STOCKHOLM SCHOOL OF ECONOMICS

Department of Economics

5350 Master's Thesis in Economics

Academic year 2015–2016

The Adult Education Initiative and Cognitive Skills: Long-Term Evidence

Arvid Hedlund (22538) & Hannes Skugghall (22540)

Abstract

We investigate the relationship between participation in the Adult Education Initiative (AEI) and cognitive skills. Using propensity score matching, we find that AEI participants do not perform better than non-participants. This provides new evidence explaining the absence of positive general earnings and employment effects in the existing literature. The results are in particular driven by significantly lower performance among male participants. Female participants do not perform significantly better or worse, relative to non-participating females. This sheds light on the gender gap in terms of labor market outcomes found in the recent literature. It also has important implications for the understanding of the heterogeneity of the human capital production function in adult education and for future policy design.

Keywords: Skills, Adult Education, PIAAC

JEL Classification: I26, I28, H52, H75

Supervisors: Erik Mellander & Kelly Ragan

Date submitted: May 16th, 2016

Date examined: May 25th, 2016

Discussants: David Rydén & Rianne den Toom

Examiner: Johanna Wallenius

Acknowledgments

We would like to thank our supervisors Erik Mellander and Kelly Ragan for their invaluable help throughout the writing of this thesis. Furthermore, we are grateful for the tireless assistance of Patrik Lind. We would also like to thank Olof Lundborg and Per Sonnerby for insightful comments and encouragement at the beginning of the process. Financial support from the Institute for Evaluation of Labor Market and Education Policy (IFAU) is gratefully acknowledged.

Contents

List of Tables

List of Figures

1 Introduction

As economies evolve, the role of skills in determining economic success and labor market outcomes becomes more important. Simpler tasks become automated and the remaining, more complex tasks, put higher demands on the cognitive skills of the labor force. This paper, for the first time in the literature, investigates the effect of participation in Sweden's largest adult education program, the *Adult Education Initiative*[1] (AEI), on cognitive skills. The new OECD survey *Programme for the International Assessment of Adult Competencies* (PIAAC) allows us to quantify this relationship.

Increased expenditure on education is not inherently a productive investment. Rather, as underscored by *Hanushek and Wößmann* [2008], adequate evaluation of educational programs require two things to be ascertained: i) if and how such investments translate into skills and ii) investigating how those skills relate to economic outcomes. The previous literature on adult education, and the AEI in particular, fails to do so. It usually takes i) as a fact, and focuses on investigating ii). The reason for this is that, until the introduction of PIAAC, no adequate way of investigating i) has existed. This may lead to suboptimal recommendations, as without knowing the channels through which an educational program operates, we cannot draw as informed conclusions on its merits or shortcomings. If AEI participants do not perform better in PIAAC than their matched counterparts, perhaps the program composition was wrong to start with. Until now, no such evaluation has been made of the AEI.

The most common approach to increasing skills[2] is through education and there exists a large body of evidence on the positive effects from education on labor market outcomes, health and economic growth. In theory, two main channels can explain the benefits from education in the labor market: signalling and skills. Formal, quantitative measures of education, such as diplomas and degrees that signal the completion of educational programs, are easily examined. Such qualifications may through signalling be rewarded in the labor market. Meanwhile, measuring skills has been more complicated. This is problematic

[1] The Swedish name of the AEI is Kunskapslyftet.

[2] Skills and cognitive skills are sometimes used interchangeably throughout this paper, if we refer to a markedly different type of skill, such as non-cognitive skills, this is made explicit.

because the actual skill level an employee brings to a company ought to be the main concern, although concealed in a principal-agent relationship characterized by asymmetric information.

General training programs are less frequently implemented than specific training programs. The effectiveness of training programs, and those of more general character in particular, is still an open question in the literature. Recently, economists have argued that educational programs of a more general character can be particularly effective during periods of economic downturns, when unemployment is high and opportunity costs are low.[3] As a response to the economic downturn in the beginning of the 1990s, the Swedish government introduced a program of such character, the AEI, a 42 billion SEK policy initiative. The program primarily targeted low-skilled unemployed adults, and was aimed at raising these individuals to the medium-skilled level, focusing on enhancing general skills such as Swedish, English, mathematics and computer science.[4]

This paper evaluates the effect on cognitive skills from participation in the AEI. Specifically, our research question is formulated as follows: *Do AEI participants exhibit higher skill-levels than non-participants in the long run?* In order to answer our question we study individuals that participated in both the AEI and PIAAC. The time lag between the AEI, which ran from 1997 to 2002, and PIAAC, conducted in 2011 and 2012, allows us to capture long-term effects that adult education hopes to cultivate. We use both propensity score matching and ordinary least squares to estimate the effect.

Our data combination allows us to go beyond the previous evaluation method, which focuses on wage and employment effects. Such approaches are a good starting point, but inherently incomplete. Particularly so in the Swedish setting, where both wage dispersion and the returns to education are low. Previous studies have been required to take the channel running from educational investment to skills for granted. This is a strong assumption, that we now have the possibility to clarify. Consequently, the results of this paper provide additional information for future policy recommendations. Sweden face challenges with

[3] See *Heckman and Urzua* [2008] and *Pissarides* [2011].
[4] For skill-level definitions, see section 2.1.

respect to the integration of low-skilled immigrants. New initiatives, similar to the AEI, aimed at decreasing skill inequalities, are proposed in the policy discussion. A smaller scale, updated version of the AEI, is set to be introduced in the near-term. This paper has implications for whether and how this and other programs of its ilk could be beneficial. On a more general level, we contribute to the literature on skills versus wages. Specifically, how skills are an essential but neglected part of policy evaluations, complementing income and employment analyses, and growing in importance as the labor market evolves through technological progress.

We find that AEI participants did not perform better in PIAAC relative to the matched group. This pins down an important channel that may explain the absence of effects on employment and wages from the program. Interestingly, the lower overall PIAAC scores are driven by significantly lower results for the male AEI participants. Meanwhile, there are no significant differences between participating and non-participating females. These results are in line with the previous literature on the AEI, which, as we will demonstrate in section 3.3, generally does not find significantly improved overall labor market outcomes. However, recent papers with a longer follow-up period, find evidence of better outcomes for female AEI participants. Such results can now be put into the context of our findings regarding the skill discrepancy between men and women. We also find differences in relative performance based on educational qualifications before entering the program. Those with relatively higher qualifications fared as well as their matched comparison group, whereas those with low educational qualifications did not do as well as their matched comparison group. As this was one of the main target groups of the AEI, it is an important insight in the evaluation of the AEI.

2 Background

2.1 The Adult Education Initiative

In the beginning of the 1990s, Sweden experienced its deepest economic recession in the post-war period. Real GDP per capita decreased by 6.4 % between 1990 and 1993. During the same period, unemployment increased from 1.7 % to 8.2 %, the highest recorded rate since the Great Depression in the 1930s. Amongst the government's policies to fight unemployment was the AEI, a fully funded five-year program of investment and development in adult education. The AEI ran between July 1997 and December 2002 and is by far the largest adult education program undertaken in Sweden. Politically, the main objectives of the program were:

- A renewal of both labor market policy and education policy
- Reduced wage inequality
- Increased economic growth

The implementation of the AEI was delegated to the municipal authorities which applied for funding based on the estimated number of participants, the relative share of unemployment and the historic size of their adult secondary education system (Komvux), where the vast majority of all AEI courses were carried out. More than 100 000 seats were made available annually, and it was followed in 2002 by a permanent expansion of the overall number of seats available at Komvux. Thus, as noted by *Albrecht et al.* [2005], the AEI may be viewed as a major quantitative upscaling of the adult education system.

The focus of the AEI was to enhance general skills, such as Swedish, English, mathematics and computer science. Courses typically lasted for one semester, running from August to December in the fall and from January to June in the spring. Of the total participants during 1997-2002, a slight majority were above 30 years of age, about two thirds were female and a fifth were born outside of Sweden. Furthermore, two thirds were low-skilled and more than a quarter of the participants were medium-skilled. Both politically and in previous research on the AEI, low-skilled is defined as having an educational attainment below the level of a three-year upper secondary school degree. Medium-skilled is defined as having attained this level, but not any levels beyond that. Throughout the

AEI, the courses most frequently studied were, in falling order, computer science, mathematics, Swedish, English and business economics. Table 1 presents a detailed overview of the program.

The AEI primarily targeted low-skilled unemployed adults and secondarily low-skilled employed adults.[5] The latter were often working part-time or full-time alongside the program. Others were accepted into the program subject to availability. The program aimed at raising low-skilled workers to the medium-skilled level, thereby strengthening their position on the labor market. With the previously mentioned definition, the number of low-skilled workers in the total population aged 25-54 fell from around 2 million to 1.6 million between 1998 and 2002. However, defining skills in terms of the low- and medium-skilled definitions above, which are merely quantitative measures of years of schooling, is insufficient. This paper explores a more precise measurement instrument, PIAAC, in order to investigate whether the AEI increased long-term skill levels.

A special grant for education and training (UBS[6]), roughly corresponding to existing unemployment benefits, was tied to the AEI. Importantly, and similar to the existing body of research on the AEI (see e.g. *Axelsson and Westerlund* [2005], *Stenberg* [2005], *Stenberg and Westerlund* [2015]), this paper uses the UBS to identify AEI participants. Section 4 provides a further discussion on the UBS and also shows key data on UBS recipients. Other sources of financing participation in the program included partial financing (SVUX), regular student funding and other sources, including students who were entirely self-financed. With SVUX, students borrowed an amount corresponding to 35 % of the unemployment benefits, and were granted the remaining 65 %. With regular funding, students received 30 % and borrowed 70 % of the total amount.

[5] See section 9.2 in the appendix for official guidelines regarding target groups.

[6] *Särskilt utbildningsbidrag.*

Table 1: The Adult Education Initiative

Variable	F97[7]	S98	F98	S99	F99	S00	F00	S01	F01	S02	F02	**Average**
No. of full-time seats [8]	113475	142529	144630	156815	129693	141730	124430	132300	116599	123164	96539	**129264**
Avg no. courses per student	4.5	4.9	4.2	4.3	4.0	3.9	3.9	4.1	3.7	3.8	3.4	**4.1**
Characteristics (%)												
Male	32.8	33.4	32.0	32.5	31.9	33.3	32.0	33.0	32.4	33.8	32.0	**32.6**
Female	67.2	66.6	68.0	67.5	68.1	66.7	68.0	67.0	67.6	66.2	68.0	**67.4**
< 30 years old	51.0	45.0	44.0	40.0	42.0	40.0	42.0	41.3	42.5	39.4	44.3	**42.9**
Foreign born	18.0	18.0	18.5	19.7	20.3	19.0	19.5	19.0	20.9	23.0	23.1	**19.9**
PES[9] eligible					52.0	52.0	48.0	46.0	43.0	42.0	40.0	**46.1**
Education (%)												
≤ 2 years secondary school	63.0	65.0	70.0	68.0	69.0	64.0	67.0	58.0	60.4	52.0	60.0	**63.3**
> 2 years secondary school	26.6	24.9	21.2	24.0	23.0	26.0	23.0	31.0	28.0	33.0	28.0	**26.2**
< 3 years post secondary school	7.3	7.1	6.3	6.0	6.0	7.0	7.0	7.0	6.0	7.0	7.0	**6.7**
> 3 years post secondary school	2.7	2.5	2.4	2.0	3.0	3.0	3.0	3.0	4.0	4.0	4.0	**4.0**
Course details (%)[10]												
Administration	4.4	4.5	4.4	4.3	3.7	3.8	3.7	3.7	2.8	3.0	2.5	**3.7**
Computer science	14.7	17.1	16.4	14.9	12.1	11.8	11.4	13.2	10.2	9.9	13.9	**13.2**
English	8.8	8.2	7.4	6.9	6.6	6.0	5.7	6.6	5.6	5.4	5.5	**6.6**
Business economics	5.8	5.3	5.6	5.3	5.8	5.4	5.8	5.5	6.1	5.1	4.4	**5.5**
Physics	2.0	1.6	1.5	1.3	1.5	1.2	1.2	1.3	1.1	0.9	1.1	**1.3**
History	2.0	2.1	2.0	2.1	1.9	1.9	1.6	1.9	1.6	1.8	1.8	**1.9**
Chemistry	1.6	1.9	1.2	1.4	1.2	1.3	1.0	1.2	1.0	0.8	0.9	**1.2**
Mathematics	12.1	11.7	10.2	9.9	9.5	9.3	8.3	9.6	8.1	8.2	8.6	**9.6**
Human science	2.4	2.7	3.2	3.6	4.4	4.8	5.1	2.9	2.5	0.6	0.1	**2.9**
Nature studies	3.3	3.0	2.7	2.5	2.5	2.2	2.1	2.6	3.5	1.8	1.9	**2.6**
Health care	1.9	1.8	2.3	2.4	3.1	3.1	3.5	3.9	4.2	7.4	9.5	**3.9**
Psychology	1.8	1.9	1.9	2.1	2.2	2.2	2.2	2.2	2.5	2.7	2.8	**2.2**
Religious studies	1.8	1.9	1.5	1.7	1.4	1.5	1.2	1.5	1.2	1.5	1.3	**1.5**
Social science	5.2	5.0	4.2	4.1	3.6	3.5	3.0	3.8	2.9	3.1	3.0	**3.8**
Social welfare studies	2.0	2.2	2.6	2.9	3.5	3.7	3.9	3.6	4.7	5.2	5.7	**3.6**
Swedish	10.0	8.7	7.9	7.5	7.4	6.6	6.4	7.0	5.3	5.0	4.9	**7.0**
German	1.5	1.2	1.0	0.9	0.8	0.7	0.6	0.8	0.5	0.5	0.5	**0.8**
Other languages	1.7	1.4	0.9	1.3	1.5	1.4	1.5	1.7	2.3	2.5	2.7	**1.7**
Orientational courses	1.9	1.8	2.2	2.7	3.8	3.8	4.4	4.0	6.2	6.3	7.1	**4.0**
Folk high school courses	0.2	0.1	0.2	0.2	0.2	0.2	0.1	0.2	0.1	0.1	0.2	**0.2**
Other courses	15.0	16.1	20.6	22.0	23.7	25.5	27.0	22.8	27.8	28.2	21.8	**22.8**
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: This data is based on eleven separate municipality reports. The data collection was done by Statistics Sweden on behalf of Skolverket, that was assigned by the government to conduct six-month follow-ups of the development. The authors have compiled, distilled and organized these reports in order to create the table above.

[7] F stands for fall, S stands for spring, the numbers refer to the year.

[8] The scope of municipal activity in the program is measured in activity points. In adult upper secondary school education this is measured based on the upper secondary school points that they are worth. For orientation courses and folk high school courses 21 activity points per week are counted for full-time students. In the case of part-time studies the activity points are adjusted in the corresponding way. The number of full-time seats are calculated by dividing the activity points with 378, which is the number of activity points that correspond to six months of full-time studies.

[9] Public Employment Service - *Arbetsförmedlingen.*

[10] The data on courses taken is for the full sample of adult education students, and not specific for AEI participants. However, if different, AEI participants are more likely to have taken a further share of courses in mathematics, Swedish and computer science.

2.2 Education in Sweden

Swedish children attend school during nine years between the ages of seven and sixteen. Compulsory school qualifies the students for upper secondary school, which until 1994 was either a two- or a three-year program. In the two-year programs, students could choose between fifteen, mainly vocational, educations, many with strongly skewed gender distributions among participants, such as construction, electronics and nursing. The three-year programs offered theoretical studies in fields such as social science, human science, natural science or technical studies, with the intention of preparing and qualifying the students for higher education. Following the reform in 1994, all programs were required to be three years long and provide the students with the necessary qualifications for higher education. For all intents and purposes of this paper the pre-1994 system is the relevant one.[11]

In this context, it is important to highlight that compared to continental Europe, the difference between the educational content in the theoretical and vocational programs in upper secondary school is, and has been, relatively small in Sweden. Thus, some individuals enroll in adult education intending to change the direction of their studies, for example in order to become eligible for a specific university program or degree. Other participants include dropouts from compulsory or upper secondary school. Moreover, adult education targets both those already in the labor force, who seek to take courses for additional training in their profession, and those aiming to change their occupation.

Swedish municipalities have since 1969 been obliged by law to provide education for adults wishing to re-enroll at lower or upper secondary level. Anyone aged 20 or above is eligible, although individuals holding the shortest education usually have priority for all courses. Adult secondary education is part of the Swedish public sector school system and offers courses at three different levels: basic adult education, upper secondary schooling and post secondary schooling. Basic adult education aims at providing a qualification equivalent to the nine-year compulsory school, focusing on the core subjects (Swedish, English,

[11]Anyone graduating from the reformed system, earliest in 1997 and turning 20 years old, would not have been eligible for the UBS grant during the AEI, because one of the formal eligibility requirements were that an individual must be above 25 years old.

mathematics and social studies). The second level is similar but not identical to the regular upper secondary education; courses may differ somewhat in terms of content and scope.

In addition to the adult secondary education system, there are other forms of adult education, most notably labor market training (LMT), which is generally oriented towards teaching specific skills and towards a given profession. Some 87 000[12] individuals enrolled in a range of LMT programs in 1998 (*SOU [2007:18]*).[13] Historically, male dominated programs such as construction, computer science, manufacturing, machine operator and transport, have had significantly higher enrollment rates than female dominated programs, leading to a larger share of men in LMT programs in general. Importantly, the male dominated programs were also the types of programs that to a higher extent led to employment after completion, compared to the female dominated programs such as health care and services, which to a lower extent led to employment (*Jans et al.* [2014]). This has important implications for the choice of training program individuals faced, general (the AEI) or specific (LMT), and we will return to this in our results discussion in section 7.1.

[12]This figure declined steadily to 45 000 in 2000, 28 000 in 2002 and 17 000 in 2004.

[13]In Government reports, the different programs are usually categorized as follows: Business and social sciences, computer science, construction, culture and media, customer services, health and care services, machine operator, manufacturing, mechanics and natural sciences, office and warehousing, pedagogy, restaurant, transport, other occupations.

3 Previous Research

The effect of education on economic outcomes is one of the perennial questions of economics. The dominant approach to this question can be traced back to *Mincer* [1970, 1974], and attempts to estimate how different amounts of schooling affects individual earnings. Economists have continued investigating this relationship, all concluding that more schooling is associated with higher individual earnings.[14] However, and as we will return to in section 3.3, this effect has often been found to be small or negligible for the AEI.

Education may affect economic growth through at least three important mechanisms. First, education increases the human capital inherent in the labor force, which in turn increases labor productivity and leads the economy to a higher level of output (*Mankiw et al.* [1992]). Second, the new knowledge this generates promotes growth (*Lucas* [1988], *Romer* [1990]). Third, education may facilitate diffusion and transmission of knowledge needed to understand and process new information in order to successfully implement technologies developed by others (*Nelson and Phelps* [1966], *Benhabib and Spiegel* [2005]). The positive relationship between quantitative measures of schooling and economic growth is well established and robust (*Sianesi and Reenen* [2003], *Doppelhofer et al.* [2004]). However, studies based on quantitative measures of schooling neglect the qualitative differences in the ensuing knowledge (*Hanushek and Wößmann* [2008]).

The remainder of the literature review is organized as follows. In section 3.1 we focus on the role of skills, and specifically outline how and why skills are a necessary yet often overlooked dimension in the evaluation of training programs. Section 3.2 looks closer at the relationship between skills and wages, while section 3.3 is focused on adult education, and findings related to the AEI in particular. Finally, we specify our research question in section 3.4.

[14] See e.g. *Psacharopoulos* [1994], *Card* [1999], *Harmon et al.* [2003] and *Heckman et al.* [2006a].

3.1 The Role of Skills

The traditional focus on quantitative measures of educational attainment in the academic literature is in stark contrast to the policy discussion that, even in the poorest areas, involves elements of "quality" of schooling (*Hanushek and Wößmann* [2008]).[15] There is a lively policy discussion in most countries about the quality of schooling, the role of teachers and students, and how to provide the best possible education. Test scores in evaluations such as the Programme for International Student Assessment (PISA) are scrutinized and declining results lead to calls for national investigations and action plans. Some of the usual policy questions related to school quality are teacher salaries, class sizes, help tools such as computer tablets and funding. These debates rest on the presumption that there is a high rate of return to schooling and to quality in particular (*Hanushek and Wößmann* [2008]). But not all educational investments can be presumed worthwhile. Two recent papers investigate whether increased school spending leads to improvements, and although often true, it is not always the case, and the type of investment matters (*Lafortune et al.* [2016], *Jackson et al.* [2016]). As laid out in the introduction, two things are essential to investigate: how various investments translate into skills, and how those skills translate to economic returns. Up until now, most investigations of adult education, including those on the AEI, have focused on the latter. This approach, combining data on AEI participation with PIAAC scores, enables us to also examine the former. This is essential as better general skills makes the population more adaptable and flexible to labor market adjustments, and is a potential argument for implementing general adult education programs such as the AEI.

The literature on education has recently begun to emphasize that focus has been too narrowly constrained to quantitative measures of education (*Hanushek and Wößmann* [2008]). Formal attainment and enrollment rate measures are frequently used, likely due to the fact that they are more easily available, compared with data on skills. As is the case in many fields of economics, much research comes from the United States, where the rate of return to schooling is around 10 % per additional year of education. In Sweden, estimates are usually lower,

[15]See *Hanushek* [2011], *Chetty et al.* [2014a], *Chetty et al.* [2014b] for studies examining the economic impact of higher quality teachers - the academic interest in "quality" of schooling has increased in recent years.

closer to 5 % (*Isacsson* [1999], *Harmon et al.* [2003], *Meghir and Palme* [2005]). This discrepancy is important to keep in mind, as it suggests that when evaluating educational policies from a wage perspective it may be more difficult to distinguish differences in the Swedish setting, compared to other countries. The estimates above are on the private returns to schooling. However, there are also social returns to consider, and these are generally deemed to be in excess of the private returns (*Hanushek and Wößmann* [2008]). Social returns to schooling stem from positive effects on, for example, crime (*Lochner and Moretti* [2004], *Machin et al.* [2011], *Groot and Van Den Brink* [2010]), measures of health (*Currie and Moretti* [2003], *Cutler et al.* [2006]), financial market participation (*Cole and Shastry* [2009]) and citizen participation (*Dee* [2004], *Milligan et al.* [2004]).

Quantitative measures of education are important determinants when employers make hiring decisions. However, the longer time a person spends at a job, the more important and apparent becomes the actual skills of the employee. An increasing number of companies are also moving beyond traditional measures of evaluation such as degrees and grade points averages, by also integrating personality and skill evaluations.[16]

The majority of studies evaluating the impact of skills on labor market outcomes focus on outcomes relatively early in workers' careers, primarily due to data availability issues. Evidence on the effects over the entire working life is more scant, but *Altonji and Pierret* [2001] find that the impact of skills on earnings grows with experience, relative to formal measures of education. This is consistent with the intuitive hypothesis that employers are able to better judge the actual skills of employees the longer they interact with them. It should however be noted that the evidence on how returns to cognitive skills varies throughout life is not conclusive. A more recent study by *Hanushek and Zhang* [2009] shows that the above pattern does indeed hold for the United States, but not for a wider set of countries.

As is the case with quantitative measures of education, there are substantial differences between countries in terms of labor market returns to skills. Sweden,

[16]Two recent articles discussing these developments are "'Big Four' look beyond academics", Financial Times, January 28th, 2016, and "Today's Personality Tests Raise the Bar for Job Seekers", Wall Street Journal, April 14th, 2015.

along with countries such as Norway and the Czech Republic, exhibit some of the lowest returns to skills. On the other end of the scale are countries such as the United States, Germany and Ireland (*Hanushek et al.* [2015]). Also when it comes to the returns to skills as measured through the IALS[17], Sweden exhibits low rates of return, while the United States has the highest return (*Hanushek and Zhang* [2009]). There is an interesting and burgeoning literature on the role of non-cognitive skills (*Bowles and Gintis* [1976], *Bowles et al.* [2001], *Heckman et al.* [2006b]). The results show that in addition to cognitive skills, non-cognitive skills play an important role in determining labor market outcomes. Disentangling the effects of cognitive and non-cognitive skills and the channels through which they operate is important when thinking about new policies and laws, but outside the scope of this paper.

Cognitive skills can also account for differences in growth across countries in the OECD (*Hanushek and Wößmann* [2011]). Empirical growth research demonstrates that consideration of cognitive skills alters the assessment of the role of education and knowledge in the process of economic development dramatically. *Hanushek and Kimko* [2000] use international test data in order to build a measure of labor force quality and find a statistically and economically significant positive effect of cognitive skills on economic growth between 1960 and 1990. The effect is stronger than the association between quantity of education and growth. Similar results have been found in *Lee and Lee* [1995] and *Barro and Lee* [2001]. *Coulombe et al.* [2004] and *Coloumbe and Tremblay* [2006] use data from IALS, and find that test score measures outperform quantitative measures of schooling across 14 OECD countries. *Jamison et al.* [2007] focus on the mathematics component of IALS and add a larger set of countries than previous studies. Their findings suggest that cognitive skills increase income levels primarily through speeding up technological progress, rather than shifting the level of the production function or increasing the impact of additional years of schooling.

[17]The International Adult Literacy Survey, see section 4.2 for a detailed description of the IALS.

In conclusion, the empirical evidence suggests that student performance as measured in tests is a more important factor in explaining economic growth than quantitative measures of schooling. This holds for both developed and developing countries, controlling for factors such as institutions, openness, fertility and geography. Furthermore, and of particular relevance to this paper, it suggests that educational policies being evaluated on their effect on the participants' cognitive skills is of the utmost importance, and affects national growth prospects. Earlier we discussed the role that better education has on other measures of well-being, such as crime, health and civic participation. Similar relationships can be found when looking at skill levels in the form of PIAAC data. In all the countries in the survey, individuals who score in the lower levels of proficiency in literacy are more likely to report poor health, believe that they have little impact on the political process, and less likely to participate in associative or volunteer activities. In the majority of the countries, they are also more likely to exhibit lower levels of trust in others (*OECD* [2013]). Numeracy skills are also associated with better financial outcomes, even after controlling for differences in education, risk preferences, beliefs about future income, financial knowledge and other factors that might influence wealth (*Estrada-Mejia et al.* [2016], *Cole et al.* [Forthcoming]).

3.2 The Relationship Between Skills and Wages

Research has shown that good performance on standardized tests leads to higher individual earnings.[18] These studies show that measured achievement has a clear impact on earnings also after allowing for differences in schooling, experience, and other factors related to earnings. Three studies conducted in the United States show direct and rather consistent estimates of the impact of test performance on earnings (*Mulligan* [1999], *Murnane et al.* [2000], *Lazear* [2003]). They utilize nationally representative longitudinal data sets and find that a one standard deviation increase in performance in mathematics leads to approximately 12 % higher annual earnings.

Income inequality has been increasing in many OECD countries over the last years, also in Sweden. It is increasingly becoming a pressing matter for policy

[18] See e.g. *Bishop* [1989], *O'Neill* [1990], *Blackburn and Neumark* [1993], *Murnane et al.* [1995] and *Murnane et al.* [2001].

makers, and attempts to halt the development are high on the agenda. Skills play an important role in determining the distribution of income. The supply and demand for different types of skills within countries impact the wage-setting dynamics. Thus, it also explains trends of income inequality, although the extent of the impact is largely unexplored, mainly due to a lack of suitable data prior to the introduction of PIAAC. With PIAAC comes better knowledge about both skill levels and its distribution, which enables research on the relation between technological change and domestic wage structures (*Rinawi and Backes-Gellner* [2015]).

Juhn et al. [1993] illustrate that skills have had an increasingly important impact on the distribution of income over time. Income differences have also been shown to be increasing within schooling groups in the United States, indicating that skills drive income inequality, controlling for education (*Levy and Murnane* [1992]). Technological change decreases the demand for certain types of skills, while increasing the demand for other (*Katz et al.* [2006], *Autor and Dorn* [2013]). This might increase wage inequality and low-skilled individuals, such as those targeted by the AEI, are generally those most at risk from these developments. Therefore, understanding how these programs translate into skills is essential in order to better understand the dynamics of inequality in society, and how it can be dealt with.

There is no constant relationship between wage inequality and skills, neither when looking at cross-country comparisons, nor when examining specific countries. As demonstrated in *OECD* [2016a], there are countries where both skills and wage inequality are high (United States), where both are low (Czech Republic, Slovak Republic), where the former is low and the latter is high (France) and the other way around (Korea). The relationship is complex and labor market structures and policies are likely to impact the channels through which skill inequality translates into wage inequality. Arguably, this could be of even greater importance in the Swedish setting, where wage dispersion is low and the bargaining power of unions is strong. As mentioned previously, countries differ in their tendency to reward skills in the labor market - another fundamental factor behind why skill inequality is difficult to map into wage inequality. In PIAAC, skills are rewarded the highest in the United States, England and Germany, and the lowest in Sweden (*OECD* [2016a]). On average, the return to skills is marginally convex, implying that wages increase faster at higher levels. All in all, this points towards expecting modest, if any, positive effects on wages from a program such as the AEI in Sweden.

3.3 Adult Education and the AEI: Empirical Evidence

A great deal of research evaluates the effects of education on labor market outcomes. However, only a small fraction concentrates on adult education. Findings from regular education may rarely be generalized to adult education, due to two main reasons. First, adult education is more flexible regarding study pace and at what age individuals start and end their studies. Second, participants are generally more experienced and therefore, arguably, basing their studying choices on superior information. This complicates the selection mechanism when studying adult education, and little is known about how it has influenced existing results (*Stenberg and Westerlund* [2008]).

In general, the education programs that in the literature are thought to provide the best return are early childhood interventions (see *Heckman and Carneiro* [2003], *Burgess* [2016] and *Cullen et al.* [2013]). This view is now being nuanced, and high-dosage tutoring of adolescents seems to be even more effective than early childhood investments. This argues against the view that there is a point beyond which investments in education are unlikely to yield significant returns.[19] *Card et al.* [2010] conduct a meta-analysis of active labor market policy evaluations. The authors report and categorize the impacts on earnings and employment from 97 studies conducted between 1997 and 2007. Notably, a third of the studies stem from Germany, Austria or Switzerland and a quarter stem from the Nordic countries. A significant share of the analyzed studies look at the impact of training programs, including programs of more general character. Such programs appear ineffective in the short-term, but are often associated with positive impacts after 3 years or more. Interestingly, screening all studies, no differences between men and women are found with respect to the impact of labor market programs.

[19]See *Fryer Jr.* [Forthcoming] that does a meta-study of 196 randomized field experiments designed to better understand the human capital production function.

In analyzing the effects of adult secondary education on unemployment and earnings using Danish data from 1981 to 1991, *Holm et al.* [1995] find a positive effect on both accounts in comparison to a control group of non-participants. The effect is especially large for those who had been long-term unemployed before participation. One of the arguments in favor of adult education is that it may stimulate lifelong learning, create a group of individuals more able to adapt to a changing labor market and develop corresponding new skills. These arguments gain further support by *Jenkins et al.* [2003], that study the key determinants to whether someone undertakes lifelong learning[20] in the United Kingdom. They model the effect of lifelong learning on wages and the likelihood of being employed and find little evidence of positive wage effects but significant positive effects on employment. Moreover, undertaking one episode of lifelong learning is found to increase the probability of undertaking more lifelong learning.

Jacobson et al. [2005] evaluate the impact from an academic year of community college training. They find participation to be associated with a long-term earnings increase of some 10 % looking at quantitative and technically orientated courses. For other courses, the impact is estimated at 3-5 %. *Alm Stenflo* [2000] use a sample of Swedish adult education graduates from 1992 and 1993, and tracks the labor market outcomes of the group until 1997. Each person in the sample is then matched against three randomly chosen individuals with similar background and characteristics. In 1993, the year following graduation, there were 16 % more unemployed amongst the adult education graduates. This difference gradually decreased and in 1997 it had been reduced to 0.4 %.

Beyond its low wage dispersion and strong union power, Sweden is a case of specific interest for labor market policy evaluation. Partly due to political tradition, and partly responding to the sharp increase in unemployment during the early 1990s, its focus on active labor market policies is arguably unrivaled (*Calmfors et al.* [2002]). The set-up of a program like the AEI could moreover easily be recreated[21], making it especially interesting to look at from a pol-

[20]The authors define lifelong learning as learning between the ages of 33 and 42 that results in a qualification.

[21]One such example is the Facila Program in Portugal, modeled after the AEI with similar intention and structure.

icy perspective. The majority of the educational programs that governments initiate are labor market oriented. Previous empirical evaluations of the AEI have usually been done by comparing labor market outcomes from participation with a reference group of individuals in LMT. In part, the two programs targeted similar individuals and both programs made its participants eligible for a compensation level corresponding to unemployment benefits. Thus, such comparisons are important in terms of policy evaluation.

Stenberg [2002] compares earnings in 1999 between individuals participating in the AEI or LMT during the fall of 1997. As in this paper, he uses the UBS grant to identify AEI participants and a one-to-one propensity score matching approach. In total, the wage effect is found to be relatively lower for AEI participants, although more beneficial for females and in Stockholm county, compared with males and the sparsely inhabited inland of Norrland, respectively. Using the same data, *Stenberg* [2003] extends the study to include data on earnings for 1999 and 2000 and the wage effects are again found to be relatively lower for the AEI. However, the results suggest a time-lag on the improvement of earnings from participation in the AEI relative to LMT. This is argued to stem from the greater degree of targeting specific occupations in LMT. Regarding job mobility, the findings suggest, perhaps surprisingly, that the AEI participants had a relatively lower probability of changing industry. Continuing, *Stenberg* [2005] compares AEI with LMT using unemployment incidence and unemployment duration as outcome variables, both measured immediately upon completion of the programs. The paper finds that the AEI lowered the incidence of unemployment but increased the duration. These results regarding unemployment incidence following the two programs are in line with *Axelsson and Westerlund* [2001], but contrasts those regarding duration from *Westerlund* [2000].

In *Albrecht et al.* [2005], the development of earnings and the probability of employment is investigated using data for the period between 1991 and 2000. Participants are required to have completed a minimum of one and a maximum of two sustained Komvux semesters between 1997 and 1998. Importantly, although restricting their data to low-skilled individuals between the ages of 25-55, *Albrecht et al.* [2005] do not identify participants by the UBS grant. Instead, anyone attending adult education at Komvux, and thus not only AEI participants, are examined. Therefore, the sample used would likely contain a lower share of unemployed and low-skilled individuals, than the UBS sample. Effects are estimated by comparing the change in earnings and employment with a control group consisting of persons chosen from a sample of 200 000 Nordic individuals between the ages of 25 to 55. They report no significant effects on wages, although male AEI participants increase their probability of being employed. For females no significant effects are found with regards to employment and income.

Importantly, both *Jacobson et al.* [2003] and *Stenberg and Westerlund* [2015] argue that a follow up period of around 10 years is required in order to give a fair estimate of effects from general education for adults. Furthermore, one intention with the AEI was to enable its participants to change the direction of their careers and educations. Consequently, the short follow up period of the program in the papers covered so far tells little about the medium or long-term outcomes, which the program is perhaps more likely to result in. *Stenberg and Westerlund* [2015] therefore make an important contribution in comparing the effects of LMT and AEI on wages with a follow-up period of thirteen years. The paper finds that LMT is indeed initially relatively more beneficial. Interestingly though, the difference converges after 5 to 7 years. Beyond that time horizon, in the long-term, the AEI is found to be associated with relatively higher earnings for both females and for participants working in Stockholm. Finding it difficult to balance the samples using one-to-one matching, *Stenberg and Westerlund* [2015] uses four-to-one matching. However, while lowering the expected variance of the treatment effect estimates, increasing the prospect of significant results, adding additional matches to each participant also increases the risk of greater bias. This is because four-to-one matching increases the risk of making poor matches (*Morgan and Winship* [2007]).[22]

A key source of concern in evaluating programs such as the AEI, and of similar ilk, is that there are unobservable characteristics that differ between those that selected into the program, and those that they are compared to. These differences could stem from, for example, IQ and motivation. This is an issue that we will discuss further in the empirical strategy and method section. Although we

[22]This choice will be discussed in more detail in the empirical strategy, section 5.2.2.

do not have access to IQ data on participants before they entered the program, other studies do. *Stenberg and Westerlund* [2015] have access to data on cognitive test scores from the mandatory military enlistment for males born in 1953 or later. Importantly, this information alleviates some of the concerns surrounding unobservable differences between the two groups. *Stenberg and Westerlund* [2015] compare participants in the AEI with participants in LMT and find no statistically significant difference in the cognitive abilities of the two groups.[23]

Previous AEI evaluations have due to lack of adequate data sources not investigated whether, and, if so, how, participation in the program actually translated into cognitive skills. Subsequently, nor has the literature enhanced our understanding of how such (potential) skill increases are related to economic outcomes. Instead, previous AEI evaluations have been constrained to estimating the effect of participation on wage development and employment incidence, either assuming a perfect transition between wages and skills, or ignoring the skill dimension. Although not to be interpreted as a causal relationship, *OECD* [2016a] finds a strong positive correlation between participation in adult education and skills proficiency in PIAAC. However, it is also argued that for individuals whose skills are already at a very low level, adult education and training is unlikely to boost their skills.

3.4 Research Question

Existing evidence on the return to wages for adult education programs indicate that the returns are lower than for regular education. This holds true also when it comes to the AEI, as *Stenberg* [2002, 2003, 2005] find no discernible general effects on earnings and employment from participation in the AEI. *Albrecht et al.* [2005] compare labor market outcomes between Komvux participants and a matched group of individuals with similar characteristics, and find no discernible effects on earnings, neither for women, nor for men. But as recent research indicates (*Stenberg and Westerlund* [2015]), it may take time for effects from general education programs to surface. Therefore, it is especially interesting to conduct studies with a long-term perspective, which the PIAAC

[23] These measurements are based on test scores of inductive, verbal, technical, and spatial skills. The tests are completed at the age of 18 or 19 and measured on a scale of 1 to 9, where 9 is the highest score. The scores are available for a sub sample of 97 027 males born 1953 or later, and the p-value of the difference between the two groups is 0.530, thus insignificant.

data allows. *Stenberg and Westerlund* [2015] furthermore find that the positive effects from AEI participation were driven primarily by women, and that they were larger for individuals in Stockholm. Such heterogeneous treatment effects are important in policy design and motivates subgroup analyses. So do the recent results from *OECD* [2016a], which indicate that the outcomes of skill development programs in adult education vary strongly based on educational qualifications prior to enrollment. Both these subgroup analyses, on gender and education prior to enrollment, will be performed in this paper.

Much of the literature on education has, due to simplicity and availability, focused on quantitative measures of schooling. There are two important factors to consider when evaluating educational investments: how the investments translate into skills, and how those skills translate to economic returns. We fill a gap in the literature by evaluating the AEI on the skill dimension for the first time, investigating whether participants perform better when measuring cognitive skills. Improving the skills of the participants was one of the goals of the program. But previous evaluations of the AEI have focused on labor market outcomes, being unable to examine the channel through which these educational programs operate. Combining PIAAC and AEI data, we can evaluate whether AEI participants perform better on PIAAC compared to similar non-participants. Our research question is formulated as follows:

Do AEI participants exhibit higher skill-levels
than non-participants in the long run?

4 Data

4.1 The UBS Grant

Using Swedish registry data on UBS recipients, the special grant for education and training, this paper identifies individuals in the Swedish PIAAC sample that also participated in the AEI. The data is carefully inspected and handled by Statistics Sweden, ensuring high quality and reliability. The UBS grant furthermore distinguishes participants in the AEI from students potentially studying the same courses in the same classrooms, but within the regular adult education system Komvux. This is the main reason why the identification method has been dominant in previous research on the AEI.[24] Applicants for the UBS had to be between 25 and 55 years of age and be entitled to unemployment insurance.

Some characteristics regarding the recipients of the UBS are worth noting. First, as shown in Table 2, females financed their studies with UBS to a somewhat greater extent than males. In part, this explains why our data on participants contain 122 women and 48 men. Second, throughout the program an average of 85 % of UBS recipients had maximum two years of upper secondary schooling. Finally, relatively fewer participants from larger cities, including Stockholm, financed their studies with UBS. Notably, Swedish-born individuals financed their studies with UBS to a greater extent than foreign-born individuals. This is due to the UBS requirement of being eligible for unemployment insurance. A further discussion on the UBS grant and the implications of using it as an identification strategy can be found in section 7.2.

[24] Some notable examples are: *Stenberg* [2002, 2003, 2005]; *Stenberg and Westerlund* [2015], and *Westerlund* [2000].

Table 2: Characteristics of UBS Recipients

Variable	F97	S98	F98	S99	F99	S00	F00	S01	F01	S02	F02	**Average**
No. of seats[25]	113475	142529	144630	156815	129693	141730	124430	132300	116599	123164	96539	**129264**
UBS[26] (% of seats)	27.3	33.3	30.3	27.8	17.1	15.6	13.5	12.7	11.4	10.0	9.7	**19**
UBS (males, % of seats)	22.7	27.7	23.5	22.5	14.1	13.1	11.5	10.2	10.4	8.0	7.4	**16**
UBS (females, % of seats)	29.6	36.2	33.5	31.1	18.5	16.9	14.5	14.0	15.0	11.2	10.8	**21**
Upper secondary ≤2 years			87.0	88.0	88.0	87.0	86.0	83.0	83.0	81.0	81.0	**85**
Education[27] (%)												
Compulsory <9 years			2	2	3	3	3	3	2	2	2	**2**
Compulsory 9-10 years			19	23	25	27	26	20	21	21	22	**23**
Upper secondary ≤2 years			66	63	60	57	57	61	60	58	57	**60**
Upper secondary >2 years			8	7	8	9	10	13	13	14	14	**11**
Higher education <3 years			4	3	3	3	3	3	3	3	3	**3**
Higher education ≥3 years			1	1	1	1	1	1	1	1	1	**1**
UBS share in area (%)												
City > 200 000	19	27	23	22	14	13	12	11	10	7	8	**15**
Suburban municipalities	29	27	28	27	14	13	11	11	11	9	7	**17**
City 50-200 000	24	31	28	27	16	14	12	12	12	10	8	**18**
City < 50 000	33	38	34	32	19	18	15	14	15	12	9	**22**
Industrial municipalities	33	35	32	38	18	17	14	13	14	11	11	**21**
Rural municipalities	36	41	38	35	21	19	15	16	17	12	11	**24**
Back country municipalities	44	43	41	35	24	22	20	18	21	13	13	**27**
Other larger municipalities	37	43	38	34	21	18	16	15	17	12	12	**24**
Other smaller municipalities	35	40	37	34	21	19	16	14	16	11	11	**23**

Source: This data is based on half-year municipal reports. The data collection and analysis was done by Statistics Sweden on behalf of Skolverket that was tasked by the government with conducting six-month follow-ups of the development. The authors have compiled, distilled and organized these eleven reports in order to create the table.

[25] The number of seats refers to full-time seats, including both AEI and Komvux participants.

[26] For definition, see section 4.1 above.

[27] Data for education levels were not reported in the first two half-year municipal reports.

4.2 PIAAC

The Survey of Adult Skills is an international survey conducted in 33 countries[28] as part of the Programme for International Assessment of Adult Competencies (PIAAC). It is conducted by the OECD and designed to measure the key cognitive and workplace skills needed for individuals to participate in modern societies and for economies to prosper. The first data collection was done between August 2011 and March 2012. Around 166 000 adults, representing 724 million adults, were surveyed in 24 countries. In Sweden, the total number of participants was 4469 and the survey was conducted by the government agency Statistics Sweden that carried out home visits to the participants. The initial selection was determined through a random stratified selection of 10 000 people from a register data base covering the entire Swedish population called *Registret över totalbefolkningen*.[29] For an outline of the response rates of the original RTB sample, broken down by subgroups, see section 9.9 in the appendix.

The original requirement for countries to participate in the survey was that at least 5000 individuals were surveyed. However, due to the substantial efforts taken by Statistics Sweden to ascertain representativeness of the collected sample, Sweden was allowed to remain in the study. Formally, the objectives[30] of the survey are to:

- provide policymakers in each participating country with a baseline profile of adults in their country in terms of the knowledge, skills and competencies that are thought to underlie both personal and societal success;
- assess the impact of these competencies on a variety of social and economic outcomes at the individual and aggregate levels;
- gauge the performance of education and training systems in generating the required competencies; and
- help clarify some of the policy levers that could contribute to enhancing competencies.

[28] The participating countries in round 1 (2011-2012) are: Australia, Austria, Belgium (Flanders), Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, Poland, Russian Federation, Slovak Republic, Spain, Sweden, United Kingdom (England and Northern Ireland) and the United States. The countries participating in round 2 (2014-2015) are: Chile, Greece, Indonesia, Israel, Lithuania, New Zealand, Singapore, Slovenia and Turkey.

[29] See *Larsson and Eriksson* [2013].

[30] See *OECD* [2013].

The idea behind the survey is to provide the possibility for countries to evaluate and better understand how education and training systems can nurture these skills. In each of the participating countries, a representative sample of adults between the ages of 16 and 65 were interviewed in their homes. The mean proficiency scores for the countries participating in PIAAC are presented in Table 17, located in section 9.9 in the appendix. Furthermore, for a detailed outline of the study design using PIAAC, see Figure 4 in the appendix section 9.9.

Before the introduction of PIAAC, the primary source (*Hanushek et al.* [2015]) for international comparisons of the returns to cognitive skills was the International Adult Literacy Survey (IALS).[31] It was conducted in Sweden in 1994, before the introduction of the AEI, but the overlap between IALS and PIAAC participants is not available for this paper. PIAAC builds on the previous international skills surveys that have been conducted, which in addition to IALS also includes the Adult Literacy and Lifeskills Survey (ALL) conducted in 2003 and 2006.[32] Sweden was, however, not one of the participating countries in the ALL, and it is thus not applicable to our research question. PIAAC is designed to assess the current state of the skills of individuals and nations in the information age. The survey has several advantages over IALS. It is a more recent study, and the returns to skills from over two decades ago is likely to be less informative as economies have undergone significant technological change (*Autor et al.* [2003], *Goldin and Katz* [2009] and *Acemoglu and Autor* [2011]). PIAAC furthermore benefits from more participating countries and larger sample sizes. Also, whereas IALS included tests that examined very basic skill competencies, PIAAC substantially extends the depth and range of measured skills and also attempts to measure problem solving skills in technology rich environments. One example of this is that the PIAAC design broadens the definition of literacy to make it more relevant for the information age.

The questions in the survey were designed to measure cognitive skills in three domains: numeracy, literacy and problem solving in technology rich environments. The PIAAC questions aim to mimic real-world problems such as maintaining a driver's logbook (numeracy) or reserving a meeting room on a particular time

[31]See *Darcovich et al.* [2000] and *Kirsch* [2001] for more information on the IALS.
[32]See *Satherley et al.* [2008] for more information on the ALL.

slot in an online booking system (problem solving). More formally, the three cognitive skill dimensions are defined as:

- **Literacy:** ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential;
- **Numeracy:** ability to access, use, interpret, and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life;
- **Problem solving in technology rich environments:** ability to use digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks.

The literacy component in PIAAC excludes writing or producing text. Still, literacy is defined in a broader way than reading. The intention is for the component to encompass the range of cognitive strategies adults must use to react suitably to texts read in different scenarios or contexts. Moreover, a unique feature is the testing of the ability to read digital texts as well as traditional print-based texts. Regarding numeracy, literacy skills are acknowledged to be an enabling factor for numerate behavior. Basically, when numerate questions involve text, the respondents performance is also to some extent dependent on his or hers literacy skills. But the PIAAC numeracy component is designed to cover more than using arithmetical skills on information embedded in text. Specifically, it relates to a wide range of skills and knowledge, where answers involve more than numbers and questions involve more than just numbers in text.

PIAAC also covers the specific class of problems people deal with when using information and computer technology (ICT). These problems share the following characteristics:

- The existence of the problem is primarily a consequence of the availability of new technologies.
- The solution to the problem requires the use of computer-based artifacts (applications, representational formats, computational procedures).

- The problems are related to the handling and maintenance of technology rich environments themselves (e.g., how to operate a computer, how to fix a settings problem, how to use the Internet browser in a technical sense.)

4.3 PIAAC Methodology

PIAAC is based on a complex survey design and statistical methods designed to ensure high reliability and comparability of the data. This thus means that there are several considerations that must be taken into account when analyzing the data (*Pokropek and Jakubowski* [2013]). Two important adjustments must be made due to the complex sample designs used in the survey. First, all point estimates must be computed using sample weights. Second, it is necessary to use special procedures to calculate correct standard errors. The Jackknife replicate procedure is used in the analysis of PIAAC data (see *Efron* [1982], *Levy and Lemeshow* [2013] and section 9.8 in the appendix).

PIAAC, like most modern large-scale surveys, uses plausible values. Plausible values were first developed for analyses of the 1983-1984 US National Assessment of Educational Progress data and is based on the work of *Rubin* [1987].[33] The method for calculating standard errors varies depending on whether cognitive components (and plausible values method) are included in the specific estimation or not. If cognitive components are not part of the analysis the standard errors are more straightforward and based on computations that summarize the variability of the estimates in subsequent subsets of samples called replicates, calculated as follows:

$$SE_\theta = \sqrt{f \sum_{r=1}^{R} \left(\hat{\theta_r} - \hat{\theta_0} \right)}$$

where R is the number of replicates, $\hat{\theta_r}$ represents the statistic of interest (i.e. mean, variance, regression, coefficient) that does *not* involve plausible values for replicate $r = (1, ..., R)$; $\hat{\theta_0}$ represents the statistic of interest estimated using the whole sample and final sample weight. The constant f differs depending on the country analyzed, as it depends on the sampling procedure used in the survey. There were two types of sampling used in PIAAC: non-stratified and stratified. If the sampling procedure used was non-stratified, $f = \frac{R-1}{R}$, otherwise $f = 1$. In

[33] Other important contributors include *Mislevy* [1991], *Mislevy et al.* [1992] and *Beaton and Gonzalez* [1995].

Sweden, stratified sampling was used, thus $f = 1$ for the purpose of this paper.

Calculation of standard errors becomes more complicated when cognitive components are included in the analysis. The standard errors involving plausible values are calculated as follows:

$$SE_{\varepsilon} = \sqrt{\left[\sum_{p=1}^{P} \left(f \sum_{r=1}^{R} (\hat{\varepsilon}_{r,p} - \overline{\varepsilon}_{0,p})^2 \right) \frac{1}{P} \right] + \left[\left(1 + \frac{1}{P} \frac{\sum_{p=1}^{P} (\hat{\varepsilon}_{0,p} - \overline{\varepsilon}_{0,P})^2}{P-1} \right) \right]}$$

where

$$\overline{\varepsilon}_{0,P} = \frac{\sum_{p=1}^{P} \varepsilon_{0,p}}{P}$$

and P is the number of plausible values, $p = (1, ..., P)$; $\hat{\varepsilon}_{r,p}$ represents the statistical estimate for replication r and the p^{th} plausible value; $\hat{\varepsilon}_{0,p}$ represents the statistical estimate using the final sample weight for the p^{th} plausible value; $\overline{\varepsilon}_{0,P}$ represents the unweighted average of the statistic over the plausible values using the whole sample and the final weight. For a more detailed description of plausible values, see section 9.12 in the appendix.

4.4 Data Processing and Final Data

To reach a data set of maximum relevance for the research question we do several important adjustments to the overall sample of 4469 PIAAC participants. This includes creating new variables and dropping observations that fail to meet certain criteria. First of all, because this study evaluates the effects from the AEI, it would make little economic sense to match participants with individuals who, potentially, might have attended the same courses in the same classrooms, but who were not part of the AEI. As we are interested in isolating the effect of the AEI, such a comparison would be meaningless. Thus, individuals participating in Komvux, but not in the AEI, are dropped. This corresponds to 516 observations.

Second, when using propensity score matching it is essential to use prior knowledge about the characteristics of the program in order to generate a relevant sample group (*Angrist and Pischke* [2008]). Towards this end, we use the knowledge about formal requirements for entry into the AEI (see the appendix section 9.2) in order to drop 592 observations that, before the launch of the AEI in 1997, had a formal educational qualification corresponding to 2 to 3 years of studying at university level or more.[34] Our final data set thus consists of 3361 individuals. Of these, 170 individuals participated in the AEI. Summary statistics for these two groups can be found in the appendix section 9.3. Furthermore, we construct a range of variables providing information on the characteristics of the individuals before they entered the program. These include categorical variables for educational attainment and whether the individuals had any children living at home at the time of the program.

5 Empirical Strategy and Method

This study deals with non-experimental economic data. Observing participation, but absent random assignment, the key concern is finding participants and non-participants that are similar in their characteristics. If such concerns are not properly alleviated, estimates are unlikely to be informative. In this paper, concerns of that nature could be motivated, as self-selection into the AEI might stem from baseline differences with respect to motivational, educational and cognitive skill levels between the two groups. To mitigate these concerns, we control for a range of variables and use empirical methods in order to make the two groups as similar as possible.

This paper estimates the effect on skills from participation in the AEI. Two strategies are used. First and foremost, propensity score matching (PSM) is implemented. Second, we use ordinary least squares (OLS). Both are data heavy and rely on the assumption that all relevant differences between participants and non-participants can be captured with the observable variables in the data. In the context of the research question, and considering its rich individual data of some 400 variables covering, for instance, educational history, professional experience and family background, PIAAC is well suited to both approaches.

[34] International standard classification of education 5B or higher.

In short, PSM creates an artificial control group of individuals (henceforth called matched individuals) based on their similarities to actual AEI participants for a chosen range of characteristics. This reduces the matching problem to a single dimension called the propensity score, which is defined as the probability of an individual receiving treatment given the observed characteristics. Matching on the propensity score then allows for estimating treatment effects by comparing the performance on PIAAC between matched individuals and the actual participants. With OLS, we instead control for the same covariates used for matching in the PSM. As noted by *Morgan and Winship* [2007], there has been a sizable debate as to which of these two strategies of confounder control ought to be favored, and in what context. While OLS is biased when the outcome regression model is misspecified, PSM is biased when the propensity model is misspecified.

With large differences in the covariate mean values between participants and non-participants, which is often the case in evaluations of labor market policies, OLS results are very sensitive to the assumption of linearity. In such situations, PSM is arguably the favoured strategy, as it does not require the imposing of a specific functional form. This is a valuable feature of PSM, as specific functional forms are rarely suggested by economic theory or justified by the data (*Dehejia and Wahba* [1999] and *Smith* [2000]). Partly for this reason, PSM has been a widely applied tool in evaluations of labor market policies (see e.g. *Dehejia and Wahba* [1999] and *Heckman et al.* [1997]). Furthermore, the covariate mean differences aspect is essential for our evaluation, because the two groups, as expected, to a large extent[35] differ at baseline (see Table 3 below).[36]

Implementing PSM will moreover inevitably highlight the issue of common support. With poor overlap between treated and non-treated, doubts could be raised regarding the robustness of traditional, parametric methods (*Bryson et al.* [2002]). For this reason, implementing PSM could be valuable on its own merits, simply because of the potential side benefit of enhanced understanding regarding the extent of the overlap and, thus, also of how sensitive any estimates would be to the choice of functional form, should one also use a linear regression model.

[35] In our covariate vector X_i of pre-treatment characteristics, 13 out of 15 covariates are statistically significantly different at the one percent level.

[36] See appendix section 9.10 for a more detailed description of the variables.

The advantages and disadvantages of the two approaches will be further discussed, both throughout this section and in the limitations section. Meanwhile, the remaining part of the section will unfold as follows. First, the foundations of PSM, as well as its key identifying assumptions, are presented and discussed. Second, the implementation of the PSM analysis is outlined and motivated according to the steps suggested in *Caliendo and Kopeinig* [2008]. Lastly, the OLS model is described.

Table 3: Covariate Comparison - Participants and Non-participants

	AEI		Non-AEI	
Variable	Mean	S.E.	Mean	S.E.
Gender***	1.718	0.035	1.459	0.009
Highest education - mother***	1.364	0.051	1.818	0.014
Highest education - father***	1.442	0.055	1.786	0.014
Parents qualification***	1.547	0.067	1.969	0.016
Education level***	0.941	0.081	0.810	0.018
Low education***	0.700	0.068	0.611	0.015
Immigrant	1.700	0.045	1.671	0.010
Children at home***	0.400	0.038	0.178	0.007
Age category 1***	0.094	0.022	0.564	0.009
Age category 2***	0.165	0.029	0.066	0.004
Age category 3***	0.171	0.029	0.067	0.004
Age category 4***	0.153	0.028	0.068	0.004
Age category 5***	0.129	0.026	0.065	0.004
Age category 6***	0.182	0.030	0.067	0.004
Age category 7	0.106	0.024	0.103	0.005

For covariate definitions, see section 9.10 in the appendix

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

5.1 Propensity Score Matching

The omitted-variable bias formula states that coefficients on included variables are unaffected by the omission of variables when the omitted variables are uncorrelated with those included. The propensity score theorem (*Rosenbaum and Rubin* [1983]) extends this idea to estimation strategies that rely on matching instead of regression, where the causal variable of interest is a categorical treatment indicator. The theorem states that if potential outcomes are independent of treatment status conditional on the covariate vector of pre-treatment characteristics X_i, then potential outcomes are also independent of treatment status conditional on a scalar function of covariates, the propensity score:

$$p(X_i) \equiv Pr\{D_i = 1 | X_i\} = E\{D_i | X_i\} \tag{1}$$

where, in this paper, $D = \{0, 1\}$ is the AEI participation indicator. Knowing the propensity score $p(X_i)$ then allows for estimating the average effect of treatment on the treated (ATT):

$$\begin{aligned} \mathcal{T} &\equiv E\{Y_{1i} - Y_{0i} | D_i = 1\} \\ &= E[E\{Y_{1i} - Y_{0i} | D_i = 1, p(X_i)\}] \\ &= E[E\{Y_{1i} | D_i = 1, p(X_i)\} - E\{Y_{0i} | D_i = 0, p(X_i)\} | D_i = 1] \end{aligned} \tag{2}$$

where the outer expectation is over the distribution of $(p(X_i) | D_i = 1)$ and Y_{1i} and Y_{0i} are the potential outcomes in the two counterfactual situations of participation and non-participation. Formally, the following two propositions are needed to derive (2) given (1)[37]:

Lemma 1. Balancing of pre-treatment variables given the propensity score. If $p(X_i)$ is the propensity score, then

$$D \perp X \mid p(X)$$

Lemma 2. Unconfoundedness given the propensity score. Suppose that assignment to treatment is unconfounded, i.e.

$$Y_i, \;\; Y_0 \perp D \mid X$$

[37]For proof, see section 9.4 in the appendix.

then assignment to treatment is unconfounded given the propensity score, i.e.

$$Y_1, \quad Y_0 \perp D \mid p(X)$$

If the balancing of Lemma 1 is satisfied, observations with the same propensity score must have the same distribution of observable characteristics independently of treatment status. In other words, for a given propensity score, exposure to treatment is random and therefore treated and matched individuals should on average be observationally identical (*Becker and Ichino* [2002]).

Unconfoundedness, or the conditional independence assumption (CIA), demands that there exists a set of observable covariates, such that after controlling for these, potential outcomes are independent of treatment status. Notably, the CIA is also a necessary requirement in linear regression models (*Caliendo and Kopeinig* [2008]). The assumption needs to be justified by the quality of the data and, generally, a very rich data set is required (*Caliendo and Kopeinig* [2008]). It is moreover emphasized by amongst others *Heckman et al.* [1999], that in order to credibly satisfy the CIA, data for both treated and non-treated should originate from the same source, in our case, from PIAAC.

Meanwhile, the common support assumption states that for each covariate, there is a positive probability of being both treated and non-treated. This assumption ensures the existence of sufficient overlap in the characteristics of participants and non-participants, so as to find adequate matches:

$$0 < P(D = 1 \mid X) < 1$$

When both the CIA and the common support assumption hold, a state called strong ignorability, PSM produces unbiased estimators of the treatment effect (*Rosenbaum and Rubin* [1983]).

5.2 Implementing the PSM Analysis

In practice, the PSM analysis is implemented in several steps. In the following subsections, we present, motivate and implement the different stages in PSM, following the framework suggested in *Caliendo and Kopeinig* [2008]: (1) estimating the propensity score, (2) choosing matching algorithm, (3) investigating overlap and (4) assessing the matching quality. These steps are mirrored in the structure of our results section, where we outline how our matching process worked, the extent of the overlap and the quality of our matching.

5.2.1 Estimating the Propensity Score

Estimating the propensity score $p(X_i)$ requires choosing what model of estimation to use and what variables to include in this model. This paper deals with a binary treatment case. In such situations there is a strong preference for logit or probit models (*Smith* [1997]). As noted by *Caliendo and Kopeinig* [2008] and *Zhao* [2008], the two specifications often yield very similar results. Therefore, the choice is seldom critical. However, the logit distribution has slightly more density mass in the probability bounds, which arguably better reflects our data. We experimented with both specifications and they did not significantly affect our conclusions. Thus, this paper estimates the propensity score using a logit model:

$$Pr\{D_i = 1|X_i\} = \phi(h(X_i))$$

where ϕ denotes the logistic cumulative distribution function and $h(X_i)$ includes the carefully constructed vector of covariates, X_i.

There is an extensive discussion in the literature as to which covariates to include. The key concept regarding inclusion is strong ignorability, which is demanded by most non-experimental evaluation approaches, matching included. Strong ignorability assumes that there are no observed differences between the treated and non-treated group (*Caliendo and Kopeinig* [2008]). To satisfy this assumption, variables known to be related to both participation and the outcome should be included. Or, more formally, included covariates should (*Rosenbaum and Rubin* [1983]; *Rubin and Thomas* [1996]; *Tomz et al.* [2003]):

i. Influence the decision of participation and the outcome variable

ii. Be unaffected by participation

iii. Be unaffected by the anticipation of participation

In the specification of this paper, 15 covariates are used. These are displayed in Table 4 below and as highlighted all included covariates meet the formal requirements for inclusion.[38] Excluded covariates either do not fulfill the formal requirements, contain extensive missing data or make little or no economic sense to include.

Table 4: Covariates Included in Vector X_i

Variable	**i.**	**ii.**	**iii.**
Gender	✓	✓	✓
Immigrant	✓	✓	✓
Parents qualification	✓	✓	✓
Highest education - mother	✓	✓	✓
Highest education - father	✓	✓	✓
Education level	✓	✓	✓
Low education	✓	✓	✓
Children at home	✓	✓	✓
Age category 1-7	✓	✓	✓

Included covariates are explained in detail in section 9.10 in the appendix.

5.2.2 Choosing a Matching Algorithm

There are a number of matching estimators and they differ in several ways. First, they define different neighborhoods for the treated individuals and differ with respect to how weights are assigned to these neighbors. Furthermore, they deal with the issue of common support differently. However, and as noted by *Smith* [2000], when sample sizes grow all estimators close in on comparing only exact matches, and should thus return very similar estimates. But the choice might be of greater importance with small sample sizes. In such cases, a trade-off between bias and variance generally arises and the choice then depends

[38] All covariates influence the decision of participation. Furthermore, as outlined in chapter 3 in *OECD* [2013], they are all strongly correlated with the performance on PIAAC. Hence, requirement i) is met. Lastly, because all included covariates are time-fixed, requirement ii) and iii) are met as well.

on the context (*Heckman et al.* [1997]). As described in the data section 4.5, $n = 3361$ observations remain in our sample after the necessary data adjustments, of which 170 participated in the AEI. Thus, in the context of the current paper, the choice of matching estimator is not unimportant.

In this paper, we use a $k = 1$ nearest neighbor (NN) matching method (see *Rubin* [1973]).[39] NN matching selects for each treated individual i the matched individual j with the smallest distance from individual i in terms of propensity score to form a matched pair (*Stuart* [2010]). Formally, and following the structure suggested by *Becker and Ichino* [2002], the approach unfolds as follows. Let T be the set of treated units and C the set of control units, respectively. Denote by C_i the set of control units matched to the treated unit i with an estimated value of the propensity score p_i. NN matching then sets:

$$C_i = \min j|p_i - p_j|$$

which is a singleton set unless there are multiple nearest neighbors. Denote the number of controls matched with observation $i \in T$ by N_i^C and define the weights $w_{ij} = \frac{1}{N_i^C}$ if $j \in C(i)$ and $w_{ij} = 0$ otherwise, and Y_i^T and Y_j^C be the observed outcomes of the treated and control units. Then, the formula for the matching estimator can be written as follows:

$$\mathcal{T} = \frac{1}{N^T}\sum_{i\in T}[Y_i^T - \sum_{j\in C(i)} w_{ij}Y_j^C] = \frac{1}{N^T}[\sum_{i\in T} Y_i^T - \sum_{i\in T}\sum_{j\in C(i)} w_{ij}Y_j^C]$$

and the estimation of average treatment effects based on the propensity score is finally given by:

$$= \frac{1}{N^T}\sum_{i\in T} Y_i^T - \frac{1}{N^T}\sum_{j\in C} w_j Y_j^C$$

where the weights w_j are defined by $w_j = \sum_i w_{ij}$. There are several approaches to NN matching. In essence, two choices has to be made and motivated, to match *with* or *without* replacement and how many untreated individuals j to assign to each treated unit i. This paper matches without replacement and sets $k = 1$.

[39] Where k is the number of untreated individuals matched to each treated individual.

First of all, matching with or without replacement is, as underlined by *Caliendo and Kopeinig* [2008], mainly an issue should the data demonstrate large differences in the propensity score distributions between treated and control individuals. This is, as we demonstrate in Figure 1 in section 6.1.1, not a concern for this study. But we still need to motivate our choice. Matching *with* replacement allows matched individuals j to be used as partner for more than one treated individual i, while matching *without* replacement does not. In general, the choice involves a trade-off between bias and variance. Allowing replacement will often decrease the bias and increase the variance. It is a particularly useful approach in the context of a small data set with few comparable observations to the treated individuals (*Stuart* [2010]). Because this study has almost 20 non-participants for each treated individual, this is not a priority. Instead, considering the rather small sample of AEI participants, we want to minimize the variance. From this perspective matching without replacement is preferable, should the matching also produce a well balanced set of groups, which will be evident in the results section.

Furthermore, inference is more complex when matching with replacement. Using this approach, matched individuals j are no longer independent. Rather, some will appear in the matched group more than once and this issue needs to be handled in outcome analysis. The most common way of doing so is by using frequency weights (*Stuart* [2010]). But due to the nature of PIAAC, we have certain constraints on the matching estimators we can use. This is due to the fact that PIAAC's complex survey design requires several essential weighting measures to be undertaken[40] and can not be combined with weighting measures that would be necessary to use, should we employ NN-matching with replacement. We are thus left with NN-matching without replacement as the matching method that is most suitable to our data set and analysis. A potential issue related to NN matching without replacement is, however, that estimates depend on the order in which observations get matched (*Caliendo and Kopeinig* [2008]). Hence, when using the approach, it should be ensured that ordering is randomly done, which is ascertained in our analysis.

[40]These measures were discussed more extensively in the PIAAC Methodology, section 4.3.

As to how many untreated individuals to match to each treated individual, the most common approach is $k = 1$ matching. In a nutshell, a larger k increases the size of the matched sample, which potentially enhances the precision of estimated treatment effects. However, it also results in matching increasingly dissimilar individuals, which in turn has been known to increase bias. This originates from the intuitive understanding that the second match will, in general, be of lower quality than the first. On this, *Imbens* [2004] writes: "within the class of matching estimators, using only a single match leads to the most credible inference with the least bias, at most sacrificing some precision." This view is confirmed by the extensive Monte Carlo simulations carried out by *Austin* [2010], in which the author examines the statistical performance of matched estimators with respect to the choice of k. The results from the 97 analyzed scenarios demonstrate that the bias in the estimated treatment effect is increasing in k. Furthermore, it showed that the mean square error of the estimated treatment effect was minimized in more than two thirds of the scenarios using $k = 1$ matching. In conclusion, *Austin* [2010] argues that in a majority of settings, using $k = 1$ or $k = 2$ will result in optimal estimation of treatment effects. Moreover, merely in a small minority of settings was using $k > 2$ optimal. Holding the above discussion in mind, this paper sets $k = 1$.

5.2.3 Overlap and Common Support

ATT is only defined in the region of common support. Hence, a vital step is to check the region of common support and the overlap between the treated and non-treated group. Several ways are suggested in the literature, the most straightforward one is to conduct a visual analysis of the density distribution of the propensity score in both groups. *Lechner* [2008] argues that given that the support problem can be spotted by inspecting the propensity score distribution, there is no need to implement a complicated formal estimator. Once one has defined the region of common support, individuals that fall outside this region have to be disregarded and for these individuals the treatment effect cannot be estimated. When the proportion of lost individuals is small, this poses few problems. However, if the number is too large, there may be concerns about whether the estimated effect on the remaining individuals can be viewed as representative (*Caliendo and Kopeinig* [2008]). In this study not a single individual in the treated or the matched group falls outside the region of common support.

Enforcement of common support can result in the loss of a sizable proportion of the treated population. One must bear this in mind when considering the policy relevance of results in general. This is because a policy analyst wishes to know the effect of a policy on all participants, not just a sub-sample for whom common support is enforceable. If treatment effects differ non-randomly with those unsupported characteristics, the treatment effect relevant to the supported sub-population will not provide a consistent estimate for the unsupported sub-population (*Bryson et al.* [2002]). Whether this is a problem in practice will depend upon the proportion of the treatment group lost, which in this study, as mentioned above, is zero.

In general, common support will be a problem when participants differ markedly from non-participants. For instance, if only those with a very high level of motivation are volunteering for the program. Very low take-up of a voluntary program may be an early signal of such a problem. For the purpose of this evaluation, this is not a major source of concern, as the AEI had a very high take-up rate. Furthermore, and as mentioned in section 3.3, based on the IQ test conducted at the mandatory military enlistment tests, there does not appear to be significant differences between AEI participants and other similar individuals. Importantly, the objective when using PSM is not necessarily to ensure that all matched pairs are similar in terms of their covariate values. Rather, the matched groups should on average be similar across all their covariate values. The adequacy of the model used to estimate the propensity score can thus be evaluated by looking at the balance that results on average across the matched groups (*Gelman and Hill* [2006]).

5.2.4 Assessing the Matching Quality

Two measurements designed to evaluate the quality of matching over covariates and the extent to which they balance are presented in *Rubin* [2001]. The first one is Rubin's B, the standardized difference of means of the linear index of the propensity score in the treated and non-treated group. Rubin's B should be below 0.25 for the regression adjustment to be trustworthy. The second one is Rubin's R, the ratio of treated to non-treated variances of the propensity score.[41] This variance ratio is recommended to be between 0.5 and 2.0 in order for the samples to be considered sufficiently balanced (*Rubin* [2001] and *Stuart* [2010]). When these two indicators are unsatisfactory, the propensity score model could be misspecified. Should the problems persist, even after including new or excluding old covariates, it may serve as an indication of failing the CIA. In such cases, the researcher should pursue a different evaluation strategy (*Smith and Todd* [2005]).

Another suitable indicator to assess the distance in marginal distributions of the X-variables is the standardized bias (SB) suggested by *Rosenbaum and Rubin* [1985]. For each covariate, X, it is defined as the difference of sample means in the treated and matched group as a percentage of the square root of the average of sample variances in both groups. The SB before matching is given by:

$$SB_{before} = 100 * \frac{(\bar{X}_T - \bar{X}_{\neq T})}{\sqrt{0,5 * (V_T(X_T) + V_{\neq T}(X_{\neq T}))}}$$

The SB after matching is given by:

$$SB_{after} = 100 * \frac{(\bar{X}_T - \bar{X}_M)}{\sqrt{0,5 * (V_T(X_T) + V_M(X_M))}}$$

where $V_T(X_T)$ is the mean (variance) in the treatment group before matching and $V_{\neq T}(X_{\neq T}))$ the analogue for the non-treated group. $V_T(X_T)$ and $V_M(X_M)$ are the corresponding values for the matched samples. This is a common approach used in many evaluation studies (e.g. *Lechner* [1999], *Sianesi and Reenen* [2003] and *Caliendo et al.* [2005]). In most empirical research, a bias reduction of between 3 to 5 % is considered sufficient (*Caliendo and Kopeinig* [2008]).

5.3 Ordinary Least Squares

In order to extend our analysis we also employ a standard ordinary least squares (OLS) framework.[42] As with PSM, the aim is to estimate the effect of participation in the AEI on performance in PIAAC, with respect to numeracy, literacy

[41] See *Rubin* [2001] for a more extensive discussion on Rubin's B and Rubin's R.
[42] See *Wooldridge* [2015] for a description of the methods and assumptions behind OLS.

and the problem solving skills. We estimate three models:

$$pvlit = \alpha_{lit} + \beta_{lit_i} X_i + \varepsilon_{lit}$$

$$pvnum = \alpha_{num} + \beta_{num_i} X_i + \varepsilon_{num}$$

$$pvpsl = \alpha_{psl} + \beta_{psl_i} X_i + \varepsilon_{psl}$$

Where *pvlit*, *pvnum* and *pvpsl* is the score on the literacy-, numeracy- and problem solving section of PIAAC, respectively. While α is the intercept, β is the coefficient of interest on the specific $i : th$ covariate in the vector X_i, and ε is the error term. The covariates included are the same as in the propensity score analysis and consist of seven age-interval dummies, a gender dummy, a variable that captures an individual's immigration status, variables for the highest education of the individual's mother, father and a combined measure of parental educational qualifications. Furthermore, we include variables for the highest level of education prior to the introduction of AEI and a special education dummy for people with especially low level of education before the AEI. Lastly, we also control for the number of children in the home at the time of the AEI. Due to the nature of the survey data, and the fact that we are estimating plausible values, we have to employ special measures in order to calculate the standard errors correctly, the exact formula for these can be found in the PIAAC Methodology in section 4.3.

6 Results

In this section, we first demonstrate the results from the PSM analysis. The structure follows the framework outlined in section 5.2. After assessing common support and the quality of the matching, the estimated treatment effects are presented, both for the entire sample and a number of subgroups. Following this, we present a range of recently measured outcome variables, also over different subgroups. Finally, the OLS estimates are presented.

6.1 Propensity Score Matching

6.1.1 Overlap and Common Support

The histogram over the propensity scores by treatment shown in Figure 1 illustrates that the common support stretches throughout all propensity score values. As outlined in section 5.1, the propensity score may assume any value between $0 < P(X) < 1$. In Figure 1, the propensity scores are plotted along the x-axis, while the frequency of individuals with particular scores are plotted along the y-axis. In this analysis, the mean propensity score for the treated group (upper bars) is 0.161. Meanwhile, the mean propensity score for the matched group (lower bars) is 0.168. This high quality of the overlap is important, as it allows us to use the whole sample. Thus, we do not need to disregard units outside the area of common support, because there are none. This is partially facilitated by the fact that the propensity score estimates extends between 0 and 0.4, with a longer span it generally becomes more difficult to attain common support. However, the key concern is that common support is attained within the estimated propensity score span, and this is the case in our analysis.

The density plots in Figure 2 display the effectiveness of the matching procedure. The graphs present the distribution of propensity scores over (top-left) treated and untreated individuals and (top-right) treated and matched individuals respectively. Before matching, the propensity scores of treated and untreated individuals varies widely. After matching, as visualized by the zoomed in plot (bottom), the propensity score distributions overlap very closely. Thus, relative to the grouping before matching, the procedure has successfully identified and ordered individuals into more comparable groups.

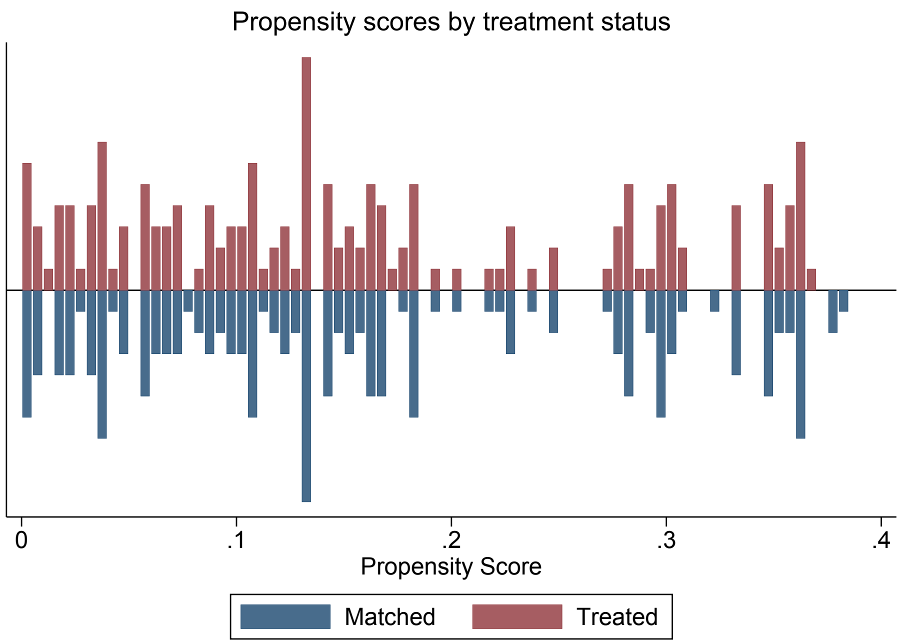

Figure 1: Propensity Score Histogram by Treatment Status

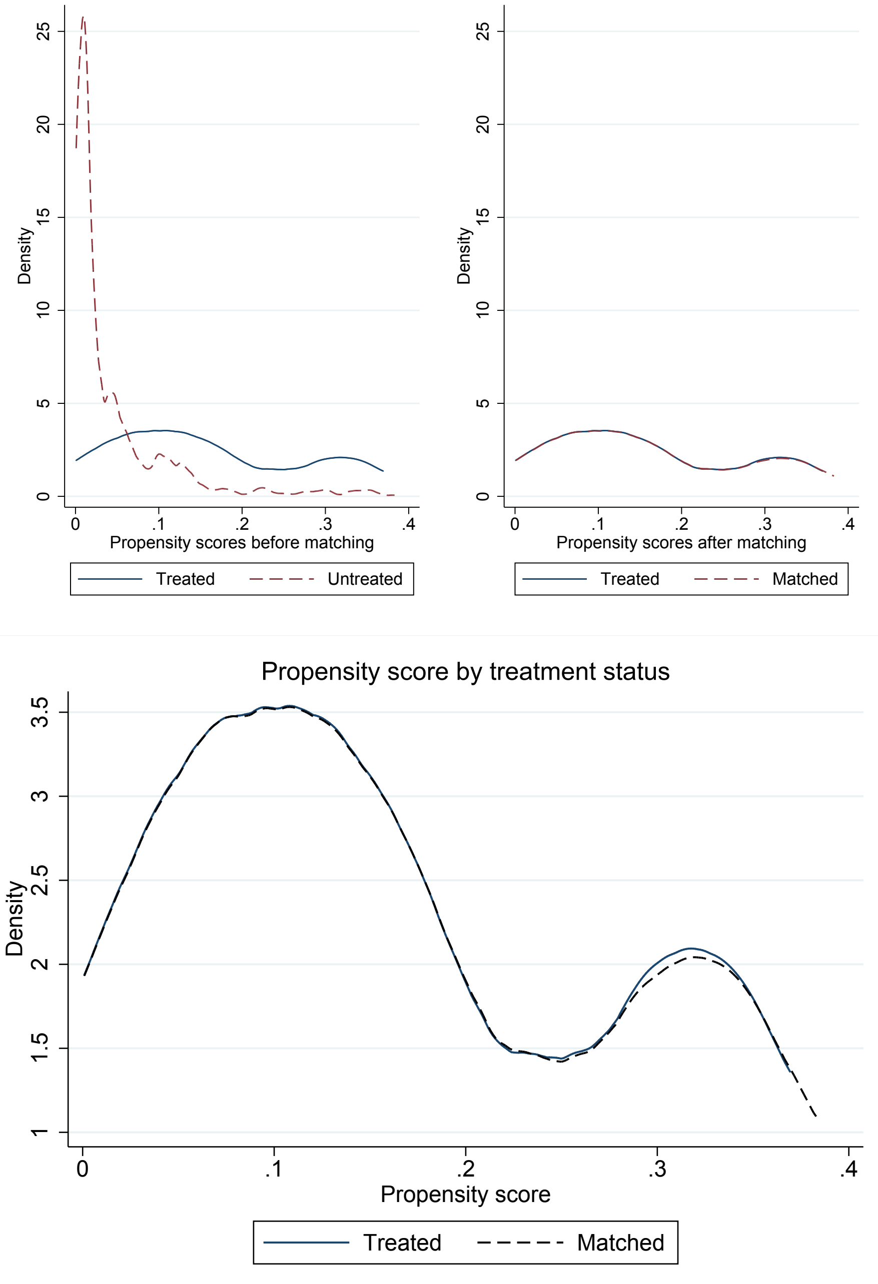

Figure 2: Propensity Score Overlap - Before and After Matching

6.1.2 Matching Quality

Figure 3 shows a before and after comparison of the standardized bias measure across covariates, which was described in section 5.2.4. Ideally, after matching, the SBs should be as close to zero as possible, although a bias reduction to between 3 to 5 % is usually sufficient. Importantly, as we go from unmatched to matched, all of the covariates, excluding age category 7, move towards the zero line, demonstrating bias reduction from the matching procedure.

More specifically, Table 5 shows how the mean covariate values differ between the treated and the matched group before and after the matching, enabling an overview comparison of the matching quality. Ten out of fifteen covariates are well within the recommended limit of introducing at maximum 5 % bias after matching. Of the remaining five covariates, all being binary age categories, three are very close to the limit (at 5.60, -5.50 and -5.80 % respectively), while two still introduce -10.00 and 11.30 % bias respectively. On the right-hand side of Table 5 are the results from a t-test on mean differences between the two groups, and none of the covariates are statistically significantly different between the two groups, which is ideal.

The two formal measures of how successful the balancing across covariates is, Rubin's B and Rubin's R, are estimated at 0.188 and 1.12 respectively. Both are well within the values stipulated to have a well-balanced sample (see the limits defined in section 5.2.4). Section 9.5 in the appendix provides a more basic version of the same results, displaying only how the values differ between the treated and the matched group after matching.

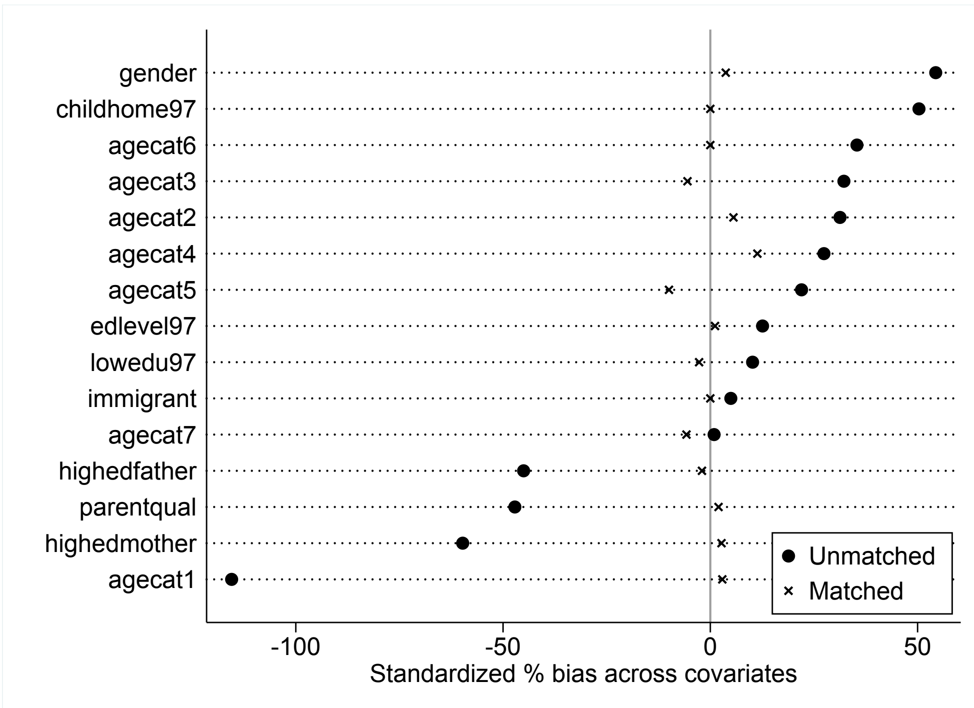

Figure 3: Standardized Bias Across Covariates - Before and After Matching

Table 5: Balancing Comparison - Before and After Matching

		Mean			t-test		
Variable	Sample	Treated	Non-treated	SB (%)	t	p-value	$\frac{V(t)}{V(c)}$
Gender	U	1.72	1.46	54.40	6.62	0.00	0.72
	M	1.72	1.70	3.70	0.36	0.72	0.94
Highest education - mother	U	1.36	1.82	-59.70	-6.94	0.00	0.83
	M	1.36	1.34	2.70	0.28	0.78	0.99
Highest education - father	U	1.44	1.79	-45.00	-5.42	0.00	0.84
	M	1.44	1.46	-2.00	-0.20	0.85	0.92
Parents qualification	U	1.55	1.97	-47.20	-5.86	0.00	1.07
	M	1.55	1.53	2.00	0.19	0.85	1.07
Education level	U	0.94	0.81	12.60	1.62	0.11	0.97
	M	0.94	0.93	1.10	0.11	0.92	1.11
Low education	U	0.70	0.61	10.20	1.31	0.19	0.93
	M	0.70	0.72	-2.70	-0.25	0.80	1.07
Immigrant	U	1.70	1.67	4.90	0.62	0.54	0.94
	M	1.70	1.70	0.00	0.00	1.00	0.97
Children at home	U	0.40	0.18	50.30	7.24	0.00	1.91
	M	0.40	0.40	0.00	0.00	1.00	1.00
Age category 1	U	0.09	0.56	-115.50	-12.25	0.00	0.51
	M	0.09	0.08	2.90	0.38	0.70	1.15
Age category 2	U	0.17	0.07	31.30	4.91	0.00	2.41
	M	0.17	0.15	5.60	0.45	0.66	1.14
Age category 3	U	0.17	0.07	32.20	5.07	0.00	2.42
	M	0.17	0.19	-5.50	-0.42	0.67	0.92
Age category 4	U	0.15	0.07	27.40	4.20	0.00	2.15
	M	0.15	0.12	11.30	0.95	0.34	1.23
Age category 5	U	0.13	0.06	22.0	3.28	0.00	2.00
	M	0.13	0.16	-10.00	-0.77	0.44	0.85
Age category 6	U	0.18	0.07	35.40	5.66	0.00	2.57
	M	0.18	0.18	0.00	0.00	1.00	1.00
Age category 7	U	0.11	0.10	0.90	0.12	0.91	1.02
	M	0.11	0.12	-5.80	-0.51	0.61	0.87
			Statistics				
Sample	Pseudo R^2	LR chi-2	p>chi-2	MeanBias	MedBias	B (%)	R
Umatched	0.20	262.99	0.00	36.60	32.20	137.70	0.89
Matched	0.00	3.01	0.99	3.70	2.70	18.80	1.12

MeanBias/MedBias is the mean/median absolute standardized bias. B and R stands for Rubin's B and Rubin's R.

6.2 Propensity Score Matching: Treatment Effects

Several interesting observations emerge when examining the estimates on PIAAC-scores in literacy, numeracy and problem solving (see Table 6). In the full sample, the AEI participants are not exhibiting higher scores on any of the components. Instead, they are statistically significantly worse at the ten percent level in both numeracy and problem solving, compared to the matched group, while the results in literacy are not statistically different.

In the subgroup analysis by gender, we find that the results for female participants are not statistically different, compared to non-participants. A different picture emerges when studying male participants, whom performed significantly worse across all three test components at the one percent significance level in numeracy and problem solving, and at the five percent significance level in literacy. The results for the full sample are thus driven by the lower performance of men, compared to non-participating men. This discrepancy will be discussed extensively in section 7.1.

It is also possible to distinguish differences in performance based on the education level attained before the start of the program. Those with lower education are, like men, performing worse than their matched group of counterparts. Meanwhile, those with better educational qualifications at the start of the program do not distinguish themselves across the three components relative to their matched counterparts. As outlined in the variable definitions in the appendix section 9.10, low education is here defined as an upper secondary qualification or less[43] from the pre-reformed system, and could arguably be described as a target group for the AEI (see section 9.2 in the appendix).

[43]ISCED 3A-B or less.

[44]High education is here defined as having studied at levels higher than upper secondary qualification (ISCED 4 or more) as measured before 1997.

[45]Low education is here defined as an upper secondary qualification or less (ISCED 3A-B or less) as measured before 1997.

Table 6: PSM: Estimated Treatment Effects

	Mean_T	Mean_M	Diff	σ_T	σ_M	MSE	t-statistic
Main specification							
Literacy	264.20	268.68	-4.48	3.84	3.62	3.73	-1.20
Numeracy	262.36	269.65	-7.29	4.05	3.82	3.94	-1.85*
Problem Solving	268.26	275.52	-7.26	3.97	4.65	4.31	-1.68*
Female							
Literacy	265.00	263.95	1.05	4.15	4.17	4.16	0.25
Numeracy	260.91	262.01	-1.10	4.47	4.94	4.71	-0.23
Problem Solving	266.66	268.50	-1.84	4.23	5.79	5.01	-0.37
Male							
Literacy	262.26	280.17	-17.91	7.59	6.91	7.25	-2.47**
Numeracy	265.85	288.23	-22.38	7.77	7.03	7.40	-3.02***
Problem Solving	272.81	293.09	-20.28	8.37	6.15	7.26	-2.79***
High Education[44]							
Literacy	273.29	271.44	1.85	7.14	6.69	6.92	0.27
Numeracy	272.88	273.13	-0.25	6.62	7.73	7.18	-0.03
Problem Solving	274.24	274.93	-0.69	8.48	7.87	8.16	-0.08
Low Education[45]							
Literacy	261.62	268.01	-6.39	4.05	4.50	4.28	-1.49
Numeracy	259.38	268.81	-9.43	4.51	4.68	4.59	-2.05**
Problem Solving	266.46	275.67	-9.21	4.35	5.58	4.97	-1.85*

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

T = Treated, M = Matched

6.3 Outcome Comparison

The PIAAC data enables examination of how the treated and matched groups differ over a wide range of outcomes[46], as measured between 2011 and 2012 when PIAAC was conducted. There is thus a follow-up period of some 9 to 15 years, depending on when an individual participated in the AEI, and the year the individual participated in PIAAC. The estimates for the full sample, and the male- and female samples separately are found in Table 7 to 9 respectively. All critical values are Bonferroni corrected due to the large number of comparisons and more arbitrary nature of the outcome variables (see section 9.6 in the appendix).

Examining first the labor market outcomes, no differences are found with respect to unemployment incidence, hours worked per week, nor regarding having attained paid work over the past five years. Interestingly though, the treated group reports a significantly higher likelihood of working physically for long periods at their current jobs, an effect driven by the female subgroup differences. Meanwhile, all the matched groups display significantly higher earnings, both per hour and accumulated over a year.[47] The higher wage levels, as measured per hour, of the matched groups may be partially explained by marginal differences in the number of years of professional work experience. There are no differences between the two groups with respect to educational outcomes. Neither the level of the groups average top qualification, nor the average age for when this qualification was obtained, is statistically different. Similarly, the aggregated total years of schooling appear almost equal over the different groups. Regarding social outcomes, an important variable is health, where the treated group in the full sample (Table 7) reports significantly lower levels of self-reported health. This effect is driven by the male sample, which reports lower levels in health, statistically significant at the one percent level.

The preliminary[48], findings based on the earnings data, will be explored further in the results discussion. It will also be related to the more general question of evaluative measures in education programs, and the importance of skills and wages as outcome measures.

[46]For a detailed description of the outcome variables, see section 9.11 in the appendix.

[47]We only have access to earnings data in deciles, which is the reason why we refrain from more thorough wage analyses of the type found in previous literature on the AEI.

[48]Due to only having access to deciles data on earnings.

Table 7: Outcome Comparison

Variable	AEI Mean	AEI S.E.	Matched Mean	Matched S.E.	t-stat
Labor market outcome					
Employed	1.38	0.06	1.29	0.05	1.27
Paid work last 5 years	0.90	0.02	0.92	0.02	-1.00
Working physically***	3.39	0.14	2.83	0.15	3.86
Hourly earnings interval***	4.47	0.20	5.80	0.23	-6.19
Hours work	36.77	0.68	38.16	0.81	-2.00
Yearly earnings rank***	3.31	0.10	3.91	0.12	-5.45
Years with paid work*	25.84	0.77	28.22	0.84	-2.96
Job satisfaction*	1.85	0.07	1.66	0.06	2.92
Education outcome					
Education level	3.57	0.11	3.72	0.14	-1.20
Top qualification	7.11	0.25	7.38	0.29	-1.00
Age finishing top qualification	30.80	0.90	29.72	0.99	1.14
Total years of education	12.08	0.15	12.11	0.23	-0.16
Learning activities last year	2.29	0.06	2.17	0.05	2.18
Need more training	1.73	0.04	1.68	0.04	1.25
Social outcome					
Health***	2.74	0.09	2.45	0.08	3.65
Trust few people	2.51	0.10	2.63	0.10	-0.80
No government influence	3.04	0.10	3.08	0.09	0.42
Partner	1.20	0.04	1.13	0.03	2.00
Children	2.40	0.10	2.50	0.10	-1.00
Born in country	1.16	0.03	1.15	0.03	0.33
Read books	3.37	0.11	3.35	0.11	0.18
Read news	4.51	0.07	4.56	0.06	-0.77
Volunteer work	1.72	0.09	1.66	0.08	0.71

The critical values are Bonferroni corrected, see the appendix section 9.6.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8: Outcome Comparison - Males

Variable	AEI Mean	AEI S.E.	Matched Mean	Matched S.E.	t-stat
Labor market outcome					
Employed	1.46	0.12	1.22	0.08	2.40
Paid work last 5 years	0.85	0.05	0.93	0.04	-1.78
Working physically	3.22	0.28	2.74	0.25	1.81
Hourly earnings interval**	5.38	0.51	6.81	0.42	-3.08
Hours work	37.47	1.71	40.70	1.89	-1.79
Yearly earnings rank***	3.59	0.22	4.42	0.24	-3.61
Years with paid work	26.48	1.49	27.96	1.71	-0.93
Job satisfaction	2.03	0.12	1.77	0.10	2.36
Education outcome					
Education level	3.29	0.20	3.55	0.25	-1.16
Top qualification	6.58	0.46	6.94	0.54	-0.92
Age finishing top qualification	26.68	1.62	24.42	1.43	1.49
Total years of education	11.67	0.30	11.71	0.42	-0.11
Learning activities last year	2.40	0.10	2.18	0.11	2.10
Need more training	1.75	0.07	1.66	0.07	1.29
Social outcome					
Health***	2.96	0.17	2.20	0.16	4.61
Trust few people	2.73	0.19	2.51	0.17	1.22
No government influence	2.83	0.19	2.86	0.17	-0.17
Partner	1.10	0.05	1.17	0.06	-1.27
Children	2.32	0.20	2.41	0.20	-0.45
Born in country	1.17	0.05	1.14	0.05	0.60
Read books	2.85	0.20	2.65	0.21	0.98
Read news	4.67	0.11	4.49	0.13	1.50
Volunteer work	1.85	0.18	1.88	0.19	-0.16

The critical values are Bonferroni corrected, see the appendix section 9.6.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9: Outcome Comparison - Females

Variable	AEI Mean	AEI S.E.	Matched Mean	Matched S.E.	t-stat
Labor market outcome					
Employed	1.35	0.06	1.32	0.06	0.50
Paid work last 5 years	0.92	0.03	0.92	0.03	0.00
Working physically**	3.45	0.17	2.87	0.18	3.31
Hourly earnings interval***	4.16	0.20	5.40	0.27	-5.28
Hours work	36.50	0.68	37.04	0.81	-0.72
Yearly earnings rank***	3.20	0.10	3.67	0.12	-4.27
Years with paid work*	25.60	0.90	28.34	0.94	-2.98
Job satisfaction	1.78	0.08	1.61	0.07	2.27
Education outcome					
Education level	3.68	0.14	3.80	0.17	-0.77
Top qualification	7.31	0.29	7.57	0.35	-0.81
Age finishing top qualification	32.28	1.05	32.06	1.22	0.19
Total years of education	12.25	0.18	12.29	0.27	-0.18
Learning activities last year	2.25	0.07	2.17	0.06	1.23
Need more training	1.73	0.05	1.68	0.05	1.00
Social outcome					
Health	2.65	0.11	2.55	0.10	0.95
Trust few people	2.43	0.12	2.69	0.12	-2.17
No government influence	3.11	0.11	3.17	0.11	-0.64
Partner***	1.24	0.04	1.10	0.03	4.00
Children	2.43	0.11	2.53	0.11	-0.91
Born in country	1.16	0.03	1.15	0.03	0.33
Read books	3.57	0.13	3.66	0.11	-1.13
Read news	4.42	0.09	4.60	0.07	-2.25
Volunteer work	1.66	0.10	1.56	0.08	1.11

The critical values are Bonferroni corrected, see the appendix section 9.6.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

6.4 Ordinary Least Squares

This section presents the results from estimating the effect of participation in the AEI on performance in PIAAC using the OLS approach outlined in section 5.3. There are three specifications with controls corresponding to the covariate vector X_i. The dependent variable is *literacy* (LIT), *numeracy* (NUM) and *problem solving* (PSL) respectively. The results for the full sample are shown in Table 10 below. The independent variable of primary interest is AEI participation and while the coefficient is negative for all three specifications, it is not significant. Thus, both when using PSM and OLS, AEI participants are not found to have performed better in any of the three skill dimensions, compared with non-participants.

Table 10: OLS Regression Estimates on PIAAC Scores

Variable	**LIT**	**NUM**	**PSL**
AEI-Participation	–5.32 (4.24)	–5.94 (4.21)	–5.21 (4.26)
Gender	–1.79 (1.87)	–11.90*** (1.80)	–2.85 (1.76)
Highest education - mother	6.13*** (1.86)	5.62*** (2.08)	6.53*** (1.90)
Highest education - father	1.75 (1.63)	1.80 (1.83)	2.59* (1.33)
Parents qualification	7.82*** (2.18)	7.83*** (2.30)	4.25** (1.96)
Education level	4.96*** (1.39)	6.80*** (1.45)	5.94*** (1.21)
Low education	–5.58*** (1.34)	–8.24*** (1.48)	–9.27*** (1.14)
Immigrant	17.77*** (1.45)	18.56*** (1.70)	9.93*** (1.31)
Children at home	0.05 (2.47)	0.23 (2.62)	–0.35 (2.75)
Age category 1	11.02 (18.35)	8.16 (15.31)	33.05 (36.06)
Age category 2	11.22 (18.90)	9.72 (15.23)	25.20 (36.09)
Age category 3	10.12 (18.66)	8.97 (15.54)	17.41 (36.70)
Age category 4	8.88 (18.10)	8.66 (15.04)	15.45 (36.33)
Age category 5	0.00 (17.72)	0.00 (14.85)	5.90 (36.03)
Age category 6	–1.89 (18.37)	0.78 (14.75)	0.00 (35.30)
Age category 7	–12.48 (18.23)	–5.55 (15.44)	–7.71 (36.06)
R-squared	0.16	0.14	0.27
Number of observations	3361	3361	3361

Standard errors reported in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

As is done in the PSM analysis, the OLS regression is also done for the two subgroups. The results are located in the appendix section 9.7. First, in Table 14, the results are presented for the female and male sample respectively. Second, in Table 15, the sample is split over high and low education level[49] as attained before the AEI.

The results with respect to gender are similar to the full specification, where the estimated effects for AEI participants are not statistically significant in any of the three skill dimensions. Noteworthy though, is that all estimates are negative[50] and the estimates for the male sample tend to be lower than for the female sample. Also when dividing the sample over high and low education level, results in line with the patterns found earlier emerge. Again for the OLS estimates, there are no statistically significant estimates for the three skill dimensions. However, as with the PSM-estimates, the low education sample appears to produce lower estimates than the high education sample. For both subgroup analyses, a clear majority of the remaining independent variables, the age categories excluded, are highly significant. A further discussion on the merits and shortcomings of the two estimation strategies in the AEI context is provided in section 7.1 and 7.3.

[49] As explained in the data section 4.5.

[50] The t-statistics range from -0.62 to -1.49.

7 Discussion

7.1 Results Discussion

Previous evaluations on the effects from participation in the AEI, discussed extensively in section 3.3, have focused on earnings and employment incidence, without finding substantial positive effects. In cases where positive effects have been found, most notably in the recent long-term study by *Stenberg and Westerlund* [2015], the relative wage improvements from participation have been limited to female participants. For men, previous research has not found significant positive effects on earnings or employment for AEI participants. Within this context, and coupled with the knowledge of how rewards in the labor market are driven by two factors, signalling and skills, we would expect to see no markedly better performance in skill evaluations for AEI participants on the whole, and if we would see any effect, it would likely be for women.

The results from the previous section show that AEI participants did not perform better in PIAAC compared to non-participants. Rather, the results for participants are significantly lower at the ten percent level in both numeracy and problem solving, and lower, but not statistically significantly so, in literacy. This is thus in line with expectations, based on the previous research not showing general positive effects on earnings and employment. This new evidence regarding skills is highly informative for the interpretation of the previous economic studies examining the effects from participation in the AEI. We may now contribute with an explanation for the absence of considerable labor market rewards for participants - they do not exhibit higher skills than non-participants.

Considering that recent literature on the AEI has showcased positive effects on earnings for females, it is motivated to investigate potential differences between genders. This subgroup analysis on the skill dimension does indeed provide interesting results. Female participants are on par with their matched counterparts on all skill dimensions, the results of the treated and matched groups are not statistically significantly different. The lower scores in the full sample are instead driven by the performance of male AEI participants. On all three skill dimensions, men participating in the AEI perform statistically significantly worse, at the one percent level in two cases, and at the five percent level in one case, compared to the matched group of non-participating males with similar characteristics. These two findings illustrate that women have fared relatively better than men from participation, and suggests that the improved labor market outcomes recently found in the literature could be explained by the relatively higher skill levels that women exhibit. As our results for the male subgroup show that their skill levels are below other males in the matched group, the absence of labor market rewards for participants is not surprising.

The strong discrepancy between male participants and non-participants merits further discussion. The matched group of male non-participants are doing well on all three PIAAC components, specifically so in numeracy and problem solving. One concern that then arises is whether this result could be driven by the matching process itself, namely, that the procedure worked better for the female subgroup. Moreover, women are in majority both as AEI participants and UBS recipients, leading to a larger share of females also in the overlap studied in this paper[51]. Statistically, this should lead to a more representative sample of females. In evaluating if the matching worked less well for males, it could be useful to revisit the outcome comparisons in section 6.3.

An indicator of worse matching quality for the male subgroup would be that there are relatively larger differences in the male outcome comparison. But this is not the case. In the female subgroup, the participants and non-participants are statistically different in five variables (Table 9) whereas the male participants and non-participants differ only in three variables (Table 8). This must however be cautioned with the fact that the larger female sample also leads to more precise standard errors. Both the male and female participants have lower earnings at the time of PIAAC compared to their relative matched groups, with the difference for the female group being even stronger. Furthermore, the female participants differ in two variables in a way that does not indicate that the male matching worked less well. These are the variables capturing the degree of working physically and the number of years with paid work, where the female AEI-participants are relatively more likely to work physically, and report relatively fewer years of paid work. If anything, this would arguably make our

[51] As stated in the Data section, the overlap studied in this paper consists of 122 women and 48 men.

findings discussed above more robust, as the male participants and their counterparts are not statistically significantly different.

Interestingly, male AEI participants report lower levels of health than non-participants. Meanwhile, females are not statistically different relative to their matched counterparts. As this study only match on time-fixed observables, this might be a variable that not only tells a story in itself, where worse health affects performance negatively, but also may serve as a proxy for other performance influencing unobservables. Such unobservables could, for instance, be more severe health issues, concentration issues, motivational factors or other troubles that are not captured in the data set.

It is important to remember that the AEI was not introduced into a vacuum, it was a policy added to a labor market with other options and structural elements that resulted in idiosyncratic choice sets both before and after the introduction of the program. The discrepancy discussed in above paragraphs could thus be driven by the alternative decisions that non-AEI participating individuals made, and this could help explain the differences in our results. As outlined in section 2, females entered into the AEI to a higher degree than men, and men entered into labor market orientated programs to a higher degree than women. That the clear majority of AEI participants were female, and that LMT programs were generally skewed towards male dominated sectors, with a similar but inverted ratio of participants by gender as the AEI, with men being in the majority, shows that this was the case.

The male dominated LMT programs, available at the time of the AEI, were more effective at leading to employment, compared to the female dominated programs. The choice of program individuals faced, general or specific, is arguably closely connected to the model by *Heckman et al.* [1999], in which individuals choose to enroll in a program only if the expected returns are positive. If taken into account by the participants, it may have led to differences in terms of the relative share of individuals with behavior characterized by moral hazard for the men and women that entered into the AEI. This selection mechanism for men might have worked in a two-fold manner, in that it could also be a factor leading to higher results in the matched group of males. That group is likely to be the type of individuals of similar characteristics to AEI participants before, as we have tried to make sure with our matching, but with better health and, perhaps, other differences in unobservables, such as motivation.

One potential explanation for the results discrepancy is then that the males that selected into the AEI were, to a relatively higher degree than for the female sample, a subgroup of males, perhaps with lower health and motivation, and that may have been more prone to the issue of moral hazard.[52] Because of the differences in efficiency of the LMT programs, men arguably had better choices regarding which program to enter. This may then, to a relatively higher degree, have attracted men with certain characteristics into the AEI, compared to women, that entered into the AEI to a larger extent.

Demonstrating subgroup heterogeneity of treatment effects is important as it may result in useful insights for policy makers wishing to evaluate the efficiency and effectiveness of the program. It provides the opportunity to leverage the additional information, enabling identification of subgroups for which treatments are effective and provide knowledge of how programs should be constructed so to maximize positive effects. With this in mind, we furthermore find a relevant discrepancy when we perform subgroup analyses based on the education with which AEI participants entered the program. Those entering with lower education are not performing as well as non-participants with similar characteristics. Meanwhile, those that entered the program with relatively higher education are performing as well as their counterparts.

The original idea with the AEI was to target the group with lower educational qualification.[53] The results of this study show, however, that members of this target group were the ones that benefited the least from participation in terms of increasing their skills. This is essential knowledge in a policy evaluation and it raises the question of whether the program was less well suited to the group with lower educational qualifications. Moreover, in highlighting whether the program was structured in an ideal way for accomplishing the goals that were set out, such findings may improve future policies. In particular, it should have

[52]Moral hazard is discussed more extensively in relation to the UBS grant in the general discussion.

[53]This is discussed in the background section, and in appendix section 9.2.

the potential to enhance the structure of the updated version of the AEI, which is set to be implemented in the near future, and similar initiatives.

Matching as a strategy to control for covariates is motivated by the conditional independence assumption (*Angrist and Pischke* [2008]), and, ultimately, matching amounts to covariate-specific treatment-control comparisons, weighted together to produce a single overall average treatment effect. The causal interpretation of the OLS regression coefficient is based on the same conditional independence assumption. Matching and OLS are thus both control strategies and because the core assumption underlying causal inference is the same, it is worth asking to what extent the two approaches actually differ.

Angrist and Pischke [2008] argue that regression can be thought of as a particular kind of weighted matching estimator, and therefore the differences between regression and matching estimates are unlikely to be of major empirical importance. The regression estimate differs from the matching estimate only in the weights used to combine the covariate-specific effects, δ_X, into a single average effect. In particular, while matching uses the distribution of covariates among the treated to weight covariate-specific estimates into an estimate of the effect of treatment on the treated, regression produces a variance-weighted average of these effects. *Heckman and Vytlacil* [2005] state that "all average causal effect estimators can be interpreted as weighted averages of marginal treatment effects whether generated by matching, regression, or local instrumental variable estimators."

As in the PSM, our OLS estimates show no significant differences for the full sample of AEI participants, further strengthening the robustness of the finding that AEI participants in general do not exhibit better results on PIAAC than non-participants, in the long run. This also holds true with respect to the female subgroup, that also when using OLS, do not perform significantly better or worse than their matched counterparts. Although the point estimates are in the same direction, and demonstrating similar patterns with respect to heterogeneous treatment effects by gender and pre-program education levels, the OLS results yield no significant results for the subgroup analyses based on males and those with low education level. However, the results for the high education level participants are the same as in the PSM analysis.

The discrepancy between the PSM and OLS results could originate in either model being misspecified. However, PSM has considerable merits when used correctly, and particularly in the context of this study. Even though there is no asymptotic efficiency gain from the use of estimators based on the propensity score, there will often be a gain in precision in finite samples (*Angrist and Hahn* [2004]), which is of empirical significance. Importantly, when using propensity score matching, it is essential to use prior knowledge for dimension reduction. This is what drives the improvement in finite-sample behavior, and it requires reducing the dimensionality of the matching problem in a manner that has empirical consequences (*Angrist and Pischke* [2008]). We have aimed to do so in our evaluation strategy, through the extensive knowledge about the AEI that is available.

Furthermore, *Morgan and Winship* [2007] show that matching significantly outperform OLS models should a linear relation be employed when the true functional form is nonlinear. This is especially the case when there is a poor overlap in covariates, something the common support analysis highlighted as being the case before matching, as seen in Figure 2 (the top left graph), a situation when traditional parametric methods such as OLS struggle (*Bryson et al.* [2002]). The OLS results in this paper are thus likely to be sensitive to the linearity assumption. Our OLS specifications still remain linear, however, due to two reasons. First, we want the results from the two approaches to be as comparable as possible. Second, economic theory is ambiguous about whether any nonlinear OLS forms are to be imposed, and if so on which variables. We thus argue that the PSM estimates are likely to be more relevant for this paper.

7.2 General Discussion

A skilled and broadly educated population enables higher productivity. It also creates a more adaptable labor force for the future. Sweden, like many developed nations, is seeing a shift in the demographic structure towards a higher share of elderly people. This is driven both by fertility rates that are lower than the replacement rate and increased longevity. The new demographic structure puts an increased strain on the social welfare systems, particularly in countries such as Sweden, that have strong support systems for the unemployed and generous leave policies. The age-dependency ratio, the number of working age people supporting the rest of the population, is worsening. It is not far fetched to consider that adjustments such as increased or indexed retirement age will be introduced. Indeed, these proposals are being put forward by economists and, more tentatively, politicians. Such changes, coupled with the effect that technological development has had on the labor market, will require people to learn new skills. Especially skills of a cognitive nature, making them more adaptable to different types of jobs and tasks.

A well functioning adult education system, and programs aimed at improving the human capital of the entire population, are thus essential for making sure that the country has a workforce capable of sustaining its social systems. Wages may be an insufficient indicator to evaluate these programs, as the future restructuring of the labor market is not always taken into account in hiring decisions focused on the near-term, making it important to also examine actual skills, and especially those of more general character. Our findings regarding differences in wage outcomes (see section 6.3) between the groups further emphasize the importance of including measures of skills in evaluations of this type. The female AEI participants in our analysis perform as well as their counterparts on the measured skill dimensions, but report significantly lower earnings. This points to potential heterogeneous effects also depending on the evaluation measure used, skills or wages. The caveat here, as mentioned in section 6.3, is that we only have access to earnings data in deciles, but this is something that would be interesting to explore with more precise data in future research.

Policies aimed at affecting the skills of the population can reduce wage inequality and enable people to better contribute to the productivity of societies. In some countries, gains can be made from making the skills distribution more equal, but that requires human capital investments. In a world where high-level skills are increasingly in demand, it is especially important to invest in skills of certain subgroups that might otherwise be left trailing, such as people from lower socio-economic backgrounds, and migrants. In Sweden, with the high influx of low-skilled migrants in recent years, this is of particular relevance.[54] It would

[54]This is highlighted in a recent report, see *OECD* [2016b].

be interesting to investigate how the immigrant subgroup fared in the AEI. About twenty percent of AEI participants had immigrant backgrounds. Therefore, with our overlap of 170 individuals, we do not have a big enough sample, as the immigrant subgroup is not homogeneous, and would need to be analyzed over subgroups of pre-program qualifications, origin and gender, for meaningful conclusions to be drawn. Still, this is an interesting area for future research, and a question which should be easier to study as more PIAAC waves are conducted.

It is important to remember that most evaluations of the AEI identifies program participants with data on the UBS grant. UBS recipients were, however, not the only type of AEI participants (see Table 2 in section 4.1 for more information on the characteristics of UBS recipients) and differ from other participants in two important dimensions. First, the UBS recipients were most likely unemployed to a larger extent before the program[55], as compared to other participants. Second, those who were awarded UBS grants are those that received the most funding, corresponding to full unemployment benefits and with no requirements in terms of repayment. The UBS recipients are thus the group of individuals that took the smallest financial risk[56], and this may have affected motivation from a moral hazard perspective.[57] This can be contrasted with the types of funding that other AEI participants had, such as SVUX, regular student funding and self-financed individuals, which required individuals to pay more out of their own pocket in order to finance their studies. At the same time, that the UBS recipients to a higher extent are unemployed might also serve as a motivating factor, which could potentially bias the estimated treatment effects upwards.

Differences in the way that individuals finance their studies may affect motivation and performance in the classroom. This, and discrepancies regarding the education that individuals entered the AEI with, which would affect performance, are interesting to consider from the point of ability grouping. In adult education, little is known about the effects from ability grouping, or lack

[55] As this was an eligibility requirement for the UBS grant, see section 9.2 in the appendix.

[56] At least in terms of the funding that they received, it could be argued that these are individuals that live with smaller margins in terms of income shortfall, and with lower savings.

[57] As outlined in the previous section, *Heckman et al.* [1999] present a model in which the choice of enrolling or not enrolling in a program in period k is assumed to depend on expected future economic returns. In theory, any individual will enroll if, and only if, such returns are expected to be positive. What distinguishes the AEI is, however, that the UBS grant ensures positive returns from day one, and without financial risk.

thereof. Even in regular education, the findings regarding its effectiveness are inconclusive. Theoretically, one might think that the people targeted primarily and initially by the AEI, low-skilled individuals that have not performed well previously in schooling environments, may be especially sensitive to classroom dynamics, and become adversely affected in, for example, situations where progress goes too fast, causing certain individuals to fall behind. On the other hand, it is also possible that classroom dynamics worked well, and that better performing individuals made the learning experience more beneficial for everyone. It is worth remembering that due to the significant budget allocation, no individuals were turned away from participating in the AEI, and about half a million people entered the program, also those with higher skills than the main target group. The within-classroom composition, and how it affected group- and learning dynamics, is unknown. But it is worth taking into consideration, especially due to the heterogeneity in treatment effects found in this paper.

7.3 Limitations

The estimation methods used in this paper ignore the impact a program may have had on the outcomes and behavior of non-participants. These effects, known as general equilibrium effects, are of two kinds in the context of the AEI: i) an increase in labor supply through making previously inactive individuals active on the labor market, and ii) a change in the composition of the labor supply where the relative share of medium-skilled[58] workers, to the detriment of low-skilled non-participants, increased. Considering the scale of the AEI, it is possible that it was indeed accompanied by such general equilibrium effects. If so, it would strengthen our conclusions from the previous section, as non-participants' performance on PIAAC would, if anything, be biased downwards. This would imply that the conclusions regarding the differences between participants and non-participants are more robust.

One limitation in this paper is the size of the overlap between the AEI and PIAAC, which consists of 170 individuals. As new waves of PIAAC are implemented, the number of people that participated in the AEI and PIAAC will increase. Ideally, we would have wanted a larger sample so as to increase the

[58]Under the common definition discussed in section 2.1., where medium-skilled corresponds to having completed upper secondary education.

external validity and precision of our findings. Due to the statistical methods involved in selecting participants for PIAAC, and the absence of self-selection into the survey, the group size in itself is, however, unlikely to lead to a biased sample. In total, although being a relatively small sample, it may be considered a random sample of AEI participants with UBS.

A more minor caveat is that the response rate for the original RBT sample[59], which is used in the stratified selection from the overall population, is marginally lower for participants from Stockholm, where estimated treatment effects from participation in the AEI on labor market outcomes has previously been found to be slightly higher.[60] This could in theory bias the treatment effects, but the difference in response rate is less than 10 percentage points, which would on average result in a few more respondents from other regions relative to Stockholm, in comparison with the actual AEI sample. Furthermore, as can be seen in Table 2 in section 4.1, a somewhat lower degree of participants financed their studies using UBS grants in large cities, compared to other parts of Sweden. Due to that the effect found in previous research is relatively small, and the fact that the difference in response rate is below 10 percentage points, this is unlikely to have resulted in significant bias.

Another important consideration is how well PIAAC measures the intended learning outcomes of the AEI, which in turn depends on the courses studied. There is a very good overlap in terms of the types of courses that were most common in the AEI (computer science, Swedish and mathematics, see Table 1 in section 2.1) and the skill dimensions measured by PIAAC (literacy, numeracy and problem solving in technology rich environments). However, this study lacks data about the actual coursework for the treated individuals in our sample. This, combined with the aforementioned sample size issue could bias the treatment estimates downwards if the treated individuals in our sample do not adequately represent the overall distribution in terms of studied courses.

Similar to regression analysis, matching is a heavily data-dependent technique, and the precision of our estimates could be improved with enhanced data on

[59]See Table 16 in section 9.9 in the appendix.

[60]See both *Stenberg* [2002] and *Stenberg and Westerlund* [2015], no effect is found in *Stenberg* [2003].

the individuals. Although PIAAC is an exceptionally rich data source in some aspects, it lacks some, arguably important, information, for the purposes of this paper. First and foremost, this concerns one of the formal requirements regarding eligibility for the AEI, found in section 9.2 in the appendix. Although containing thorough information on individuals' professional history, the data provides no detail as to participants employment situation in the years prior to the AEI. And while using the UBS grant to identify participants, ensuring that the treated group was indeed unemployed before the start of the program, we lack information on this key aspect for non-participants. If this is not captured by the vector of covariates X_i used for matching, the estimated treatment effects are likely to be downward biased for AEI participants, relative to the matched group.

Furthermore, the lack of unemployment data prior to program enrollment, may bias treatment effects if skills measured in PIAAC were also improved in the types of jobs that were undertaken, and if the nature and extent of these effects vary. A similar argument can be made with respect to the absence of data on the grades individuals attained in primary and upper secondary school, although this is more likely to be better captured by the covariates. Data on both employment history and grades is available in Sweden, which collects micro data of high quality, but which we did not have access to in this study.

Similarly, even after matching, there may still be additional differences in the characteristics between treated individuals and their matched counterpart, that PIAAC does not allow us to capture with our empirical strategy. The situation would have been enhanced with data on, for instance, motivation and health, as measured before the AEI. Differences in such characteristics may lead to either over or underestimation of treatment effects. In order to causally interpret the estimated effects, one has to assume that mechanisms of that sort would not induce systematic differences with respect to the performance in PIAAC. This would arguably be a strong assumption to make. Even after considering the quality of the matching analysis, and the similarities between the two groups holding employment status, education- and social outcomes in mind, not including such covariates could have a significant impact on the estimates. Most likely, the estimates for the treated group presented in this paper are in such case downward biased relative to the matched group.

Lastly, while our data provides information about whether an individual participated in the AEI, it does not specify the exact year of enrollment between 1997 and 2002. Constructing all pre-treatment variables as measured in 1997, an issue of precision could arise for those individuals who in fact entered the program in, for instance, 2001. However, as can be seen in Table 2 in section 4.1, the number of UBS recipients entering the program is heavily skewed towards the period between 1997 and the fall of 1999. Thus, although remaining an issue, it is unlikely to be of major significant importance. Again, this is information that exists, but that is out of reach for this paper due to data availability.

8 Conclusion

We investigate whether AEI participants exhibit better skill-levels than non-participants in the long run. Adequate evaluation of educational programs require two important parts: i) understanding if and how such investments translates into skills and ii) investigating how those skills relate to economic outcomes. The current literature on adult education, and the AEI in particular, fails to do so. It usually takes i) as a fact, and focuses on ii). This leads to incomplete recommendations. Without knowing the channels through which an educational program operates, we cannot draw as informed conclusions on its shortcomings or merits.

Our results relate to the existing literature, which has found no overall effects on wages and employment, in that it finds that AEI participants do not perform better in the skills measured in PIAAC than non-participants. We can now, for the first time, pin down a possible reason behind the previous results in the literature regarding wages and employment incidence: the participants do not perform better in the the literacy-, numeracy- and problem solving skills components of PIAAC. Furthermore, the heterogeneity of treatment effects with respect to gender outcomes, with women doing better in terms of labor market outcomes, is through the results presented in this paper now potentially attributable to the heterogeneity with respect to cognitive skills. Moreover, the performance discrepancy found on educational qualifications before entering the program has important implications for future policy design, especially in terms of how the main target group fared. Lastly, this evaluation enhances the understanding of the human capital production function, with respect to adult education. In order to solidify the policy implications, future research would do well to incorporate more of the variables discussed in the limitations section.

References

Acemoglu, D., and D. Autor, Skills, Tasks and Technologies: Implications for Employment and Earnings, in *Handbook of Labor Economics*, vol. 4, edited by O. Ashenfelter and D. Card, pp. 1043–1171, Elsevier North-Holland, Amsterdam, 2011.

Albrecht, J., G. van den Berg, and S. Vroman, The Knowledge Lift: Adult Education in Sweden, *IZA Discussion Paper Series*, *No. 1503*, 2005.

Alm Stenflo, G., Inkomst-och sysselsättningseffekter av kommunal vuxenutbildning, *Temarapport från prognosinstitutet, SCB*, *1*, 2000.

Altonji, J. G., and C. R. Pierret, Employer Learning and Statistical Discrimination, *Quarterly Journal of Economics*, *116*(1), 313–350, 2001.

Angrist, J., and J. Hahn, When to Control for Covariates? Panel Asymptotics for Estimates of Treatment Effects, *Review of Economics and Statistics*, *86*(1), 58–72, 2004.

Angrist, J. D., and J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2008.

Austin, P. C., Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-To-One Matching on the Propensity Score, *American Journal of Epidemiology*, *172*(9), 1092–1097, 2010.

Autor, D., F. Levy, and M. Richard, The Skill Content of Recent Technological Change: An Empirical Exploration, *Quarterly Journal of Economics*, *118*(4), 1279–1333, 2003.

Autor, D. H., and D. Dorn, The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market, *American Economic Review*, *103*(5), 1553–97, 2013.

Axelsson, R., and O. Westerlund, Inflöde och varaktighet som arbetssökande efter avslutad utbildning: deltagare i kunskapslyftet och arbetsmarknadsutbildning höstterminen 1997, *Umeå Economic Studies*, *No. 566*, 2001.

Axelsson, R., and O. Westerlund, Kunskapslyftets effekter på årsarbetsinkomster - nybörjare höstterminen 1997, *Umeå Economic Studies*, *No. 647*, 2005.

Barro, R. J., and J.-W. Lee, International Data on Educational Attainment: Updates and Implications, *Oxford Economic Papers*, *53*(3), 541–563, 2001.

Beaton, A. E., and E. Gonzalez, *The NAEP Primer*, Chestnut Hill, MA: Boston College, 1995.

Becker, S. O., and A. Ichino, Estimation of Average Treatment Effects Based on Propensity Scores, *Stata Journal*, *2*(4), 358–377, 2002.

Benhabib, J., and M. M. Spiegel, Human Capital and Technology Diffusion, in *Handbook of Economic Growth*, vol. 1, edited by P. Aghion and S. N. Durlauf, pp. 935–966, Elsevier North-Holland, Amsterdam, 2005.

Bishop, J. H., Is the Test Score Decline Responsible for the Productivity Growth Decline?, *American Economic Review*, *79*(1), 178–197, 1989.

Blackburn, M. L., and D. Neumark, Omitted-Ability Bias and the Increase in the Return to Schooling, *Journal of Labor Economics*, *11*(3), 521–544, 1993.

Bland, J. M., and D. G. Altman, Multiple Significance Tests: The Bonferroni Method, *British Medical Journal*, *310*(6973), 170, 1995.

Bowles, S., and H. Gintis, *Schooling in Capitalist America: Educational Reform and the Contradictions of American Life*, New York: Basic Books, 1976.

Bowles, S., H. Gintis, and M. Osborne, The Determinants of Earnings: A Behavioral Approach, *Journal of Economic Literature*, *34*(4), 1137–1176, 2001.

Bryson, A., R. Dorsett, and S. Purdon, The Use of Propensity Score Matching in the Evaluation of Active Labor Market Policies, *Department for Work and Pensions*, *Working Paper Series No. 4*, 2002.

Burgess, S., Human Capital and Education: The State of the Art in the Economics of Education, *IZA Discussion Paper Series*, *No. 9885*, 2016.

Caliendo, M., and S. Kopeinig, Some Practical Guidance for the Implementation of Propensity Score Matching, *Journal of Economic Surveys*, *22*(1), 31–72, 2008.

Caliendo, M., R. Hujer, and S. L. Thomsen, Individual Employment Effects of Job Creation Schemes in Germany with Respect to Sectoral Heterogeneity, *IAB Discussion Paper Series*, *No. 13*, 2005.

Calmfors, L., A. Forslund, and M. Hemström, Does Active Labor Market Policy Work? Lessons from the Swedish Experiences, *IFAU Working Paper Series*, *No. 4*, 2002.

Card, D., The Causal Effect of Education on Earnings, in *Handbook of Labor Economics*, vol. 3, edited by O. Ashenfelter and D. Card, pp. 1801–1863, Elsevier North-Holland, Amsterdam, 1999.

Card, D., J. Kluve, and A. Weber, Active Labor Market Policy Evaluations: A Meta-Analysis, *Economic Journal*, *120*(548), 452–477, 2010.

Chetty, R., J. N. Friedman, and J. E. Rockoff, Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates, *American Economic Review*, *104*(9), 2593–2632, 2014a.

Chetty, R., J. N. Friedman, and J. E. Rockoff, Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood, *American Economic Review*, *104*(9), 2633–2679, 2014b.

Cole, S., A. Paulson, and G. K. Shastry, High School Curriculum and Financial Outcomes: The Impact of Mandated Personal Finance and Mathematics Courses, *Journal of Human Resources*, Forthcoming.

Cole, S. A., and G. K. Shastry, *Smart Money: The Effect of Education, Cognitive Ability, and Financial Literacy on Financial Market Participation*, Harvard Business School, Boston, MA, 2009.

Coloumbe, S., and J. Tremblay, Literacy and Growth, *Topics in Microeconomics*, *6*(2), 2006.

Coulombe, S., J.-F. Tremblay, and S. Marchand, Literacy Scores, Human Capital and Growth Across Fourteen OECD Countries, *Statistics Canada*, 2004.

Cullen, J. B., S. D. Levitt, E. Robertson, and S. Sadoff, What Can Be Done to Improve Struggling High Schools?, *Journal of Economic Perspectives*, *27*(2), 133–152, 2013.

Currie, J., and E. Moretti, Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings, *Quarterly Journal of Economics*, *118*(4), 1495–1532, 2003.

Cutler, D. M., A. S. Deaton, and A. Lleras-Muney, The Determinants of Mortality, *National Bureau of Economic Research*, *Working Paper Series No. 11963*, 2006.

Darcovich, N., et al., *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*, OECD Publishing, 2000.

Dee, T. S., Are There Civic Returns to Education?, *Journal of Public Economics*, *88*(9), 1697–1720, 2004.

Dehejia, R. H., and S. Wahba, Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs, *Journal of the American Statistical Association*, *94*(448), 1053–1062, 1999.

Doppelhofer, G., R. I. Miller, and X. Sala-i Martin, Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach, *American Economic Review*, *94*(4), 813–835, 2004.

Efron, B., *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM Publishing, 1982.

Estrada-Mejia, C., M. de Vries, and M. Zeelenberg, Numeracy and Wealth, *Journal of Economic Psychology*, *54*, 53–63, 2016.

Fryer Jr., R. G., The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments, in *In Handbook of Field Experiments*, vol. 1, edited by E. Duflo and A. Banerjee, Elsevier North-Holland, Amsterdam, Forthcoming.

Gelman, A., and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, 2006.

Goeman, J. J., and A. Solari, Multiple Hypothesis Testing in Genomics, *Statistics in Medicine*, *33*(11), 1946–1978, 2014.

Goldin, C. D., and L. F. Katz, *The Race Between Education and Technology*, Harvard University Press, 2009.

Groot, W., and H. M. Van Den Brink, The Effects of Education on Crime, *Applied Economics*, *42*(3), 279–289, 2010.

Hanushek, E. A., The Economic Value of Higher Teacher Quality, *Economics of Education Review*, *30*(3), 466–479, 2011.

Hanushek, E. A., and D. D. Kimko, Schooling, Labor-Force Quality, and the Growth of Nations, *American Economic Review*, *90*(5), 1184–1208, 2000.

Hanushek, E. A., and L. Wößmann, The Role of Cognitive Skills in Economic Development, *Journal of Economic Literature*, *46*(3), 607–668, 2008.

Hanushek, E. A., and L. Wößmann, How Much Do Educational Outcomes Matter in OECD Countries?, *Economic Policy*, *26*(67), 427–491, 2011.

Hanushek, E. A., and L. Zhang, Quality-Consistent Estimates of International Schooling and Skill Gradients, *Journal of Human Capital*, *3*(2), 107–143, 2009.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Wößmann, Returns to Skills Around the World: Evidence From PIAAC, *European Economic Review*, *73*, 103–130, 2015.

Harmon, C., H. Oosterbeek, and I. Walker, The Returns to Education: Microeconomics, *Journal of Economic Surveys*, *17*(2), 115–156, 2003.

Heckman, J., and P. Carneiro, Human Capital Policy, *National Bureau of Economic Research*, *Working Paper Series No. 9495*, 2003.

Heckman, J. J., and S. Urzua, The Option Value of Educational Choices and the Rate of Return to Educational Choices, *Unpublished manuscript, University of Chicago. Presented at the Cowles Foundation Structural Conference, Yale University*, 2008.

Heckman, J. J., and E. Vytlacil, Structural Equations, Treatment Effects, and Econometric Policy Evaluations, *Econometrica*, *73*(3), 669–738, 2005.

Heckman, J. J., H. Ichimura, and P. E. Todd, Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *Review of Economic Studies*, *64*(4), 605–654, 1997.

Heckman, J. J., R. J. LaLonde, and J. A. Smith, The Economics and Econometrics of Active Labor Market Programs, in *Handbook of Labor Economics*, vol. 3, edited by O. Ashenfelter and D. Card, pp. 1865–2097, Elsevier North-Holland, Amsterdam, 1999.

Heckman, J. J., L. J. Lochner, and P. E. Todd, Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, in *Handbook of the Economics of Education*, vol. 1, edited by E. A. Hanushek and F. Welch, pp. 307–458, Elsevier North-Holland, Amsterdam, 2006a.

Heckman, J. J., J. Stixrud, and S. Urzua, The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior, *Journal of Labor Economics*, *24*(3), 411–482, 2006b.

Holm, A., J. Rasmussen, and P. Thagesen, Effekter på indkomst og beskæftigelse af de amts-kommunale enkeltfagskurser, *Nationaløkonomisk Tidsskrift*, *133*, 138–147, 1995.

Imbens, G. W., Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review, *Review of Economics and Statistics*, *86*(1), 4–29, 2004.

Isacsson, G., Estimates of the Return to Schooling in Sweden from a Large Sample of Twins, *Labour Economics*, *6*(4), 471–489, 1999.

Jackson, C. K., R. C. Johnson, and C. Persico, The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms, *Quarterly Journal of Economics*, *131*(1), 157–218, 2016.

Jacobson, L., R. LaLonde, and D. G. Sullivan, Should We Teach Old Dogs New Tricks? The Impact of Community College Retraining on Older Displaced Workers, *FRB of Chicago Working Paper Series*, *No. 25*, 2003.

Jacobson, L., R. LaLonde, and D. G. Sullivan, Estimating the Returns to Community College Schooling for Displaced Workers, *Journal of Econometrics*, *125*(1), 271–304, 2005.

Jamison, E. A., D. T. Jamison, and E. A. Hanushek, The Effects of Education Quality on Income Growth and Mortality Decline, *Economics of Education Review*, *26*(6), 771–788, 2007.

Jans, A.-C., M. Taslimi, and Y. Thorsén, Jämställdhetsanalys av två arbetsmarknadspolitiska program, *Bilaga till arbetsmarknadsrapport 2014*, 2014.

Jenkins, A., A. Vignoles, A. Wolf, and F. Galindo-Rueda, The Determinants and Labor Market Effects of Lifelong Learning, *Applied Economics*, *35*(16), 1711–1721, 2003.

Juhn, C., K. M. Murphy, and B. Pierce, Wage Inequality and the Rise in Returns to Skill, *Journal of Political Economy*, *101*(3), 410–442, 1993.

Katz, L. F., M. S. Kearney, and D. H. Autor, The Polarization of the US Labor Market, *American Economic Review*, *96*(2), 189–194, 2006.

Kirsch, I., The International Adult Literacy Survey (IALS): Understanding What Was Measured, *ETC Research Report Series*, *2*, 1–61, 2001.

Lafortune, J., J. Rothstein, and D. W. Schanzenbach, School Finance Reform and the Distribution of Student Achievement, *National Bureau of Economic Research*, *Working Paper Series No. 22011*, 2016.

Larsson, A.-C., and A. Eriksson, Den internationella undersökningen av vuxnas färdigheter, *Statistics Sweden Tema: Utbildning*, *2*, 2013.

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

Lechner, M., Earnings and Employment Effects of Continuous Off-The-Job Training in East Germany after Unification, *Journal of Business & Economic Statistics*, *17*(1), 74–90, 1999.

Lechner, M., A Note on the Common Support Problem in Applied Evaluation Studies, *Annales d'Economie et de Statistique*, *91*, 217–235, 2008.

Lee, D. W., and T. H. Lee, Human Capital and Economic Growth Tests Based on the International Evaluation of Educational Achievement, *Economics Letters*, *47*(2), 219–225, 1995.

Levy, F., and R. J. Murnane, U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations, *Journal of Economic Literature*, *30*(3), 1333–1381, 1992.

Levy, P. S., and S. Lemeshow, *Sampling of Populations: Methods and Applications*, John Wiley & Sons, 2013.

Lochner, L., and E. Moretti, The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports, *American Economic Review*, *94*(1), 155–189, 2004.

Lucas, R. E., On the Mechanics of Economic Development, *Journal of Monetary Economics*, *22*(1), 3–42, 1988.

Machin, S., O. Marie, and S. Vujić, The Crime Reducing Effect of Education, *Economic Journal*, *121*(552), 463–484, 2011.

Mankiw, N. G., D. Romer, and D. N. Weil, A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics*, *107*(2), 407–437, 1992.

Meghir, C., and M. Palme, Educational Reform, Ability, and Family Background, *American Economic Review*, *95*(1), 414–424, 2005.

Milligan, K., E. Moretti, and P. Oreopoulos, Does Education Improve Citizenship? Evidence from the United States and the United Kingdom, *Journal of Public Economics*, *88*(9), 1667–1695, 2004.

Mincer, J., The Distribution of Labor Incomes: A Survey With Special Reference to the Human Capital Approach, *Journal of Economic Literature*, *8*(1), 1–26, 1970.

Mincer, J. A., *Schooling, Experience and Earnings*, National Bureau of Economic Research, 1974.

Mislevy, R. J., Randomization-Based Inference About Latent Variables from Complex Samples, *Psychometrika*, *56*(2), 177–196, 1991.

Mislevy, R. J., A. E. Beaton, B. Kaplan, and K. M. Sheehan, Estimating Population Characteristics from Sparse Matrix Samples of Item Responses, *Journal of Educational Measurement*, *29*(2), 133–161, 1992.

Morgan, S., and C. Winship, *Counterfactuals and Causal Inference*, Cambridge University Press, 2007.

Mulligan, C. B., Galton Versus the Human Capital Approach to Inheritance, *Journal of Political Economy*, *107*(6), 184–224, 1999.

Murnane, R. J., J. B. Willett, and F. Levy, The Growing Importance of Cognitive Skills in Wage Determination, *National Bureau of Economic Research, Working Paper Series No. 5076*, 1995.

Murnane, R. J., J. B. Willett, Y. Duhaldeborde, and J. H. Tyler, How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?, *Journal of Policy Analysis and Management*, *19*(4), 547–568, 2000.

Murnane, R. J., J. B. Willett, M. J. Braatz, and Y. Duhaldeborde, Do Different Dimensions of Male High School Students' Skills Predict Labor Market Success a Decade Later? Evidence from the NLSY, *Economics of Education Review*, *20*(4), 311–320, 2001.

Nelson, R. R., and E. S. Phelps, Investment in Humans, Technological Diffusion, and Economic Growth, *American Economic Review*, *56*(1), 69–75, 1966.

OECD, *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, 2013.

OECD, *OECD Employment Report*, OECD Publishing, 2016a.

OECD, *Working Together: Skills and Labor Market Integration of Immigrants and their Children in Sweden*, OECD Publishing, 2016b.

O'Neill, J., The Role of Human Capital in Earnings Differences Between Black and White Men, *Journal of Economic Perspectives*, *4*(4), 25–45, 1990.

Pissarides, C. A., Regular Education as a Tool of Counter-Cyclical Employment Policy, *Nordic Economic Policy Review*, *1*, 209–232, 2011.

Pokropek, A., and M. Jakubowski, Piaactools: Stata® Programs for Statistical Computing Using PIAAC Data, *OECD Publishing*, 2013.

Psacharopoulos, G., Returns to Investment in Education: A Global Update, *World Development*, *22*(9), 1325–1343, 1994.

Rinawi, M., and U. Backes-Gellner, Occupational Skills and the Evolution of Wages, *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung*, 2015.

Romer, P. M., Endogenous Technological Change, *Journal of Political Economy*, *98*(5), 71–102, 1990.

Rosenbaum, P. R., and D. B. Rubin, The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, *70*(1), 41–55, 1983.

Rosenbaum, P. R., and D. B. Rubin, Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score, *American Statistician*, *39*(1), 33–38, 1985.

Rubin, D. B., Matching to Remove Bias in Observational Studies, *Biometrics*, *29*(1), 159–183, 1973.

Rubin, D. B., *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, 1987.

Rubin, D. B., Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation, *Health Services and Outcomes Research Methodology, 2*, 169–188, 2001.

Rubin, D. B., and N. Thomas, Matching Using Estimated Propensity Scores: Relating Theory to Practice, *Biometrics*, *52*(1), 249–264, 1996.

Satherley, P., E. Lawes, and S. Sok, The Adult Literacy and Life Skills (ALL) Survey: Overview and International Comparisons, *Comparative Education Research Unit, Research Division, Australian Ministry of Education*, 2008.

Sianesi, B., and J. V. Reenen, The Returns to Education: Macroeconomics, *Journal of Economic Surveys*, *17*(2), 157–200, 2003.

Smith, H. L., Matching With Multiple Controls to Estimate Treatment Effects in Observational Studies, *Sociological Methodology*, *27*(1), 325–353, 1997.

Smith, J., A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies, *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, *136*(3), 247–268, 2000.

Smith, J. A., and P. E. Todd, Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?, *Journal of Econometrics*, *125*(1), 305–353, 2005.

SOU [2007:18], *Arbetsmarknadsutbildning för bristyrken och insatser för arbetslösa ungdomar*, Statens Offentliga Utredningar, 2007.

Stenberg, A., Short Run Effects on Wage Earnings of the Adult Education Initiative in Sweden, *Umeå Economic Studies*, *No. 593*, 2002.

Stenberg, A., The Adult Education Initiative in Sweden–Second Year Effects on Wage Earnings and the Influence on Branch Mobility, *Umeå Economic Studies*, *No. 608*, 2003.

Stenberg, A., Comprehensive Education for the Unemployed - Evaluating the Effects on Unemployment of the Adult Education Initiative in Sweden, *Labour*, *19*(1), 123–146, 2005.

Stenberg, A., and O. Westerlund, Does Comprehensive Education Work for the Long-Term Unemployed?, *Labour Economics*, *15*(1), 54–67, 2008.

Stenberg, A., and O. Westerlund, Kunskapslyft för arbetslösa genom generell utbildning i stället för yrkes-inriktade program, *Ekonomisk Debatt*, *43*(1), 24–31, 2015.

Stuart, E. A., Matching Methods for Causal Inference: A Review and a Look Forward, *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *25*(1), 1, 2010.

Tomz, M., G. King, and L. Zeng, ReLogit: Rare Events Logistic Regression, *Journal of Statistical Software*, *8*(2), 2003.

Von Davier, M., E. Gonzalez, and R. Mislevy, What Are Plausible Values and Why Are They Useful, *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, *2*, 9–36, 2009.

Wallenstein, S., C. L. Zucker, and J. L. Fleiss, Some Statistical Methods Useful in Circulation Research, *Circulation Research*, *47*(1), 1–9, 1980.

Westerlund, O., Inskrivningstid vid arbetsförmedlingen efter avslutade studier: deltagare i kunskapslyftet och arbetsmarknadsutbildning 1997, *Umeå Economic Studies*, *No. 526*, 2000.

Wooldridge, J., *Introductory Econometrics: A Modern Approach*, Nelson Education, 2015.

Wu, M., The Role of Plausible Values in Large-Scale Surveys, *Studies in Educational Evaluation*, *31*(2), 114–128, 2005.

Zhao, Z., Sensitivity of Propensity Score Methods to the Specifications, *Economics Letters*, *98*(3), 309–319, 2008.

9 Appendix

9.1 Project Review

This project has been conducted with financial support from IFAU to pay for the data processing that Statistics Sweden conducted and for the two binary variables regarding participation in the AEI and Komvux to be added to the PIAAC data. The written application in order to attain this data was filed by our supervisor Erik Mellander at IFAU, and was approved on March 2nd, 2016. The data was made accessible remotely through Statistics Sweden's MONA database. In writing this thesis, we wrote our code for the analysis based on the PIAAC public use file, with the addition of two noise-induced dummies related to AEI and Komvux participation. This code was then sent to IFAU for running against the restricted use file, which included the same two binary variables, but without noise. This is similar to the approach that non-American researchers must undertake when analyzing U.S. data. Non-Americans are allowed to access a U.S. PIAAC synthetic restricted use file to prepare code for the analysis, which can then be sent in for analysis on the actual U.S. RUF file.

9.2 Eligibility for Adult Secondary Education and the AEI

Applicants for adult secondary education were accepted based on two statues with partially different orders of priority. The order of priority is stated in the statue of municipal adult education SFS 1992:403:

> *The selection should particularly consider if an applicant*
>
> – *has a short previous education,*
> – *needs the education for planned or ongoing professional duty or is faced by a career choice,*
> – *needs the education to supplement a reduced program from the upper secondary school or for other supplementing eligibility, or*
> – *can complete the studies that the applicant has started according to an established study plan.*

The statue of state funds for specific investments in adult education SFS 1998:276 (Kunskapslyftsförordningen) gives priority to some applicants:

> *§ 9 ...priority should be given in the order specified in point 1-3 to an applicant missing or with insufficient skills in such knowledge that is required for a leaving certificate from adult upper secondary education or corresponding education, if the applicant:*
>
> *1 is unemployed or participates in a labor market program and is a registered job seeker within the public employment office according to the rules set down by the National Labor Market Board,*
>
> *2 is employed and the employer by an arrangement with the local labor organization has committed to employing someone who, in at least the same proportion as the applying employed, is long-term unemployed and who has been assigned by the public employment office, or*
>
> *3 is laid off from his or her work place or is threatened by unemployment.*
>
> *10 § Regarding selection between applicants within any of the groups stated in § 9 or between other applicants the regulations of selection in the statue of municipal adult education or in the statue of adult state schools should apply.*

9.3 Summary Statistics

This section provides summary statistcs for the covarite vector X_i. For definitions of the included covariates, see section 9.10.

Table 11: Summary Statistics for X_i - Full Sample Excluding AEI Participants

Variable	Observations	Mean	Standard Deviation
Gender	3191	1.459	0.498
Highest education - mother	3191	1.812	0.838
Highest education - father	3191	1.786	0.812
Parents qualification	3191	1.969	0.917
Education level	3191	0.810	1.023
Low education	3191	0.611	0.856
Immigrant	3191	1.176	0.380
Child at home	3191	0.178	0.383
Age category 1	3191	0.564	0.496
Age category 2	3191	0.066	0.248
Age category 3	3191	0.067	0.251
Age category 4	3191	0.068	0.251
Age category 5	3191	0.065	0.246
Age category 6	3191	0.067	0.250
Age category 7	3191	0.103	0.304

Table 12: Summary Statistics for X_i - AEI Participants

Variable	Observations	Mean	Standard Deviation
Gender	170	1.718	0.451
Mother - highest education	170	1.364	0.671
Father - highest education	170	1.442	0.713
Parents qualification	170	1.547	0.871
Education level	170	0.941	1.053
Low education	170	0.700	0.883
Immigrant	170	1.700	0.584
Child at home	170	0.400	0.491
Age category 1	170	0.094	0.293
Age category 2	170	0.165	0.372
Age category 3	170	0.171	0.377
Age category 4	170	0.153	0.361
Age category 5	170	0.129	0.337
Age category 6	170	0.182	0.387
Age category 7	170	0.106	0.309

9.4 Proof: Propensity Score Theorem

Proof: Show that $P[D_i = 1|Y_{ji}, p(X_i)]$ does not depend on Y_{ji} for $j = 0, 1$:

$$\begin{aligned} P[D_i = 1|Y_{ji}, p(X_i)] &= E[D_i|Y_{ji}, p(X_i)] \\ &= E\{E[D_i|Y_{ji}, p(X_i), X_i]|Y_{ji}, p(X_i)\} \\ &= E\{E[D_i|Y_{ji}, X_i]|Y_{ji}, p(X_i)\} \\ &= E\{E[D_i|X_i]|Y_{ji}, p(X_i)\}\text{, by the CIA.} \end{aligned}$$

However, $E\{E[D_i|X_i]|Y_{ji}, p(X_i)\} = E\{p(X_i)|Y_{ji}, p(X_i)\}$, which is $= p(X_i)$ (*Angrist and Pischke* [2008]).

9.5 Balancing Test after Matching

Table 13: Balancing Comparison - After Matching

	Mean			t-test		
Variable	Treated	Matched	Bias (%)	t	p-value	$\frac{V(t)}{V(c)}$
Gender	1.72	1.70	3.70	0.36	0.72	0.94
Highest education - mother	1.36	1.34	2.70	0.28	0.78	0.99
Highest education - father	1.44	1.46	-2.00	-0.20	0.85	0.92
Parents qualification	1.55	1.53	2.00	0.19	0.85	1.07
Education level	0.94	0.93	1.10	0.11	0.92	1.11
Low education	0.70	0.72	-2.70	-0.25	0.80	1.07
Immigrant	1.70	1.70	0.00	0.00	1.00	0.97
Children at home	0.40	0.40	0.00	0.00	1.00	1.00
Age category 1	0.09	0.08	2.90	0.38	0.70	1.15
Age category 2	0.17	0.15	5.60	0.45	0.66	1.14
Age category 3	0.17	0.19	-5.50	-0.42	0.67	0.92
Age category 4	0.15	0.12	11.30	0.95	0.34	1.23
Age category 5	0.13	0.16	-10.00	-0.77	0.44	0.85
Age category 6	0.18	0.18	0.00	0.00	1.00	1.00
Age category 7	0.11	0.12	-5.80	-0.51	0.61	0.87
			Statistics			
Pseudo R^2	LR chi-2	p>chi-2	MeanBias	MedBias	B	R
0.006	3.01	0.99	3.70	2.70	18.80	1.12

9.6 Bonferroni Correction

The Bonferroni correction is a multiple comparison correction which applies if several dependent or independent statistical tests are performed simultaneously. In such cases, the significance level α is lowered to account for the number of comparisons being done, to avoid spurious positives. As outlined in *Wallenstein et al.* [1980], the conservative critical value for the t-statistics is then attained from the t-distribution, but using a significance level of α/n, where n is the number of comparisons between groups to be done. Deriving the Bonferroni correction follows from Boole's inequality, see e.g. *Goeman and Solari* [2014]).

For an intuitive explanation of why Bonferroni correction needs to be applied, consider the following example outlined in *Bland and Altman* [1995]. If performing a large number of significance tests, interpretation will become increasingly difficult, because if continuing testing long enough, something will eventually be found "significant". Thus, one should be cautious in attaching too much consideration into a single significant result among a range of non-significant results. It could simply be the one in twenty which we anticipate merely by chance. For instance, consider testing a null hypothesis which is, in fact, true, and using 0.05 as the critical significance level. The probability of ending up with a correct conclusion (in this case not significant) is then 0.95. Next, testing two independent nulls, neither test will be significant at the probability of $0.95 * 0.95 = 0.90$. Ultimately, in the case of testing twenty such hypotheses, neither test will be significant at $0.95^{20} = 0.36$ probability. Thus, the probability of getting at least one significant result is $1 - 0.36 = 0.64$. With other words, in the end we are more likely to find a significant result, than we are not.

9.7 OLS Subgroup Estimates

Table 14: OLS Estimates on PIAAC Scores - Gender Differences

	Female			Male		
Variable	**LIT**	**NUM**	**PSL**	**LIT**	**NUM**	**PSL**
AEI-participation	−3.59 (4.54)	−3.08 (4.98)	−6.06 (4.51)	−8.41 (7.24)	−11.05 (7.43)	−5.13 (8.00)
Highest education - mother	6.21 * (3.17)	6.78 * (3.52)	7.37 ** (2.95)	6.11 *** (2.40)	5.01 * (2.74)	6.27 *** (2.46)
Highest education - father	−0.08 (2.17)	0.32 (2.44)	−0.84 (1.71)	3.65 (2.33)	3.42 (2.67)	5.86 *** (2.13)
Parents qualification	10.56 *** (3.30)	9.33 *** (3.31)	5.81 ** (2.74)	5.09 * (2.86)	5.86 * (3.50)	2.36 (2.71)
Education level	3.57 (2.58)	3.66 (2.85)	4.90 ** (2.07)	5.86 *** (1.61)	8.69 *** (1.69)	6.41 *** (1.57)
Low education	−4.47 * (2.55)	−5.99 ** (2.71)	−7.18 *** (2.36)	−6.48 ** (1.72)	−9.09 *** (1.86)	−10.15 *** (1.76)
Immigrant	19.87 *** (2.46)	21.19 *** (2.67)	11.25 *** (1.85)	16.00 *** (2.10)	16.42 *** (2.29)	8.86 *** (1.81)
Children at home	−4.79 (4.35)	−3.74 (4.22)	−0.68 (4.42)	2.64 (3.11)	2.33 (3.71)	−0.97 (3.43)
Age category 1	13.12 (136.47)	6.20 (111.81)	43.00 (189.28)	10.47 (24.73)	8.28 (18.52)	25.02 (46.16)
Age category 2	16.90 (136.42)	9.06 (111.78)	38.94 (189.06)	8.61 (25.47)	8.87 (19.13)	14.50 (46.58)
Age category 3	14.34 (136.70)	9.34 (112.34)	27.85 (188.10)	8.03 (24.79)	6.89 (19.04)	8.87 (46.65)
Age category 4	21.67 (136.51)	17.67 (111.76)	28.51 (188.20)	1.02 (24.81)	0.89 (18.61)	5.47 (46.20)
Age category 5	1.71 (135.96)	−2.00 (111.29)	13.49 (188.84)	0.00 (24.38)	0.00 (18.01)	0.00 (46.41)
Age category 6	0.00 (135.33)	0.00 (110.69)	9.90 (188.77)	−2.22 (24.31)	−0.29 (17.90)	−7.40 (46.68)
Age category 7	−11.44 (136.50)	−6.91 (112.08)	0.00 (188.32)	−12.13 (24.60)	−13.55 (45.89)	−5.21 (4.26)
R-squared	0.20	0.15	0.27	0.14	0.11	0.28
Number of observations	1774	1774	1774	1587	1587	1587

Standard errors reported in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 15: OLS Estimates on PIAAC Scores - Education Differences

	High			Low		
Variable	**LIT**	**NUM**	**PSL**	**LIT**	**NUM**	**PSL**
AEI-participation	−0.08	−0.65	1.17	−6.30	−7.22	−6.97
	(7.76)	(7.27)	(8.18)	(4.66)	(4.94)	(4.94)
Gender	−2.70	−15.48***	0.99	−0.65	−9.95***	−3.32
	(4.01)	(4.48)	(3.48)	(2.06)	(2.08)	(1.95)
Highest education - mother	1.14	1.80	7.92**	7.00***	6.37***	6.06***
	(4.17	(4.96)	(3.63)	(2.06)	(2.22)	(2.26)
Highest education - father	−0.79	−0.01	4.01	2.14	2.12	2.12
	(3.82)	(4.42)	(3.62)	(1.85)	(1.95)	(1.60)
Parents qualification	4.41	2.49	−2.13	8.50***	8.75***	5.80**
	(4.23)	(4.73)	(4.15)	(2.47)	(2.57)	(2.36)
Low education	−10.42***	−14.82***	−13.39***	12.63***	17.09***	−9.27***
	(2.16)	(2.35)	(1.94)	(5.10)	(5.72)	(1.34)
Immigrant	10.97***	10.79***	2.99	18.99***	19.98***	11.13***
	(3.29)	(4.06)	(3.01)	(1.72)	(2.02)	(1.52)
Children at home	3.80	4.28	5.98*	−3.02	−3.32	−4.72
	(4.14)	(4.47)	(3.55)	(3.10)	(3.54)	(3.63)
Age category 1	18.88	10.91	34.10	−1.73	−4.48	32.71
	(119.67)	(71.47)	(210.78)	(68.42)	(59.11)	(142.95)
Age category 2	19.01	14.60	24.36	−0.94	−2.74	26.23
	(119.79)	(71.50)	(210.78)	(68.12)	(58.58)	(142.96)
Age category 3	11.24	4.89	12.61	1.97	2.24	21.27
	(120.22)	(72.13)	(210.98)	(69.24)	(59.84)	(143.27)
Age category 4	10.90	7.59	8.52	0.00	0.00	21.42
	(120.18)	(72.18)	(210.46)	(68.15)	(58.55)	(142.67)
Age category 5	0.58	−2.60	2.75	−6.35	−6.13	9.29
	(120.18)	(71.89)	(209.98)	(68.07)	(58.36)	(142.70)
Age category 6	0.00	0.00	0.00	−11.14	−9.28	0.00
	(119.77)	(71.60)	(210.05)	(68.97)	(59.45)	(142.12)
Age category 7	−5.09	−1.96	−5.02	−22.93	−15.60	−8.74
	(120.50)	(72.68)	(211.23)	(67.97)	(58.51)	(142.77)
R-squared	0.13	0.15	0.23	0.18	0.15	0.27
Number of observations	3361	3361	3361	3361	3361	3361

Standard errors reported in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

9.8 Jackknife Resampling

PIAAC uses a statistical method called Jackknife resampling. It is similar to bootstrapping in that it involves resampling, but instead of sampling with replacement, the method samples without replacement. The Jackknife samples are selected by taking the original data vector and deleting one observation from the set. Thus, there are n unique Jackknife sample and the ith Jackknife sample vector is defined as:

$$X_{[i]} = \{X_1, X_2, ..., X_{i+1}, ..., X_{n-1,X_n}\}$$

The ith Jackknife Replicate is defined as the value of the estimator s(.) evaluated at the ith Jackknife sample.

$$\hat{\theta}_{(i)} := s(X_{[i]})$$

The Jackknife Standard Error is defined as:

$$SE(\hat{\theta}_{jack}) = \left\{ \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}^2_{(.)} \right\}^2,$$

where $\hat{\theta}_{(.)}$ is the empirical average of the Jackknife replicates:

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}$$

9.9 Further Notes on PIAAC

Table 16: Participation Rates for Original RTB Sample

Group	Share responding (%)
Total	46.5
Female	46.9
Male	46.1
16-25 years	46.9
26-35 years	44.3
36-45 years	44.0
46-55 years	47.5
56-65 years	50.0
Lives in Stockholm	38.9
Lives in other regions	48.7
Pre upper-secondary school education	41.8
Upper-secondary school education	42.5
Post upper-secondary school education	55.9
Born in Sweden	47.0
Born abroad	44.2

Source: *Larsson and Eriksson* [2013]

Table 17: Mean Proficiency in Literacy and Numeracy, and the Percentage Scoring at Level 2 or 3 in Problem Solving in Technology Rich Environments

Country	Literacy	Numeracy	Problem-solving
Australia	280	268	38
Austria	269	275	32
Canada	273	265	37
Cyprus	269	265	-
Czech Republic	274	276	33
Denmark	271	278	39
England/N. Ireland (UK)	272	262	35
Estonia	276	273	28
Finland	288	282	42
Flanders (Belgium)	275	280	35
France	262	254	-
Germany	270	272	36
Ireland	267	256	25
Italy	250	247	-
Japan	296	288	35
Korea	273	263	30
Netherlands	284	280	42
Norway	278	278	41
Poland	267	260	19
Slovak Republic	274	276	26
Spain	252	246	-
Sweden	**279**	**279**	**44**
United States	270	253	31
Average	**273**	**269**	**34**

Source: *OECD* [2013]

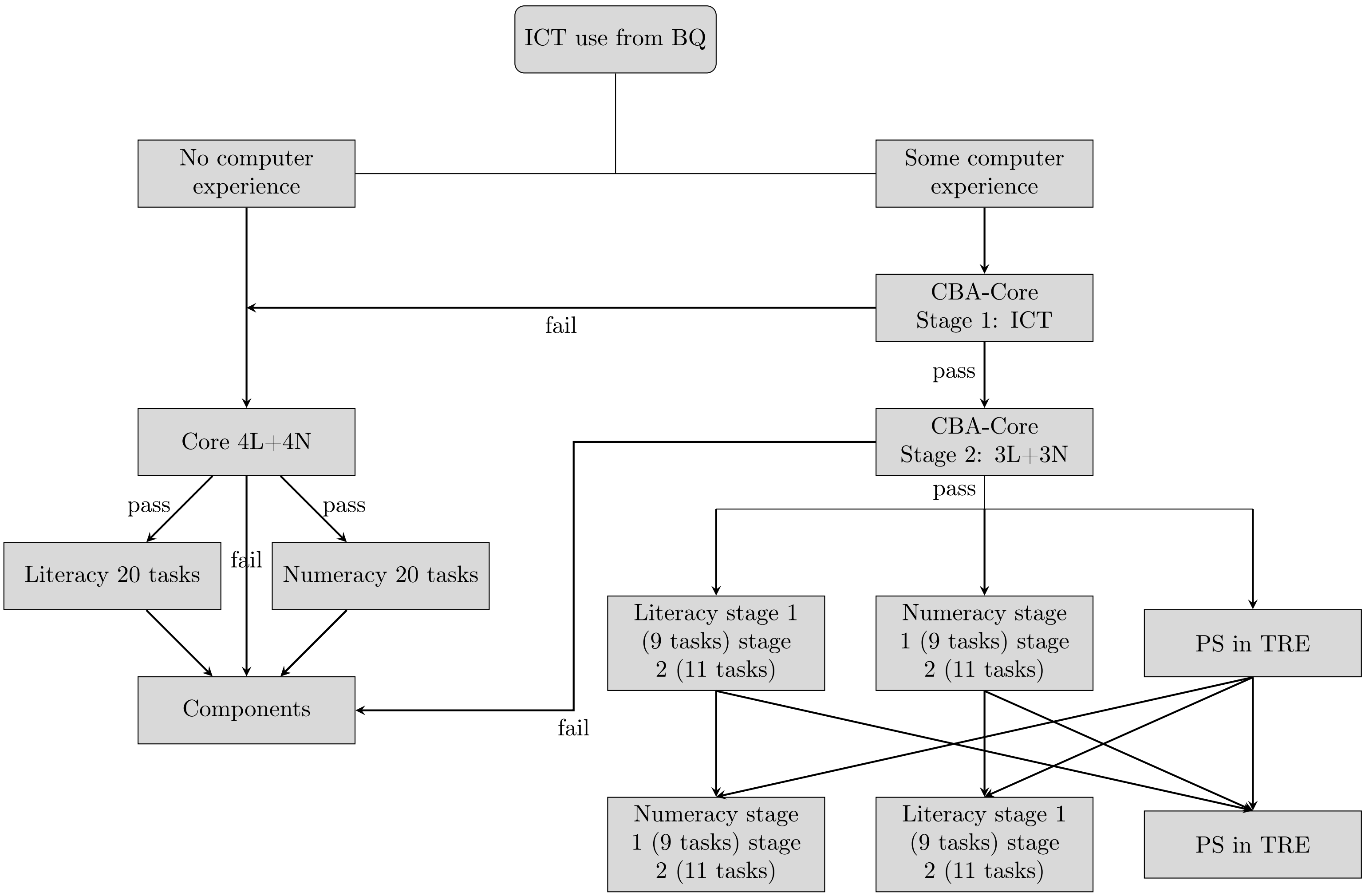

Figure 4: PIAAC Main Study Assessment Design

9.10 Covariate Definitions

Table 18: Detailed Description of Covariates Included in Vector X_i

Variable	Definition
Gender	Binary variable: If male then 1, if female then 2
Immigrant	Binary variable: If first or second generation immigrant, then 1, otherwise 0
Parents qualification	Categorical variable: If neither parent has attained upper secondary, then 1, if at least one parent has attained secondary and post-secondary (non-tertiary), then 2, if at least one parent has attained tertiary, then 3
Highest education - mother	Categorical variable: If observation's mother or female guardian's highest educational qualification corresponds at maximum to ISCED 2, then 1, to ISCED 4, then 2, to ISCED 6, then 3
Highest education - father	Categorical variable: If observation's father or male guardian's highest educational qualification corresponds at maximum to ISCED 2, then 1, to ISCED 4, then 2, to ISCED 6, then 3
Education level	Categorical variable: If observation by the year 1997 has attained 0 to 9 years of schooling, then 1, 10-12, then 2, more than 12, then 3, otherwise 0
Low education	Binary variable: If observation by the year 1997 has attained no more than an upper secondary school degree, then 1, otherwise 0
Children at home	Binary variable: If observation by the year 1997 has a child between the age of 0 to 18, then 1, otherwise 0
Age category 1	Binary variable: If $35 < age <= 39$, then 1, otherwise 0
Age category 2	Binary variable: If $39 < age <= 43$, then 1, otherwise 0
Age category 3	Binary variable: If $43 < age <= 47$, then 1, otherwise 0
Age category 4	Binary variable: If $47 < age <= 51$, then 1, otherwise 0
Age category 5	Binary variable: If $51 < age <= 55$, then 1, otherwise 0
Age category 6	Binary variable: If $55 < age <= 60$, then 1, otherwise 0
Age category 7	Binary variable: If $60 < age <= 65$, then 1, otherwise 0

9.11 Outcome Variables Definitions

Table 19: Detailed Description of the Outcome Variables

Variable	Definition
Labor market outcome	
Employed	Employment status: 1 if employed, 2 if unemployed, 3 if out of labor force
Paid work last 5 years	Binary: 0 if no, 1 if yes
Working Physically	Incidence of working long hours: Increasing scale from 1 to 5
Hourly earnings interval	In deciles: Increasing scale from 1 to 10
Hours work	Reported in intervals: Increasing scale from 1 to 6
Yearly earnings	Percentile rank categories: Increasing scale from 1 to 6
Years with paid work	Reported in integers
Job satisfaction	Decreasing scale from 1 to 5
Education outcome	
Education level	Highest formal education obtained: Increasing scale from 1 to 9
Top qualification	Highest qualification (ISCED levels): Increasing scale from 1 to 15
Age finishing top qualification	Reported in integers
Total years of education	Reported in integers
Learning activities last year	Number of learning activities: Increasing scale from 1 to 5
Need more training	Binary: 1 if yes, 2 if no
Social outcome	
Health	Decreasing scale from 1 to 5
Trust few people	Decreasing scale from 1 to 5
No government influence	Decreasing scale from 1 to 5
Partner	1 if yes, 2 if no
Children	Number of children: Reported in integers
Born in country	1 if yes, 2 if no
Read books	Increasing scale from 1 to 5
Read news	Increasing scale from 1 to 5
Volunteer work	Increasing scale from 1 to 5

9.12 Plausible Values

To understand plausible values, one may consider a simple example first introduced by *Wu* [2005][61]:

"[...] suppose that there is an interest in knowing the percentage of people who are over 60 living in a community. If a random sample of people is selected from the community, the percentage p, of people over 60 in the sample will be an unbiased estimate of the percentage of over 60s in the population. The standard error of the estimate can be computed using the formula $\sqrt{\frac{p(1-p)}{n}}$, which is referred to as sampling error. Now, suppose there is an interest in knowing what percentage of people in this community cannot speak English. While it is relatively easy to determine a person's age, it is not straightforward to determine if each person can speak English, since people's English proficiency is on a continuum: some can speak a few words; some can speak a few sentences; some can speak well enough to be understood, but with many mistakes. So, unlike the measure of people's age, a clear definition of 'the ability to speak English' will need to be made. Based on this definition, some assessment will have to be carried out to determine if a person 'can speak English'. For practical reasons, the assessment can only sample a small part of the English language so as not to place too much burden on each person's time. Consequently, the result of the assessment will contain some uncertainty. This uncertainty is referred to as measurement error. The percentage of people who cannot speak English in the community can be estimated by the percentage p, of a random sample from the community who failed the assessment. The standard error of this estimate, however, is expected to be larger than $\sqrt{\frac{p(1-p)}{n}}$, since there is uncertainty associated with the measurement of each person, in addition to sampling error. One way to express the degree of uncertainty of measurement at the individual level is to provide several scores for each individual to reflect the magnitude of error of the individual's estimate. If measurement error is small, then multiple scores for an individual will be close together. If measurement error is large, then multiple scores for an individual will be far apart. These multiple scores for an individual, sometimes known as multiple imputations, are *plausible values*."

More formally, plausible values can be thought of as representing the range of

[61] The quote is from page 115.

abilities that an individual might have, given the individual's item responses. Instead of providing a point estimate for the individuals' ability parameter, θ, a range of possible values for an individuals θ, with associated probabilities for each of these values, are estimated. Plausible values are random draws from this estimated distribution for an individual's θ. This distribution is referred to as the posterior distribution for an individual (*Wu* [2005]).

Given an item response pattern x, and ability θ, let $f(x|\theta)$ be the item response probability. Further, assume that θ comes from a normal distribution $g(\theta) \sim \mathcal{N}(\mu, \sigma^2)$. The function $f(x|\theta)$ is referred to as the item response model, and $g(\theta)$ the population model. It can be shown that, the posterior distribution, $h(\theta|x)$, is given by:

$$h(\theta|x) = \frac{f(x|\theta)g(\theta)}{\int f(x|\theta)g(\theta)d\theta}$$

If the item response pattern of an individual is x, then the posterior θ distribution is given by $h(\theta|x)$. Plausible values for an individual with item response pattern x are random draws from the probability distribution with density $h(\theta|x)$. Therefore, plausible values not only provide information about an individual's ability estimate, but also about the uncertainty associated with the estimate. If many plausible values are drawn from a student's posterior distribution $h(\theta|x)$, these plausible values will form an empirical distribution for $h(\theta|x)$. Plausible values can be used to estimate population characteristics, and they do a better job than point estimates of abilities (*Wu* [2005] and *Von Davier et al.* [2009]).