teacher, there are career benefits from *being seen* as a highly productive teacher. This can explain teachers' high participation rate in training programs, in spite of the fact that most programs are evaluated as being ineffective. In the next section, we formalize our argument by developing a model on teachers' decisions to enter professional development programs.

3. Teacher training model

In this section we describe our model. First we introduce the single program model and state our assumptions. Later we describe the situation where teachers can choose between different types of available training programs.

3.1 Single training program model

We assume that teaching ability (a) is a single united set of skills[1] necessary to improve students' performance, uniformly distributed on $[0,1]$.[2] The assumption of a continuous distribution is used in order to allow for heterogeneity in initial ability.[3] The training program improves all skills in the bundle homogeneously at the same rate. We assume that the total added value of training ($\alpha(a)$) is

[1] Teaching ability can also be assumed to be multidimensional (i.e. include more than one skill). In this case a teacher makes a decision about every specific skill independently, and the initial model can be applied for each dimension separately.
[2] To check for the sensitivity of our results to the choice of a uniform distribution, we simulated the model using several different distributions of initial ability and got qualitatively similar results.
[3] In this, we follow the evidence from previous work (e.g. Aaronson, Barrow & Sander 2007; Staiger & Rockoff, 2010) showing considerable heterogeneity in teacher value added.

linear in ability and consists of the fixed benefit (α_0) and marginal returns to training (α_1). Those values are determined by the program characteristics:

(1) $$\alpha(a) = \alpha_0 + \alpha_1 a.$$

We assume that all the available training programs are advanced or basic by design, depending on which type of teacher benefits more. The main difference between these types of training is the relationship between returns to training and the participants' ability. The added value of basic training is decreasing with teaching ability ($\alpha_1 < 0$), so low ability teachers benefit more from this training. Conversely, for advanced training programs the added value is increasing with ability ($\alpha_1 > 0$), as teachers with higher ability can learn more from these programs. That is to say that added value of basic (advanced) training is higher for teachers with low (high) ability. We restrict the values of the parameters between -1 and 1 to have correspondence in magnitude with the ability variable.

By construction, fixed added value in (1) defines the gains of teachers at the bottom of the ability distribution, since it outweighs marginal added value for low values of a. Analogously, marginal added value determines the gains of teachers with high ability, and we assume that fixed and marginal added values are negatively correlated. An increase in marginal added value makes the training more beneficial for higher ability teachers, and makes it less useful for low ability teachers simultaneously. The same logic holds for an increase in fixed added value: by adjusting to serve the needs of teachers with low ability, the training becomes less useful for high ability teachers.

Since teacher productivity is not perfectly observable, individual teaching ability and added value of training are private information. The information the market receives about an individual teacher is whether or not he went into training, and general information about the program he took. Based on the information about the program, market agents form beliefs about the level of training. For example, programs that lead to a master's degree will likely be inferred to be an advanced form of training, while a program labeled a 'subject knowledge refreshment course' is likely to be judged as basic. Using the available information, market agents make assumptions about the average level of ability of teachers who did and did not participate in training. Here we analyze the extreme case when training participation is the only available source of information regarding teaching ability.[4] Market agents (e.g. school principals) infer a teacher's ability from the type of training he takes. For example, a teacher who signs up for advanced training, is expected to be of higher initial ability than a basic program participant. Therefore, teachers take into account market beliefs when deciding to take the training.

The decision to participate in the training is a single period voluntary decision. Teachers are rational to the extent that they base their decision on a cost and benefit analysis, i.e. the teacher makes a decision (Y) to participate in the training if the utility of participating is higher than the utility of not participating ($U_p(a) \geq U_{np}(a)$).[5] This decision could be modelled binary:

$$Y = \begin{cases} 1, if\ U_p(a) \geq U_{np}(a) \\ 0, if\ U_p(a) < U_{np}(a) \end{cases}.$$

To distinguish between ability before and after training we refer to ability after the participation decision as knowledge:

$$(2) \qquad K(a) = \begin{cases} \alpha_0 + (1+\alpha_1)a,\ if\ Y = 1 \\ a,\ if\ Y = 0 \end{cases}.$$

This means that for teachers who do not take the training, knowledge is equal to their initial ability.

The utility a teacher gains from the training includes net added value and the signal of the training:

$$(3) \qquad U_P(a) = \alpha(a) - c + EK(a^*)_p - K(a),$$

[4] In practice, school principals form beliefs about the productivity of their teachers over time (Rockoff, Staiger, Kane & Taylor, 2012). However, for the results of our model to hold, productivity needs only be partially unobserved, as we show in Section 3.3.1.

[5] Without loss of generality we assume that if the teacher is indifferent between taking and not taking the training ($U_p(a) = U_{np}(a)$), he will participate.

where a^* stands for a participation threshold, and is defined as the ability level for which the utility from training is equal to the utility from not taking the training.

The net added value of training is expressed as the difference between total added value and total costs of training ($\alpha(a) - c$), and represents the teacher's intrinsic valuation of the net gains from training.

The total costs of training (c) include monetary costs and time investment. Following Spence (1973) we can assume that costs decrease with ability, as the smarter individuals require less time to master the same material. However, the decreasing nature of the costs can be incorporated in (3) in the coefficient multiplying the ability, i.e. if $c = c_0 - c_1 a$:

$$U_P(a) = \alpha_0 + a(\alpha_1 + c_1) - c_0 + EK(a^*)_p - K(a).$$

Therefore, without loss of generality we can make a simplifying assumption that the costs are equal for all teachers ($c = c_0$).

We assume that all teachers value their teaching ability in the same way, so there are no differences in the weights they put on the net gains in the utility function.[6]

$EK(a^*)_p - K(a)$ is defined as the signal of the training program. This is the difference between expected knowledge of the group and the individual teacher's knowledge after training. As the best guess the market can make about a person's knowledge is the average knowledge of people (not) taking training, every teacher signing up for training is assumed to be $EK(a^*)_p$. Notice that the signal of the same program can be positive or negative for different teachers depending on their individual ability. Individuals with below average knowledge will have additional (signaling) benefits from the training, since they will be assumed to be more able than they actually are, while for teachers with above average knowledge it will be a penalty. Following the same logic, the decision to not participate in the training can also be treated as a signal of certain knowledge, and these teachers gain

[6] Results do not change if valuation of teaching ability is assumed to be heterogeneous.

(lose) the difference between the expected knowledge of the group not taking training and their individual level of knowledge:

(4) $U_{np}(a) = EK(a^*)_{np} - a.$

Everyone for whom the benefits from training exceed the benefits from not participating will take the training in equilibrium ($U_p(a) \geq U_{np}(a)$). The teacher with marginal ability (a^*), for whom the net benefits from going and not going into training are equal, is indifferent.

Since the solutions differ for different types of training, we look separately at both situations. First we predict participation in the ideal situation when productivity is perfectly observable, and training is only taken by those who benefit from it in terms of ability, then show how the results change once productivity becomes unobservable and signals enter the decision making process.

3.1.1 Basic training

We first consider a situation where teacher productivity is perfectly observable. Training cannot be used for signaling and is only valued for the skills increase it provides. In this case only net added value is present in the utility function (3), while utility of not taking the training (4) is zero. Consequently, all teachers who derive positive utility from this training sign up. Using the threshold definition and equating (3) to zero, we find the participation threshold ability (a^*):

$$U_p(a^*) = \alpha(a^*) - c = 0,$$

(5) $a^* = \frac{c-\alpha_0}{\alpha_1}.$

As in this case total added value only decreases with ability, all teachers with ability lower than the threshold derive positive utility from the training and sign up for it. Thus, the participation rate equals the threshold.

Now let’s look at the situation where ability is not observable.

Proposition 1. In the equilibrium all teachers are indifferent between taking and not taking the training.

We find the participation threshold by equating utility from training (U_p) to the utility from not signing up (U_{np}) and plugging (2) into (3):

(6) $$EK(a^*)_p - c = EK(a^*)_{np}.$$

Since (6) does not depend on a directly, the benefits from going (not-going) into training are the same for everyone and boil down to the expected knowledge of each group. This means that, in equilibrium, the utility of taking the training is the same for the entire teacher pool and is equal to the utility of not taking the training. The equilibrium is determined by the number of teachers in each group and their individual abilities. This situation is similar to the perfect competition theory in Microeconomics, where the profits of an individual firm depend on the number of firms in the market. This mechanism regulates the number of operating firms. Similarly to our case, in the equilibrium all active firms receive zero profits (equal to the profits they would receive by not operating).

Proposition 2. Participation rate of a basic training increases with an increase in added value and decreases with an increase in costs.

As $EK(a^*)_p$ ($EK(a^*)_{np}$) is the expected knowledge of the group taking (not taking) the training, $K(0)$ is the knowledge level of the lowest ability teacher, and 1 is the maximum possible ability, we can calculate the average knowledge of each group:

(7) $$EK(a^*)_p = \frac{K(0)+K(a^*)}{2},$$

and

(8) $$EK(a^*)_{np} = \frac{a^*+1}{2}.$$

Taking (2) into account we solve (6) for a^*:

(9) $a^* = \frac{2c - 2\alpha_0 + 1}{\alpha_1}$.

Taking into account that $\alpha_1 < 0$, we see from (9) that the participation rate depends positively on added value parameters and negatively on the costs. When α_0 is increasing, participants gain relatively more from the training. The same is true for an increase in α_1 (decrease in absolute value). As the added value curve becomes flatter, on average total gains increase. Therefore both changes lead to increased participation in the program.

Proposition 3. Basic training should be extremely effective or appealing to a wide audience in order to attract participants.

We also notice that in order for a basic training program to attract participants (a^* to be positive), its net fixed added value should be higher than $\frac{1}{2}$. Signing up for a basic training of this kind will signal low initial ability, for which the training should compensate by being effective enough to bring the lowest ability teacher to an above average level. This extreme result is explained by our assumption that training participation is the only available source of information about teacher quality. Once we allow for more sources of information, the magnitude of the result becomes smaller (see Section 3.3.1).

The negative signal associated with basic training implies that this type of training will attract few (if any) participants. This poses a problem for policy interventions aimed at low quality teachers. For example, let's assume that the aim is to improve the teachers at the bottom 10% of the ability distribution and to bring them to the level just above this 10th percentile. Our model shows that nobody will participate in the training due to the negative signal even if the training program is proven to be effective. Policy makers that aim to target the bottom of the distribution will therefore necessarily have to design their program so that it is appealing to a broader audience in order to induce teachers to participate. However, generalizing the program will decrease its efficiency in raising the quality at the low end of the distribution. While somewhat counterintuitive, decreasing the knowledge gained from training for those at the bottom of the distribution increases their willingness to participate. The other, less realistic, option would be to make the training so effective that low ability teachers are better off after the training than not participating at all and being assumed to be average.

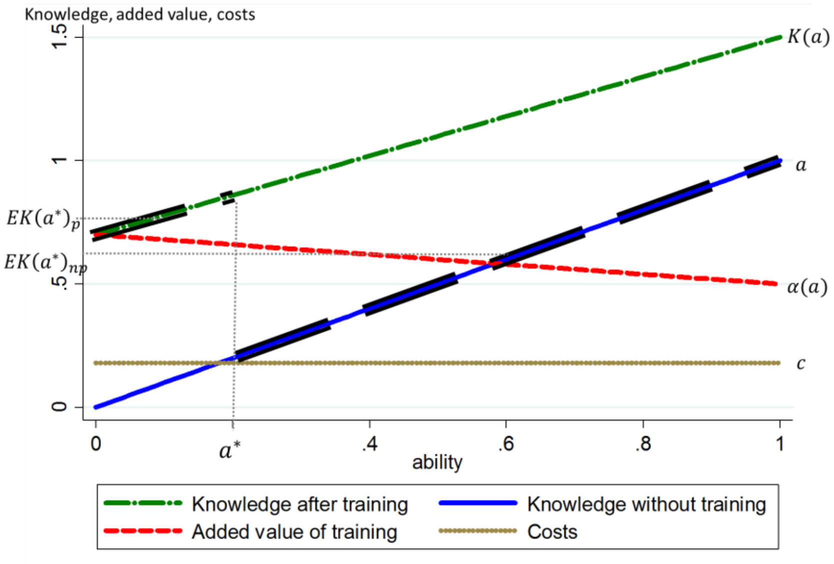
Proposition 4. Imperfect information decreases the participation rate of the basic program.

Now we compare this result with hypothetical situation (5) without signaling discussed earlier. If we rewrite (9) we can treat the second term as a signaling stigma, reducing participation ($c - \alpha_0 > -1$):

$$a^* = \frac{2c-2\alpha_0+1}{\alpha_1} = \frac{c-\alpha_0}{\alpha_1} + \frac{c-\alpha_0+1}{\alpha_1}.$$

A graphical example of a basic training program is provided in Figure 1. The added value and costs of the training are: $\alpha_0 = 0.7,\ \alpha_1 = -0.2,\ c = 0.18$. We see that the line of gross added value ($\alpha(a)$) is always above the costs (c), therefore in the situation where productivity is observed and the training does not have any signaling value, all teachers sign up for training, and the participation rate is 100%. Everyone improves their initial ability and shifts from a (solid) to $K(a)$ (dash-dot) line. The knowledge after training curve ($K(a)$) is the sum of knowledge without training (α) and added value of training ($\alpha(a)$). If productivity is not observable, using the formulas derived earlier, we calculate that the stigma of the basic training reduces participation to 20% (a^*). Teachers with initial ability below a^* take the training and their knowledge shifts to the $K(a)$, while teachers with initial ability higher than a^* stay on a. The thick dashed black line is used to show the knowledge distribution in the market in equilibrium.

Figure #1. Basic course equilibrium



3.1.2 Advanced training

Similar to the previous section we first analyze a situation where signaling is not applicable. Since we use the same formulas, the participation threshold is the same as for the basic program:

$$a^* = \frac{c - \alpha_0}{\alpha_1}.$$

In this case teachers with initial ability higher than a^* take the training, therefore the participation rate equals $1 - a^*$.

Now we consider the situation where ability is not observable and teachers can sign up for training to signal their ability. Since the program is assumed to be advanced and highly able teachers benefit from them the most, smart individuals are expected to sign up. Therefore, if for a particular teacher the actual net benefit from the training is negative (added value is lower than the costs), but at the same time his knowledge after training is lower than the average knowledge of teachers attending that training, he can pretend to have higher knowledge just by subscribing to that training. Therefore, because of the positive signal, more teachers are expected to take the training program than just those that benefit from it in terms of increased ability.

Proposition 5. Participation rate of an advanced training increases with increases in added value and decreases with an increase in costs.

Again the threshold is derived from (6), but this time representations for expected knowledge of the groups are different due to the fact that the tails of the distribution for which the training is most and least beneficial change places:

$EK(a^*)_p = \frac{K(1)+K(a^*)}{2}$ and $EK(a^*)_{np} = \frac{a^*+0}{2}$.

$K(1)$ is the knowledge level of the teacher with the highest ability and 0 is the lowest possible ability. Then the threshold is

$$(10) \quad a^* = \frac{2(c-\alpha_0)-1-\alpha_1}{\alpha_1},$$

and the participation rate is

$$(11) \quad 1-a^* = \frac{2(\alpha_1+\alpha_0-c)+1}{\alpha_1}.$$

Here again the participation rate depends positively on added value parameters and negatively on the costs. The underlying intuition is the same as in the previous case. An increase in added value makes more teachers better off taking the training and shifts the threshold to the left, increasing the participation rate.

Proposition 6. The decision to go into training does not have to be related to program effectiveness and can have a pure signaling effect.

We notice that for low or even negative fixed added value (α_0) and low absolute value of marginal returns, in the equilibrium without signaling nobody signs up for training because the net benefits are negative. However, when teachers' productivity is unobservable and signaling is possible, teachers do go into training when the costs are higher than the added value in terms of ability. If the program is useless but assumed to be for highly able individuals, it provides an opportunity for teachers to signal high ability. Therefore, teachers at the margin will sign up for a program from which they gain no knowledge if the program is assumed to be advanced.

Proposition 7. Advanced training programs are oversubscribed.

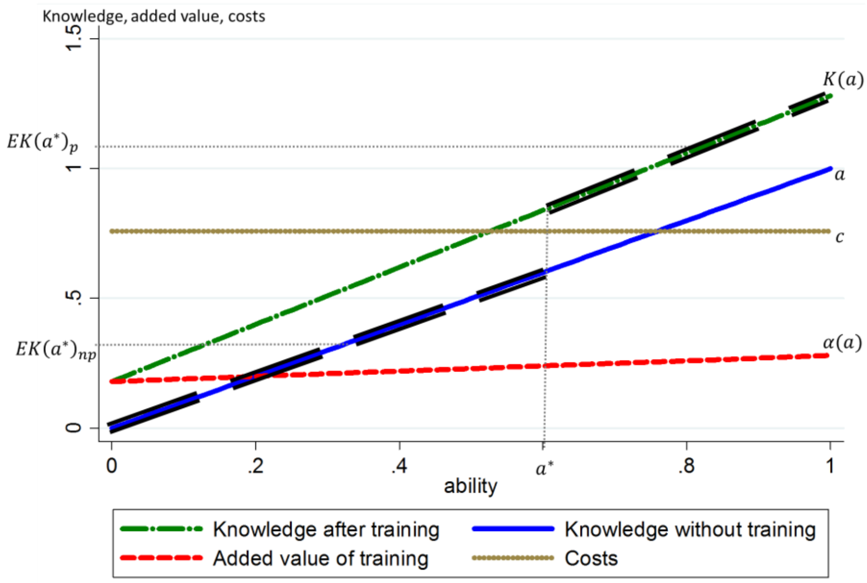
Now let's rewrite (10) to compare it with (5) and see what happens to the threshold due to signaling:

$$a^* = \frac{c-\alpha_0}{\alpha_1} + \frac{c-\alpha_0-1-\alpha_1}{\alpha_1}.$$

We see that the threshold is lower and more teachers are participating, including those for whom the training is too costly in addition to those who would take the training without signaling.

Figure 2 provides an example of an advanced program. The added value parameters and costs are: $\alpha_0 = 0.18$, $\alpha_1 = 0.1$, and $c = 0.76$. We see that for any ability level the costs are higher than total added value. Therefore in the case that teacher productivity is observable, nobody signs up for the training and the participation threshold $a^* = 1$. In the case that teacher productivity is not observable, participation rate rises to 40% ($a^* = 0.6$), entirely due to the positive signal. Again, the thick dashed black line shows the knowledge distribution in the market.

Figure #2. Advanced course equilibrium



3.2 Multiple training programs model

So far we discussed a situation where teachers can only choose one training program. Now we extend the model to a situation where teachers can choose between an advanced and a basic program. We assign additional indices $i = 1$ to the parameters corresponding to a basic training and $i = 2$ to the parameters of an advanced training. Added value and costs of training are now denoted as:

$\alpha_i(a) = \alpha_{0i} + \alpha_{1i}a$ and c_i, $i = 1,2$.

Now teachers can choose between three options: basic training, advanced training, and no training. Each option provides a signal, as described in the previous section. As the nature of the programs does not change, we assume that in equilibrium some teachers from the bottom of the distribution take basic training (with abilities below some a_1^*), and some from the top take an advanced one (with abilities above some a_2^*). Then two cases are possible. First, teachers in the middle of the distribution (between a_1^* and a_2^*) are better off not taking any training and there are two different thresholds a_1^* and a_2^* (Figure 3). Second, there is full training participation and only one threshold a^*. Teachers with ability lower than a^* take basic training and teachers with ability above a^* take an advanced one (Figure 4).

Figure #3. Multiple courses 2-threshold equilibrium

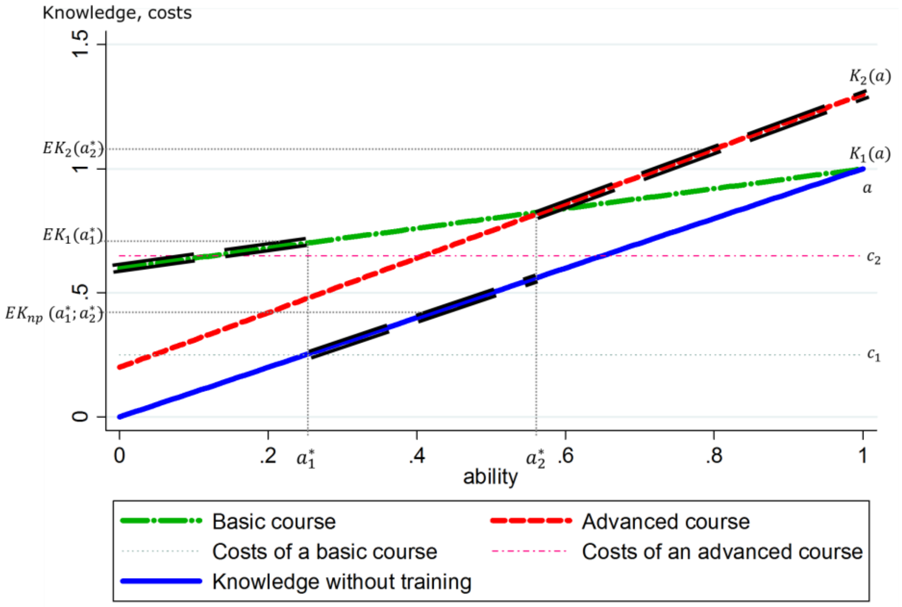
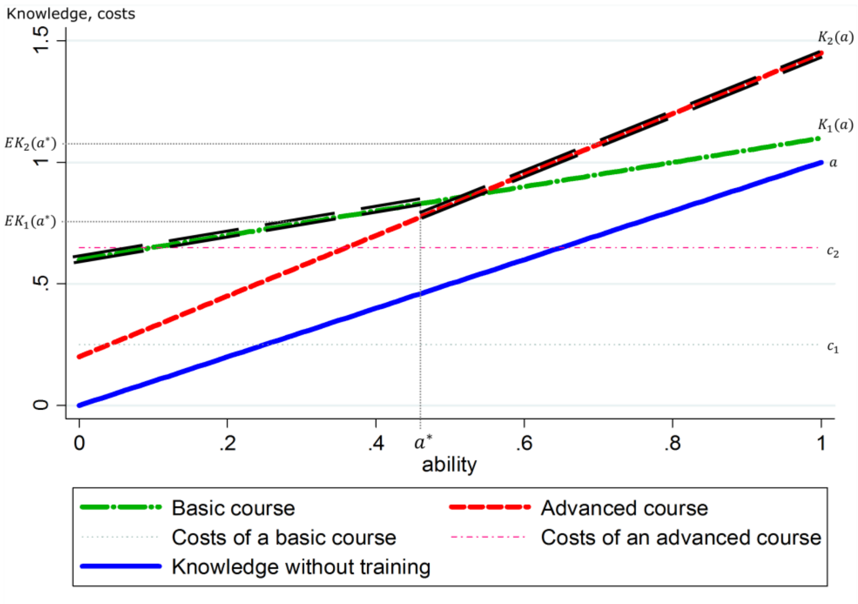


Figure #4. Multiple courses 1-threshold equilibrium



Proposition 8. Parameters of different programs affect each other's participation.

We start with the first situation. The threshold of the basic program is strictly lower than the threshold of the advanced program ($a_1^* < a_2^*$). Rewriting (6) for both programs gives a system of equations:

(12) $$\begin{cases} EK_1(a_1^*) - c_1 = EK_{np}(a_1^*; a_2^*) \\ EK_2(a_2^*) - c_2 = EK_{np}(a_1^*; a_2^*) \end{cases}.$$

Solving for a_1^* and a_2^*:

(13) $$\begin{cases} a_1^* = \frac{2\alpha_{02} + \alpha_{12}[2\alpha_{01} + 1 - 2c_1] + 1 - 2c_2}{1 - \alpha_{11}\alpha_{12}} \\ a_2^* = \frac{2\alpha_{01} + \alpha_{11}[2\alpha_{02} + \alpha_{12} + 1 - 2c_2] - 2c_1}{1 - \alpha_{11}\alpha_{12}} \end{cases}.$$

We can see from the parametrical representations of a_1^* and a_2^* that each threshold depends on the specifics of both programs.

Proposition 9. Interventions targeting advanced training have an indirect effect on demand for basic training.

For example, if the advanced training program gets better (increase in fixed or marginal added value), the share of participants of the basic program increases (together with its threshold). It also increases with a decrease in costs of the advanced program. Improvements in design of the advanced program ($\uparrow \alpha_2$) as well as availability to a wider audience ($\downarrow c_2$) boost demand for the basic program without any direct interventions.

The mechanism behind this result is the following. A rise in a_{02} or a_{12} and a drop in c_2 makes the advanced program appealing to more teachers and shifts the participation threshold to the left (a_2^*). As a_2^* is also a right margin of the group not taking the training, its average ability decreases ($\downarrow EK_{np}$). As a result, the benefits from not taking any training go down for all ability levels. Therefore for teachers with abilities close to the threshold a_1^*, who were better off not taking basic training, the benefits from taking it are now relatively higher. This shifts a_1^* to the right, which increases both the participation rate and the average knowledge of teachers who take the basic training ($\uparrow EK_1$).

Proposition 10. Increased participation of a basic training reduces the positive signal of an advanced training.

Conversely, higher added value of basic training ($\uparrow a_{01}, a_{11}$) and lower costs ($\downarrow c_1$) lead to a higher participation threshold ($\uparrow a_2^*$) and a lower participation rate of the advanced program ($\downarrow 1 - a_2^*$). However, as in previous case only teachers close to margin (a_2^*) are affected. Therefore only teachers who sign up purely because of the signal choose not to participate in the program, and sorting into training becomes more efficient.

Proposition 11. Different professional development programs mutually reduce the absolute value of each other's signal.

Comparing (9) and (11) with (13) we see that both professional development programs mutually reduce the absolute value of each other's signal. Moreover, improvements in basic or advanced training alone affect participants of both types of programs, as well as teachers not taking training at all.

As mentioned above, this holds for the situation where $a_1^* < a_2^*$. Plugging representations for a_1^* and a_2^* from (13) into this inequality we get

(14) $$2[\alpha_{01} - \alpha_{02} + c_2 - c_1] - 1 - \alpha_{12}[2\alpha_{01} + 1 - 2c_1] + \alpha_{11}[2\alpha_{02} + \alpha_{12} + 1 - 2c_2] > 0.$$

(14) can be used as a condition to distinguish between full and partial participation. If (14) holds, we are in the first case and $a_1^* < a_2^*$. Otherwise, if the left-hand side of (14) is less or equal to zero we get the second case and $a_1^* = a_2^* = a^*$. To find out participation threshold, we substitute a_1^* and a_2^* with a^* in (12) and get

(15) $$a^* = \frac{2(\alpha_{01} - \alpha_{02} + c_2 - c_1) - 1 - \alpha_{12}}{\alpha_{12} - \alpha_{11}}.$$

In this situation all teachers sign up for training: teachers with ability lower than a^*choose the basic training program, and teachers with higher ability sign up for the advanced program. It is clear from (16) that the participation threshold positively (negatively) depends on the added value of the basic (advanced) training and negatively (positively) on the costs of the basic (advanced) one. This is quite intuitive, as due to the full participation, increase in the added value (as well as decrease in costs) of the basic program attracts teachers who were better off signing up for the advanced training. Therefore the threshold just shifts to the left.

3.3 Extensions

3.3.1 Multiple signals

So far in this chapter we introduced two simple cases where a teacher's productivity is either perfectly observed or not observed at all, and the training program is the only available source of information about teacher ability. In practice, however, agents usually have additional ways to infer teachers' ability (e.g. personal observations, teacher characteristics such as work experience). Therefore it is logical to include more sources of teacher quality information into the model.

We assume that in the general case, observed teacher knowledge (OK(a)) is a combination of the signal of a training program ($EK(a^*)$) and information about teacher knowledge ($K(a)$) from other sources:

$$OK(a) = \lambda EK_i(a_i^*) + (1-\lambda)K(a), \;\; i = 1{,}2,$$

where $\lambda \in [0{,}1]$ characterizes information asymmetry in the market. If $\lambda = 1$, we are in the case of perfect asymmetry analysed in detail earlier. With a decrease in λ the importance of the signal of the training program decreases, as information from other sources becomes more reliable. Since λ can be perceived as importance of the signal of the training program, it affects the benefits of teachers from the training and therefore (3) becomes:

$$U_p(a) = \alpha(a) - c + \lambda[EK(a^*)_p - K(a)].$$

Although a decrease in λ attenuates the signal of the course compared to (3), it is clear that, while becoming smaller in magnitude, all the results of the model still hold.

3.3.2 Applications

While the model we develop is used solely to analyse teachers' decision to participate in professional development programs, some of the implications of our results can be generalized to other contexts. Particularly, in any situation where productivity is partly unobservable, and employees have an informational advantage regarding their ability, on-the-job training programs have signalling value. Because introducing a new training program attenuates the value of the signal of signing up for other programs, sorting becomes more efficient. Therefore, evaluations of these programs should take the effects of improved sorting into account. Evaluations that do not look at these spillover effects may underestimate the value of offering that particular training program.

Secondly, cost-benefit analyses on the effects of interventions to increase training participation by reducing costs should be based on the expected productivity gains of the marginal, rather than the average, participant. Our model shows that increasing the attractiveness of a training program by reducing its costs (e.g. through subsidizing participation) will induce participation of employees that are close to the participation threshold. However, those that are on the participation threshold are exactly those people for who the training adds relatively little in terms of productivity. Therefore, companies or policy makers that base their cost-benefit analysis on the average increase in productivity per participant will overestimate the benefits of subsidizing training.

4. Empirical predictions

The results from our model lead to several empirically testable predictions. In this section we will suggest some ways to validate our model. We also discuss how the predictions from our model differ from predictions following from models based on intrinsic or public sector motivation.

First we note that the gains from sending a positive signal are greater at certain points in a teacher's career. In our model we assumed for simplicity that the signal each course sends is valued equally among all teachers. However, those for whom the evaluation by their employer is more important in terms of their future employment, benefit more from signaling high quality. In terms of the model, these teachers will put relatively more weight on the value of the signal ($EK(a^*)_p$) than on the gains in actual knowledge ($K(a)$). In practice, this implies that teachers who are on fixed-term contracts are more likely to sort into advanced courses than teachers that are on permanent contracts. When the signal gains importance relative to actual gains in knowledge, it also means that sorting into training becomes less efficient. At the high participation threshold, teachers will undertake training which adds nothing in terms of knowledge for the positive signal. At the low participation threshold, teachers who would gain knowledge from basic training will not take it because of the negative signal.

A strongly related prediction is that, considering school principals form increasingly accurate beliefs of their teachers' productivity (Rockoff, Staiger, Kane & Taylor, 2012), teachers have more influence over their perceived quality early in the employment relationship. Therefore, our model predicts more participation in advanced courses with high positive signaling value for teachers in schools that have come under management of a new school principal. Again, we would expect that this effect is stronger for teachers on fixed term contracts.

Importantly, the preceding two predictions diverge from predictions made by a model that considers intrinsic or public sector motivation as the main factor in training participation. If intrinsic motivation would drive training participation, there should not be a difference in training behaviour related to contract status or length of the employment relationship. Each individual teacher would sort into the training program that increases their ability most effectively, regardless of the signal. If anything, assuming basic skills are necessary to benefit from advanced training, we would expect the opposite pattern: teachers should sort into basic training more often early in their career.

Second, because the existence of multiple courses targeted at different parts of the teaching ability distribution reduces the signaling value of each individual course, schools which offer a higher variety of training programs will have a higher training participation rate. As we have shown before, adding the option of an advanced course to a choice set consisting only of basic courses reduces the average ability of those teachers who do not sort into any training. This decreases the negative signal associated with signing up for a basic course, increasing participation in said course. A related prediction is that, because the existence of advanced courses increases participation in basic courses, schools that offer advanced training opportunities will have a higher level of average teacher quality. This holds regardless of the effectiveness of the advanced training programs, provided that the basic course is effective.

Again, these predictions differ from predictions made by an intrinsic motivation model. If intrinsic motivation drives training participation, the existence of an advanced course should not influence participation in a basic course when there are no signalling effects. If sorting into training is based only on gains in teaching ability, the existence of a course that is dominated in terms of gains should not influence the rational teacher’s sorting decision under the independence of irrelevant alternatives assumption.

5. Concluding comments

In this paper we develop a model on the training participation decision of teachers. We argue that because performance is partly unobserved and teachers face career concerns, teachers can use training participation as a signal of their teaching ability. We show that because teachers incorporate the signaling value of signing up for training into their decision making process, sorting into training is inefficient. Advanced courses are oversubscribed, and basic courses are underused. Our model can help explain the paradoxical finding that training participation in the teaching profession is high (OECD, 2014), while most programs are evaluated to be ineffective (Yoon et al., 2007; Jacob & McGovern, 2015).

Schools and policy makers can improve sorting into training by offering a variety of courses. The existence of an advanced course increases participation in basic courses, as the average ability of teachers that do not take any training decreases. The negative signal of signing up for basic courses weakens as a result. Because of these spillover effects, program evaluations that look at the effectiveness of training programs in isolation may miss potential overall positive effects on non-participants whose behaviour is changed through the introduction of the program. Furthermore, policy makers that aim to increase training participation by reducing costs (e.g. through subsidies), should base their cost-benefit analyses on the gains of the marginal participant. Because the marginal participant's benefits are below average, calculations based on the average increase in productivity per participant will lead to an overestimation of the benefits of decreased costs of training. Finally, policy makers that target the bottom of the distribution will have to design their program to appeal to a wider audience. Generalizing the course will likely decrease its effectiveness for the target population, but is a necessity to induce them to participate. This illustrates the difficulty of improving teaching quality at the bottom of the distribution.

In order to most efficiently raise the quality of education through increasing the quality of teachers, a clear understanding of the decision to enter professional development programs is invaluable. Our model provides an explanation for why sorting into training could be inefficient, and offers suggestions to decrease inefficiencies. However, empirical research should address the question whether our model can explain observed training participation patterns before policy makers take up suggestions based on theory alone. For this reason, we make several empirical predictions that diverge from predictions made by models based on intrinsic motivation. Ultimately, when teachers use training as a signal in practice, insights from our model can help policy makers in designing the most efficient implementation strategy for professional development programs.

Reference list

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*(1), 95-135.

Acemoglu, D. (1997). Training and innovation in an imperfect labour market. *Review of Economic Studies*, 64:445–464.

Acemoglu, D. and Pischke, J.-S. (1998). Why do firms train? Theory and evidence. *Quarterly Journal of Economics*, 113(1):79–119.

Acemoglu, D. and Pischke, J.-S. (1999a). Beyond Becker: Training in imperfect labor markets. *Economic Journal*, 109:F112–F142.

Acemoglu, D. and Pischke, J.-S. (1999b). The structure of wages and investment in general training. *Journal of Political Economy*, 107(3):539–572.

Autor, D. H. (2001). Why do temporary help firms provide free general skills training?. *Quarterly Journal of Economics*, 1409-1448.

Bassanini, A., Booth, A., Brunello, G., De Paola, M. and Leuven, E. (2007) Workplace training in Europe, in *Education and Training in Europe* (Eds) G. Brunello, P. Garibaldi and E. Wasmer, Oxford University Press, Oxford, pp. 143–309.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70:9–49.

Blank, R. K., & De Las Alas, N. (2009). *The effects of teacher professional development on gains in student achievement: How meta analysis provides scientific evidence useful to education leaders*. Council of Chief State School Officers. Washington, DC.

Delfgaauw, J., & Dur, R. (2008). Incentives and workers' motivation in the public sector. *The Economic Journal*, *118*(525), 171-191.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, 696-727.

European Commission (2012), "Supporting teacher competence development for better learning outcomes". Retrieved from: *http://ec.europa.eu/education/school-education/doc/teachercomp_en.pdf*.

Fouarge, D., Schils, T., & Grip, A. de (2013). Why do low-educated workers invest less in further training? *Applied Economics*, *45*(18), 2587-2601.

Frazis, H., & Loewenstein, M. A. (2005). Reexamining the returns to training functional form, magnitude, and interpretation. *Journal of Human Resources*, *40*(2), 453-476.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., ... & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement.* NCEE 2008-4030.National Center for Education Evaluation and Regional Assistance.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Walters, K., Song, M., ... & Doolittle, F. (2010). *Middle school mathematics professional development impact study: Findings after the first year of implementation.* NCEE 2010-4009. National Center for Education Evaluation and Regional Assistance.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., ... & Sepanik, S. (2011). *Middle school mathematics professional development impact study: Findings after the* *second year of implementation.* NCEE 2011-4024. National Center for Education Evaluation and Regional Assistance.

Gibbons, R., & Murphy, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *The Journal of Political Economy*, *100*(3), 468-505.

Haelermans, C., & Borghans, L. (2012). Wage Effects of On-the-Job Training: A Meta-Analysis. *British Journal of Industrial Relations*, *50*(3), 502-528.

Hansson, B. (2008). Job-Related Training and Benefits for Individuals: A Review of Evidence and Explanations. OECD Education Working Papers, No. 19. OECD Publishing (NJ1).

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30(3), 466-479.

Hashimoto, M. (1981). Firm-specific human capital as a shared investment. *American Economic Review*, 71(3):475–482.

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, *39*(1), 50-79.

Jacob, A., & McGovern, K. (2015). *The mirage: Confronting the hard truth about our quest for teacher development*. TNTP.

Killeen, K. M., Monk, D. H., & Plecki, M. L. (2002). School district spending on professional development: Insights available from national data (1992-1998).*Journal of Education Finance*, 25-49.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*,*10*(2), 179-214.

Leuven, E. (2005), The economics of private sector training: A survey of the literature. *Journal of Economic Surveys, 19*: 91–111.

Leuven, E., & Oosterbeek, H. (2004). Evaluating the effect of tax deductions on training. *Journal of Labor Economics*, *22*(2), 461-488.

Leuven, E., & Oosterbeek, H. (2008). An alternative approach to estimate the wage returns to private-sector training. *Journal of Applied Econometrics*, *23*(4), 423-434.

Ministerie van Onderwijs, Cultuur en Wetenschap (2014), *Rijksbegroting 2015 VIII Onderwijs, Cultuur en Wetenschap* (kst-34000-VIII-2 ISSN 0921 – 7371). ‘s-Gravenhage: Sdu Uitgevers.

OECD (2014), *TALIS 2013 Results: An international perspective on teaching and learning*, OECD Publishing, Paris.

Prendergast, C. (2007). The motivation and bias of bureaucrats, *American Economic Review*, vol. *97*(1), pp.180-96.

Pritchett, L. (2002). It pays to be ignorant: a simple political economy of rigorous program evaluation. *The Journal of Policy Reform*, *5*(4), 251-269.

Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *The American Economic Review*, *102*(7), 3184-3213.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355-374.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives, 24(3)*, 97-117.

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). *Professional development in the United States: Trends and challenges* (Vol. 28). Dallas, TX: National Staff Development Council.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the US and abroad. Technical Report*. National Staff Development Council.

Wilson, J. Q. (1989). *Bureaucracy: What government agencies do and why they do it*. Basic Books.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement. Issues & Answers*. REL 2007-No. 033. Regional Educational Laboratory Southwest (NJ1).