

IZA DP No. 8992

**Does Performance Information Affect Job Seekers
in Selecting Private Providers in Voucher-Based
ALMP Programs?**

Christer Gerdes

April 2015

Does Performance Information Affect Job Seekers in Selecting Private Providers in Voucher-Based ALMP Programs?

Christer Gerdes

*Swedish Public Employment Service,
SOFI, Stockholm University and IZA*

Discussion Paper No. 8992
April 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Does Performance Information Affect Job Seekers in Selecting Private Providers in Voucher-Based ALMP Programs?*

One of the services provided at the Swedish Public Employment Services is job coaching. Since 2012, that service has been organized according to a voucher-based choice system. In 2013, a performance indicator of job coaching providers was made available at an online website. This study examines to what extent such information affected job coaches' ability to attract participants, as well as how the use of provider ratings varied across different groups of participants. The results indicate that the performance indicator strengthened the positive link between "quality" of services and the number of new participants that signed up with a provider. At the same time, there is evidence showing that more vulnerable groups, such as the less educated and those born outside Europe, do not use the rating values to the same extent as others. It appears that there is a tradeoff as to efficiency and inequality with respect to services that are contracted to private companies in a system based on consumer choice.

JEL Classification: H41, J64, J68, J78

Keywords: public sector, private providers, ALMP, rating, efficiency, equality

Corresponding author:

Christer Gerdes
Swedish Public Employment Service
Department of Analysis, Research and Evaluation Unit
SE 113 99 Stockholm
Sweden
E-mail: christer.gerdes@arbetsformedlingen.se

* I would like to thank participants at the 5th Joint IZA/IFAU Conference on Labor Market Policy Evaluation, the SVS-seminar at SOFI, the staff at the Research and Evaluation Unit, as well as Susanne Vinell for helpful comments on earlier drafts of the paper. Any opinions expressed here are those of the author and not those of the Swedish Public Employment Service.

1 Introduction

Following the change of government in autumn 2006, when a liberal-conservative government gained power, Sweden underwent an overhaul of broad parts of its public sector. That policy change aimed to make way for new ways of organizing the provision of public services. In some areas of public services, such as primary and secondary education, Sweden already had been practicing similar free-choice policies since the beginning of the 1990s. One main aspects of this more recent overhaul has been an increased focus on marketization of public welfare services in areas such as public health care, as well as public employment services.¹ To facilitate the ability to choose providers, new regulations were implemented that allowed non-public providers to enter into areas that until then exclusively were served by public entities. Thus, parts of the Swedish public sector have become what are sometimes referred to as “quasi-markets.” There are various ways in which elements of market mechanisms can be introduced in the publicly funded welfare sector. According to le Grand (1991), the distinctive element of quasi-markets is the fact that the state stops being both funder and provider of services; instead private, voluntary, and public providers compete with each other. Providers may either be contracted using a procurement process or rely on consumers (or agents acting on their behalf) using vouchers earmarked for that service.²

¹ Those changes gained some attention outside Sweden, not least in the British public debate, where there was extensive media coverage of the Swedish reforms in recent years; see for example *The Economist*, “The new model: A bit more unequal, a lot more efficient,” October 13th, 2012.

² There might be other aspects motivating the introduction of choice options in organizing public welfare services, such as empowering beneficiaries to enable them to shape the form and content of certain services, such as curriculum in schools, which is “based on the view

One precondition for a system relying on consumer choice to work properly is that consumers must be able to make their choice based on useful information, such as providers' profiles and track records. Without information about how providers have performed, it becomes difficult, or even impossible, to make rational choices. The necessity of this kind of information has repeatedly been highlighted in the literature on marketization of services that have the character of being public goods; see, e.g., Shleifer (1998) and Blank (2000) for more theoretical discussions of this issue. Specifically, for a quasi-market to work properly, the following must hold: information about the quality of different providers must be freely available to those choosing the provider, and there should be no obstacles when switching from one to another provider, so that free competition will force those providers that fail to deliver high quality services out of the market. Only then will it be possible to improve services "by enhancing the 'voice' of service users and their awareness of 'exit' options," Dan and Andrews (2014), p.5.

One key aspect in this context regards the extent to which the efforts of providers actually can be measured in a consistent and reliable manner. This conditions the extent to which it is possible to track providers, as well as the possibility of coming up with a reliable and valid indicator of the quality of providers. It is by no means obvious how to design such a performance measure. In some areas, such as health care, designing a performance measure can become rather challenging, as the services provided at one single entity (e.g., a hospital)

that public organizations work better when members of their client groups get representation and can help shape the mission of the organization," Besley and Ghatak (2005) p.629.

Another, and to some extent related issue, regards the discussion on contracting out compared to the in-house provision of public services due to incomplete contracts, see Hart et al. (1997).

I will not address such issues in this study.

can vary greatly. This might have the undesirable consequence that provision of a performance indicator focusing on some areas could lead to a reduction in overall quality of services, since information regarding the quality of one single aspect/dimension might reduce a provider's incentive to improve quality in other areas. See for example Katz (2013) for a theoretical discussion on the scope of quality (dis)improvement following the introduction of quality indicators. See also Bergman et al. (2014) and references provided there, as well as Propper and Wilson (2003) for a discussion on the pros and cons of different performance measures. Regarding the usefulness and possible pitfalls of outcome-based performance measures of labor market programs, see Heinrich (2002) and Barnow and Heinrich (2010). The latter highlights the importance of considering "performance standards adjustment procedures," i.e., the question of adjustment for conditions connected to participant characteristics.³

This study presents an initial evaluation of the introduction of a quality indicator for providers of job coaching in Sweden in April 2013. That measure, called rating, is an adjusted performance measure, i.e., one that controls for the characteristics of participants taking part in job coaching. The questions addressed in this study are twofold. The first question concerns how information about the coaches' previous performance affects their chances to attract new participants. This adds to the literature on the importance of disclosing information on the quality of providers of publicly financed services that are open to client choice, and how this

³ They report that there only are few cases (in the US) where standardized performance measures have been used. One of the arguments in favor of having an adjusted measure is that it will counteract incentives to "game" the system, i.e. the tendency to focus on certain group of consumers that are less costly, so called "cream skimming." We discuss these issues in section 3.

information can be translated to increase market share for those providers that have proven capable of delivering better services. The extent to which such information is available and utilized is a vital precondition to realize the potential efficiency gains of having a system based on consumer choice of providers. In essence, quality-related information can be seen as a prerequisite for implementing incentive structures for providers of public welfare services to improve the quality of services. If “consumers” of services do not respond to performance indicators, it is hard to see how quality aspects can have an impact on the incentive structure of providers of services, and by extension on increasing the efficiency of provided services.⁴

The second question we address concerns the type of group of job seekers that actually make use of the offered information. Earlier research in related areas provides evidence that

⁴ At least this holds true when quality measures are not used for other purposes, such as to determine remuneration, or used in later tendering procedures of providers. For practical reasons, the extent to which the introduction of the ratings leads to real improvement in quality cannot be addressed within the scope of this study. To the best of my knowledge, there are only a few (credible) studies on quality improvement from providing quality information, such as Lockwood et al. (2013), which studies how the so called “Comprehensive Performance Assessment” worked as an incentive scheme for local governments in England to improve services. A related study is Besley et al. (2009), which investigates the effect of the hospital star rating regime in England on wait times for hospital treatment. While the first study did not find efficiency gains, the latter reported reduced wait times following disclosure of information on previous wait times. In some areas, there might be a will to improve quality without pure demand aspects due to “non-pecuniary incentives,” such as being better than one’s peers, as proposed by Kolstad (2013). While such mechanisms might become relevant in highly academic professions, this is probably not the case for the market of job coaching.

there is a tendency for those who are better-off to profit from having a system based on free choice, while more vulnerable groups end up with lower-quality services, since less disadvantaged groups are better equipped and motivated to make proper choices. This could be both due to better cognitive capacity to access information, as well as to understand its content.⁵ Some scholars (see, for example, Davis and Meatcalf, 2014 or Kahneman, 2011) have highlighted the fact that processing information is both time consuming and requires some amount of cognitive effort, which for some people would make it rational to abstain from engaging with such information, especially in situations where the expected gains are rather small. Connected to the possible impact of cognitive constraints on processing information, there also could be a lower propensity of risk taking along cognitive lines, in the sense that less privileged users are more risk-averse and abstain from taking a risk based on some (more or less reliable) signals of quality, as argued by Jilke (2014). As is known from the psychological literature, cognitive capacity and socioeconomic status are highly correlated; see for example Turkheimer et al. (2003). Accordingly, providing easily accessible and reliable information on the quality of service providers might help to reduce (search) costs and thus help to “level the playing field” (le Grand, 2006), as well as to counter differences in risk-preferences across different social groups. Besides mere cognitive issues in terms of processing information, socioeconomic aspects by themselves might discourage people from using information on quality. One example is the observation that low-income families place lower weights on school test scores because of lower expected returns on education for their children, as suggested by Hastings and Weinstein (2008). They find that easy access to (readily understandable) test score information as well as real options to choose

⁵ Previous research has shown that access and dissemination of information usually go via social networks, where information dissemination favors some groups more than others; see e.g. Hipp and Warner (2007).

from (i.e., schools in close proximity to their homes) can get even disadvantaged families to make choices based on performance measures. The performance indicator for job coaches studied in this paper was intended to meet such standards.⁶

To the best of my knowledge, addressing both these issues—i.e., the potential for market gains for qualitatively better providers, and the use of services with respect to socioeconomic background—in one and the same analytical framework has rarely been done before; among others things, this study will contribute to the debate on efficiency-equity aspects of introducing free choice options into public services. The prospect of conducting an evaluation of the disclosure of rating values has been complicated due to the fact that the rating values were launched simultaneously all over the country. As a result, there is no possibility of finding out what the choice of job coaches would have looked like without the publication of rating values (the counterfactual situation). However, it has been possible to produce estimates based on discontinuities brought about by regulatory thresholds, which will allow us to measure causal effects in a number of quasi-experimental settings by using a regression discontinuity framework.

The remaining part of this study is organized as follows. The next section starts with a brief account of the literature addressing efficiency-equity concerns, followed by a short review of studies aimed at the experiences of private providers in the public sector, as well as evaluations of performance indicators. Section 3 provides a description of “job coaching” as a

⁶ The use of performance indicators for the purpose of counteracting segregation has repeatedly been highlighted in education literature; see e.g. Betts and Loveless (2005). According to Hibbard et al. (2002), there are three steps involved in processing information; first, there has to be awareness of the quality information; second, there has to be an understanding of its contents; and, third, it has to be seen as useful.

service and the way rating values for that service have been calculated and implemented. Section 4 provides some exploratory statistics on the association between the numbers of new participants and the estimated quality of providers. Section 5 presents the outcome of calculations measuring causal relations between job coaches' rating value and their ability to attract new participants, followed by section 6, addressing the impact of (not) having been awarded a rating value. In section 7, we study the extent to which there are heterogeneous effects in using the published performance indicators across groups, while section 8 summarizes and discusses the results.

2 Review of related literature

2.1 Efficiency vs. equality

How to balance equality and efficiency concerns is a key question in the political policy arena. One challenging aspect of this debate relates to the definition of equality, occasionally also called equity. In its simplest form, equality/equity across citizens means equal distribution of outcomes, i.e., access to different forms of publicly provided services (such as education) or material standards (e.g., income transfers). Presumably, at least among economists such a view of equality would not find much support because of moral hazard; more agreeable would then be a definition of its antithesis, inequality, as formulated by Woessman and Schuetz (2006), saying that “inequality should be tolerated only if it is due to persons' differences in levels of effort, but not if it is due to circumstances which are beyond a person's control. [...] This concept of equity would not necessarily call for a strict equality of [...] outcomes in the sense of a perfect sameness or egalitarianism,” p. 2. In spite of the fact that equity concerns have been a long-standing issue in economic literature, most economists usually have focused on issues related to efficiency rather than equity. A standard view, which can be traced back to Okun (1975), assumes that there is a trade-off between efficiency

and equity, i.e., that more of one necessitates less of the other. In his view, the question (for policy makers) therefore becomes how much efficiency loss one is willing to accept to impose measures that try to help those persons that are less well off. Yet, there is a growing literature questioning the extent to which such an efficiency-equity trade-off should be seen as an inevitable side effect of organizing the public welfare sector. See, for example, le Grand (1990), Blank (2002), and Ostry et al. (2014) and references given there. Some authors point out that efficiency (in terms of economic growth given a certain set of resources) instead would increase by focusing on those groups that are trapped in conditions that do not allow them to raise themselves (or their children) to a better living standard without the help of some targeted measures, especially measures that focus on human capital enhancing activities aimed for socially deprived groups. According to Bourguignon et al. (2007), such policies might be able to overcome market failures (“inequality traps”), so that underprivileged groups are able to converge to “higher level of advantages,” p. 248. Questions relating to the efficiency and equity tradeoff have also been addressed in the literature on higher education admission policies, such as the consequences of entrance exams, see Black et al. (2014) and references provided there. The way these kinds of market failures should be resolved remains very much an issue of debate. One solution proposed in the debate on free school choice to raise both equity and efficiency is the introduction of voucher-based financial systems (Hoxby, 1996). To some extent, the present study might give some insights on the scope of efficiency and equity improvements following the application of a free choice voucher-based system.

2.2 Private providers in public employment services

During the last decade, a number of countries such as Australia, Denmark, England, Germany, and the Netherlands have experimented with voucher-based systems in active labor market policy (ALMP). The results of studies evaluating these systems have been rather mixed. Evaluations conducted for Sweden that looked at the effect of privatization of parts of ALMP have focused on the question of whether private actors are better at getting job seekers to work than the PES's own case workers; see e.g. Gartell (2011) and Liljeberg et al. (2012). None of these studies found significant differences between private and public job search assistance after controlling for observable background factors among participants. One problem with this type of evaluation concerns the issue of self-selection, i.e., that some choose private providers of job coaching, while others prefer using a job coach connected to the PES. There are some studies addressing that kind of selection problem. Benmarker et al. (2013) looks at the outcome of a trial scheme where participants were randomly assigned to private placement agencies. They do not find any major differences in participants' chances of finding a job compared to the control group that participated in PES's ordinary activities. Similarly, using a field experiment, Laun and Skogman Thoursie (2014) studied unemployed people that were randomly offered private or public rehabilitation. They found some minor differences on various outcomes, but overall they conclude that there are no efficiency gains from privatizing vocational rehabilitation. For studies of ALMP measures in other countries than Sweden, evidence of efficiency gains following the privatization of services also seems rather weak. A study by Winterhager et al. (2006) looks at the impact of introducing a voucher scheme earmarked for job placement activities in Germany in 2002. On the aggregate level, they found positive effects for participants. In the Netherlands, a country more prominent when it comes to privatization of its employment services, it has been shown that privatization has not led to an improvement in the outflow to work; see, e.g., Koning and

Heinrich (2013). As a consequence, outsourcing of welfare services was partly reversed. In their review of Australian and Dutch systems, Struyven and Steurs (2005) found that privatization of parts of the employment service meant some cost savings (in Australia), but has not been accompanied by an improvement in the efficiency of services. A study by Behaghel et al. (2014) looked at job seekers in France in private and public job counseling programs. They conclude that private firms were less effective than the public program, a result that, as they argue, was caused by misguided incentive structures when designing contracts for contracted private firms for those kinds of services.

2.3 The use of performance measures by public service beneficiaries/clients

As for the use of information regarding private actors' previous performance by the unemployed, such questions have not been studied to any greater extent. Most studies on the use and usefulness of performance measures in ALMP measures have focused on public management and accounting. Partly, this follows from the fact that a system of allowing free choice of providers is still rather rare for ALMP programs across countries. In other areas of the public sector, such as choosing schools or health care providers, free choice options are more common. Regarding school choice, several studies have evaluated the use of school league tables and college rankings. One attempt to address the relation between school-cohort compositional aspects and performance indicators such as school league tables can be found in Leckie and Goldstein (2009), who question the usefulness of such measures. Other studies have criticized the usefulness of performance indicators based on the fact that they may be subjected to "teaching to the test" and "easy courses" strategies; see Muriel and Smith (2011) for a survey of the recent literature. A different kind of quality indicator was studied by Mizala and Urquiola (2013), who evaluated a quality measure that explicitly takes into account the socioeconomic composition of students. That study did not find evidence that

providing such information changed enrolment rates across schools. As the authors argue, it might be the case that there are other partly offsetting incentives for parents, i.e., that they strive for a certain peer composition, rather than making use of information on the effectiveness of schools.⁷ Evidence of the usefulness of providing information on school quality (after controlling for compositional aspects) seems thus to be rather weak. One issue that sometimes gets attention in the literature on the use of information on school performance concerns stratification (sorting/segregation) caused by high school rankings or school league tables. Most studies point out that there is an effect on stratification; see Clarke (2007) for a literature survey. One opposing view is provided by Burgess et al (2013), who claims that there is no evidence of social stratification in the context of their study. In a study for the Netherlands, Koning and van der Wiel (2013) report a positive correlation between publication of a quality index in newspapers and enrolment rates to better rated secondary schools that have a more academic curriculum. That correlation is driven by parents actively searching for relevant information.

With respect to health care, there are a number of studies examining the impact of quality indicators for consumer choice, see for example Jin and Sorensen (2006), Cooper et al. (2011), and Varkevisser et al. (2012). Those studies provide evidence of a positive connection between a positive rating and consumer choice. Similar results are found by Pope (2009). A study by Jha and Epstein (2006) presents somewhat mixed findings, where better performance did not result in increased market share for better rated providers. Propper et al. (2008) shows that more competition among hospitals (determined by local variations of competitors) leads to worse outcomes with respect to mortality rates. The authors argue that their results are

⁷ Such a tradeoff between school quality indicators and peer composition in schools are in line with a number of other studies, e.g. Ladd (2002) and Rothstein (2006).

caused by hospitals prioritizing other activities, such as reducing wait times, i.e., measures that were easier to observe by clients. As to questions of stratification along socioeconomic lines in the health sector, these have repeatedly gained attention. For example, in their survey of the literature, Faber et al. (2009) report that persons with higher incomes, as well as to some extent those who were better educated, were more likely to notice and process information and were more likely to choose better rated hospitals.

3 Background: Job coaching and rating

As mentioned in the introduction, parts of the publicly financed employment service in Sweden have been contracted out to private providers. Since the change of government in autumn 2006, there has been an opening of broader parts of publicly provided active labor market services, giving the unemployed the possibility to choose among several private providers for certain services; the regulation underlying this way of contracting private providers is called “the Act of Choice” (“Lagen om valfrihet,” LOV). For some services, LOV is chosen to facilitate competition among providers and to enable the unemployed to find providers who fit their needs best. This study focuses on job coaching, a service aimed at assisting participants in how to write CVs, motivating them to apply for jobs, and/or helping them to prepare for job interviews. Instead of regulating the operational content, the PES focused on creating incentives for the providers of job coaching in helping participants to find work as soon as possible by means of a contingent payment scheme, which looked as follows.⁸ The provider of job coaching received payment divided in three steps. First, there

⁸ As a minimum, the total time in job coaching over the three month period with the supplier was set to 16 hours (reconciliation by telephone or other electronic means of contact not included). The planned time expenditure as well as the number of meetings had to be agreed on based on the participant's need for activity.

was an up-front payment of 5000 SEK (about 700 USD) for each unemployed person that started the program. An additional payment of 5000 SEK was paid if the participant stayed in the program for the full three-month period. If the participant started a job during that period or within 30 days after the end of the program, there was an additional reward of 5000 SEK paid to the job coach. So in total, a provider of job coaching could earn 15000 SEK for one participant.⁹ There were some (not very demanding) requirements for the job coaches' documented qualifications (e.g., a job coach should have relevant experience of similar work for at least one year).

The service "job coaching" studied here was regulated according to LOV and started in 2012. The Swedish PES published a rating of providers of job coaching for the first time in April 2013. The rating system is the outcome of comparing the performance of private job coaching providers from January to November in 2012. This was achieved by looking at the extent to which participants who were assigned to a job coach during that period actually did find a job. As the employability of participants looked rather different, all participants were "profiled," which means that by using a logistic regression model the chances for each participant of getting a job (within a certain timespan) were then estimated. The estimated probabilities for those participants of getting a job were compared to the actual outcome, i.e. if participants actually did find a job during the time they were assigned to a job coach.

⁹ This final reward was conditioned on the job lasting for at least one month. The payment structure looked somewhat different than the job counseling service in France studied by Behaghel et al. (2014). The main difference is that providers were guaranteed two thirds of total payment in case the participant stayed in the program over the whole period of three months, while in France only about 30% was guaranteed, the other 70% was contingent on getting the client to work.

Weighing up these two—the estimated probabilities and the realized outcome—produced a value (the “score value”) that was used to estimate how well a job coach performed compared to other job coaches. If a supplier’s score value deviated positively from the average of all providers’ score values, this resulted in a rating of three stars. Those close to the average got two stars, while those who were below average received just one star. See Appendix for details on the calculation of score values and how these were transformed into rating values.

The main goal of contracting job coaches has been to help unemployed participants find jobs as fast as possible. Thus, the fact that the performance measure was based on just one dimension—outflow to work—should not raise any concerns in terms of leading to adverse incentives. Also, there is an intentional consistency between, on the one hand, the way rating values were constructed, and, on the other hand, the compensation scheme rewarding those providers who were successful in helping the unemployed find work.

The employment officers were by regulation not allowed to tell the unemployed who they knew or considered to be a good or bad private provider. To enable the participants to receive information about the different providers, a public website was set up at the time the program started in 2012. Initially, it showed information on where job coaching providers were located and a brief self-description of their respective profiles. In April 2013, the ratings of job coaching providers in different areas were added to the website. The design of the website is shown in Appendix (A.4). Providers of job coaching were not allowed to decline any unemployed person who wanted to use their services. The main reason for this was to avoid “cream skimming” or “cherry picking” of participants.¹⁰ If a participant did not want to

¹⁰ See, e.g., Heckman et al. (2002) for a discussion and analysis for the impact of cream skimming on evaluation of performance measures in labor market evaluation studies. See also

actively choose a provider, he or she was assigned the job coach closest to his or her home address.¹¹

Due to a change in government policy, the service “job coaching” was phased out at the end of 2013. The reason for this policy change was a declared reorientation of labor market policy on the part of the liberal-conservative government, with an increased focus on the long-term unemployed.

4 Do better rated job coaches attract more participants?

In this section, some descriptive results on rating values of providers and how these relate to market outcomes are presented and discussed. All data used here and in subsequent sections are taken from registers from the Swedish Public Employment Service. Table 1 presents descriptive results regarding providers of job coaching that were awarded a rating in April 2013. It shows that about 19 percent of job coaching providers (of those who actually received a rating) were assigned one star, about 63 percent two stars, and about 17 percent three stars. Also shown is the average number of participants within the time period when the rating values were calculated (January to November 2012). Only those providers with at least 10 participants during that period got their rating published. For those providers with two stars, the average number of participants was somewhat larger than for those with one and three stars, respectively. To some extent, this follows from larger providers being given more weight when calculating the star rating values.

the discussion in Behaghel et al. (2014) and Koning and Heinrich (2013) addressing the need to account for such aspects.

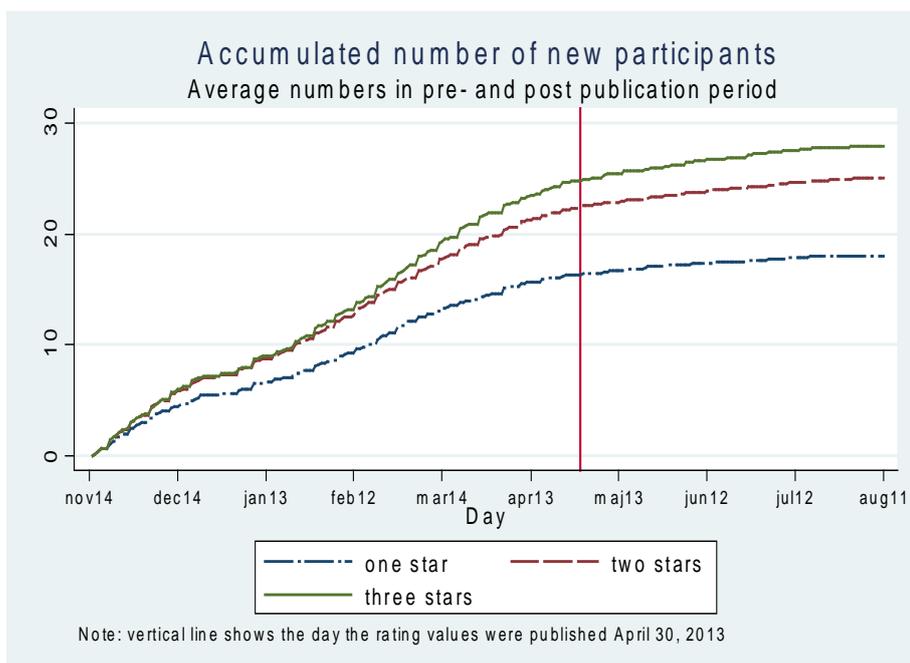
¹¹ It would be interesting to study to what extent participants made an active choice of job coach or not. Unfortunately this kind of information has not been available in the data.

Table 1
Distribution of providers of job coaching according to their rating.

Rating values	Number of providers	Average number of participants during the time rating values were calculated
One star	155	24
Two stars	506	36
Three stars	136	29
Total	797	33

Figure 1 shows that the average number of participants registered with providers with one, two, or three stars, respectively, before and after the publishing of the rating values on April 30, 2013.

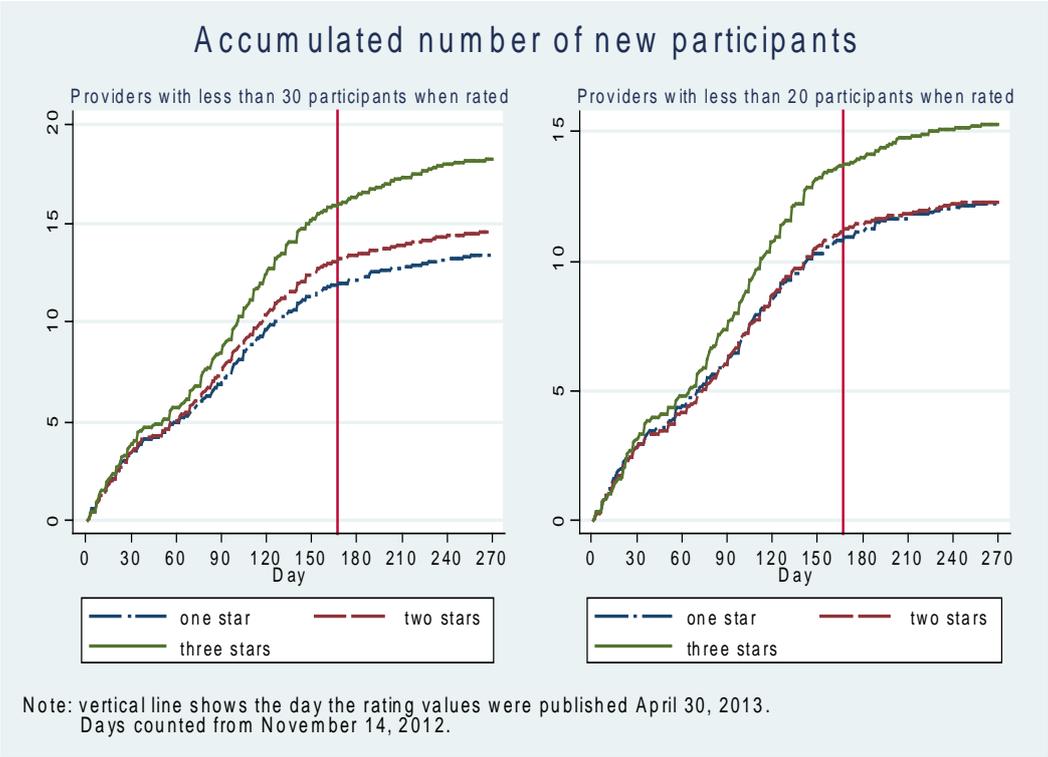
Figure 1



The relative share of persons choosing higher rated providers of job coaching (those having two or three stars) steadily increased over time in comparison with providers with one star; noteworthy is the fact that such development took place already before any rating values were published on April 30, 2013. This indicates that better providers were able to attract new

clients even without the availability of an official performance indicator. As can be seen from table 1, providers with two stars had more clients during the period that rating values were calculated. Previous size might have had a positive impact on subsequent enrollment rates, i.e., information on the number of previous clients might have been taken as proxy for better quality, leading to larger subsequent participation rates.¹² We tested the significance of the initial size of provider by showing figures where the numbers of initial participants were set to max 30 and 20 persons, respectively, see figure 2. It shows that the lower the upper bound, the less significant the difference between providers with two stars compared to those with one star. For the three star rated providers, imposing this kind of restrictions does not seem to make a difference.

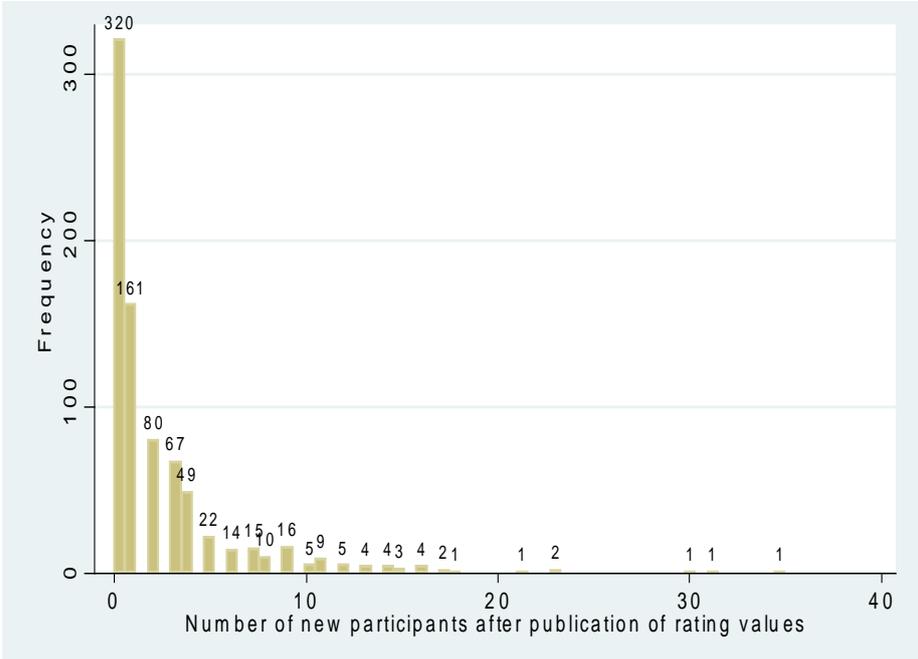
Figure 2



¹² See Hellofs and Jacobson (1999) for a discussion of mechanisms on the role of market share for perceived quality.

In figure 3, the number of new participants from April 30 to August 15, 2013—i.e., from the date the ratings were published to the date the data underlying the estimates of the current study was extracted—is shown. Only a few providers out of a total of 797 providers received more than two new participants, while 320 providers did not receive any new participants. Obviously, there was an oversupply of providers in the post-publication period.

Figure 3



Next, the relation between the chances of getting new participants and a job coach’s rating value was scrutinized by means of regression analysis. Due to the fact that the job coaches’ results (none, one, two, three, etc., new participants) are an enumeration of individual events, a negative binomial regression model was used.¹³ Larger providers of job coaching might

¹³ For a discussion of the various possibilities of analyzing this kind of data, see O’Hara and Kotze (2010). The basic idea of the model is the assumption that the data are negative binomially distributed. Here I used the standard estimation procedure offered in Stata; see Perez Truglia (2009) for details.

have been able to attract new clients because they were more well-known companies in the market for personnel services, or, as already pointed out, that mere size might have been taken as proxy for better quality. Obviously, such aspects should be controlled for if one wants to find out the value added of providing performance indicators. Furthermore, it is important to take differences across regional labor markets into account, as locally determined factors may imply rather different prospects for providers of job coaching to attract new participants. Therefore subsequent estimates include controls for different procurement areas where providers of job coaching had been operating and the "size" of the job coaches as measured by the number of participants in the period from January to November 14, 2012 (i.e., the period underlying the calculations of rating values). The intention of the model is to measure the importance of having a rating value of one, two, or three stars for the provider of job coaching to get new participants; note that only providers with a published rating are included in the estimates.

According to column (1) in table 2, the variable "Rating values," which can take on the values of 1, 2 or 3 stars, is positive and statistically significant, resulting in the conclusion that the more stars a provider of job coaching was awarded, the more participants signed up. The model in column (2) has the same structure as in column (1), except from that there are two separate indicators included: one indicator that marks a rating equal to one star, the other a rating equal to three stars. The respective coefficient estimate indicates the difference in having one and three stars compared to having two stars, respectively. According to those estimates, the provider of job coaching with three stars had a 37% "higher incidence" of receiving new participants (during the period studied) compared to a provider of job coaching

with two stars, controlling for the delivery area and initial size.¹⁴ Providers of job coaching with just one star had a 14% “lower incidence” to attract new participants; however, the latter result is not statistically significantly different from 0. Thus, it looks like that there is a comparatively larger positive impact of having three stars than having two stars, compared to what a provider of job coaching loses by having one star instead of two stars. This is in line with the conclusions drawn from looking at figure 2.

Table 2

The significance of the publication of the rating values. Actual number of participants since the publication of the rating values. Negative binomial estimates.

	(1)	(2)
Rating values (1, 2, 3 stars)	0.255 ***	
	(0.0893)	
One star		-0.137
		(0.156)
Two stars		reference
Three stars		0.367 ***
		(0.116)
Observations	797	797
Ln(alpha)	-0.103	-0.100
	(0.103)	(0.105)
Robust standard errors in parentheses *** p <0.01, ** p <0.05, * p <0.1		

Note: in addition to those above reported control variables, controls for delivery areas, the number of participants at the time of the determination of the rating value, and the total number of areas where a provider of job coaching was represented have been included.

It should be stressed that the results in table 2 cannot say anything about the causal effect of published rating values for the market outcomes of rated providers. As can be seen from figure 1, there was a relatively larger increase in enrollment numbers for providers with two

¹⁴More accurate would be to write that there was a "37% higher incidence in log differences" between providers of job coaching with two stars compared to providers of job coaching with three stars.

stars, and even more so for providers with three stars, compared to providers with one star rating already before the rating values were published. In the following sections, we seek to determine causal effects of providing information on performance of providers.

5 Estimating effects of rating on market outcomes

As stated earlier, there are some obstacles to my attempts to measure the actual impact of publicized rating values for job seekers' choice of provider. Other factors may have been important for such outcomes, e.g., some job seekers who were about to choose a provider of job coaching may have used other sources to acquire information. Such information could be spread by persons within a participant's social networks, but might also have been communicated by case workers at the local PES office, although the latter was prohibited according to PES instructions. Case workers at the PES were only allowed to inform participants about the existence of the website. Thus, it is not possible to clarify the actual processes underlying each individual's choice of job coach, so we have to stick with "revealed preferences."

The way rating values were introduced in April 2013 does not allow the estimation of more general effects of its impact for the participants' decisions as there is no possibility of quantifying the counterfactual outcome, i.e., what would have happened if the PES did not publish rating values. However, it is possible to use a quasi-experimental approach to distinguish causal effects of the introduction of rating values. This can be done for the suppliers of job coaching that were just on the verge of getting one vs. two stars, and three vs. two stars, respectively. As shown in Figure A1 in the Appendix, there are two thresholds determining the number of stars a job coach was awarded. The underlying assumption that will allow us to comment on the causal effects of having one, two, or three stars in the rating is that it was pure coincidence that determined whether a provider of job coaching close to the

thresholds ended up just to the left or right of it. No one knew beforehand where the thresholds for the different stars would be set, which made it practically impossible to “game the system” and to end up with a certain (i.e., better) rating.

We conduct the analysis by studying the closer environment surrounding the threshold value using a regression discontinuity (RD) approach.¹⁵ The way star ratings are generated allows us to make use of a “sharp” RD-design, which means that “treatment”, i.e., getting a certain rating, is a deterministic function of a covariate X , which in the present setting is the score value. Using the notation in Imbens and Lemieux (2008), p. 617, that relation can be described as follows:

$$R_i = 1 \text{ if } X_i \geq c, \tag{5.1}$$

where R_i denotes the rating of a provider and X_i its score value, while c is the predefined cut-off point. The (causal) effect on outcome Y_i at the discontinuity point c is determined as

$$\lim_{x \downarrow c} E[Y_i|X_i] - \lim_{x \uparrow c} E[Y_i|X_i]. \tag{5.2}$$

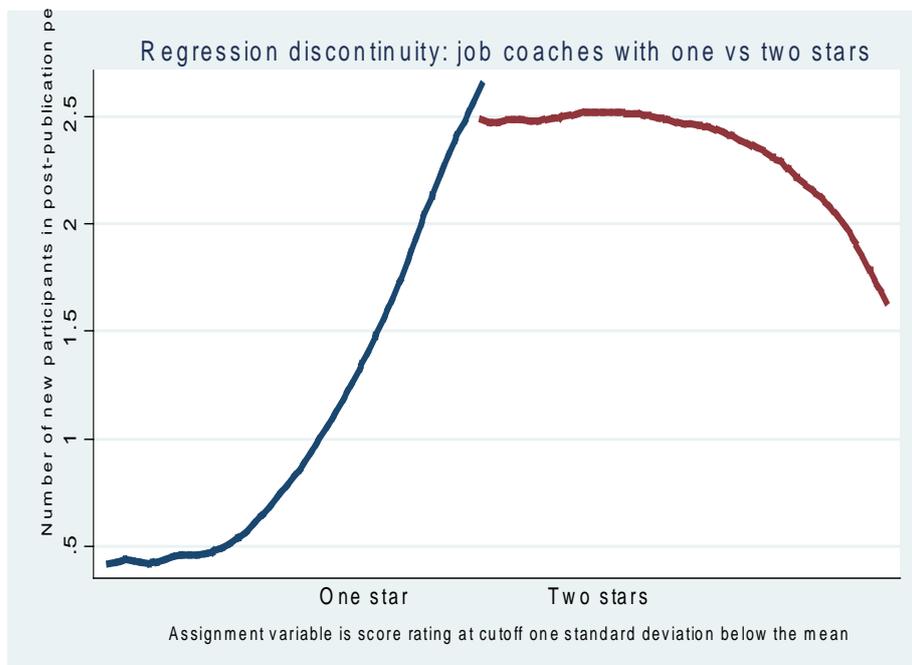
Note that there are two cut-off values; i.e., equations 5.1 and 5.2 are applied separately for the cut-off values that demark the increase both from one to two stars, and from two to three stars, respectively. The outcome variable studied is the number of new participants in the period following the publication of rating values on April 30, 2013.

¹⁵ See, e.g., Imbens and Lemieux (2008) for a detailed discussion. In the calculations reported here, I used the "rd" command in STATA; see Nichols (2011).

5.1 RD-estimation results

Figure 4 shows results at the threshold that determines whether a provider of job coaching received one or two stars. The figure suggests that there are signs of discontinuity at the cut-off point, but the difference between having one vs. two stars is rather small; the precise estimates are shown in Appendix A2. The corresponding Wald estimations shown there include controls for the 52 delivery areas, the number of participants at the time of the determination of the rating value, and score values squared to allow for a more flexible underlying model.

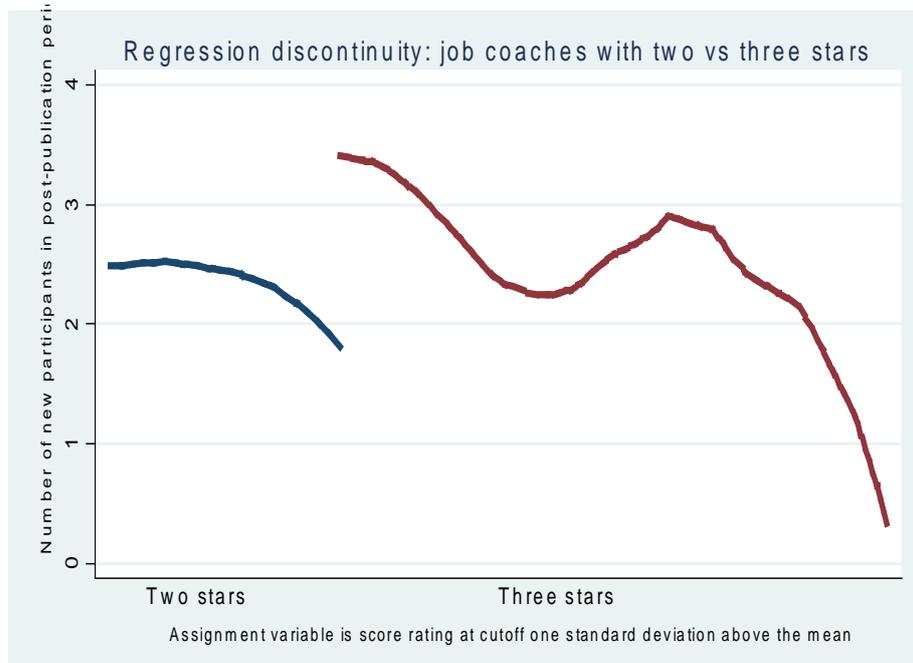
Figure 4



The situation looks quite different, however, for those providers of job coaching just to the left and right of the threshold that determined if a job coach would get two or three stars, respectively. Figure 5 shows a clear difference, which can be measured to about 1.5 additional participants, i.e., a difference between 3.5 new entrants for a provider with three stars, compared with just 2 new entrants for a provider with two stars. This should be read as follows: providers of job coaching that had a calculated score value that was slightly larger

than the cut-off (0.068) and thus received three stars received approximately 1.5 more participants than providers of job coaching that lay just to the left of the threshold and thus received two stars.

Figure 5

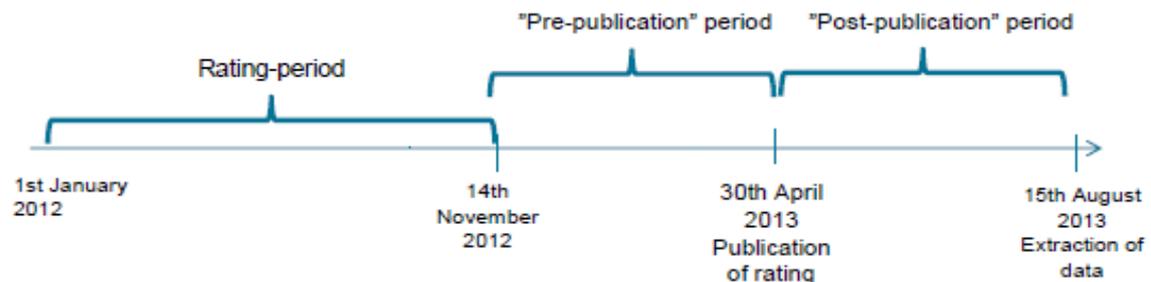


When interpreting the results, it is important to keep in mind that these are effects for providers of job coaching close to the thresholds. Accordingly, the results should not be interpreted as (average) differences for all providers of job coaching with three stars compared to all providers of job coaching with two stars. Rather they indicate the “average treatment effects at the discontinuity points” (Imbens and Lemieux, 2008), i.e., the effects for providers that approximately had the same quality, but where some were granted two stars, while others were awarded three stars. Accordingly, the measured effect can be interpreted as the impact of signaling higher quality for providers that were about similar in terms of quality with respect to their provided services.

5.2 Sensitivity estimations

The RD-estimations provide rather clear evidence; still, there might be concerns that there could be other, hidden factors driving the results. To address such concerns, we can exploit the fact that during a period prior to the publication of the rating values there were participants who chose a provider of job coaching without having access to public information about the job coach's past performance. See Figure 6 for the time frames that apply to those persons included in the calculation of the rating values, and the persons who define the comparison group in subsequent estimations.

Figure 6

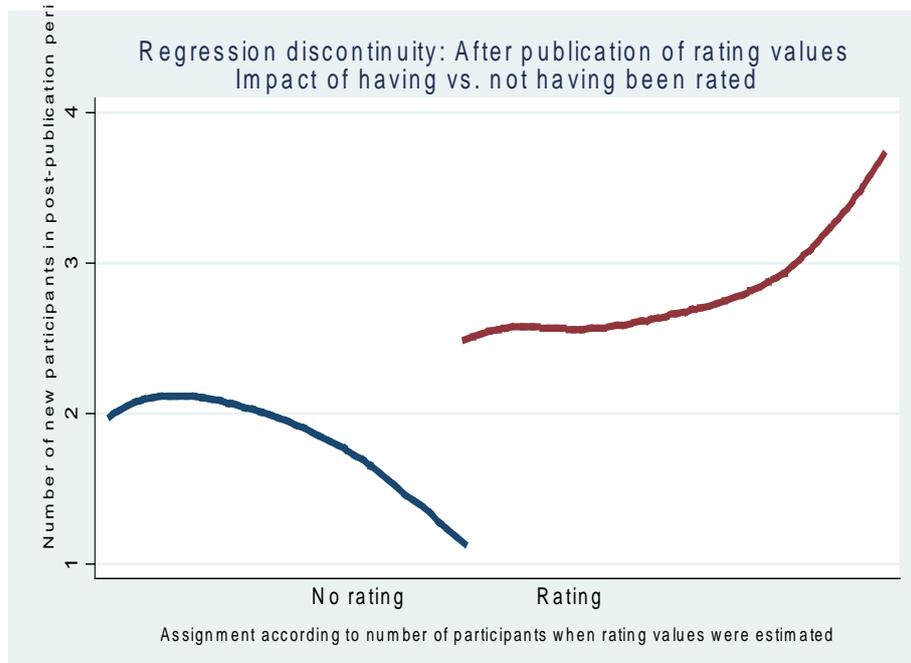


We test the reliability of the estimation procedure by performing similar calculations for the period preceding the publication of the rating values. In that period, participants chose a provider of job coaching without having knowledge of the job coaches' past performance as measured by the rating values, i.e., the "pre-publication" period shown in figure 6 above. Since there was no published rating available during that time, there should be no "discontinuity" at the thresholds of the score values that conditioned the respective rating value. Figures in Appendix A3 show that this is indeed the case. It supports the conclusion that the discontinuities found around the thresholds determining a star rating of two vs. three stars *after* publication of the rating values were caused by its publication and are not the result of some hidden factors that would imply spurious correlations.

6 The importance of (not) having a rating value

One issue that has not been addressed so far is whether it makes a difference for a provider of job coaching having received a rating value, compared to not having received a rating value. This section examines this aspect. This is done by exploiting the fact that a provider of job coaching would receive a rating value in April 2013 only if that provider had at least ten participants during the period from January to November 2012. Thus, one can compare how those with nine or fewer participants were able to attract new participants compared with providers who had ten or more participants. For that purpose, we again make use of a regression discontinuity approach. According to the outcomes shown in figure 7, a supplier who had at least ten participants in the period January-November 2012 received about 1.4 additional participants compared with a supplier who had nine participants or less in the same period. Based on a Wald-test, the difference is statistically significant at the 1% level.

Figure 7



It is not obvious what mechanisms drive these results. To some extent, a published rating could have resulted in higher visibility of that provider, which then transmitted to a larger propensity to attract new clients.¹⁶ Alternatively, it might be the case that participants saw “no-rating” as a kind of “zero”-rating, in the sense of an ordinal scale variable with values zero, one, two, or three (stars). To get a better understanding of such processes, we conducted separate RD-estimations where only those providers were included that had either one, two, or three stars, respectively. Here providers with a certain rating were compared with other providers who had no rating published, but who also would have received a similar rating if there were no requirements on the minimum number of participants. The results from those estimations, shown in table 3, indicate that the largest effects can be found for those providers who had one star (panel A), somewhat smaller effects for those with two stars (panel B), and essentially non-existent differences for those providers with three stars (panel C).¹⁷ Thus, having a rating published is better than not having a rating published, especially for those providers that had a weak quality rating. Such patterns suggest that participants looked at those providers with no rating as being worse providers. The alternative explanation—i.e., that it would have been a question of visibility—does not fit with the systematic differences of patterns in relation to increasing rating values. For the “visibility story” to be credible,

¹⁶ As shown in Appendix A.4, the actual star rating for providers in a given area were shown on a separate tab on the website, where only providers with a rating occurred.

¹⁷ The used “rd estimates” are local linear (or kernel) regression models that are estimated on both sides of the cutoff. Estimates are sensitive to the choice of what bandwidth is chosen, so by default several estimates are constructed using different bandwidths. The estimate denoted `lwald` is the default bandwidth, while `lwald50` (`lwald200`) uses half (double) the bandwidth compared to the default. See Nichols (2007) for explanations.

differences for a provider with a published vs. a non-published rating should be about similar no matter the actual rating value; apparently this is not the case.¹⁸

Table 3
RD estimations on having or not having received a rating. Outcome: Number of new participants

(Panel A)						
Providers with one star						
	Coef.	Std. Err.	z	P> z	[95%Conf.interval]	
lwald	1.834279	.7269969	2.52	0.012	.4093915	3.259167
lwald50	1	.4233532	2.36	0.018	.170243	1.829757
lwald200	2.258157	.8782427	2.57	0.010	.5368328	3.979481

(Panel B)						
Providers with two stars						
	Coef.	Std. Err.	z	P> z	[95%Conf.interval]	
lwald	1.372479	.6494346	2.11	0.035	.0996103	2.645347
lwald50	1.293451	.8376058	1.54	0.123	-.3482257	2.935129
lwald200	.672341	.4571015	1.47	0.141	-.2235614	1.568244

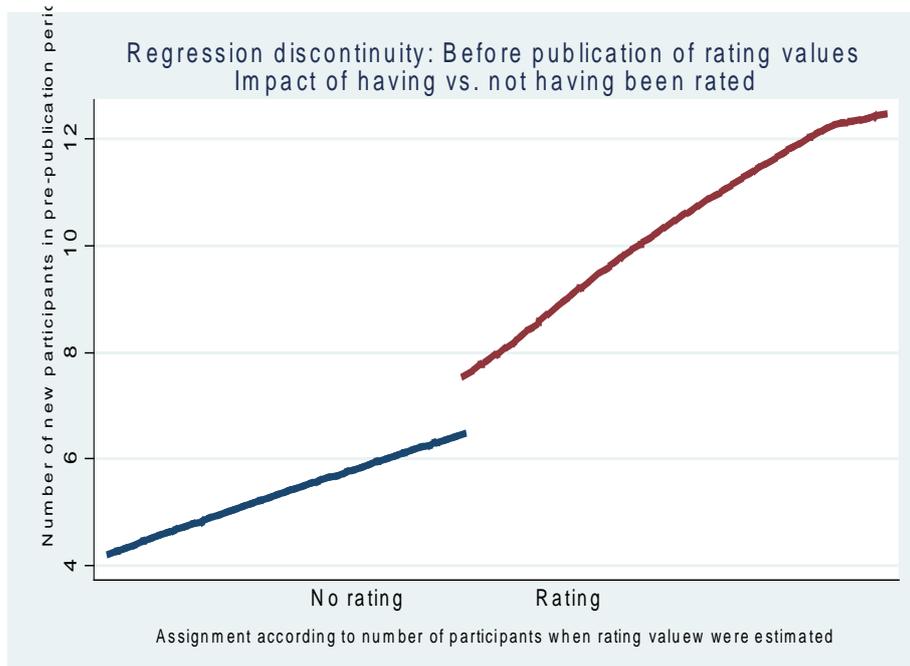
(Panel C)						
Providers with three stars						
	Coef.	Std. Err.	z	P> z	[95%Conf.interval]	
lwald	.1108577	1.513588	0.07	0.942	-2.85572	3.077436
lwald50	-.327764	2.583336	-0.13	0.899	-5.391009	4.735481
lwald200	-.1037502	1.007347	-0.10	0.918	-2.078115	1.870614

6.1 Sensitivity estimations

For the purpose of testing the RD estimations above, we run estimations covering the period preceding the publication of the rating values, shown in figure 8. The relative difference is noticeably smaller around the threshold, and in contrast to the outcome shown in figure 7, it is not statistically significant (with a p-value of .340). This confirms that there is indeed a positive effect from receiving a rating, compared to not being awarded a rating for those providers close to the cut-off in the post-publication period.

¹⁸ This reasoning follows Hilger et al. (2011), who studied consumer responses to expert opinion labels for wine.

Figure 8



7 Heterogeneous effects across participants

The results in section 5 have shown that there is a positive impact of having a better rating than average for the propensity to attract new clients. Here we address the question of whether the information was used similarly by different groups of participants. In table 4, participant characteristics in job coaching *after* the rating values were published in April 2013 are shown. The share of female participants was about 44 percent. As to the age distribution, about 55 percent of the participants were in the age group 25-49, while 27 percent were under the age of 25. The time from being registered unemployed to the date one becomes assigned to a job coach is shown according to 20 percentile increments. It shows that twenty percent waited for less than one month, while another twenty percent of participants were registered as unemployed for more than two years before they started working with a job coach. About 28 percent have post-secondary education. Notable is the fact that as many as 35 percent of the participants come from countries outside Europe.

Table 4
Mean values for participant population in post-publication period

VARIABLES	Mean	Std. Dev.
Female	0.44	0.50
Under 25 years	0.27	0.45
25-49 years	0.55	0.50
50-64 years	0.18	0.38
Very short waiting period (<33 days)	0.20	0.40
short waiting period (33–127 days)	0.20	0.40
longer waiting period (127–293 days)	0.20	0.40
long waiting period (293–750 days)	0.20	0.40
Very long waiting period (>750 days)	0.20	0.40
Compulsory education less than 9 years	0.12	0.32
Compulsory schooling 9 years or longer	0.15	0.36
Secondary education	0.46	0.50
Short post-secondary education	0.06	0.23
Long post-secondary education	0.22	0.41
Born outside Europe	0.35	0.48
Number of observations (participants)	1,882	

Next, we look at evidence for heterogeneous effects by dividing the participants along their educational background, their migrant status, as well as with respect to their gender. In table 5 below, we start by looking at participants with post-secondary education (panel A) and those without post-secondary education (panel B).

Table 5
RD estimations. Outcome: Number of new participants

(Panel A)						
Persons with post-secondary education						
Two stars vs. one star						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.2472322	.1528095	1.62	0.106	-.052269	.5467334
lwald50	.4046603	.1931084	2.10	0.036	.0261747	.7831458
lwald200	.2292366	.1204651	1.90	0.057	-.0068707	.4653438
Three stars vs. two stars						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.9332135	.4307574	2.17	0.030	.0889445	1.777483
lwald50	1.15575	.5972969	1.93	0.053	-.0149299	2.326431
lwald200	.75956	.3424793	2.22	0.027	.0883129	1.430807
(Panel B)						
Persons without post-secondary education						
Two stars vs. one star						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	-.3989423	.7438548	-0.54	0.592	-1.856871	1.058986
lwald50	.0581309	.8426623	0.07	0.945	-1.593457	1.709719
lwald200	-.1637898	.6130799	-0.27	0.789	-1.365404	1.037825
Three stars vs. two stars						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.7085096	.5469428	1.30	0.195	-.3634787	1.780498
lwald50	.6047292	.7104862	0.85	0.395	-.7877983	1.997257
lwald200	.4910422	.4678469	1.05	0.294	-.425921	1.408005

Note: rd-estimations without control variables

Obviously there are noteworthy differences. Participants with upper secondary education were more likely to choose providers that had a higher rating. Notice that this also held true at the cut-off determining which providers were awarded two stars vs. one star. For participants with less than upper secondary education (panel B), it seems like the introduction of performance indicators did not have a significant impact on their decisions about which provider to choose.

In table 6, we divided the groups into those who were either born in a non-European country, or persons born in Sweden or a country within Europe. Even here, there are large differences across groups. Participants born in a non-European country did not seem to make use of the star rating (see panel A), while for native born Swedes (as well as persons born in another European country) there were positive correlations (panel B), i.e., the higher the rating, the larger the chances that a provider of job coaching would get a person from that group to sign up.

Table 6
RD estimations. Outcome: Number of new participants

(Panel A)						
Persons born in non-European country						
Two stars vs. one star						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	-.9317572	.6204789	-1.50	0.133	-2.147874	.2843591
lwald50	-.8471497	.6208347	-1.36	0.172	-2.063963	.369664
lwald200	-.6438198	.5122212	-1.26	0.209	-1.647755	.3601153
Three stars vs. two stars.						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.3881376	.3752343	1.03	0.301	-.3473081	1.123583
lwald50	.6101273	.5346288	1.14	0.254	-.4377259	1.657981
lwald200	.204182	.3075176	0.66	0.507	-.3985414	.8069053
(Panel B)						
Persons born in Sweden or other European country						
Two stars vs. one star						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.6904895	.3800135	1.82	0.069	-.0543232	1.435302
lwald50	1.030563	.4896404	2.10	0.035	.0708857	1.990241
lwald200	.6360801	.3227044	1.97	0.049	.0035912	1.268569
Three stars vs. two stars						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	1.126381	.5075272	2.22	0.026	.1316459	2.121116
lwald50	.989027	.6276712	1.58	0.115	-.2411859	2.21924
lwald200	.8241711	.4410631	1.87	0.062	-.0402967	1.688639

Note: rd-estimations without control variables

Finally, in table 7, we look at differences with respect to gender. Women (panel A) were more likely to make use of information signaling three stars. For men (panel B), there were no statistically significant outcomes.

Table 7
RD estimations. Outcome: Number of new participants

(Panel A)						
Women						
Two stars vs. one star						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.0326491	.2975017	0.11	0.913	-.5504436	.6157417
lwald50	.039914	.3359277	0.12	0.905	-.6184922	.6983201
lwald200	.1799864	.251702	0.72	0.475	-.3133404	.6733133
Three stars vs. two stars						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.9748409	.3821785	2.55	0.011	.2257848	1.723897
lwald50	.8923169	.505932	1.76	0.078	-.0992916	1.883925
lwald200	.6807217	.3201229	2.13	0.033	.0532924	1.308151
(Panel B)						
Men						
Two stars vs. one star						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	-.1918485	.5527084	-0.35	0.729	-1.275137	.8914402
lwald50	.3992233	.6618217	0.60	0.546	-.8979233	1.69637
lwald200	-.1338753	.4426316	-0.30	0.762	-1.001417	.7336666
Three stars vs. two stars						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.6318068	.4645884	1.36	0.174	-.2787698	1.542383
lwald50	.6885157	.6169132	1.12	0.264	-.520612	1.897643
lwald200	.4412371	.3876157	1.14	0.255	-.3184757	1.20095

To summarize, there are notable differences in the use of publicly provided information regarding quality of providers across different groups. Female participants, those that have post-secondary education, and those born in Sweden or another European country seemed to make use of such information more than others; in contrast, those born in non-European countries, as well as the less well educated, did not seem to make use of the rating to any larger extent. One might wonder if the reported gender differences are due to systematic differences with respect to educational background, i.e., that more men using the services of a job coach would be lower educated, which then could explain the weak correlation as to the use of quality information. Doing similar estimations as above for men with and without post-secondary education does actually result in somewhat different outcomes (results not shown here), where men with post-secondary education used the ratings almost to the same extent as women. For women, there were no clear differences with respect to their educational background, i.e., their use of ratings that determined three star-rated providers is approximately similar. Thus, there seemed to be certain factors inhibiting poorly educated men from using the performance measure to the same degree as women did.

One aspect that we have not addressed so far is the importance of the proximity of providers. In some areas, the distances between the residence of participants and the provider's different locations could be rather large. Especially in northern parts of Sweden, the supply areas are rather wide and sparsely populated; see figure A3 in the Appendix showing the different areas. This might lead some groups of people to find it less convenient or more costly to choose a better rated provider of job coaching. On the other hand, the provided information could be more valuable for people living in remote areas, reassuring them that the time spent traveling wouldn't be wasted. In table 8, results for people living in more and less densely populated areas are shown.

Table 8
RD estimations. Outcome: Number of new participants

(Panel A)						
More densely populated areas*						
Two stars vs. one star						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	-2.418906	1.904836	-1.27	0.204	-6.152316	1.314503
lwald50	-2.037668	2.043606	-1.00	0.319	-6.043061	1.967726
lwald200	-1.484472	1.491909	-1.00	0.320	-4.40856	1.439616
Three stars vs. two stars						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.7341926	1.183429	0.62	0.535	-1.585286	3.053671
lwald50	1.486942	1.529918	0.97	0.331	-1.511642	4.485526
lwald200	.6601061	.8967014	0.74	0.462	-1.097396	2.417609
(Panel B)						
Less densely populated areas						
Two stars vs. one star						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	1.178965	.7400168	1.59	0.111	-.2714413	2.629371
lwald50	2.201341	1.125836	1.96	0.051	-.0052571	4.407938
lwald200	.8309161	.5904018	1.41	0.159	-.3262502	1.988082
Three stars vs. two stars						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	2.290067	1.154867	1.98	0.047	.0265698	4.553564
lwald50	2.36277	1.391732	1.70	0.090	-.3649744	5.090514
lwald200	1.544323	1.112165	1.39	0.165	-.6354804	3.724127

Note: * Areas included here are Gothenburg, Helsingborg, Huddinge, Lund, Malmö, Stockholm, Södertälje, Trelleborg, Uppsala, Västerås, and Örebo.

The results are rather clear cut, showing that persons living in less densely populated areas seem to be more likely to make use of information on ratings than persons living in more densely populated areas. This holds true both for two- vs. one-star rated providers, as well as three- vs. two-star rated providers. We also looked separately at men and women living in more or less densely populated areas (results not shown here), showing that the results based on living area are even more pronounced for men than they are for women. So it appears that better educated men living in more rural areas used the ratings information more than others.

8 Discussion and concluding remarks

The previous sections show that there are clear indications that the publication of job coach providers' performance in terms of rating values has been a relevant piece of information for at least some groups using the service. In general, publication of rating values strengthened the link between "quality" (in terms of their registered ability to get participants into jobs) and the number of new participants that providers of job coaching were able to

attract. Rating values seem to matter most for providers that are about similar in terms of quality, but differ in their rating values. However, it does not appear that providing this quality indicator has been a “game changer,” as one can discern patterns of an increase in market outcomes for better rated providers of job coaching even before the rating values were published in April 2013. This indicates that information on different kinds of (ad-hoc) quality indicators of providers were available even before rating values were published.

The results show that it is of particular importance to signal highest quality (three stars) for the purpose of attracting new participants. Its impact is significant, both statistically but also economically: 3.5 participants for the job coach who got rating of three stars, compared with barely 2 participants for an approximately equally good provider who received two stars, i.e., an 80 percent increase. Furthermore, there is evidence that job coaching providers with published ratings were able to attract more participants than those job coaches who did not receive a rating. On the margin, it seems to be more important to receive a rating for those providers that were relatively weaker in terms of provided quality. For the better providers, i.e., those who would have been awarded three stars, a published rating does not seem to be as important, suggesting that better providers were able to attract new clients irrespective of having a published rating or not; possibly, this was achieved by participants that communicated their experiences to other participants within their social networks.

From a more theoretical perspective, the results imply that a voucher-based quasi-market system in ALMP actually can result in efficiency gains, as those suppliers that performed better increased their market share. Those patterns were reinforced by providing a measurement of quality of services. One aspect that has not been addressed in this study, and which is of vital importance for the issue of efficiency gains of privatization of parts of the public sector, concerns the issue of how the choice of provider of job coaching translates to

improved opportunities to find work for those who participate. This issue is open for future research.

Regarding the issue of social stratification, i.e., the question of whether the provided information was used equally across social groups, the results show that persons that were better educated, as well as those born in Sweden or other European country, were more prone to make use of the provided information, while less well educated persons and those born in a non-European country, i.e., groups that regularly are seen as more vulnerable in the labor market and supposedly are in greater need of support, did not seem to utilize such information. From that perspective, the results are somewhat disappointing as one might have hoped that providing easily accessible information would help vulnerable groups more than those who already had access to better social networks and alternative ways of gathering information. To some extent, the results may be caused by lower expectations on the part of the more vulnerable groups that using that kind of information actually would make a difference in their situation.

To summarize, it appears that there is a tradeoff as to efficiency and inequality with respect to services in active labor market policy that are contracted to private companies in a system based on consumer choice. The question of how to look at such outcomes ultimately becomes a matter of political philosophy and preferences. One could argue that every participant at least formally had similar chances of choosing a better provider so that equality in opportunity was ensured. On the other hand, the systematic differences in actual choices across different groups suggest that many participants were unable to exercise appropriate choice. A system based on free choice and marketization of public services apparently requires competent and motivated consumers, otherwise tendencies of social stratification of services risk being reinforced. This should be taken into account by policy makers when deciding which services might move to a voucher-based choice system.

References

- Barnow, Burt S. and Carolyn J. Heinrich (2010), “One Standard Fits All? The Pros and Cons of Performance Standard Adjustments,” *Public Administration Review*, Vol. 70, 60–71.
- Behaghel, Luc, Bruno Crépon, and Marc Gurgand (2014), “Private and Public Provision of Counseling to Job Seekers: Evidence from a Large Controlled Experiment,” *American Economic Journal: Applied Economics*, Vol. 6, 142–174.
- Benmarker, Helge, Erik Grönqvist, Björn Öckert (2013), “Effects of outsourcing employment services: evidence from a randomized experiment,” *Journal of Public Economics*, Vol. 98, 68–84.
- Bergman, Mats, Per Johansson, Sofia Lundberg, and Giancarlo Spagnolo (2014), “Privatization and quality: Evidence from Elderly Care in Sweden,” Centre for Economic Policy Research Discussion Paper No. 9939.
- Besley, Timothy, and Maitreesh Ghatak (2005), “Competition and Incentives with Motivated Agents,” *American Economic Review*, Vol. 95, 616–636.
- Besley, Timothy, Gwyn Bevan and Konrad Burchardi (2009), “Naming & Shaming: The impacts of different regimes on hospital waiting times in England and Wales,” Center for Economic and Policy Research Discussion Paper 7306.
- Betts, Julian R. and Tom Loveless, (2005), “School Choice, Equity and Efficiency,” Chapter 1 in Julian R. Betts and Tom Loveless (Eds.) *Getting Choice Right: Ensuring Equity and Efficiency in Education Policy*, Washington, D.C.: Brookings Institution Press, 1–13.
- Black, Sandra E., Kalena E. Cortes, and Jane Arnold Lincove (2014), “Efficacy vs. Equity: What Happens When States Tinker with College Admissions in a Race-Blind Era?” NBER Working Paper No. 20804.
- Blank, Rebecca M. (2000), “When Can Public Policy Makers Rely on Private Markets?” *Economic Journal*, Vol. 110: C34–C49.
- Blank, Rebecca M. (2002), “Can equity and efficiency complement each other?” *Labour Economics*, Vol. 9, 451–468.
- Bourguignon, François, Francisco H. G. Ferreira and Michael Walton (2007), “Equity, efficiency and inequality traps: A research agenda,” *Journal of Economic Inequality*, Vol. 5, 235–256.
- Burgess, Simon, Deborah Wilson, Jack Worth (2013), “A natural experiment in school accountability: The impact of school performance information on pupil progress,” *Journal of Public Economics*, Vol. 106, 57–67.

Cooper, Zack, Stephen Gibbons, Simon Jones and Alistair McGuire (2011), “Does Hospital Competition Save Lives? Evidence From The English NHS Patient Choice Reforms,” *Economic Journal*, Vol. 121, 228–260.

Clarke, Marguerite (2007), “The impact of higher education rankings on student access, choice, and opportunity. College and university ranking systems,” *Global perspectives and American challenges*. Washington, DC, Institute for Higher Education Policy.

Dan, Sorin and Rhys Andrews (2014). “Market-type mechanisms and public service equity in Europe: A review of evidence,” *status: submitted*.

Davis, Lucas W., and Gilbert E. Metcalf (2014), “Does Better Information Lead to Better Choices? Evidence from Energy-Efficiency Labels,” No. w20720. National Bureau of Economic Research.

Faber, Marjan, Marije Bosch, Hub Wollersheim, Sheila Leatherman, and Richard Grol (2009) “Public reporting in Health Care: How do Consumers use Quality-of-Care-Information? A systematic Review,” *Medical Care*, Vol. 47, 1–8.

Gartell, Marie (2011), ”Interna och externa coachningstjänster- en utvärdering av resultatet”, Arbetsförmedlingen URA 2011:2.

Hart, Oliver, Andrei Shleifer and Robert W. Vishny (1997), “The Proper Scope of Government: Theory and an Application to Prisons,” *Quarterly Journal of Economics*, Vol. 112, 1127–1161.

Hastings, Justine S. and Jeffrey M. Weinstein (2008), “Information, School Choice, and Academic Achievement: Evidence from two Experiments,” *Quarterly Journal of Economics*, Vol. 123, 1373–1414.

Heckman, James, Caroline J. Heinrich and Jeffrey Smith (2002), “The Performance of Performance Standards,” *Journal of Human Resources*, Vol. 37, 778–811.

Heinrich, Caroline J. (2002), “Outcomes–Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness,” *Public Administration Review*, Vol. 62, 712–725.

Hellos, Linda L. and Robert Jacobson (1999), “Market Share and Customer’s Perceptions of Quality: When Can Firms Grow Their Way to Higher Versus Lower Quality?” *Journal of Marketing*, Vol. 63, 16–25.

Hibbard, Judith H., Nancy Berkman, Lauren A. McCormack and Elizabeth Jael (2002), “The Impact of a CAHPS Report on Employee Knowledge, Beliefs, and Decisions,” *Medical Care Research and Review*, Vol. 59, 104–116.

Hilger, James, Greg Rafert and Sofia Villas-Boas (2011), “Expert Opinion and the Demand for Experience Goods: An Experimental Approach in the Retail Wine Market,” *Review of Economics and Statistics*, Vol. 93, 1289–1296.

Hipp, Lena and Mildred E. Warner (2007), “Market Forces for the Unemployed? Training vouchers in Germany and the U.S.,” *Social Policy & Administration*, Vol. 42, 77–101.

Hoxby, Caroline M. (1996), “Are Efficiency and Equity in School Finance Substitutes or Complements?” *Journal of Economic Perspectives*, Vol. 10, 51–72.

Imbens, Guido W. and Thomas Lemieux (2008), “Regression discontinuity designs: A guide to practice,” *Journal of Econometrics*, Elsevier, Vol. 142, 615–635.

Jilke, Sebastian (2014), “Choice and Equality: Are Vulnerable Citizens Worse-Off after Liberalization Reforms?” *Public Administration*, in Press.

Jin, Ginger and Alan Sorensen (2006), “Information and Consumer Choice: The Value of Publicized Health Plan Ratings,” *Journal of Health Economics*, Vol. 26, 248–275.

Jha, Ashish and Arnold M. Espstein (2006), “The Predictive Accuracy Of The New York State Coronary Artery Bypass Surgery Report-Card System,” *Health Affairs*, Vol. 25, 844–855.

Kahneman, Daniel (2011), “Thinking Fast and Slow,” New York: Farrar, Strauss and Giroux.

Katz, Michael (2013), “Provider competition and healthcare quality: More bang for the buck?” *International Journal of Industrial Organization*, Vol. 31, 612–625.

Kolstad, Jonathan T. (2013), “Information and Quality When Motivation is Intrinsic: Evidence from Surgeon Report Cards”, *American Economic Review*, Vol. 103, 2875–2910.

Koning, Pierre and Caroline J. Heinrich (2013), “Cream-Skimming, Parking and Other Intended and Unintended Effects of High-Powered, Performance-Based Contracts,” *Journal of Policy Analysis and Management*, Vol.32, 461–483.

Koning, Pierre and Karen van der Wiel (2013), “Ranking the Schools: How School-Quality Information affects School Choice in the Netherlands,” *Journal of the European Economic Association*, Vol. 11, 466–493.

Ladd, Helen F. (2002), “School Vouchers: A Critical View,” *Journal of Economic Perspectives*, Vol. 16, 3–24.

Laun, Lisa and Peter Skogman Thoursie (2014), “Does privatisation of vocational rehabilitation improve labour market opportunities? Evidence from a field experiment in Sweden,” *Journal of Health Economics*, Vol. 34, 59–72.

le Grand, Julian (1990), "Equity Versus Efficiency: The Elusive Trade-Off," *Ethics*, Vol. 100, 554–568.

le Grand, Julian (1991), "Quasi-markets and social policy," *Economic Journal*, Vol. 101, 1256–1267.

le Grand, Julian (2006), "Equality and Choice in Public Services," *Social Research*, Vol. 73, 695–710.

Leckie, George and Harvey Goldstein (2009), "The limitations of using school league tables to inform school choice," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 172, 835–851.

Liljeberg, Linus, Sara Martinson och Jonas Thelander (2012), "Vad innebär det att bli coachad? En utvärdering av jobbcoachningen vid Arbetsförmedlingen", IFAU-Rapport 2012:24.

Lockwood, Ben and Francesco Porcelli (2013), "Incentive Schemes for Local Government: Theory and Evidence from Comprehensive Performance Assessment in England," *American Economic Journal: Economic Policy*, Vol. 5, 254–286.

Mizala, Alejandra and Miguel Urquiola (2013), "School markets: The impact of information approximating schools' effectiveness," *Journal of Development Economics*, Vol. 103, 313–335.

Muriel, Alastair and Jeffrey Smith (2011) "On Educational Performance Measures," *Fiscal Studies*, Vol. 32, 187–206.

Nichols, Austin 2011th rd 2.0: Revised Stata module for regression discontinuity estimation. <http://ideas.repec.org/c/boc/bocode/s456888.html>

Nichols, Austin (2007), "Causal Inference with Observational Data," *Stata Journal*, Vol.7, 507–541.

Ostry, Jonathan D., Andrew Berg and Charalambos G. Tsangarides (2014), "Redistribution, inequality, and growth," *IMF Staff Discussion Note*, SDN/14/02.

O'Hara, Robert B. and D. Johan Kotze (2010), "Do not log-transform count data," *Methods in Ecology and Evolution*, Vol. 1, 118–122.

Perez Truglia, Ricardo (2009), "Applied Econometrics using Stata," Department of Economics Harvard University, Early draft.

Pope, Devin G. (2009), "Reacting to rankings: Evidence from 'America's Best Hospitals,'" *Journal of Health Economics*, Vol. 28, 1154–1165.

Propper, Carol and Deborah Wilson (2003), “The Use and Usefulness of Performance Measures in the Public Sector,” *Oxford Review of Economic Policy*, Vol. 19, 250–267.

Propper, Carol, Simon Burgess and Denise Gossage (2008), “Competition and Quality: Evidence From The NHS Internal Market 1991–9,” *Economic Journal*, Vol. 118, 138–170.

Rothstein, Jesse M. (2006), “Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions,” *American Economic Review*, Vol. 96, pp. 1333–1350.

Shleifer, Andrei (1998), “State versus Private Ownership,” *Journal of Economic Perspectives*, Vol. 12, 133–150.

Struyven, Ludo and Geert Steurs (2005), “Design and redesign of a quasi-market for the reintegration of job seekers: empirical evidence from Australia and the Netherlands,” *Journal of European Social Policy*, Vol. 15, 211–229.

Turkheimer, Eric, Andreana Haley, Mary Waldron, Brian D’Onofrio and Irving I. Gottesman (2013), “Socioeconomic Status Modifies Heritability of IQ in Young Children,” *PSYCHOLOGICAL SCIENCE*, Vol. 14, 623–628.

Varkevisser, Marco, Stephanie A. van der Geest, and Frederik T. Schut (2012), “Do Patients Choose Hospitals with High Quality Ratings? Empirical Evidence from the Market for Angioplasty in the Netherlands,” *Journal of Health Economics*, Vol. 31, 371–378.

Winterhager, Henrik, Anja Heinze and Alexander Spermann (2006), “Deregulating job placement in Europe: A microeconomic evaluation of an innovative voucher scheme in Germany,” *Labour Economics*, Vol. 13, 505–517.

Woessmann Ludger and Gabriela Schuetz (2006), “Efficiency and Equity in European Education and Training Systems,” Analytical Report Prepared for the European Commission, Bruxelles: EU Commission.

A.1 Calculation of rating values

Rating values regarding the job coach's past performance were obtained by calculations based on register data of participants who were assigned to a job coach in the period January to November 14, 2012. A supplier could operate in one of the 52 different supply areas where companies (i.e. providers) could sign up by contacting the local PES. When calculating the rating value, each provider of job coaching within an area has been treated as an isolated provider, which means that if a supplier had at least ten participants in that area this would result in a unique rating for that provider for that area. This implies that larger providers that had been engaged in several supply areas did receive several, possibly different, ratings across the country. See the map (Figure A3) showing how supply areas were distributed across Sweden.

More in detail the rating is achieved as follows. For every job seeker in job coaching that was registered between January and November 2012 a value is calculated that indicates the likelihood that he/she will get a job within a certain period of time. The calculation is done based on statistical profiling that take into account the individual's age and sex, country of birth, unemployment record, job skills, education, etc. By adding the different expected values for all job seeker that were registered with a supplier of job coaching, a measure of the expected outcome (i.e., the number of participants expected to get a job) is estimated for that job coach. This number is then compared with the actual outcome, i.e. how many participants that were registered with that supplier that did actually find a job during the period measured. Example: Suppose that a supplier had three participants were estimated to have a probability of getting a job that is equal to 0.1, 0.2, and 0.4, respectively. Putting these three values together this adds to 0.7 ($= 0.1+0.2+0.4$). If it turned out that one of these three people got a job, the net difference becomes equal to 1 ($= 1+0+0$). According to this example the difference between the actual and expected value becomes 0.3 ($= 1-0.7$). Dividing that result

by the number of participants returns $0.1 (= 0.3/3)$. For ease of presentation, let call this number "score value". This score value is the basis for determining how well a supplier has been doing in helping participants to find a job in comparison to other suppliers of job coaching. In the estimations a minimum of ten participants for a job coach in each respective supply area under the observation period January to November 2012 were required to result in publication of a rating value. The reason to apply a lower bound of ten participants was to guarantee that the estimated ratings to a larger extent were the result of a provider's performance rather than the outcome of random events.

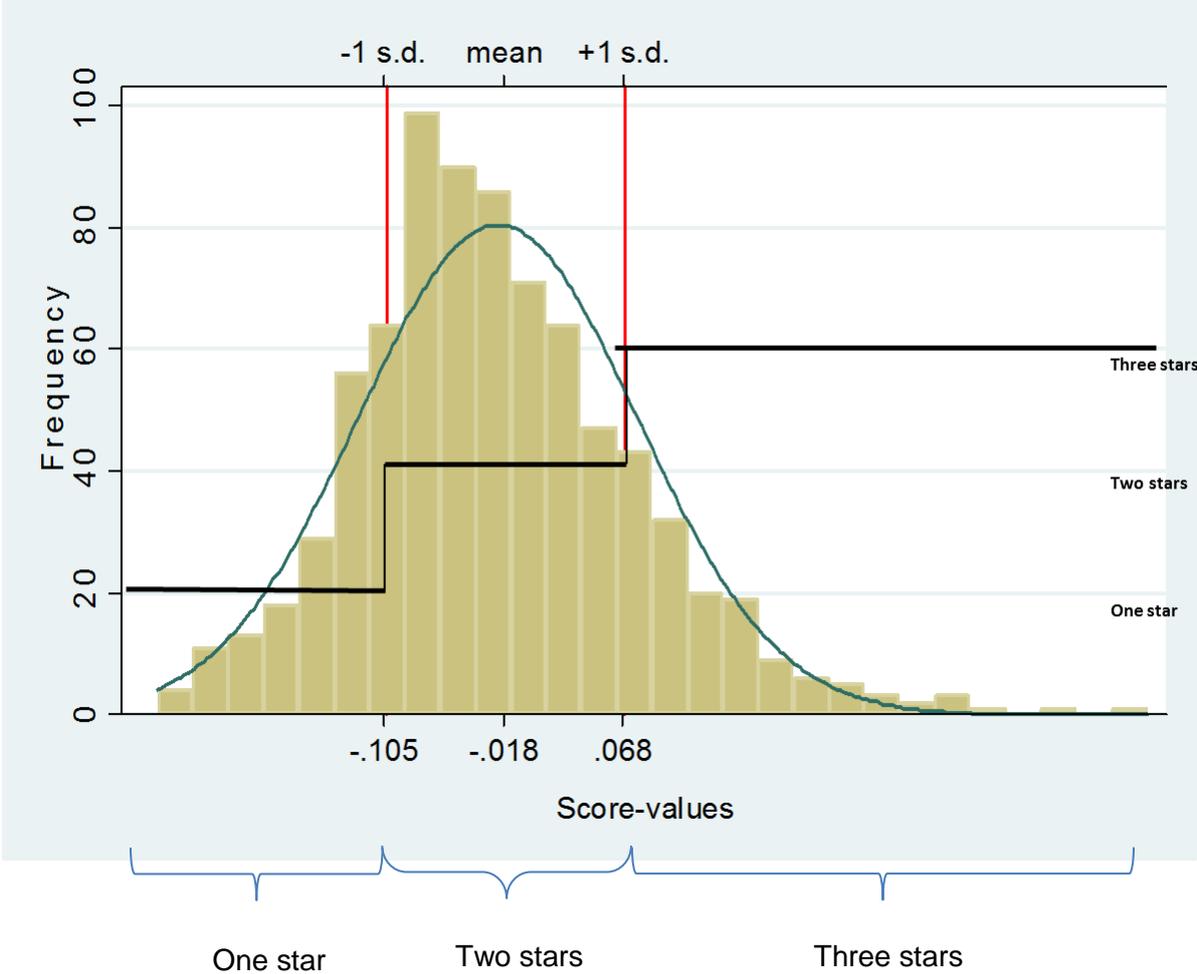
Each score-value is related to all other providers' score-values. If a supplier's score-value deviates positively from the average of all providers' score-values, the provider will receive a rating value of three stars. Those close to the average get two stars, while those who are below average receive just one star. To increase the reliability of the calculation of the mean, values are weighted by the number of participants who were registered with a provider. Accordingly, those suppliers with larger number of participants gained more weight than providers with fewer participants in the calculation of the mean and the measured standard deviation. This implies that larger providers will more likely receive two stars as they more likely are located closer to the mean in the distribution, which by design generates two stars.

Defining threshold values for the rating

An important issue is how much a rating value needs to differ from the mean to result in a rating better or worse than average. It was decided that a threshold corresponding to one standard deviation larger or smaller than average should be applied. Figure A1 shows the spread of score values. The distribution follows approximately a normal distribution. The cut-offs, defined by one standard deviation larger or smaller than the mean value, determine which rating a supplier is attributed. When the rating was published in late April 2013 those

providers with score values lower than $-.105$ received one star, those with score values between $-.105$ and $+.068$ received two stars, and those who were above $+.068$ got three stars.

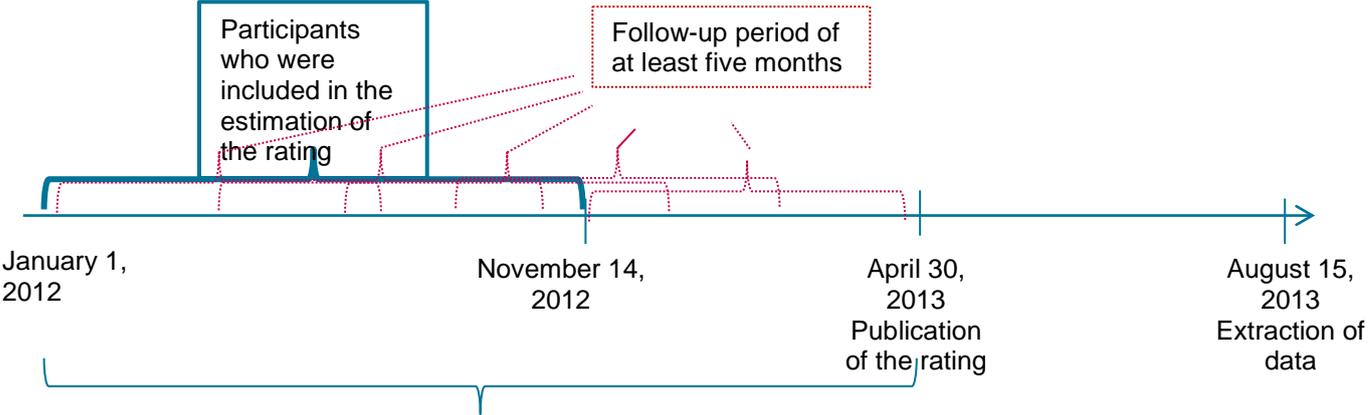
Figure A1



In defining transition to work there is a window of time that begins with the date of the recorded start for a participant using the services of a job coach and ends 120 days later (the time within the program plus an additional period of 30 days). If the registered participant within this interval switches to a registration code that marks that she has obtained work, it counts as 1 ("Successful match"), if not as 0 ("No match"). One restriction was that the registered work should have lasted for at least 30 days. To determine if there has been a successful match to job or not it is therefore necessary to apply a follow-up period of 150 days (120+30). This explains why no individuals that were registered with a job coach after the

November 14, 2012 were included in the calculation of the rating values (the data underlying the calculations of score values were extracted the April 14, 2013): The follow-up timeframe would not have been sufficiently long (i.e. at least five months). See figure A2 for a clarification on what time periods that form the basis for calculating the rating values.

Figure A2



Changes in the registration of results are noted and taken into consideration in the calculation of a job coach's score value

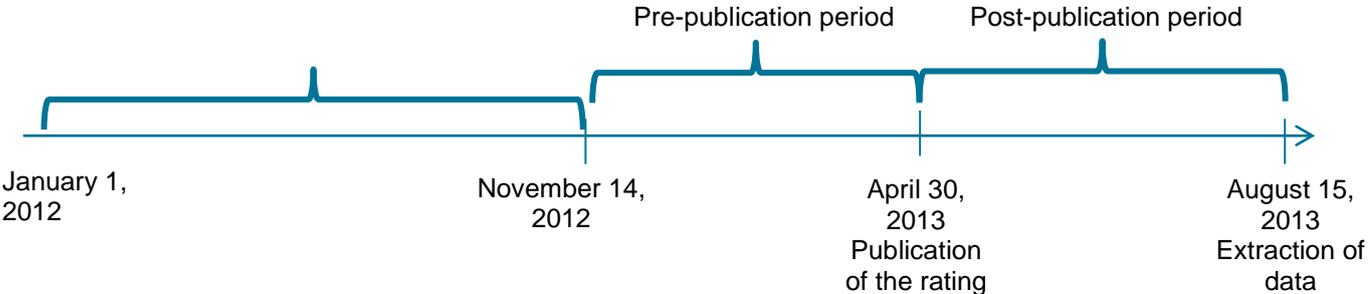
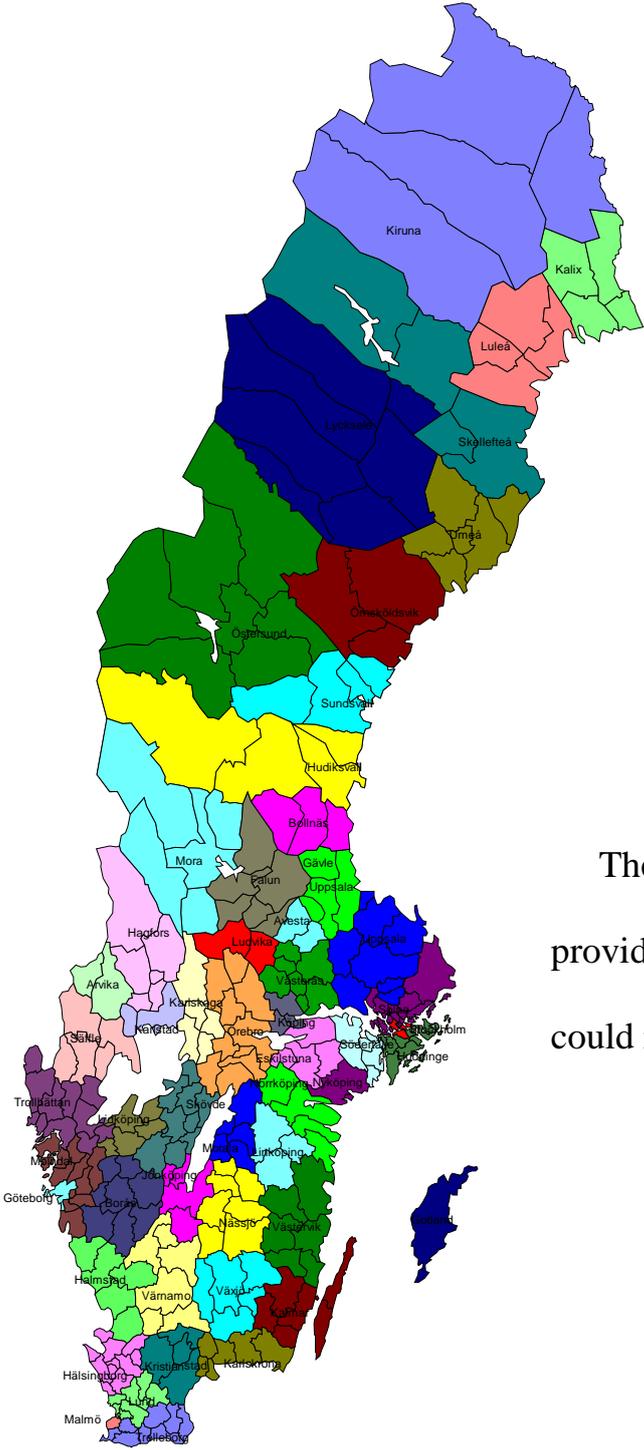


Figure A3



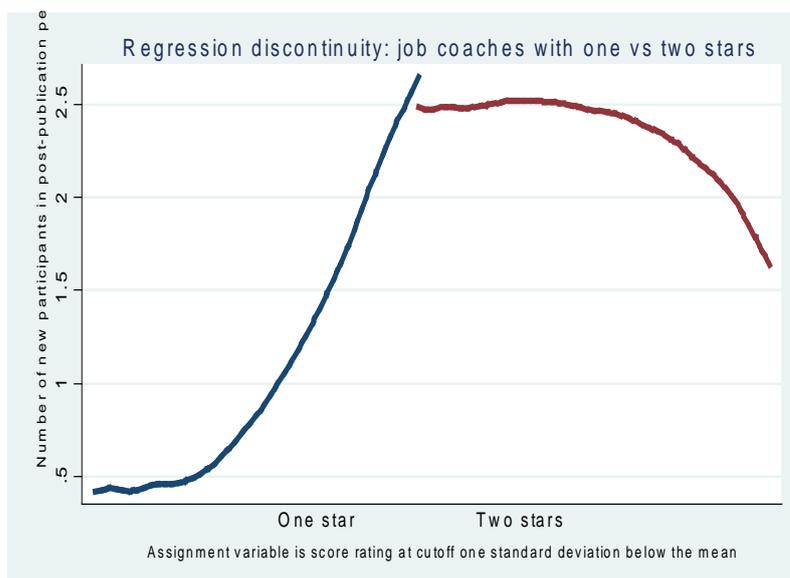
The supply areas where providers of job coaching could register

A.2 REGRESSION DISCONTINUITY

One star vs. two stars. Outcome: Number of new participants

Estimations on the data *after* publication of the rating values

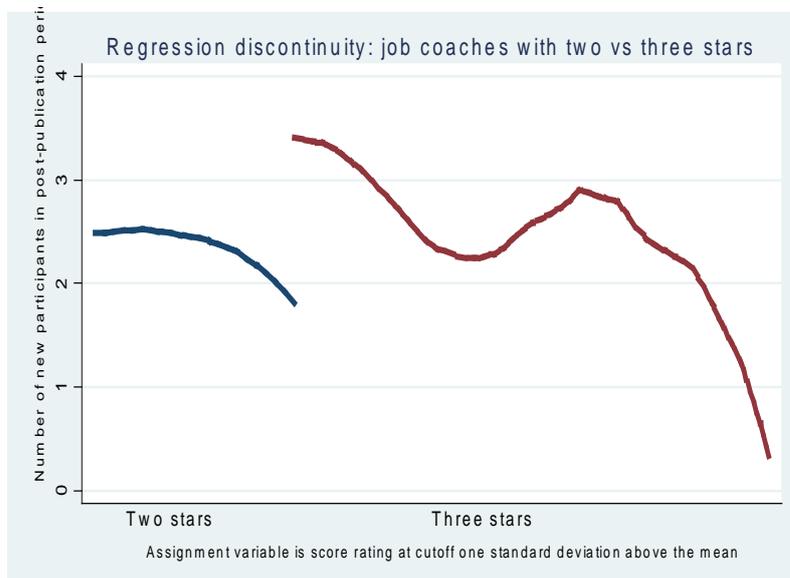
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.4655113	.8912675	0.52	0.601	-1.281341	2.212363
lwald50	.739329	1.108242	0.67	0.505	-1.432786	2.911444
lwald200	.532121	.6867992	0.77	0.438	-.8139807	1.878223



Three stars vs. two stars. Outcome: Number of new participants

Estimations on the data *after* publication of the rating values

	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	1.830836	.6013351	3.04	0.002	.6522406	3.009431
lwald50	1.591916	.8526606	1.87	0.062	-.0792684	3.2631
lwald200	1.534553	.5095715	3.01	0.003	.5358116	2.533295

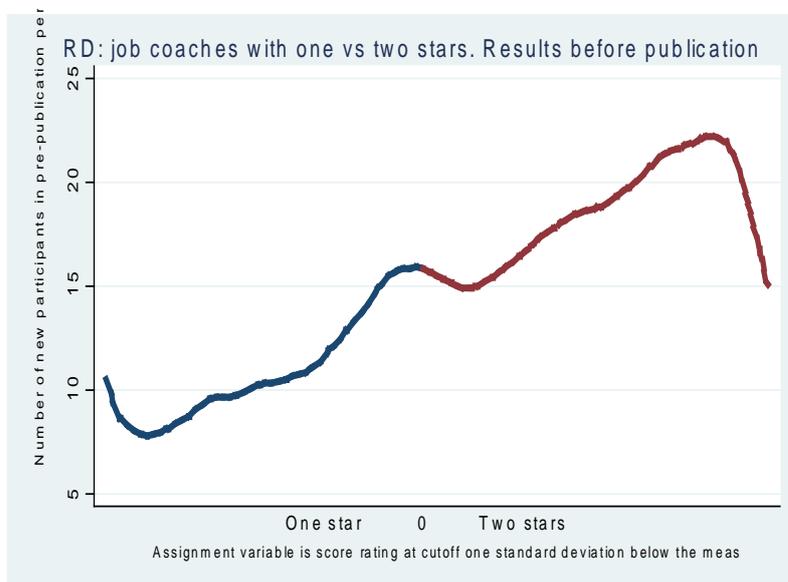


A.3 REGRESSION DISCONTINUITY: SENSITIVITY ANALYSIS

One star vs. two stars. Outcome: Number of new participants

Estimations on the data *before* publication of the rating values

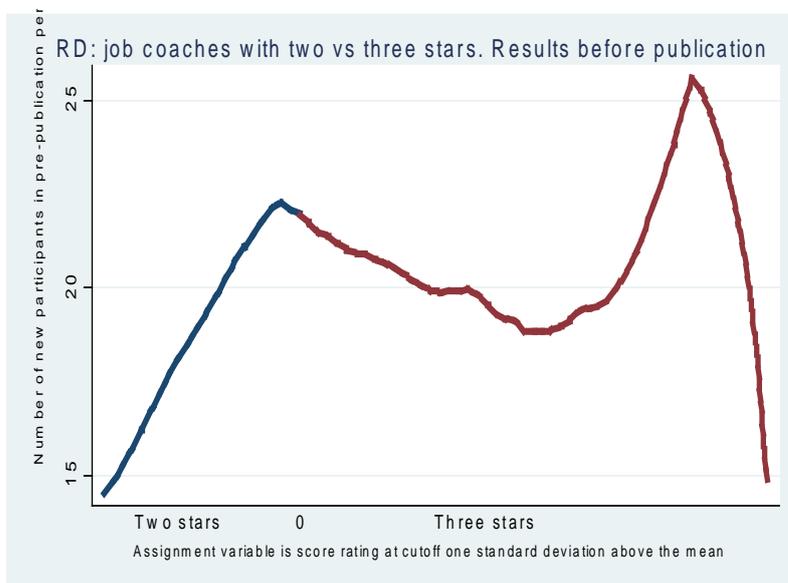
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	1.157894	3.35875	0.34	0.730	-5.425136	7.740923
lwald50	8.149067	6.608037	1.23	0.217	-4.802448	21.10058
lwald200	-1.278655	2.216747	-0.58	0.564	-5.6234	3.066089



Three stars vs. two stars. Outcome: Number of new participants

Estimations on the data *before* publication of the rating values

	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	-5.991643	4.358306	-1.37	0.169	-14.53377	2.55048
lwald50	-3.825027	5.513708	-0.69	0.488	-14.6317	6.981642
lwald200	-4.124864	3.287986	-1.25	0.210	-10.5692	2.31947



A.4 *The website for those to choose a job coach*

Jobbcoachning

Via län och kommun

Hitta närmaste

Gäller detta dig

Rating

Denna modul visas endast online

Denna modul visas endast online

Välj område



Så här fungerar rating av leverantörer

I Arbetsförmedlingens ratingsystem för leverantörer av jobbcoachning jämförs coachningsföretag med varandra och tilldelas en till tre stjärnor. Antalet stjärnor baseras på hur många av deltagarna som fick arbete inom fyra månader efter påbörjad jobbcoachning.

Människor har olika lätt att få ett arbete, ett faktum som systemet tar hänsyn till. Beroende på faktorer som exempelvis ålder, arbetslösheten i hemkommunen och arbetslivserfarenhet räknas det ut en statistisk sannolikhet för varje deltagare att få ett jobb. Sedan jämförs personens sannolikhet med hur det gick i verkligheten, om personen fick jobb.

Till sist slås alla värden för deltagarna hos ett coachningsföretag ihop. Företaget jämförs sedan med andra jobbcoachföretag i hela landet och ges ett betyg som motsvarar:

- * under medel
- ** medel
- *** över medel

Ratingvärdena är senast uppdaterade i september 2013.

Tänk på att...

Det finns fler företag att välja på än de som listas under fliken rating. Vi publicerar bara ratingsiffror för företag som har haft fler än tolv deltagare mellan januari 2012 till februari 2013. Företag som är nya eller har haft få deltagare i jobbcoachning går att hitta under fliken "Via län och kommun".

 [Mer om hur rating fungerar](#) 

Jobbcoachning

Via län och kommun

Hitta närmaste

Gäller detta dig

Rating

Denna modul visas endast online

Denna modul visas endast online

Välj område

Arvika

Arvika, Eda

* under medel
** medel
*** över medel

Företag med rating

Företag	Rating	KA-nummer
Adecco Sweden AB	*	10010021
Miroi I-Learning AB	**	10005623
Studentconsulting Sweden AB (Publ)	***	10012334

För att få ett ratingvärde behöver ett företag ha haft minst tolv deltagare. De företag som finns med i listan har haft tolv deltagare eller fler. De företag som har haft färre deltagare finns under "Via Län och kommun" eller "Hitta närmaste".

 [Så här fungerar rating](#) 