*[preliminary draft - December 2014]*

# The returns to foreign education and the effects of a local signal on early labour market outcomes

Massimiliano Tani

*School of Business, UNSW Canberra and IZA Bonn*

**Abstract**

This paper studies the puzzling fact that even countries adopting selective migration policies appear to systematically under-utilise the human capital of immigrants after they settle. The analysis sets this puzzle as a case of statistical discrimination, whereby host country employers have an imperfect understanding of foreign education as a productivity signal. It develops a theoretical model that first traces the effects of this uncertainty in determining the wage and the quality of the education-occupation match in one's job. It then explores the consequences of adding an institutional 'local' signal to reduce the employers' informational asymmetry, and presents the case of an official assessment of foreign qualifications implemented in Australia on migrants settling in the mid- and late 1990s. The assessment was voluntary up to the early 2000s, after which it became mandatory. Using data from the Longitudinal Survey of Immigrants to Australia (LSIA) and controlling for the endogeneity of the choice of undertaking the assessment, the empirical analysis finds marked improvements in both wage and probability of a correct education-occupation match in the early experience of Australian labour market. The results are robust to several specifications and instruments controlling the endogeneity of the assessment choice. The findings support that the local signal eases the realization of immigrants' true productivity by effectively speeding up domestic employers' learning. They also suggest that a relatively simple tool such as the assessment of foreign qualification may indeed fasten and improve the international transferability of human capital, with positive spillovers on both private and social returns to immigration.

JEL classification: C34, J24, J61
Keywords: immigration, over-education, immigration policy, statistical discrimination

School of Business, UNSW Canberra

Northcott Drive
Campbell, ACT 2612
Australia
Email: m.tani@adfa.edu.au
Tel: +61 2 6268 8831

## 1. Introduction

Despite a fiercer international competition for skilled workers the returns to foreign education are, for many immigrants, below those of comparably educated natives even in host countries where selective immigration policies are in place. As an example, the incidence of skill mismatch (over-education) amongst foreign workers in Australia, Canada and New Zealand is as high as 40-50% versus 10-20% amongst comparable natives (e.g. Green *et al*, 2007; Wald and Fang, 2008; Poot and Stillman, 2007). This evidence is puzzling. On the one side immigration policies select potential immigrants on the basis of their human capital. On the other, the destination country's employers appear to systematically under-utilise the skills of those who have been selected.

Existing work has highlighted that the initial penalty in the returns to foreign education tends to reduce over-time thanks to the acquisition of local experience or further educational investments in the host country, and suggested a number of potential causes of this common fact. For instance, host country employers' may have preferences skewed in favour of domestic workers ('taste' discrimination, as in Battu and Sloane, 1999). Or immigrants may have an inadequate knowledge of the host country language (Dustmann, 1999), preventing them to access jobs for which they would otherwise qualify, or a well-developed social network to provide information about available jobs (Piracha et al, 2014). Alternatively, human capital may simply be imperfectly transferable across countries as economic, historical and cultural conditions differ around the globe (Chiswick and Miller, 2009; Basilio and Bauer, 2010).

A common feature of these explanations is the assumption that employers have always complete information about the productivity of their prospective employees, who, as a result, have no incentives, besides personal preferences, to provide additional signals of their ability. This situation may accurately portray employer-sponsored ('demand-driven') migration.

However, in the common case of supply-driven migration[1], employers' information is likely to be incomplete, and a different scenario can emerge. An employer may not know how to interpret the productivity signal of a degree completed abroad, especially if it is from a country with different language and culture for his/her own. As a result, s/he would weigh the relevance of education in forming a job offer with other observable indicators of productivity (e.g. physical characteristics such as height – Wang, 2015) and the average educational characteristics of group the to which the immigrant applicant belongs. The job offer proposed would then sub-optimally reflect the ability of the immigrant, and be a case of 'statistical' rather than 'taste' discrimination, as the employer's incomplete information underpins the penalty attributed to the immigrant's human capital vis-à-vis that of a domestic, but otherwise equally productive, job seeker.

While the true productivity of the foreign educated immigrant may eventually emerge and be correctly rewarded (Altonji and Pietter, 2001; Lange, 2007), the opportunity cost of such delayed recognition can be high because, in the meanwhile, the immigrant receives a lower return to his/her human capital, the employer under-uses skills that are already available, and the host country's society receives less tax and consumption spending from the immigrant's lower earnings.

Unfortunately the literature has been silent about what host country policymakers can do to reduce the opportunity cost of the delayed recognition beyond the usefulness of introducing non-discriminatory laws (Lundberg and Startz, 1983) and effectively leave the rest to market forces. This paper advances the status quo on possible policy initiatives by illustrating the case of a tool reducing the informational asymmetry between an immigrant's foreign qualifications and their evaluation by a host country's employer. The policy is the official

---

[1] This includes migration initiated by an individual optimising his/her choice set and moving independently or to reunify with family, temporarily or permanently, and refugee migration.

assessment of an immigrant's foreign qualifications by the host country's public institutions. In particular, I present the case of an Australian initiative, which until the early 2000s gave an admitted immigrant the option to have his/her foreign qualifications officially assessed by the Overseas Qualifications Units (OQU) of the Department of Immigration and Citizenship (DIAC) and its authorized agents. The OQU's objective was "to assist migrants to obtain recognition of their overseas gained skills and qualifications" especially with respect to "statements of educational comparison for qualifications obtained overseas; and information on where and how to obtain specific occupational assessments and which occupations have licensing and regulatory requirements[2]." The OQU covered both education and relevant work experience, but it primarily focused on the completion of formal education. Since the early 2000s, Australia has made such assessment mandatory.

The paper focuses on the difference that such an instrument made to the returns to foreign education within the first few months from settlement in Australia using a database that covers the crucial first 2-3 years post-migration at that time (Longitudinal Survey of Immigrants to Australia – LSIA). The results suggest that when foreign qualifications are officially assessed, there are marked improvements in both wage and probability of a correct educational-occupational job match. The results are robust to several specifications and different instruments controlling for the endogeneity of the assessment choice.

The results support that the availability of an official local signal of human capital content in the host country raises migrants' wages and reduces their incidence of over-education, with positive effects on the private and social returns of immigration. Undertaking the assessment may lend initial jobs with wages about a third higher than those accessed without it, and cut in half the probability of being over-educated. The results also suggest that the lower penalty

[2] http://www.immi.gov.au/asri/os-qual-units.htm - accessed 14 January 2014

associated with undertaking the assessment becomes less valuable over time, after migrants' productivity has been observed, in line with what predicted by the statistical discrimination model.

If there is persistence in employment, the magnitude of the marginal effects obtained over an individual's life cycle is staggering. Instituting a relative simple mechanism to formally recognize foreign qualifications, as was the case in Australia, emerges as an effective tool to improve and speed up the international transferability of human capital, with positive economic and wellbeing spillovers amongst both migrants and host societies.

The paper is organised as follows: Section 2 provides a brief literature background. Section 3 presents a simple economic model. Section 4 presents the instrument and the data. Section 5 discusses the empirical strategy. Section 6 presents the results. Section 7 concludes.

## 2. Literature

Early work on statistical discrimination in the labour market originates in the early 1970s and 1980s, and it is mostly concerned about wage differences related to employers' perceptions of productivity amongst individuals of different race or gender (Arrow, 1973; Phelps, 1972), especially in presence of noisy signals of ability (Cain and Aigner, 1977), and their negative and persistent consequences on human capital investment (Lundberg and Startz, 1983). The emergence of statistical discrimination as a rational response to imperfect information is mitigated by the fact that a dynamic setting *per se* ensures future convergence in wages between statistically discriminated groups, as employers eventually learn their employees' true productivity, though the time lag for this to occur may be take months or years.

Although theoretically appealing, the literature on statistical discrimination in the labour market had little empirical applications until the problem of disentangling statistical discrimination from other confounding explanations was resolved. The typical test nowadays focuses on the nature of the interaction between characteristics observed by firms, those observed only by the econometrician, and time (Farber and Gibbons, 1996; Altonji and Pierret, 2001). As employers learn, by repeatedly observing the productivity of their workers, the effect of formal schooling should decline over time, while that of productivity-related variables that are unobserved by employers, such as previous ability test scores, should correspondingly increase.

Applications of this methodology have been carried out in a variety of context and their results have been used to explain lower wages in presence of race (Oettinger, 1996), skill (Mansour, 2012), and age differences (Altonji and Pierret, 2001), as well as differences in educational level (Arcidiacono, Bayer and Hizmo, 2009), university prestige (Bordon and Braga, 2013), and use of health services (Balsa and McGuire, 2001).

To date, there is only rare work studying statistical discrimination in the case of immigrants' human capital despite the large interest surrounding the topic (Artuc et al 2015; Docquier et al 2013). Even scarcer is evidence of the labour market effect of complementing the informational value of education completed abroad with an official signal acquired in the host country.

The closest study to this paper is Siniver (2011), who tests the statistical discrimination of immigrants' physicians in Israel by focusing on the exogenous introduction of a mandatory 'accuracy test' in 1999 for those who intend to work in this profession. By comparing the earnings' difference between natives and foreign physicians entering Israel pre- and post-1999, Siniver finds evidence of statistical discrimination, with statistically significantly higher earnings for those undertaking the test.

Differently from Siniver, I analyse the effect of the choice to obtain an official assessment of their foreign qualifications made by a heterogeneous group of immigrants holding various levels of human capital arriving to Australia at approximately the same time. To better frame why the assessment may affect the labour market outcome of the immigrant, it is useful to develop a simple theoretical approach, based on Lundberg and Startz (1983) before presenting the empirical analysis.

### 3. A theoretical model

Consider an immigrant $i$ completing his/her education in country $j$. Migrants have innate productive abilities $a_{ij}$, which are distributed randomly in the population and acquired productive abilities $X_{ij}$, which include education and labour market experience. Each immigrant worker has productivity $MP_{ij}$ equal to:

$$MP_{ij} = \left(a_{ij} + bX_{ij}\right) \qquad (1)$$

where for simplicity $a_{ij} \sim N(\bar{a}, \sigma_a^2)$ and $X_{ij}$ is defined as:

$$X_{ij} = \rho_0 + \rho_a a_{ij} + \mu_{ij} \qquad (2)$$

where $\rho_0, \rho_a$ are parameters independent of $\mu_{ij} \sim N(0, \sigma_\mu^2)$.

*Case 1: no local signalling available (normal case)*

In the traditional case of statistical discrimination, host country employers use the average and variance of the distribution of innate and acquired characteristics in the population as a reference to assign migrants to jobs, but they cannot observe the individual marginal product $MP_{ij}$. However, they observe an imperfect indicator of foreign education:

$$\tilde{X}_{ij} = X_{ij} + \varepsilon_{ij} \qquad (3)$$

where the i.i.d. error term $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. The observed imperfect signal of schooling implies that the employer offers a wage equal to the workers' expected marginal productivity conditional on observed education, which can be estimated by Ordinary Least Squares (OLS) by adding a random a normally distributed error $\tau_{ij}$ accounting for measurement errors using the functional form:

$$w_{ij} = E\left(\left(a_{ij} + bX_{ij}\right)|\tilde{X}_{ij}\right) = E\left(a_{ij} + bX_{ij}\right) + \frac{Cov\left(\left(a_{ij}+bX_{ij}\right),\tilde{X}_{ij}\right)}{Var\left(\tilde{X}_{ij}\right)}\left(\tilde{X}_{ij} - \overline{\tilde{X}}\right) \qquad (4')$$

The OLS estimate of the return to observed education is:

$$\hat{\beta}_{\tilde{X}} = \frac{\sigma_a^2\rho_a(1+b\rho_a)+\sigma_\mu^2 b}{\sigma_a^2\rho_a^2+\sigma_\mu^2+\sigma_\varepsilon^2} \qquad (4'')$$

Statistical discrimination arises because the higher is the noise of the human capital indicator (high $\sigma_\varepsilon^2$ in (3)), the lower is $\hat{\beta}_{\tilde{X}}$: $\frac{\partial\hat{\beta}_{\tilde{X}}}{\partial\sigma_\varepsilon^2} < 0$.

*Case 2: local signalling available (Australian case)*

If each individual, knowing his/her productivity, has the possibility of undertaking an assessment of his/her foreign qualifications then he/she will undertake the assessment if deemed to bring sufficient benefits to his/her labour market outcome post-migration. Such belief is likely the result of evaluating the information collected on labour market prospects with and without the assessment in the various host country places where the migrant considers moving into in light of his/her own ability and circumstances: a migrant from a place that is culturally and linguistically close to the host country will have limited or no incentive to undertake the assessment to facilitate a prospective employer's understanding of his/her qualifications, unless the occupation he/she intends to carry out in the host country requires a licence or the approval of a domestic professional body. Similarly, there is little incentive to undertake the assessment if the migrant is of low ability or wants to either change his/her career or delay his/her occupational decision. Conversely, there are strong incentives to undergo the assessment if the migrant intends to immediately all use his/her human capital in the host country, or if he/she views it as an insurance policy to optimise the chances of undertaking a wide set of occupations.

These considerations make it impossible to narrow the choice of undertaking the assessment to a single purpose. However, they frame the assessment as a mechnism whose principal aim is to overcome informational or regulatory asymmetries between the host country and the migrants' places of origin and education. The migrant is informed about the purpose and mechanics of the assessment and its potential benefits and costs, which s/he then relates to personal circumstances and aspirations. The cost-benefit analysis can be formalised in the random variable $I_{ij}$, which determines the choice of undertaking the assessment, modelled as a (normalised) latent variable:

$$D_{ij}^{*} = \theta_0 + \theta_1 I_{ij} + \pi_{ij} \qquad (5)$$

where $\theta_0$ can be viewed as a parameter representing the fixed cost (e.g. time), $\theta_1$ is a parameter, and $\pi_{ij}$ is an i.i.d. error term. What is observed is the binary outcome:

$$D_{ij} = \begin{cases} 1 \; if \; D_{ij}^{*} > 0 \\ 0 \; if \; D_{ij}^{*} \leq 0 \end{cases} \qquad (6)$$

which is assumed to change the noise in the educational signal (3) into:

$$\varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon}^{2} + \left(1 - D_{ij}\right)\Gamma\right) \qquad (7)$$

where $\Gamma > 0$, under the assumption that the local assessment reduces the noise attached to foreign education. As in Case 1, the employer pays the log wage described by (4') but now the return to foreign education reflects whether or not the assessment was undertaken:

$$\hat{\beta}_{\tilde{X}} = \frac{Cov(MP_{ij}, \tilde{X}_{ij})}{Var(\tilde{X}_{ij})} = \frac{\sigma_a^2 \rho_a (1 + b\rho_a) + \sigma_\mu^2 b}{\sigma_a^2 \rho_a^2 + \sigma_\mu^2 + \sigma_\varepsilon^2 + (1 - D_{ij})\Gamma} \qquad (7')$$

In particular, $\frac{\sigma_a^2 \rho_a (1 + b\rho_a) + \sigma_\mu^2 b}{\sigma_a^2 \rho_a^2 + \sigma_\mu^2 + \sigma_\varepsilon^2 + (1 - D_{ij})\Gamma}$ raises when $D_{ij}$= 1, implying that the log wage increases if migrant undertakes the assessment, as this reduces the employer's uncertainty about foreign education as a signal of productivity. The migrant may nevertheless find it rational not to undertake it if s/he plans not to use it immediately, or because it may reveal his/her low-productivity type.

Equations (4") and (7') make it possible to quantify the premium that employers attribute to a better understanding of foreign qualifications, which is the ratio $\frac{\hat{\beta}_{\tilde{X},D=1}}{\hat{\beta}_{\tilde{X},D=0}} = \frac{\sigma_a^2 \rho_a^2 + \sigma_\mu^2 + \sigma_\varepsilon^2 + \Gamma}{\sigma_a^2 \rho_a^2 + \sigma_\mu^2 + \sigma_\varepsilon^2} > 1$.

*Effect on job match*

The effect of observing an imperfect signal of education is also likely to affect the probability of being assigned to a job for which a migrant is over-/under-qualified. Consider the latent variable:

$$y^*_{ij} = w_{ij} - MP_{ij} \qquad (8)$$

which represents the difference between rewarded productivity as perceived by the employer ($w_{ij}$) and the migrant's true productivity ($MP_{ij}$). The latent variable $y^*_{ij}$ is observed separately for over- and under-education education as well as for those who are correctly matched as described by:

$$y_{1ij} = \begin{cases} 1\ if\left(w_{ij} - MP_{ij}\right) < -k^* \\ 0\ otherwise \end{cases} \qquad \text{(8' over-education)}$$

$$y_{2ij} = \begin{cases} 1\ if - k^* < w_{ij} - MP_{ij} < k^* \\ 0\ otherwise \end{cases} \qquad \text{(8'' correct match)}$$

$$y_{3ij} = \begin{cases} 1\ if\left(w_{ij} - MP_{ij}\right) > k^* \\ 0\ otherwise \end{cases} \qquad \text{(8''' under-education)}$$

where $k^* > 0$ is a threshold[3]. The outcome $y_{1,2,3ij}$ is the probability of an observed mismatch between the actual educational level of an immigrant carrying out a job and the minimum level identified for that occupation.

The distribution of the latent variable $y^*_{ij}$ is assumed to depend on that of the observable variable $y_{ij}$. As a result, the probability of over-education can be modelled as:

$$\Pr\left(y_{1ij} = 1 \middle| \tilde{X}_{ij}, D_{ij}\right) = \Pr\left(w_{ij} - MP_{ij} < -k^* \middle| \tilde{X}_{ij}, D_{ij}\right)$$

---

[3] For example it describes an individual with an education ISCED 5+ but occupation ISCO 4-9 as in OECD (2014), or 1 standard deviation above/below the average level of education for a given occupation, or a measure of excess education relative to what required by the job as measured 'objectively' by institutions. Three broad approaches have been used in the literature to measure the incidence of under-/over-education. One approach, which is typically based on survey data, uses the workers' self-assessment about the minimum education level needed for the job they perform or their understanding of the average education level for a particular job and whether they possess that or not (Sicherman 1991; Dolton and Vignoles 2000). A second approach, developed by Verdugo and Verdugo (1989), uses the mean education level required across a range of occupations. Under this approach an individual is considered over- or under-educated if his education level is, respectively, one standard deviation above or below the mean education level required for that particular job. A third way to analyse the level of over-/under-education is the "objective" measure based on methods used by different countries/labour organizations to assess the average required education for a particular job (Rumberger 1987 and Green *et al.* 2007).

$$= \Phi\left(-k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right) \quad (9')$$

where $\Phi(.)$ is the normal cumulative distribution function. Similarly, the probability of a correct match can be written as:

$$\Pr\left(y_{2ij} = 1|\tilde{X}_{ij}, D_{ij}\right) = \Pr\left(-k^* < w_{ij} - MP_{ij} < k^*|\tilde{X}_{ij}, D_{ij}\right) =$$

$$\Phi\left(k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right) - \Phi\left(-k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right) \quad (9'')$$

Finally, the probability of under-education can be written as:

$$\Pr\left(y_{3ij} = 1|\tilde{X}_{ij}, D_{ij}\right) = \Pr\left(w_{ij} - MP_{ij} < -k^*\middle|\tilde{X}_{ij}, D_{ij}\right)$$

$$= 1 - \Phi\left(k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right) \quad (9''')$$

The corresponding likelihood functions are:

$$L_1 = \prod_{j=1}^{m}\prod_{i=1}^{n}\left\{\Phi(-k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right\}^{y_{3ij}}\left\{1 - \Phi(-k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right\}^{1-y_{3ij}} \quad \text{(10' over-education)}$$

$$L_2 = \prod_{j=1}^{m}\prod_{i=1}^{n}\left\{\Phi\left(k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right) - \Phi\left(-k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right)\right\}^{y_{2ij}}$$
$$\left\{1 - \Phi\left(k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right) + \Phi\left(-k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right)\right\}^{1-y_{2ij}} \quad \text{(10'' correct match)}$$

$$L_3 = \prod_{j=1}^{m}\prod_{i=1}^{n}\left\{1 - \Phi(-k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right\}^{y_{3ij}}\left\{\Phi(-k^* - (b_0 + b_1\tilde{X}_{ij} + b_2 D_{ij})\right\}^{1-y_{3ij}} \quad \text{(10''' under-education)}$$

These equations can be estimated by including a random error term $u_{ij}$ to account for measurement errors.

*Employers' learning*

As time passes, host country employers have opportunities to observe a migrant's productivity and hence to adjust both his/her wage and quality of the job match, if necessary. Employers' learning can be modelled using the theoretical framework developed by Altonji and Pierret (2001) and Lange (2007), in which each new measurement of productivity updates the existing set available to employers, and where wages can be expressed as a weighted average of the initial schooling level and its relevance at different points in time or labour market experience (see the Appendix for a formal treatment). This approach enables one to identify learning with the coefficient of the interaction term between schooling, and the local signal is the assessment is undertaken, and time.

## 4. Data

The data used in the analysis is the Longitudinal Surveys of Immigrants to Australia (LSIA), a survey designed to document the early adjustment of a representative group of off-shore[4] applicant immigrants. I restrict the analysis to the primary applicants of LSIA 1, the first cohort of the LSIA, which includes only migrants arrived between September 1993 and August 1995; and LSIA 2, which includes only immigrants arrived between September 1999 and August 2000. Overall, there are 5,192 primary applicants in LSIA (Cohort) 1, and 3,124 primary applicants in LSIA (Cohort) 2. The sample is further restricted to those aged 20-65. Figure 1 presents the incidence of over-education for both cohorts shortly after settlement by type of education and geographic areas where foreign education was completed.

[Figure 1]

The incidence of over-education is remarkable, being multiplies of that characterising natives (dotted line). On average over-education is lower for those possessing a vocational education degree vis-à-vis those with university education. This may reflect specific labour market shortages in Australia, as several of the occupations in high demand for which applicants would have received additional points have a vocational or trade nature. It could also reflect

[4] The LSIA was explicitly designed to exclude potential immigrants applying onshore, such as international students in Australia. However, it includes about 350 observations about individuals completing their highest education in the country prior to returning home and, presumably, re-applying for permanent settlement. These observations are omitted from the sample used in the analysis.

the need to quickly enter the labour market, hence taking up any job offer, to limit the financial burden associated with migration and current living expenses regardless of the type of education. There is also significant regional variation, with substantially lower rates of over-education for those with degrees from an English-speaking country (Ireland, United Kingdom, United States, Canada) and highest rates amongst those graduating in South East Asia and Latin America.

Figure 2 presents the uptake rates of the assessment of foreign qualifications by type of education and geographic areas.

[Figure 2]

The decision to undertake the assessment is generally lower for migrants with vocational relative to university education, aside for the English-speaking places of education and South Asia (India, Pakistan, Bangladesh, Sri Lanka), where English is widely spoken and where several institutions are similar to Australia's, thanks to common historical roots. There is however substantial heterogeneity in assessment uptake, which *prima facie* does not seem to correspond to over-educational outcomes[5].

Table 1 summarises the main variables used in the empirical analysis.

[Table 1]

[5] Table A1 reports some descriptive statistics about those undertaking the assessment, by type of education. More than 50% of respondents having their qualifications assessed work in a licensed occupation (first line), and there is wide heterogeneity in uptake by type of visa (highest amongst those in the skilled independent category) and region of education (highest amongst English-speaking group and South Asia). The assessment is associated with higher wages and lower over-education for those with vocational education but not for those with a tertiary degree. Choosing to undertake the assessment appears to be a complex decision that reflects labour supply variables such as preferences, risk aversion and expectations, as well as Australia's labour demand features, like barriers to enter particular occupations and local labour market characteristics. The lack of a clear-cut relationship between undertaking the assessment and labour market outcomes supports its formal treatment as the outcome of a latent variable depending on a personal cost-benefit analysis.

The majority of immigrants in the sample are in prime working (about 35 years old), with about 15 years of work experience. They are mostly married, with one or two children. Just over a third are women, possibly reflecting the higher score obtained by younger and highly educated wives in the case of applicant couples. About 40% of the immigrants in the sample have a vocational education degree as highest completed educational level. This corresponds to about a third in the original LSIA database.

Immigrants' schooling is completed in a wide variety of geographic areas. Those using English as the official language (UK/Ireland, US and Canada) account for about a quarter of the observations in the sample (New Zealanders are not part of the LSIA as they can enter/leave Australia with no restrictions on labour market access). In about a third of cases, the foreign qualification has been formally assessed in Australia by approved agencies. These include professional accreditation bodies (e.g. institute of Chartered Accountants, Institutes of Engineers, Australian medical Council, Australian Nursing Council, Australian Computer Society…) as well as state and federal government departments (e.g. State Medical Board, State Department of Education, Department of Immigration, Department of Employment…). Most immigrants were interviewed in English confirming their high level of language skills.

## 5. Empirical strategy

The empirical analysis presents two sets of results. The first focuses on the returns to foreign education and the effects of acquiring the local signal in the months immediately after settlement. The second discusses the evolution of migrants' labour market outcomes over time, and host country employers' learning of migrants' productivity.

*The returns to foreign education*

*Wage regression*

Equation (4') provides the bases for performing the wage regression, which is carried out using the functional form:

$$w_{ij} = \beta_0 + \beta_1 \tilde{X}_{ij} + \beta_2 D_{ij} + \tilde{X}_{ij} D_{ij} \beta_3 + Z_{ij}\gamma + \tau_{ij} \qquad (11)$$

where $\tau_{ij}$ is an i.i.d. error term and $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$ and $\gamma$ is the set of parameters to be estimated. In particular:

$w_{ij}$ is the logarithm of the gross weekly wage reported as the mid-point of each of the nine wage interval categories expressed in Australian dollars[6] reported in the LSIA;

$\tilde{X}_{ij}$ is a set of dummy variables of the regions where the migrant completed his/her highest level of education. The reference is the group of English-speaking countries, namely the United Kingdom, Ireland, the United States, Canada and South Africa;

$D_{ij}$ is a dummy variable indicating if the migrant has undertaken the local assessment. To account for the likely endogeneity of undertaking the assessment of foreign qualifications the variable $D_{ij}$ is instrumented by $\widehat{D_{ij}}$ and so is the interaction term $\tilde{X}_{ij} D_{ij}$ (by $\tilde{X}_{ij}\widehat{D_{ij}}$). The approach and choice of instrument are discussed in more details in a separate sub-section.

$Z_{ij}$ is a set of exogenous covariates that include experience and experience squared, gender, marital status, number of household members, proficiency in spoken English (three categories), year of arrival, state of residence, employer-sponsored or refugee visa[7], time, and two other covariates. The first is the hazard (inverse Mills' ratio) taking care of self-selection into participating in the labour market. The exclusion restrictions, reflecting Green *et al.*

---

[6] The variable reporting the number of working hours has fewer data points than that reporting the weekly wage. As the empirical results performed on weekly or hourly wages are effectively identical, I focus on those based on weekly wages as the higher number of observations provides more power to some of the tests carried out.

[7] These two visa categories are associated with demand-driven and non-economic migration, respectively, and characterize people whose settlement circumstances substantially differ from those migrating as skilled independent migrants or reuniting with family in Australia.

(2007), are car ownership, the presence of children in the household, and whether the immigrant had own funds at the time of arrival[8].

The second covariate takes into account that some occupations require a licence to be carried out, and is constructed as the share of jobs potentially requiring a licence by each Australian state and year of observation. Australia does not have a national system of licensed occupations, as these tend to be managed at a State level[9]. The licensing requirement is explicitly reported in the description of some occupations, such as those in the medical profession. However, the description of several other professions only highlights that a licence "may be required", implying different regulations based on the type and location where a profession is carried out. I use this broader definition to construct a variable equal to the share of occupations that may require licensing for a given state and year, which I add as exogenous covariate in all regressions.

Equation (16) is separately estimated by OLS IV (two-stage least squares) on vocational and tertiary educational levels, respectively, on data pooled across waves and standard errors clustered at individual level. Random effects estimation (RE) is also used as a robustness check to limit the influence of unobserved individual heterogeneity, exploiting the panel nature of the LSIA.

*Endogeneity of qualifications' assessment and interaction term*

Undertaking the local assessment was a choice at the time when the LSIA was conducted, and it is likely to be simultaneously determined with labour market outcomes, especially if the migrant works in a licensed occupation where formal qualifications must be assessed by a

---

[8] The hazard is estimated from an initial wage equation simultaneously estimated with the probability of participating in the labour market using the exclusion restrictions highlighted in the main text. This approach follows Green et al (2007).

[9] The introduction of a national occupational licensing system for certain occupations was initially tackled by by the Council of Australian Governments (COAG) in 2008 to remove "duplicate and inconsistent regulation for specific occupations between states and territories". The proposed regulatory change however faced substantial local opposition and was abandoned in 2013, following a change in federal government. For more information, see http://www.coag.gov.au/node/516 (accessed 26 December 2014).

professional body or association, or the regulator. As a result, the variable reporting whether one has completed the recognition of foreign qualifications is instrumented. The exclusion restriction is the spatial heterogeneity of the respondents receiving official information about the recognition of foreign qualifications[10] across Australia, which is contained in both LSIA1 and LSIA2. In particular, an individual's assessment choice is instrumented with the proportion of those undertaking the assessment in the same broad settlement area. I aggregate the 83 'interview statistical subdivisions' used in the LSIA into 41 broader geographic areas. The logic behind this instrument is that spatial differences in the use of information are likely to reflect local conditions that influence an individual's benefits and costs of migrating to the area where s/he plans to initially settle. However there is no *a priori* reason for wages to be positively or negatively related to the use of information[11]. Table 2s reports the key correlations between labour market outcomes, endogenous and instrumental variables for those holding a vocational and a tertiary qualification.

[Table 2a,b]

The endogeneity of the decision to undertake the assessment is tested using Wooldridge's approach (2010 – eq.15.51 p.528), in which the residual of an OLS regression of the endogenous variable on all exogenous and instrumental variables is included in the regressors of a second OLS regression. If the residual from the first stage regression is statistically significantly different from zero, then the variable measuring the local assessment is deemed

[10] After applying for a visa, an applicant is sent official information about labour market conditions in Australia, job opportunities as well as the recognition of foreign qualifications. Such information is sent by the Department of Immigration in a booklet, but additional information can be sent by State governments and industry associations. In theory each migrant receives the same information from DIAC.

[11] The F-test of joint significance of the interactions between the 40 dummy variables representing aggregate areas and the instrument obtained from an OLS wage regression is not significant: $F(33, 2987)=1.27$ (p-value: 0.1355).

endogenous. The results reported in Table 2b support this conclusion for in both cases of vocational and tertiary education[12].

Both wage and over-education models include interaction terms between endogenous and region of education variables to allow possibly different values of the local signal across the regions where education was completed. In the regressions using pooled data these interaction terms are instrumented with the corresponding interactions between instrumented variable and (exogenous) regions of education, as suggested by the literature (Wooldridge, 2011). This approach is however questioned (Bun and Harrison, 2014), and the analysis also presents the results of separate regressions by region of education.

*The probability of over-education*

The analysis of the quality of the job match focuses on over-education, as it is the most costly mis-match in terms of skill under-use. Equation (10') leads to the functional form:

$$\Pr\left(y_{1ij} = 1 | \tilde{X}_{ij}, D_{ij}\right) = b_0 + b_1\tilde{X}_{ij} + b_2\widehat{D_{ij}} + \tilde{X}_{ij}\widehat{D_{ij}}b_3 + Z_{ij}\lambda + u_{ij} \quad (12)$$

where $\tilde{X}_{ij}$, $\widehat{D_{ij}}$ and their interaction are described as above, $\lambda$ is the vector of returns to the exogenous covariates $Z_{ij}$, and $u_{ij}$ is an i.i.d. error term.

Over-education, $y_{1ij}$, is obtained using the objective method: namely, as the difference between the immigrant's actual educational level and the minimum educational level necessary to perform that occupation, as reported in the classification of occupations compiled by the Australian Bureau of Statistics (2006).

Equation (12) is also estimated separately on vocational and tertiary education. As advocated in Agrist and Pischke (2009: 198-204), equation (12) is estimated using a linear probability model with IV to account for the endogeneity of $\widehat{D_{ij}}$ and its interaction with $\tilde{X}_{ij}$. As the IV

[12] As an additional check of the potential endogeneity of the choice of undertaking the assessment, I use instrument that variable with the quality of education as measured by a new comparable indicator of secondary schooling quality developed by Altinok et al (2013) using data from PISA and several other data sources. The results obtained in the main estimations are similar to those obtained using the average proportion of those claiming to have received information on the recognition of foreign qualifications, by broad geographic area.

OLS model is biased and its performance may be inferior to a correctly specified maximum likelihood model[13] (Nichols, 2011), the results obtained may be viewed as a lower boundary of the causal effect estimated.

*Employers' learning*

The LSIA covers only up to 41 months since arrival, or at most three waves, and only for the fist cohort interviewed. Hence it has only limited information about the evolution of a migrant's wage and quality of the job match in Australia. Yet, because it covers the period at the very outset of each migrant's new life cycle in the host country, it captures employers' initial expectations of a migrant's productivity, and subsequent expectations once productivity is observed. A test of employers' learning relies on the comparison of the returns to foreign human capital ($\beta_1$ and $b_1$ in equations (11) and (12), respectively) and to the local signal ($\beta_2$ and $b_2$) between the first time when a migrant is surveyed and subsequent records. Employers' learning arises if the returns to foreign schooling are lower, and correspondingly the returns to the local signal are higher, as time passes. This prediction is tested by performing separate IV OLS regressions on vocational and tertiary educational data in the first and subsequent waves of the panel, as well as applying a random effects model to all waves under the assumption that the endogenous covariates are independent of the error term (Baltagi's two-stage least squared random effects estimator, "EC2SLS" in Stata), and including an interaction between the variables of interest and time.

In particular, I perform the wage regression:

[13] The problem of a dichotomous dependent variable with an endogenous dichotomous regressor is not entirely solved by the literature (Nichols, 2011). While Angrist and Pischke (2008) highlight that the linear probability model performs well, this may not be the case if the response to a treatment varies across individuals. IV estimates are consistent but biased and inefficient. A better estimator may be obtained using a maximum likelihood model, though this generally imposes strict conditions on the structure of the error distributions. In the case of a single endogenous regressor, with homoskedastic bivariate normal errors in both outcome and selection equations it is possible to use the maximum likelihood biprobit model (Heckman, 1978). In presence of heteroskedasticity and non-normal errors, possible solutions apply the general method of moments or semiparametric solutions. For an overview, see Nichols (2011).

$$w_{ijt} = a_0 + a_1\tilde{X}_{ijt} + a_2\tilde{X}_{ijt}t + a_3\widehat{D_{ijt}} + a_4\widehat{D_{ijt}}t + \tilde{X}_{ijt}\widehat{D_{ijt}}a_5 + Z_{ijt}A + \theta_{ijt} \quad (13)$$

where the covariates are as previously described, $a_0, \ldots, a_5$ and $A$ are parameters/vectors of parameters to be estimated, and the error term $\theta_{ijt} = \delta_{ji} + \eta_{ijt}$, which contains a time-invariant individual component $\delta_{ij}$, and an i.i.d. error term $\eta_{ijt}$. Employers' learning occurs if $a_2$ is positive and statistically significantly different from zero and, correspondingly, $a_4$ is zero or negative. Regressions are performed separately on the sub-sample of those who stay with the same employer between surveys (as asked in the LSIA) to control for employers' heterogeneity, and on everyone regardless of whom they work for.

Similarly, employers' learning in the quality of job match equation is estimated using:

$$\Pr\left(y_{1ijt} = 1 | \tilde{X}_{ijt}, D_{ij}\right) = b_0 + b_1\tilde{X}_{ijt} + b_2\tilde{X}_{ijt}t + b_3\widehat{D_{ijt}} + b_4\widehat{D_{ijt}}t + \tilde{X}_{ijt}\widehat{D_{ijt}}b_5 + Z_{ijt}B + \zeta_{ijt} \quad (14)$$

where the error term $\zeta_{ijt} = \delta_{ij} + \vartheta_{ijt}$contains a time-invariant individual component $\delta_{ij}$ and an i.i.d. error term $\vartheta_{ijt}$.

## 6. Results

*Wage regressions*

Table 3 reports the estimates of the wage equation obtained by two-step two-stage least squares using Stata's ivreg2 command (Baum *et al*, 2010), by type of education. All estimates are obtained with standard errors clustered at individual level. The top part of the table reports the results obtained on vocational education while the bottom part report those obtained on tertiary education.

The first column of Table 3 shows the marginal effect of undertaking the assessment (instrumented) on observations pooled across geographic areas: this is positive and statistically significantly different from zero for both educational groups (vocational: +0.326, p-value: 0.000; tertiary: +0.383, p-value: 0.000), implying that the assessment is highly valued by Australian employers, especially in the case of university and higher education graduates. This is perhaps not surprising, given the higher average proportion of over-education amongst tertiary-educated migrants illustrated in Figure 1. Based on average wages for these two groups, the effect of the assessment translates into an additional 201 A$/week (+38%) for a migrant with vocational education and in 315 A$/week (+46%) for a migrant with tertiary education. Using a tax rate of 25%, each immigrant undertaking the assessment contributes an additional 50-79 A$/week in gross revenue to Australia's public coffers, or 130-205 million A$ based on a conservative figure of 50,000 immigrants entering the labour market and constant 1995 A$.

For each type of education, Table 3 also reports key statistical tests: namely (i) the goodness of fit of the regression, (ii) the number of observations, (iii) the F-test of significance of the instrument (Angrist and Pischke, 2009); and (iv) the Wooldridge test of endogeneity performed by Stata using the ivregress command. The results support that the regression explains about 22% of the variation in the logarithm of weekly wages, the instrument works well as it is correlated with the endogenous variable but not to the outcome variable, and the assessment choice is indeed an endogenous variable.

The remaining columns of Table 3 show the results obtained when the same regression is performed on a subgroup of observations, based on the geographic dispersion of the place of highest completed education. In particular, the second and third columns show the results obtained when on migrants completing their education in a country with an English speaking (ESB) or non-English speaking background (NESB). Here the results differ: undertaking the assessment has favorable and statistically significant consequences on the log weekly wages of ESB migrants with vocational education (+.444; p-value: 0.000) and less for those with a tertiary degree (+.247; p-value: 0.09). This result supports the hypothesis that ESB migrants with vocational education may have to undertake the assessment in order to enter occupations where a license is required.

ESB migrants with tertiary education may instead view the local assessment as a positive negative signal of their ability (especially if graduating from internationally well-known universities) and shun away from it. Conversely, NESB migrants receive significantly better wage outcomes when taking up the assessment, regardless of education types (vocational: +.245; p-value: .045; tertiary: +.651; p-value: .000). As the regressions control for previous labour market experience and how well migrants speak English, these results suggest that Australian employers find it more difficult to assess the productivity linked to foreign education without the addition of a 'local' signal, in line with the hypothesis of statistical discrimination.

Columns 4-7 of Table 3 further subdivide the observations covering NESB migrants by main geographic region. Substantial differences arise in the behavior by education type and by geography. South Asia-vocational educated migrants who did the assessment receive substantially higher wages than comparable migrants who did not (+.922), but the difference is not statistically significantly different from zero for all other regions. This finding reflects that vocationally-trained NESB migrants have similarly paid job opportunities besides occupations requiring a license, as highlighted by the prevalence of vocational jobs in the list of occupations in high demand, which provided additional points to prospective migrants in the 1990s and early 2000s.

In the case of tertiary education, the NESB results are driven by migrants from Russia and Central Asia (+1.023), South East Asia (+.669), and Latin America (+1.745). In contrast there is no assessment effect amongst Europeans and South Asians holding a university degree.

Table 4a reports the full sets of estimates, including those obtained by OLS when the residual of the first stage regression is included. When the instrument is interacted with the region of highest education, the coefficients estimated are positive but hardly statistically significantly different from zero. However, when the interaction term is included to the gradient of foreign education (Table 4b) the wage penalty associated with foreign education is substantially lower for each group.

*Over-education regressions*

The marginal effects of the regressions performed on the probability of over-education are reported in Table 5, which reproduces the structure of Table 3. The recognition of foreign education has substantial positive effects on the education-occupation match for all migrants, and especially those with tertiary education. The probability of having a job under-utilising a migrant's human capital drops by about 15% (p-value: 0.000) in the case of vocational education but by about 53% (p-value: 0.000) in the case of university education. These are indeed very large effects, and probably the most significant in terms of early economic assimilation in the host country's labour market and potential 'policy value' for those immigration countries that currently do not have an assessment choice one but receive substantial flows of immigrants outside demand-driven and refugee migration. Undertaking the assessment appears to considerably improve the chance of a good education-occupation match in the labour market post-migration, and the returns to human capital acquired abroad. The persistence of over-education, as documented by the literature, compounds the relevance of introducing such a policy in destination countries in order to fasten employers' learning of a migrant's true productivity and the positive spill-overs that this has on the host society (e.g. through higher taxable income and possibly well-being).

Table 5 also reveals that undertaking the assessment in the case of vocational education has relevance only for NESB (-.255) but not for ESB migrants (marginal effect is zero). This result is consistent with the hypothesis that the assessment eases entry into occupations where licensing may be required. Further geographical subdivision indicates that this outcome is driven by the behavior of those educated in South East Asia and Russia and Central Asia, which have an institutionalized system of vocational training (Eichhorst et al, 2013) whose quality may be unfamiliar to Australia-based employers.

The most striking result arises in the case of tertiary education, as both ESB and NESB migrants appear to significantly reduce their chances of over-education by undertaking the assessment (-.302; p-value: 0.000; -.564: p-value: 0.000, respectively). Further investigation reveals that in many cases those undertaking the assessment fill professional jobs that are identical to their last occupation before migrating. The assessment therefore seems to ease the immigrant's international transfer of human capital to Australia, especially with reference to early entry into professional occupations, with limited skill and knowledge wastage, if any, in adjusting to new labour market conditions post-migration. Further geographic analysis shows that these results are driven by NESB migrants completing their education in Asia, Russia/Central Asia, and Latin America, which have a relatively recent history of high skill migration to Australia.

Table 6a reports the full sets of estimates, including those obtained by OLS when the residual of the first stage regression is included. Table 6b shows the effect on the probability of over-education of including the local signal to foreign education. As for the case of wages, the local signal substantially reduces the probability of a poor quality education-occupation match.

*Employers' learning*

Tables 7 and 8 report the wage penalty associated with vocational and tertiary foreign education, respectively, based on equation (13). The first two columns of each table report separate estimates for regressions performed only on the first and subsequent waves on the sub-set of respondents working wave with the same employer. The results clearly show that the wage penalty arises in the first period after settlement, when employers have not yet observed a migrant's productivity. Over time (waves 2 and 3, though just for the first cohort), the wage penalty reduces and disappears, relative to those educated in an English-speaking region, in most cases.

The next two columns of each table report the results obtained when foreign education and signal are interacted with time since the survey interview on the sub-group of respondents working with the same employer in all LSIA waves. The coefficients of the interaction are generally no different from zero. However, when combined with the variables of interest they reduce the gradient of the local signal and correspondingly increase those of foreign education.

The last two columns in each table replicate the earlier wage regressions using interaction terms with time, but performed on all observations. They therefore cover migrants working with the same employer as well as migrants who change employers across waves. The results show that in this case the wage penalty persists over time, and the local signal continues to have a strong influence in reducing the wage penalty. These results, as well as those obtained on previous regressions, provide support to the hypothesis of statistical discrimination.

With reference to the probability of over-education the results obtained in the case of vocational education (Table 9) suggest that the effect of the local signal in reducing the migrant's education-occupation mismatch arises only when interacted with time, with the only exception of those educated in South Asia. This implies that migrants undertaking the assessment of their vocational education enjoy a faster entry to correctly-matched jobs (aside from those educated in South Asia). Such effect remains even when the analysis is extended to migrants changing employers between waves.

In the case of tertiary education (Table 10), the signal has an immediate effect on lowering the probability of over-education, and such effect becomes less relevant over time, as implied by the hypothesis of statistical discrimination.

## 7. Conclusions

The well-known fact that migrants suffer lower returns to education when this is obtained in a different country from that of destination arises also in Australia, for both vocational and tertiary qualifications. This paper explored the possibility that such penalty is not only due to taste discrimination but also by the statistical discrimination arising from poor knowledge of the productivity signal of foreign education, especially when acquired in a NESB country.

The paper uses data from the LSIA, a unique database capturing the early settlement of migrants to Australia, to study whether offering a choice whereby a migrant can get his/her foreign qualifications assessed in the host country can improve his/her labour market outcomes in the early period post-migration. Given that labour markets outcomes tend to persist, a start 'with the right foot' in the host country's labour market can not only improve a migrant's personal returns to human capital but also his/her spill-over effects on the host society, via higher wages (hence consumption and taxable income) and possibly well-being.

I find substantial improvements in wages and, most importantly, a better education-occupation match in the job when employers value a migrant's productivity through the additional local signal, in light with the hypothesis that Australian employers may statistically discriminate migrants. On average a migrant undertaking the assessment can enjoy a 30-40% increase in gross weekly wages, and half the probability of getting a job for which s/he is over-qualified. Using an average inflow of 50,000 migrants into Australia's labour market each year, the additional contribution of the assessment is about 130-205 million 1995-constant A$ for the migrants, and 32-51 million A$ for the government (based on a 25% tax rate).

These potential benefits have a relevant policy value. The issue of imperfect international transferability of human capital and its reduction as years since migration increase has been addressed to date with a lasses-faire approach relying on the efficiency of market forces in eventually recognizing individual productivity. This paper however shows that introducing an assessment (or making it mandatory) may be a relatively simple and low-cost way to effectively fasten migrants' human capital transfer to the host country. Such measure can benefit migrants and the host society alike vie a faster economic assimilation and better utilization of their skills. Given the large volumes of skilled migrants resettling each year, the efficiency gains from introducing this type of instrument are substantial.

## References


Aigner, D. J., and G.G. Cain, (1977). “Statistical theories of discrimination in labor markets”. *Industrial and Labor Relations Review*: 175-187.

Altinok, N., Diebolt, C., and J.L. Demeulemeester (2014). „A new international database on education quality: 1965–2010”. *Applied Economics*, 46(11): 1212-1247.

Altonji, J. and C. Pierret (2001). “Employer learning and statistical discrimination”. *The Quarterly Journal of Economics*, 116(1): 313-350

Angrist, J. D., and J.S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Arcidiacono,P., Bayer, P., and A. Hizmo (2010). “Beyond signaling and human capital: Education and the revelation of ability”. *American Economic Journal: Applied Economics* 2(4): 76–104.

Arrow, K. (1973). “The theory of discrimination”. *Discrimination in Labor Markets*, *3*(10): 3-33.

Artuc, E., Docquier, F., Özden, Ç., and C. Parsons (2015). A global assessment of human capital mobility: the role of non-OECD destinations. *World Development*, 65: 6-26.

Australian Bureau of Statistics (2006). *ANZSCO: Australian and New Zealand Standard Classification of Occupations*. Australian Bureau of Statistics/Statistics New Zealand. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/B4B626DEB4A0C558CA2571E600092D5A/$File/12200_2006.pdf

Balsa, A. I., and T.G. McGuire (2001). “Statistical discrimination in health care”. *Journal of Health Economics*, 20(6): 881-907.

Basilio, L., and T.K. Bauer (2010). Transferability of Human Capital and Immigrant Assimilation-An Analysis for Germany. *Ruhr Economic Paper*, 164.

Battu, H. and P.J. Sloane (2004). “Over-education and ethnic minorities in Britain”. *The Manchester School*, 72: 535-559.

Baum, C.F., Schaffer, M.E., and S. Stillman (2010). ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. http://ideas.repec.org/c/boc/bocode/s425401.html

Bloom, D.E., G. Grenier and M. Gunderson (1995). “The changing labour market position of Canadian immigrants”. *The Canadian Journal of Economics*, 28: 987-1005.

Bordón, P., and B. Braga (2013). Employer Learning, Statistical Discrimination and University Prestige. Unpublished paper, *University of Wisconsin-Madison*.

Bun, M. J., and T.D. Harrison (2014). *OLS and IV estimation of regression models including endogenous interaction terms*. Discussion paper No. 2014-3, LeBow College of Business, Drexel University.

Chiswick, B. R. and P.W. Miller (2009). “The international transferability of immigrants’ human capital”. *Economics and Education Review*, 28: 162-169.

Docquier, F., Ozden, Ç., and G. Peri (2013). “The labour market effects of immigration and emigration in OECD Countries”. *The Economic Journal*, 124(579): 1106-1145.

Dolton, P., and A. Vignoles (2000). “The incidence and effects of overeducation in the U.K. graduate labour market”. *Economics of Education Review*, 19: 179–198.

Dustmann, C., and F. Fabbri (2003). “Language proficiency and labour market performance of immigrants in the UK”. *The Economic Journal*, 113(489): 695-717.

Eichhorst, W., Rodrigues-Planas, N., Schmidl, R., and K.F. Zimmermann (2013). A roadmap of vocational education and training around the world. IZA Discussion paper series No. 7110.

Farber, H.S., and R. Gibbons (1996). “Learning and Wage Dynamics”. *Quarterly Journal of Economics*, CXI: 1007-1047.

Ferrer, A. and W. Riddell (2008). “Education, Credentials and Immigrant Earnings”. *The Canadian Journal of Economics*, 41: 186-216.

Green, C., P. Kler, and G. Leeves (2007). “Immigrant overeducation: Evidence from recent arrivals to Australia”. *Economics of Education Review*, 26(4): 420–432.

Lange, F. (2007). “The speed of employer learning”. *Journal of Labor Economics*, 25(1): 1-35.

Lundberg, S.J., and R. Startz, (1983). “Private discrimination and social intervention in competitive labor market”. *The American Economic Review*, 73(3): 340-347.

Mansour, H. (2012). “Does employer learning vary by occupation?”. *Journal of Labor Economics*, 30(2): 415-444.

Nichols, A. (2011). Causal inference for binary regression with observational data. In *CHI11 Stata Conference* (No. 6 - July). Stata Users Group.

Oettinger, G.S. (1996). “Statistical discrimination and the early career evolution of the black-white wage gap”. *Journal of Labor Economics*, 14(1): 52-78.

Phelps, E.S. (1972). “The statistical theory of racism and sexism”. *The American Economic Review*, 62(4): 659-661.

Piracha, M., Tani, M. and M. Vaira (2014). “'Social capital and immigrants’ labour market performance”, *Papers in Regional Science* (forthcoming)

Poot, J., and S. Stillman (2010). *The importance of heterogeneity when examining immigrant education-occupation mismatch: evidence from New Zealand*. IZA Discussion paper series No. 5211.

Rumberger, R.W. (1987). „The impact of surplus schooling on productivity and earnings". *Journal of Human Resources*, 22(1): 24–50.

Sicherman, N. (1991). "Overeducation' in the labour market". *Journal of Labour Economics*, 9(2): 101–122.

Siniver, E. (2011). "Testing for statistical discrimination: The case of immigrant physicians in Israel". *Labour*, 25(2): 155-166.

Verdugo, R.R., and N.T. Verdugo (1989). "The impact of surplus schooling on earnings: Some additional findings". *Journal of Human Resources*, 24(4): 629–643.

Wald, S., and T. Fang (2008). "Overeducated immigrants in the Canadian labour market: Evidence from the workplace and employee survey". *Canadian Public Policy*, 34(4): 457-479.

Wang, S.Y (2015). "Statistical discrimination, productivity and the height of immigrants". *Industrial and Labor Relations Review* (forthcoming).

Wooldridge, J.M. (2001). *Econometric Analysis of Cross-Section and Panel Data*, Cambridge: MIT Press.

Wooldridge, J.M. (2011). Endogenous interaction terms in 2SLS (Stata forum: http://www.stata.com/statalist/archive/2011-03/msg00188.html). Accessed 14 December 2014.

## Appendix

**A model of employers' learning, based on Altonji and Pierret (2001) and Lange (2007).**

Productivity is measured with error so that at time *t* only:

$$q_{ijt} = MP_{ij} + \upsilon_{ijt} \qquad \text{(A1)}$$

is observed, where $MP_{ij}$ is as described in equation (1) in the main text, and $\upsilon_{ijt}$ is an i.i.d. error term independent of all other variables. Host country employers use the past realisations of $q_{ijt} = \{q_{ij1}, q_{ij2}, \dots q_{ijt-1}\}$, to form their expectation of a migrant's productivity and will offer:

$$w_{ijt} = E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}, q_{ijt}\right) \qquad \text{(A2)}$$

which expands equation (4') of the main text to include the conditioning on $q_{ijt}$. As a result, at time 1:

$$w_{ij1} = E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}, q_{ij1}\right) = E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}\right) + \frac{Cov\left(MP_{ij}, q_{ij0}\right)}{Var\left(q_{ij0}\right)}\left(q_{ij0} - \overline{q_{ij}}\right)$$

$$= E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}\right) + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\upsilon^2}\left(q_{ij0} - E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}\right)\right)$$

where $\sigma_1^2 = Var\left(MP_{ij}\right)$ and $\sigma_\upsilon^2 = Var\left(\upsilon_{ij}\right)$. The learning parameter is $\frac{\sigma_1^2}{\sigma_1^2+\sigma_\upsilon^2}$ reflects the relevance given by host country employers to the new productivity measure, as this rises when the error $\upsilon_{ijt}$ is smaller. At time 2:

$$w_{ij2} = E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}, q_{ij2}\right) = E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}\right) + \frac{Cov\left(MP_{ij}, q_{ij1}\right)}{Var\left(q_{ij1}\right)}\left(q_{ij1} - \overline{q_{ij}}\right)$$

$$= E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}\right) + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\upsilon^2}\left(y_{ij1} - E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}\right)\right)$$

and at any time *t*, by iteration, as in Lange (2005):

$$w_{ijt} = E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}, q_{ijt}\right) = (1 - \theta(t-1))E\left(MP_{ij} | \tilde{X}_{ij}, D_{ij}\right) + \theta(t-1)\frac{1}{t-1}\sum_{r=1}^{t-1} q_{ijr}$$

(A3)

where $\theta(t-1) = \frac{(t-1)\frac{\sigma_1^2}{\sigma_1^2+\sigma_v^2}}{1+(t-1)\frac{\sigma_1^2}{\sigma_1^2+\sigma_v^2}}$ is the parameter capturing employers' learning. As the LSIA data do not contain information about what data is available to employers, one needs to derive the linear projections of these data on educational measures (and other variables that may be observable by employers but not by the LSIA) at different times. As a result, using the assumption of independence of the error term in (A1) it is possible to write:

$$E\left[w_{ijt}|\tilde{X}_{ij}, D_{ij}, t-1\right] =$$

$$= E\left[\left[\left(1-\theta(t-1)\right)E\left(MP_{ij}|\tilde{X}_{ij}, D_{ij}\right) + \theta(t-1)\frac{1}{t-1}\sum_{r=1}^{t-1} q_{ijr} + \Delta(t-1)\right]|\tilde{X}_{ij}, D_{ij}, t-1\right]$$

$$= \left(1-\theta(t-1)\right)E\left[E\left(MP_{ij}|\tilde{X}_{ij}, D_{ij}\right)|\tilde{X}_{ij}, D_{ij}\right] + \theta(t-1)E\left(MP_{ij}|\tilde{X}_{ij}, D_{ij}\right) + \Delta(t-1) =$$

$$\left(1-\theta(t-1)\right)\Theta(0)\tilde{X}_{ij} + \theta(t-1)\Theta(t-1)\tilde{X}_{ij} + \left(1-\theta(t-1)\right)\xi(0)D_{ij} + \theta(t-1)\xi(t-1)D_{ij} + \Delta(t-1) \qquad \text{(A4)}$$

which specifies wages as the weighted average of a component reflecting the link between formal education and wages, and the signal and wages, respectively, before the migrant's productivity is observed, and a second component reflecting that link after productivity is observed. Equation (A4) can be estimated using the empirical model:

$$w_{ijt} = \varrho_1\tilde{X}_{ij} + \varrho_2\tilde{X}_{ij}t + \varrho_3 D_{ij} + \varrho_4 D_{ij}t \qquad \text{(A5)}$$

after adding other control variables and an i.i.d error term.

With reference to the quality of the job match, the panel version of equation (8) of the main text is the latent variable:

$$y^*_{ijt} = w_{ijt} - MP_{ij} \qquad \text{(A6)}$$

which is observed, in the case of over-education, when:

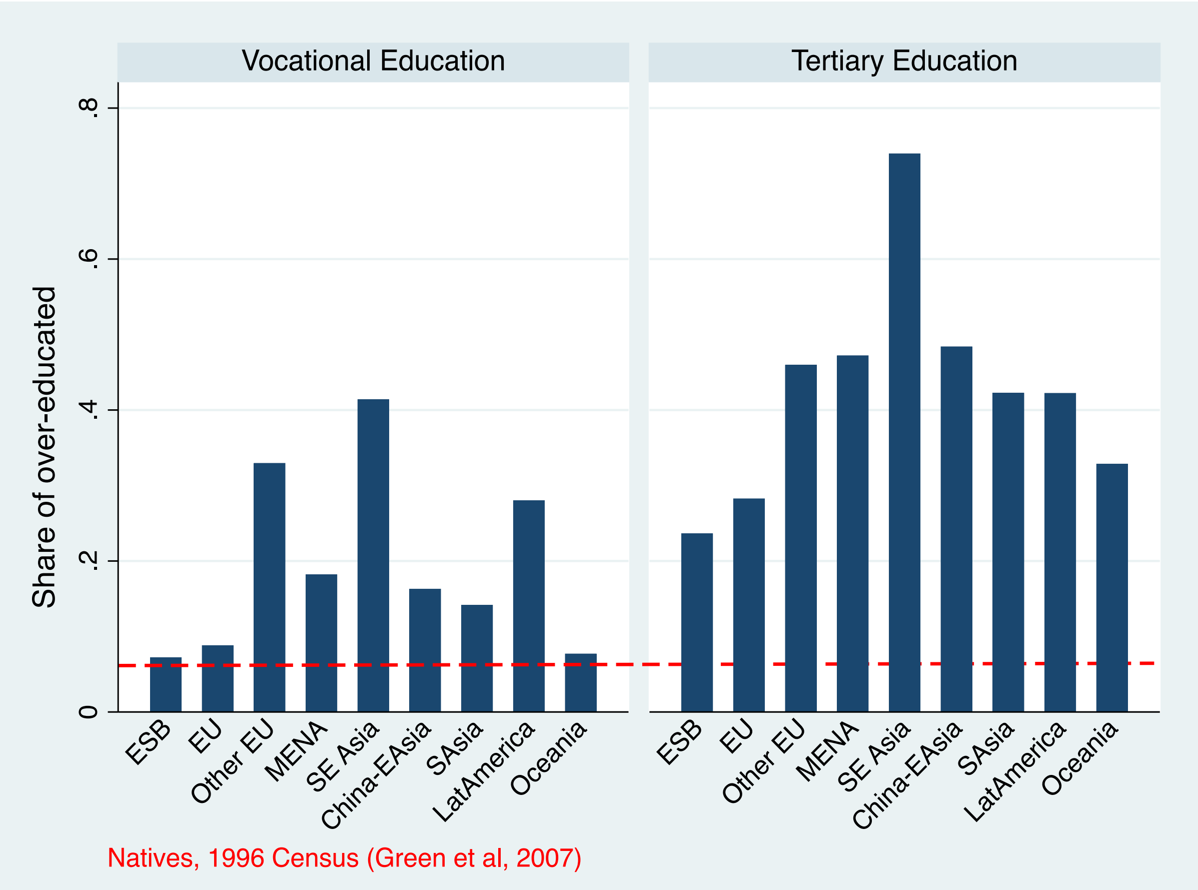
$$y_{1ijt} = \begin{cases} 1\ if\left(w_{ijt} - MP_{ij}\right) < -k^* \\ 0\ otherwise \end{cases} \qquad \text{(A7)}$$

As the latent variable is by assumption normally distributed of, the probability of over-education over time is modelled as:

$$\Pr\left(y_{1ijt} = 1 \middle| \tilde{X}_{ijt}, D_{ij}, ij, q_{ijt}\right) = \Pr\left(w_{ijr} - MP_{ij} < -k^* \middle| \tilde{X}_{ijr}, D_{ij}, ij, q_{ijt}\right) = \Phi\left(-k^* - (b_0 + b_1\tilde{X}_{ijt} + b_2 D_{ij} + b_3 ij + b_4 q_{ijt})\right) \quad \text{(A8)}$$

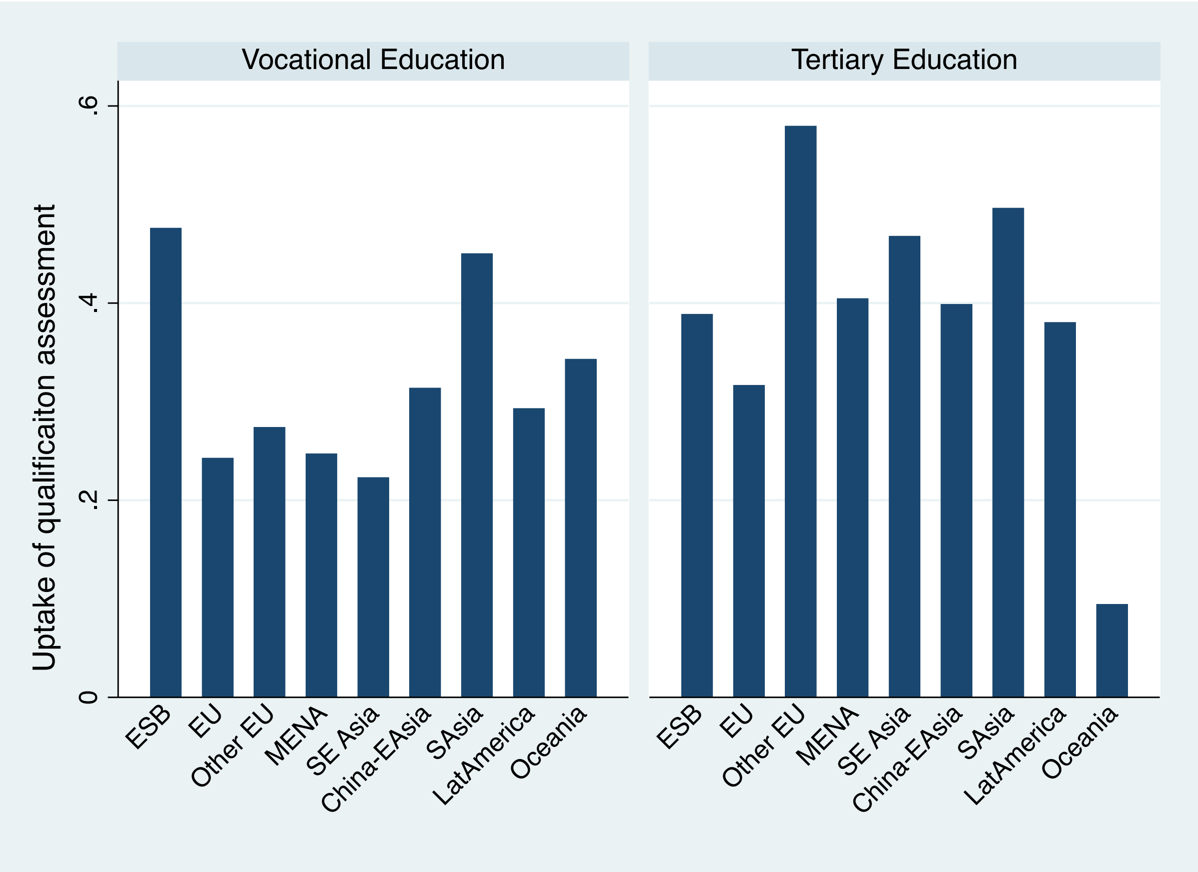
where $ij$ represents unobservable individual-specific traits.

**Figure 1: Incidence of over-education by type on settlers at 6 months post-migration**




Source: cohort 1, LSIA 1 and cohort 1, LSIA 2.

**Figure 2: Assessment of foreign qualifications: uptake rates**




Source: cohort 1, LSIA 1 and cohort 1, LSIA 2.

**Table A1: Selected summary statistics about those undertaking the assessment**

| | Vocational Education | Tertiary Education |
|---|---|---|
| **Licensed occupation** | 60.1% | 53.3% |
| **Skilled independent visa** | 76.7% | 73.1% |
| **Family reunification visa** | 14.3% | 24.0% |
| **Employer nomination visa** | 21.4% | 20.1% |
| **Education completed in** | | |
| *ESB* | 48.6% | 39.0% |
| *EU/EEA* | 25.1% | 35.2% |
| *Russia/Other EU* | 21.8% | 51.8% |
| *MENA* | 23.1% | 46.1% |
| *South East Asia* | 27.6% | 42.6% |
| *East Asia* | 28.2% | 39.9% |
| *South Asia* | 68.2% | 59.6% |
| *Latin America* | 34.7% | 56.3% |
| **Age (years)** | 34.8 | 35.0 |
| | (not assessed: 35.8) | (not assessed: 36.5) |
| **Female** | 20.5% | 39.7% |
| **Male** | 42.7% | 47.9% |
| **Gross weekly wage (ln)** | 6.27 | 6.35 |
| | (not assessed: 6.06) | (not assessed: 6.47) |
| **Over-educated** | 38.4% | 52.1% |
| | (not assessed: 45.5%) | (not assessed: 45.2%) |

**Table 1: Sample characteristics – Primary applicant aged 20 to 65**

| | Cohort 1 Wave 1 | | Cohort 2 Wave 1 | |
|---|---|---|---|---|
| Age | 34.9 | (9.4) | 35.5 | (9.8) |
| Experience | 18.2 | (10.4) | 18.7 | (11.2) |
| Experience 2 | 439.4 | (531.0) | 474.4 | (564.7) |
| Gender (Female) | .420 | (.493) | .458 | (.498) |
| Married | .731 | (.440) | .712 | (.453) |
| No. of members in household | 3.47 | (1.70) | 3.49 | (1.72) |
| Education: Bachelor or higher | .398 | (.489) | .406 | (.491) |
| Education: Vocational diploma/Certificate | .293 | (.455) | .274 | (.446) |
| Education: 12 or less years of education | .309 | (.462) | .320 | (.466) |
| Qualification assessed | .293 | (.455) | .192 | (.394) |
| Foreign Education*: UK, IRE, US, CAN, SA (ref) | .275 | (.446) | .276 | (.428) |
| Foreign Education: EU/EEA | .093 | (.291) | .104 | (.273) |
| Foreign Education: Russia, fmr USSR, Oth Europe | .165 | (.371) | .159 | (.283) |
| Foreign Education: MENA | .109 | (.312) | .075 | (.090) |
| Foreign Education: SE Asia and Pacific Islands | .102 | (.303) | .118 | (.370) |
| Foreign Education: China, East Asia | .117 | (.322) | .136 | (.308) |
| Foreign Education: South Asia | .090 | (.286) | .097 | (.283) |
| Foreign Education: Latin America | .049 | (.215) | .035 | (.194) |
| Visited Australia before immigration | .464 | (.499) | .538 | (.499) |
| Visa type: Preferential Family/Family Stream | .312 | (.464) | .426 | (.495) |
| Visa type: Concessional Family/Austr. Link | .216 | (.412) | .128 | (.334) |
| Visa type: Business Skills & Empl. Nom. Scheme | .134 | (.341) | .151 | (.358) |
| Visa type: Skilled Independent | .227 | (.419) | .153 | (.360) |
| Visa type: Humanitarian | .109 | (.312) | .141 | (.348) |
| HH owns car | .810 | (.392) | .638 | (.480) |
| Funds at time of immigration (log) | 6.33 | (4.18) | 6.51 | (4.49) |
| Currently employed | .568 | (.495) | .600 | (.490) |
| Average wage (log) | 6.28 | (.788) | 6.38 | (.817) |
| Spoken English*: very good or excellent (ref) | .453 | (.498) | .515 | (.500) |
| Spoken English: good | .270 | (.444) | .230 | (.421) |
| Spoken English: poor or do not speak | .277 | (.448) | .251 | (.451) |
| Educ. Mismatch AU: Over-educated | .288 | (.453) | .269 | (.443) |
| Educ. Mismatch AU: Correctly matched | .633 | (.482) | .627 | (.484) |
| Educ. Mismatch AU: Under-educated | .078 | (.269) | .104 | (.305) |
| Region of residence: New South Wales, ACT | .449 | (.497) | .427 | (.495) |
| Region of residence: Victoria | .233 | (.422) | .223 | (.416) |
| Region of residence: Queensland | .112 | (.316) | .104 | (.305) |
| Region of residence: S Australia, Tasmania | .069 | (.253) | .077 | (.267) |
| Region of residence: Western Australia, NT | .136 | (.343) | .170 | (.375) |
| Max Number of observations | 8,205 | | 3,166 | |

**Table 2a: Correlations between endogenous variable (Qual_as) and instrument (Avg-info), by type of education.**

| | Vocational Education | | Tertiary Education | |
|---|---|---|---|---|
| | *Wage* | *Over-education* | *Wage* | *Over-education* |
| **Qual_as** | .1433 | -.0532 | -.0754 | .0672 |
| **Avg_info** | .1611 | -.0832 | -.0335 | .0283 |
| | **Correlation with Qual_as** | | | |
| **Avg_info** | .5447 | | .4754 | |

**Table 2b: Tests of endogeneity, by type of education.**

| | Vocational Education | | Tertiary Education | |
|---|---|---|---|---|
| | *All visa classes* | *Exclude ENS and Refugees* | *All visa classes* | *Exclude ENS and Refugees* |
| **Qual_as (OLS, no residual from 1st stage)** | .050<br>(.047) | .072<br>(.052) | .023<br>(.039) | -.010<br>(.050) |
| **Qual_as** | .361***<br>(.107) | .423***<br>(.109) | .599***<br>(.115) | .564***<br>(.121) |
| **Coefficient of 1st stage with residual** | -.147***<br>(.048) | -.166***<br>(.049) | -.258***<br>(.048) | -.266***<br>(.051) |

**Table 3: Regression results: wage equation without interactions[A], by type and region of education**

| | Pooled | ESB | NESB | EU/EEA | Other Europe | MENA | South E Asia | East Asia | South Asia | Latin America |
|---|---|---|---|---|---|---|---|---|---|---|
| **Vocational Education** | | | | | | | | | | |
| ***Qual-as IV*** | **.326***** | **.444***** | **.245**** | **-.318** | **.238** | **.462** | **.434** | **.472** | **.922**** | **.003** |
| | **(.126)** | **(.133)** | **(.124)** | **(.294)** | **(.273)** | **(.492)** | **(.283)** | **(.431)** | **(.416)** | **(.383)** |
| ***Adj $R^2$*** | .2182 | .0887 | .1653 | .1861 | .2057 | .2114 | .2594 | .2085 | .0737 | .4502 |
| ***N Obs*** | 2,456 | 900 | 1,556 | 321 | 311 | 153 | 186 | 235 | 235 | 115 |
| ***F*** | 97.2 | 83.3 | 120.9 | 13.5 | 35.1 | 12.6 | 37.0 | 13.0 | 10.6 | 6.0 |
| ***Endog (p-value)*** | .0224 | .0043 | .2056 | .0647 | .1915 | .1821 | .5781 | .5485 | .0484 | .5236 |
| **Tertiary Education** | | | | | | | | | | |
| ***Qual-as IV*** | **.383***** | **.247*** | **.651***** | **.586** | **1.023**** | **.997** | **.669***** | **.628*** | **.251** | **1.745***** |
| | **(.126)** | **(.132)** | **(.148)** | **(.464)** | **(.522)** | **(1.34)** | **(.206)** | **(.352)** | **(.322)** | **(.619)** |
| ***Adj $R^2$*** | .2276 | .2020 | .1055 | .1858 | .0133 | B | .1376 | .1433 | .2078 | B |
| ***N Obs*** | 3,499 | 1,400 | 2,099 | 331 | 362 | 193 | 434 | 366 | 304 | 109 |
| ***F*** | 120.3 | 94.0 | 123.3 | 10.8 | 18.9 | 1.7 | 54.8 | 16.8 | 16.8 | 10.6 |
| ***Endog (p-value)*** | .0000 | .0609 | .0000 | .1022 | .0103 | .3848 | .0092 | .0342 | .6431 | .0002 |

[A]: interactions are instrumented in regressions on pooled data.
[B]: negative adjusted $R^2$
Notes: The dependent variable is the log of the weekly wage. The symbol * means that the estimate is significantly different from zero at 10% significance level; ** at 5%; *** at 1%. Pooled includes dummies using ESB as reference group. Standard errors are clustered at individual level and reported in parenthesis. Regression performed using the General Methods of Moments (GMM). The F test tests the null hypothesis that the coefficients of the exclusion restrictions in the first stage regression are statistically equal to zero. A value lower than 10 supports that the instruments are weak and hence the estimates are biased towards what obtained using OLS. The endogeneity test is the difference-in-Sargan, or C, statistics discussed by Hayashi (2000). The null hypothesis is that the variable qual_as is exogenous.
Source: cohort 1-3, LSIA 1 and cohort 1-2, LSIA 2.

## Table 4a: Wage Equations with interactions, by region and type of education

| Variables | Vocational Education OLS | Vocational Education IV OLS | Tertiary Education OLS | Tertiary Education IV OLS |
|---|---|---|---|---|
| **Qualifications** | 0.328*** | 0.326** | 0.589*** | 0.383*** |
| **Assessed** | (0.112) | (0.127) | (0.118) | (0.126) |
| ***Country of highest education*** | | | | |
| EU/EEA | -0.126* | 0.0118 | -0.0817 | -0.127 |
| | (0.0697) | (0.108) | (0.0601) | (0.124) |
| Russia/Central | -0.254*** | -0.228** | -0.431*** | -0.689*** |
| Europe | (0.0844) | (0.114) | (0.0976) | (0.210) |
| MENA | -0.364*** | -0.275* | -0.746*** | -0.870** |
| | (0.105) | (0.158) | (0.120) | (0.347) |
| South East Asia | -0.238** | -0.111 | -0.474*** | -0.480*** |
| | (0.0991) | (0.139) | (0.0713) | (0.132) |
| China/E Asia | -0.343*** | -0.405*** | -0.422*** | -0.427*** |
| | (0.0935) | (0.140) | (0.0759) | (0.135) |
| South Asia | -0.485*** | -0.626*** | -0.527*** | -0.467*** |
| | (0.161) | (0.195) | (0.0824) | (0.146) |
| Latin America | -0.564*** | -0.644*** | -0.541*** | -1.259*** |
| | (0.138) | (0.240) | (0.135) | (0.397) |
| ***Interactions assessment with country of highest education*** | | | | |
| Qual x EU/EEA | 0.113 | -0.332 | -0.0034 | 0.146 |
| | (0.103) | (0.267) | (0.0916) | (0.358) |
| Qual x Russia/ | -0.165 | -0.175 | 0.0710 | 0.466 |
| Central Europe | (0.128) | (0.234) | (0.113) | (0.295) |
| Qual x MENA | -0.147 | -0.360 | 0.0521 | 0.253 |
| | (0.168) | (0.361) | (0.148) | (0.583) |
| Qual x South East | 0.0970 | -0.221 | 0.0798 | 0.0852 |
| Asia | (0.115) | (0.218) | (0.0883) | (0.217) |
| Qual x China/E | 0.166 | 0.413 | -0.0892 | -0.0550 |
| Asia | (0.117) | (0.304) | (0.0970) | (0.278) |
| Qual x South Asia | 0.265 | 0.411* | 0.0466 | -0.0037 |
| | (0.165) | (0.239) | (0.106) | (0.238) |
| Qual x Latin | 0.413** | 0.623 | 0.258 | 1.282** |
| America | (0.161) | (0.472) | (0.166) | (0.522) |
| ***Controls*** | | | | |
| Female | -0.351*** | -0.346*** | -0.254*** | -0.266*** |
| | (0.0479) | (0.0477) | (0.0336) | (0.0367) |
| Experience | 0.0217** | 0.0209** | 0.0283*** | 0.0286*** |
| | (0.0094) | (0.0096) | (0.0080) | (0.0087) |
| Experience2 | -0.0005** | -0.0005** | -0.0009*** | -0.0010*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| Married | -0.0145 | -0.0124 | 0.100*** | 0.103*** |
| | (0.0361) | (0.0381) | (0.0334) | (0.0366) |
| Nr Household | -0.0396*** | -0.0423*** | -0.0348*** | -0.0365*** |
| members | (0.0112) | (0.0117) | (0.0113) | (0.0116) |
| English: good | -0.232*** | -0.215*** | -0.253*** | -0.277*** |
| | (0.0539) | (0.0565) | (0.0435) | (0.0473) |
| English: poor | -0.253*** | -0.252*** | -0.481*** | -0.507*** |
| | (0.0609) | (0.0633) | (0.0726) | (0.0781) |
| Arrived 1994 | -0.0404 | -0.0267 | 0.0703 | 0.0975* |
| | (0.0456) | (0.0478) | (0.0475) | (0.0505) |
| Arrived 1995 | -0.0293 | -0.0364 | 0.0503 | 0.0707 |
| | (0.0490) | (0.0507) | (0.0536) | (0.0575) |
| Arrived 1999 | 0.126* | 0.126* | 0.292*** | 0.314*** |

| | | | | |
|---|---|---|---|---|
| | (0.0660) | (0.0671) | (0.0591) | (0.0642) |
| Arrived 2000 | 0.139** | 0.138** | 0.232*** | 0.261*** |
| | (0.0551) | (0.0557) | (0.0560) | (0.0594) |
| Victoria | -0.122*** | -0.134*** | -0.136*** | -0.143*** |
| | (0.0425) | (0.0445) | (0.0379) | (0.0404) |
| Queensland | -0.200*** | -0.209*** | -0.168*** | -0.159*** |
| | (0.0593) | (0.0601) | (0.0487) | (0.0535) |
| SA and TAS | -0.377*** | -0.378*** | -0.256*** | -0.279*** |
| | (0.0967) | (0.0978) | (0.0773) | (0.0830) |
| WA and NT | -0.203*** | -0.204*** | -0.170*** | -0.175*** |
| | (0.0626) | (0.0647) | (0.0524) | (0.0562) |
| Selection into | -0.0119*** | -0.0125*** | -0.0171*** | -0.0194*** |
| participation | (0.0042) | (0.0043) | (0.0036) | (0.0037) |
| Visa: Empl. Nom | 0.368*** | 0.374*** | 0.499*** | 0.481*** |
| Scheme | (0.0614) | (0.0617) | (0.0450) | (0.0504) |
| Visa: refugee | -0.207** | -0.229** | -0.0171 | 0.0164 |
| | (0.0906) | (0.0916) | (0.101) | (0.112) |
| Residuals 1st stage | -0.130*** | | -0.252*** | |
| | (0.0504) | | (0.0493) | |
| Constant | 6.305*** | 6.268*** | 6.532*** | 6.515*** |
| | (0.184) | (0.184) | (0.152) | (0.149) |
| | | | | |
| Observations | 2,456 | 2,456 | 3,499 | 3,499 |
| R-squared | 0.262 | 0.224 | 0.315 | 0.228 |

Notes: Models are estimated by OLS. Standard errors are clustered at individual level and reported in parenthesis. Conventional significance notation is used: * p<.10, ** p<.05, *** p<.01.
Source: cohort 1, LSIA 1 and cohort 1, LSIA 2.

## Table 4b: Test of significance of wage penalty including interaction term

| | Vocational Education | | Tertiary Education | |
|---|---|---|---|---|
| | IV OLS | IV OLS | IV OLS | IV OLS |
| | $\beta_1$ | $\beta_1 + \beta_3$ | $\beta_1$ | $\beta_1 + \beta_3$ |
| EU/EEA | 0.0118 | -.320* | -0.0817 | .018 |
| | (0.108) | (.193) | (0.0601) | (.250) |
| Russia/other Europe | -0.228** | -.403** | -0.431*** | -.223* |
| | (0.114) | (.171) | (0.0976) | (.126) |
| MENA | -0.275* | -.634*** | -0.746*** | -.616** |
| | (0.158) | (.243) | (0.120) | (.263) |
| SE Asia | -0.111 | -.332*** | -0.474*** | -.395*** |
| | (0.139) | (.114) | (0.0713) | (.114) |
| China/E | -0.405*** | -.007 | -0.422*** | -.482*** |
| | (0.140) | (.210) | (0.0759) | (.170) |
| South Asia | -0.626*** | -.214*** | -0.527*** | -.471*** |
| | (0.195) | (.084) | (0.0824) | (.120) |
| Latin America | -0.644*** | -.021 | -0.541*** | .023 |
| | (0.240) | (.277) | (0.135) | (.194) |

Standard errors are clustered at individual level and reported in parenthesis. Conventional significance notation is used: * p<.10, ** p<.05, *** p<.01.
Source: LSIA 1 and LSIA 2.

**Table 5: Over-education equations without interactions[A], by type and region of education**

| | Pooled | ESB | NESB | EU/EEA | Other Europe | MENA | South E Asia | East Asia | South Asia | Latin America |
|---|---|---|---|---|---|---|---|---|---|---|
| **Vocational Education** | | | | | | | | | | |
| ***Qual-as IV*** | **-0.145*****<br>**(.055)** | **-.071**<br>**(.058)** | **-.255*****<br>**(.078)** | **.089**<br>**(.139)** | **-.415****<br>**(.185)** | **-.504**<br>**(.372)** | **-.317***<br>**(.168)** | **-.108**<br>**(.210)** | **-.059**<br>**(.174)** | **.067**<br>**(.289)** |
| ***Adj $R^2$*** | .1057 | .0358 | .931 | .1232 | .0423 | .0772 | .1759 | .0613 | .1321 | .3219 |
| ***N Obs*** | 2,457 | 920 | 1,537 | 340 | 282 | 143 | 196 | 251 | 220 | 105 |
| ***F*** | 103.5 | 86.2 | 105.7 | 11.9 | 28.6 | 8.5 | 46.9 | 11.5 | 10.9 | 7.0 |
| ***Endog (p-value)*** | .0025 | .1174 | .0072 | .5991 | .0710 | .2576 | .2461 | .7078 | .5041 | .6610 |
| **Tertiary Education** | | | | | | | | | | |
| ***Qual-as IV*** | **-.532*****<br>**(.101)** | **-.302*****<br>**(.109)** | **-.564*****<br>**(.094)** | **-.343**<br>**(.234)** | **-.964*****<br>**(.332)** | **-.809**<br>**(.664)** | **-.272****<br>**(.144)** | **-1.096*****<br>**(.327)** | **-.621*****<br>**(.228)** | **-.578****<br>**(.280)** |
| ***Adj $R^2$*** | .0978 | .0140 | .0839 | .1734 | [B] | [B] | .3069 | [B] | [B] | .1818 |
| ***N Obs*** | 3,439 | 1,372 | 2,067 | 332 | 351 | 187 | 423 | 388 | 284 | 102 |
| ***F*** | 105.2 | 81.4 | 119.7 | 10.0 | 12.1 | 3.3 | 42.4 | 14.2 | 28.1 | 8.8 |
| ***Endog (p-value)*** | .0000 | .0082 | .0000 | .1902 | .0041 | .0600 | .1419 | .0000 | .0079 | .1330 |

[A]: interactions are instrumented in regressions on pooled data.
[B]: negative adjusted $R^2$
Notes: The dependent variable is the probability of over-education. The symbol * means that the estimate is significantly different from zero at 10% significance level; ** at 5%; *** at 1%. Pooled includes dummies using ESB as reference group. Standard errors are clustered at individual level and reported in parenthesis. Regression performed using the General Methods of Moments (GMM). The F test tests the null hypothesis that the coefficients of the exclusion restrictions in the first stage regression are statistically equal to zero. A value lower than 10 supports that the instruments are weak and hence the estimates are biased towards what obtained using OLS. The endogeneity test is the difference-in-Sargan, or C, statistics discussed by Hayashi (2000). The null hypothesis is that the variable qual_as is exogenous.
Source: cohort 1-3, LSIA 1 and cohort 1-2, LSIA 2.

**Table 6a: Over-education equations with interactions, by region and type of education**

| | Vocational Education | | Tertiary Education | |
|---|---|---|---|---|
| Over-education | OLS | IV OLS | OLS | IV OLS |
| Qualifications Assessed | -0.177*** | -0.145*** | -0.565*** | -0.532*** |
| | (0.0615) | (0.0551) | (0.0817) | (0.102) |
| ***Country of highest education*** | | | | |
| EU/EEA | -0.0301 | -0.0441 | -0.0037 | -0.0214 |
| | (0.0299) | (0.0495) | (0.0353) | (0.0778) |
| Russia/Central Europe | 0.132** | 0.167** | 0.184*** | 0.264** |
| | (0.0520) | (0.0692) | (0.0555) | (0.118) |
| MENA | 0.138** | 0.215** | 0.137** | 0.172 |
| | (0.0630) | (0.106) | (0.0673) | (0.192) |
| South East Asia | 0.175*** | 0.153** | 0.335*** | 0.127** |
| | (0.0521) | (0.0695) | (0.0406) | (0.0632) |
| China/E Asia | -0.0412 | -0.0688 | 0.125*** | 0.151* |
| | (0.0409) | (0.0513) | (0.0429) | (0.0866) |
| South Asia | 0.0520 | 0.0652 | 0.301*** | 0.292*** |
| | (0.0645) | (0.135) | (0.0655) | (0.111) |
| Latin America | 0.173** | 0.322** | 0.379*** | 0.546** |
| | (0.0743) | (0.152) | (0.0758) | (0.269) |
| ***Interactions assessment with country of highest education*** | | | | |
| Qual x EU/EEA | -0.0217 | 0.0141 | -0.0133 | -0.0035 |
| | (0.0466) | (0.121) | (0.0619) | (0.224) |
| Qual x Russia/ Central Europe | -0.115* | -0.225* | -0.0777 | -0.164 |
| | (0.0644) | (0.133) | (0.0671) | (0.179) |
| Qual x MENA | -0.0367 | -0.219 | 0.141* | 0.125 |
| | (0.0788) | (0.194) | (0.0830) | (0.309) |
| Qual x South East Asia | -0.0159 | 0.0483 | 0.0896 | 0.521*** |
| | (0.0805) | (0.140) | (0.0565) | (0.127) |
| Qual x China/E Asia | 0.0316 | 0.0822 | -0.0447 | -0.131 |
| | (0.0599) | (0.142) | (0.0614) | (0.191) |
| Qual x South Asia | 0.0450 | 0.0371 | -0.0437 | 0.0087 |
| | (0.0751) | (0.164) | (0.0782) | (0.168) |
| Qual x Latin America | -0.115 | -0.465 | -0.160 | -0.336 |
| | (0.0978) | (0.311) | (0.105) | (0.382) |
| ***Controls*** | | | | |
| Female | -0.0166 | -0.0196 | 0.0432* | 0.0545** |
| | (0.0219) | (0.0240) | (0.0224) | (0.0251) |
| Experience | -0.0052 | -0.0038 | -0.0050 | -0.0048 |
| | (0.0047) | (0.0049) | (0.0046) | (0.0056) |
| Experience2 | .00008 | .00006 | 0.0001 | 0.0002 |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| Married | -0.0178 | -0.0217 | -0.0634*** | -0.0545** |
| | (0.0199) | (0.0209) | (0.0205) | (0.0242) |
| Nr Household members | 0.0164** | 0.0171*** | 0.0076 | 0.00769 |
| | (0.0063) | (0.0065) | (0.0063) | (0.00734) |
| English: good | 0.0534* | 0.0567* | 0.0987*** | 0.109*** |
| | (0.0291) | (0.0308) | (0.0259) | (0.0299) |
| English: poor | 0.129*** | 0.135*** | 0.187*** | 0.235*** |
| | (0.0364) | (0.0383) | (0.0402) | (0.0447) |
| Arrived 1994 | -0.0188 | -0.0122 | -0.0118 | -0.00739 |
| | (0.0274) | (0.0281) | (0.0293) | (0.0348) |
| Arrived 1995 | -0.0375 | -0.0238 | -0.0504 | -0.0354 |
| | (0.0289) | (0.0294) | (0.0312) | (0.0382) |
| Arrived 1999 | -0.161*** | -0.166*** | -0.0933** | -0.109** |
| | (0.0307) | (0.0330) | (0.0419) | (0.0485) |

| | | | | |
|---|---|---|---|---|
| Arrived 2000 | -0.147*** | -0.139*** | -0.0513 | -0.0498 |
| | (0.0294) | (0.0306) | (0.0366) | (0.0423) |
| Victoria | -0.00135 | 0.00523 | 0.0199 | 0.0145 |
| | (0.0243) | (0.0256) | (0.0237) | (0.0269) |
| Queensland | -0.0187 | -0.00385 | 0.0030 | 0.0007 |
| | (0.0282) | (0.0291) | (0.0335) | (0.0385) |
| SA and TAS | 0.0284 | 0.0427 | 0.0514 | 0.0756 |
| | (0.0424) | (0.0445) | (0.0497) | (0.0583) |
| Selection into | 0.00117 | 0.00680 | 0.0522 | 0.0583 |
| participation | (0.0288) | (0.0298) | (0.0403) | (0.0451) |
| Visa: Empl. Nom | -0.158*** | -0.170*** | -0.446*** | -0.468*** |
| Scheme | (0.0241) | (0.0274) | (0.0292) | (0.0364) |
| Visa: refugee | 0.0571 | 0.0528 | 0.124*** | 0.131** |
| | (0.0534) | (0.0551) | (0.0458) | (0.0562) |
| Residual 1st stage | 0.0723*** | | 0.225*** | |
| | (0.0261) | | (0.0333) | |
| Constant | 0.338*** | 0.244*** | 0.821*** | 0.811*** |
| | (0.0921) | (0.0881) | (0.111) | (0.120) |
| Observations | 2,457 | 2,457 | 3,439 | 3,439 |
| R-squared | 0.158 | 0.106 | 0.305 | 0.098 |

Notes: Models are estimated by linear regression model (OLS). Standard errors are clustered at individual level and reported in parenthesis. Conventional significance notation is used: * p<.10, ** p<.05, *** p<.01.
Source: cohort 1, LSIA 1 and cohort 1, LSIA 2.

## Table 6b: Test of significance of job quality mismatch including interaction term

| | Vocational Education | | Tertiary Education | |
|---|---|---|---|---|
| | **IV OLS** | **IV OLS** | **IV OLS** | **IV OLS** |
| | $b_1$ | $b_1 + b_3$ | $b_1$ | $b_1 + b_3$ |
| EU/EEA | -0.0441 | -.030 | -0.0037 | -.024 |
| | (0.0495) | (.085) | (0.0353) | (.158) |
| Russia/other Europe | 0.167** | -.058 | 0.184*** | .100 |
| | (0.0692) | (.096) | (0.0555) | (.088) |
| MENA | 0.215** | .004 | 0.137** | .297** |
| | (0.106) | (.011) | (0.0673) | (.138) |
| SE Asia | 0.153** | .202** | 0.335*** | .648*** |
| | (0.0695) | (.097) | (0.0406) | (.082) |
| China/E | -0.0688 | .013 | 0.125*** | .020 |
| | (0.0513) | (.113) | (0.0429) | (.124) |
| South Asia | 0.0652 | .102** | 0.301*** | .301*** |
| | (0.135) | (.051) | (.0655) | (.083) |
| Latin America | 0.322** | -.142 | 0.379*** | .209 |
| | (.152) | (.187) | (.0758) | (.146) |

Standard errors are clustered at individual level and reported in parenthesis. Conventional significance notation is used: * p<.10, ** p<.05, *** p<.01.
Source: cohort 1, LSIA 1 and cohort 1, LSIA 2.

## Table 7: Vocational Education: wage equations without interactions [A], by region of education

| | Same Employer | | | | All Employers | |
|---|---|---|---|---|---|---|
| | OLS IV wave 1 | OLS IV waves 2,3 | OLS IV | Panel IV | OLS IV | Panel IV |
| Qual_as | 0.262* | 0.113 | 0.277* | 0.233* | 0.305** | 0.276** |
| | (0.153) | (0.237) | (0.161) | (0.142) | (0.140) | (0.129) |
| EU/EEA | -0.0211 | -0.0431 | -0.0031 | -0.0385 | 0.0407 | 0.0383 |
| | (0.149) | (0.160) | (0.135) | (0.120) | (0.117) | (0.109) |
| Russia/other Europe | -0.250* | -0.242 | -0.241* | -0.257** | -0.259** | -0.263** |
| | (0.146) | (0.205) | (0.144) | (0.130) | (0.122) | (0.111) |
| MENA | -0.311 | -0.184 | -0.307 | -0.312* | -0.356** | -0.362** |
| | (0.196) | (0.379) | (0.204) | (0.165) | (0.172) | (0.143) |
| SE Asia | -0.202 | 0.0198 | -0.159 | -0.203 | -0.101 | -0.121 |
| | (0.178) | (0.234) | (0.170) | (0.136) | (0.149) | (0.123) |
| China/E Asia | -0.500** | -0.450 | -0.557*** | -0.530*** | -0.467*** | -0.445*** |
| | (0.198) | (0.314) | (0.205) | (0.148) | (0.150) | (0.122) |
| South Asia | -0.472* | -0.443* | -0.425** | -0.394* | -0.620*** | -0.625*** |
| | (0.260) | (0.249) | (0.208) | (0.204) | (0.197) | (0.185) |
| Latin America | -0.840** | -0.467 | -0.824** | -0.726*** | -0.688*** | -0.692*** |
| | (0.343) | (0.537) | (0.364) | (0.222) | (0.234) | (0.162) |
| **Interactions with qualification assessment** | | | | | | |
| Qual x EU/EEA | -0.189 | -0.0373 | -0.156 | -0.0726 | -0.320 | -0.340 |
| | (0.351) | (0.343) | (0.289) | (0.271) | (0.265) | (0.248) |
| Qual x Russia/ | -0.109 | 0.0505 | -0.0793 | -0.0379 | -0.184 | -0.200 |
| Central Europe | (0.378) | (0.336) | (0.286) | (0.228) | (0.238) | (0.187) |
| Qual x MENA | -0.374 | -0.370 | -0.430 | -0.433 | -0.389 | -0.391 |
| | (0.385) | (0.808) | (0.409) | (0.311) | (0.357) | (0.253) |
| Qual x South East | -0.0314 | -0.340 | -0.176 | -0.0702 | -0.208 | -0.165 |
| Asia | (0.280) | (0.377) | (0.259) | (0.234) | (0.218) | (0.194) |
| Qual x China/E | 0.430 | 1.111 | 0.526 | 0.431 | 0.342 | 0.240 |
| Asia | (0.410) | (0.936) | (0.435) | (0.304) | (0.301) | (0.245) |
| Qual x South Asia | 0.309 | 0.175 | 0.209 | 0.175 | 0.414* | 0.428* |
| | (0.310) | (0.334) | (0.250) | (0.255) | (0.239) | (0.228) |
| Qual x Latin | 0.979 | 1.036 | 0.946 | 0.665 | 0.611 | 0.606* |
| America | (0.752) | (1.549) | (0.836) | (0.495) | (0.481) | (0.317) |
| **Interactions with time** | | | | | | |
| Time x qual_as | | | -0.0278 | -0.0392 | 0.0160 | 0.0288 |
| | | | (0.0553) | (0.0704) | (0.0423) | (0.0517) |
| Time x EU/EEA | | | -0.0068 | -0.0094 | -0.0219 | -0.0189 |
| | | | (0.0336) | (0.0474) | (0.0321) | (0.0393) |
| Time x Russia/ | | | 0.0120 | 0.0080 | 0.0218 | 0.0256 |
| Central Europe | | | (0.0497) | (0.0589) | (0.0413) | (0.0398) |
| Time x MENA | | | 0.0664 | 0.0633 | 0.0613 | 0.0638 |
| | | | (0.0723) | (0.0669) | (0.0453) | (0.0483) |
| Time x South East | | | 0.0447 | 0.0392 | -0.0094 | -0.0099 |
| Asia | | | (0.0366) | (0.0574) | (0.0370) | (0.0454) |
| Time x China/E | | | 0.136*** | 0.130** | 0.0585 | 0.0638 |
| Asia | | | (0.0521) | (0.0555) | (0.0382) | (0.0429) |
| Time x South Asia | | | -0.0215 | -0.0124 | -0.0065 | -0.0110 |
| | | | (0.0378) | (0.0601) | (0.0265) | (0.0448) |
| Time x Latin | | | 0.108 | 0.100 | 0.0343 | 0.0356 |
| America | | | (0.106) | (0.0863) | (0.0540) | (0.0576) |
| Constant | 5.869*** | 6.940*** | 5.971*** | 6.168*** | 6.312*** | 6.325*** |
| | (0.293) | (0.316) | (0.267) | (0.200) | (0.196) | (0.179) |
| N | 971 | 650 | 1,621 | 1,621 | 2,456 | 2,456 |
| $R^2$ between | | | | .2726 | | .2522 |
| Adj/overall $R^2$ | 0.251 | 0.211 | 0.255 | .2728 | 0.2290 | .2379 |

[A]: interactions are instrumented in regressions on pooled data. [B]: negative adjusted $R^2$

Notes: The dependent variable is the logarithm of the gross weekly wage. The symbol * means that the estimate is significantly different from zero at 10% significance level; ** at 5%; *** at 1%. Pooled includes dummies using ESB as reference group. Standard errors are clustered at individual level and reported in parenthesis. Regression performed using the General Methods of Moments (GMM). The F test tests the null hypothesis that the coefficients of the exclusion restrictions in the first stage regression are statistically equal to zero. The The null hypothesis of the endogeneity test is that the variable qual_as is exogenous.

Source: LSIA 1 and LSIA 2, all waves

## Table 8: Tertiary Education: wage equations without interactions [A], by region of education

| | Same Employer | | | | All Employers | |
|---|---|---|---|---|---|---|
| | OLS IV wave 1 | OLS IV waves 2,3 | OLS IV | Panel IV | OLS IV | Panel IV |
| Qual_as | 0.315** | 0.292 | 0.307** | 0.326** | 0.417*** | 0.409*** |
| | (0.146) | (0.180) | (0.139) | (0.143) | (0.138) | (0.127) |
| EU/EEA | -0.165 | -0.389 | -0.337* | -0.215 | -0.116 | -0.0955 |
| | (0.144) | (0.270) | (0.185) | (0.140) | (0.128) | (0.103) |
| Russia/other Europe | -0.643** | -0.412 | -0.608** | -0.593*** | -0.687*** | -0.615*** |
| | (0.269) | (0.289) | (0.243) | (0.178) | (0.212) | (0.136) |
| MENA | -1.601*** | -2.152 | -1.740** | -1.132*** | -0.876** | -0.844*** |
| | (0.545) | (2.934) | (0.874) | (0.353) | (0.341) | (0.233) |
| SE Asia | -0.740*** | -0.436** | -0.668*** | -0.693*** | -0.564*** | -0.571*** |
| | (0.205) | (0.176) | (0.163) | (0.136) | (0.133) | (0.102) |
| China/E Asia | -0.596*** | -0.299 | -0.467*** | -0.504*** | -0.439*** | -0.464*** |
| | (0.189) | (0.236) | (0.174) | (0.157) | (0.135) | (0.114) |
| South Asia | -0.604*** | -0.360 | -0.523** | -0.639*** | -0.468*** | -0.439*** |
| | (0.206) | (0.259) | (0.216) | (0.186) | (0.153) | (0.132) |
| Latin America | -1.380** | -0.997** | -1.393*** | -1.017*** | -1.320*** | -1.238*** |
| | (0.638) | (0.503) | (0.483) | (0.373) | (0.389) | (0.260) |
| **Interactions with qualification assessment** | | | | | | |
| Qual x EU/EEA | 0.220 | 1.297 | 0.659 | 0.362 | 0.174 | 0.110 |
| | (0.392) | (0.982) | (0.515) | (0.358) | (0.361) | (0.241) |
| Qual x Russia/ | 0.326 | 0.296 | 0.294 | 0.261 | 0.487 | 0.380** |
| Central Europe | (0.391) | (0.435) | (0.349) | (0.254) | (0.301) | (0.186) |
| Qual x MENA | 1.402 | 3.018 | 1.745 | 0.682 | 0.290 | 0.236 |
| | (0.906) | (6.424) | (1.465) | (0.595) | (0.590) | (0.388) |
| Qual x South East | 0.329 | -0.00201 | 0.173 | 0.220 | 0.0525 | 0.0751 |
| Asia | (0.335) | (0.301) | (0.255) | (0.244) | (0.217) | (0.174) |
| Qual x China/E | 0.201 | -0.242 | 0.0279 | 0.129 | -0.0400 | 0.0262 |
| Asia | (0.374) | (0.603) | (0.375) | (0.324) | (0.280) | (0.218) |
| Qual x South Asia | 0.132 | -0.196 | 0.0404 | 0.215 | 0.0133 | -0.0295 |
| | (0.318) | (0.472) | (0.317) | (0.270) | (0.238) | (0.184) |
| Qual x Latin | 1.302 | 1.358* | 1.313** | 0.808 | 1.269** | 1.158*** |
| America | (0.875) | (0.701) | (0.641) | (0.501) | (0.525) | (0.333) |
| **Interactions with time** | | | | | | |
| Time x qual_as | | | 0.00626 | -0.0166 | -0.0178 | -0.00489 |
| | | | (0.0613) | (0.0437) | (0.0472) | (0.0505) |
| Time x EU/EEA | | | 0.0544 | 0.0301 | -0.0173 | -0.0157 |
| | | | (0.0355) | (0.0297) | (0.0244) | (0.0387) |
| Time x Russia/ | | | 0.0652 | 0.0717* | -0.00877 | -0.0104 |
| Central Europe | | | (0.0656) | (0.0385) | (0.0424) | (0.0375) |
| Time x MENA | | | 0.0374 | -0.0006 | -0.00768 | -0.00890 |
| | | | (0.158) | (0.0507) | (0.0567) | (0.0488) |
| Time x South East | | | 0.0617* | 0.0821*** | 0.0685** | 0.0640* |
| Asia | | | (0.0339) | (0.0302) | (0.0299) | (0.0359) |
| Time x China/E | | | -0.0019 | -0.0177 | 0.00381 | -0.000329 |
| Asia | | | (0.0428) | (0.0318) | (0.0362) | (0.0378) |
| Time x South Asia | | | -0.0074 | 0.0024 | -0.00888 | -0.0118 |
| | | | (0.0604) | (0.0396) | (0.0411) | (0.0429) |
| Time x Latin | | | 0.153** | 0.0761 | 0.0524 | 0.0474 |
| America | | | (0.0762) | (0.0624) | (0.0837) | (0.0645) |
| Constant | 6.881*** | 6.422*** | 6.404*** | 6.639*** | 6.355*** | 6.611*** |
| | (0.233) | (0.340) | (0.243) | (0.173) | (0.200) | (0.157) |
| N | 1,421 | 1,145 | 2,566 | 2,566 | 3,499 | 3,499 |
| $R^2$ between | | | | .2517 | | .2651 |
| Adj/overall $R^2$ | 0.189 | | 0.158 | .2582 | .2250 | .2492 |

[A]: interactions are instrumented in regressions on pooled data. [B]: negative adjusted $R^2$

Notes: The dependent variable is the logarithm of the gross weekly wage. The symbol * means that the estimate is significantly different from zero at 10% significance level; ** at 5%; *** at 1%. Pooled includes dummies using ESB as reference group. Standard errors are clustered at individual level and reported in parenthesis. Regression performed using the General Methods of Moments (GMM). The F test tests the null hypothesis that the coefficients of the exclusion restrictions in the first stage regression are statistically equal to zero. The The null hypothesis of the endogeneity test is that the variable qual_as is exogenous.

Source: LSIA 1 and LSIA 2, all waves

**Table 9: Vocational Education: over-education equations without interactions [A], by region of education**

| | Same Employer | | | | All Employers | |
|---|---|---|---|---|---|---|
| | **OLS IV wave 1** | **OLS IV waves 2,3** | **OLS IV** | **Panel IV** | **OLS IV** | **Panel IV** |
| Qual_as | -0.0894 | -0.0736 | -0.0203 | -0.0188 | -0.0502 | -0.0525 |
| | (0.0641) | (0.107) | (0.0665) | (0.0755) | (0.0630) | (0.0680) |
| EU/EEA | -0.0573 | -0.208 | -0.0161 | -0.0197 | -0.0337 | -0.0379 |
| | (0.0705) | (0.250) | (0.0587) | (0.157) | (0.0490) | (0.0579) |
| Russia/other Europe | 0.141 | -0.272* | 0.115 | -0.238* | 0.158** | 0.150** |
| | (0.0873) | (0.162) | (0.0793) | (0.127) | (0.0744) | (0.0606) |
| MENA | 0.0518 | -0.778 | 0.123 | 0.0181 | 0.193* | 0.165** |
| | (0.112) | (1.128) | (0.124) | (0.226) | (0.110) | (0.0827) |
| SE Asia | 0.210** | 0.0649 | 0.189** | 0.0365 | 0.150* | 0.156** |
| | (0.0899) | (0.181) | (0.0828) | (0.115) | (0.0781) | (0.0630) |
| China/E Asia | -0.0564 | 0.202 | -0.0704 | 0.180 | -0.0763 | -0.0776 |
| | (0.0727) | (0.333) | (0.0708) | (0.148) | (0.0543) | (0.0599) |
| South Asia | 0.104 | -0.146 | 0.0731 | 0.0690 | 0.0233 | 0.0061 |
| | (0.172) | (0.358) | (0.165) | (0.155) | (0.131) | (0.106) |
| Latin America | 0.340* | -0.302 | 0.325 | -0.355 | 0.297** | 0.230*** |
| | (0.203) | (0.868) | (0.212) | (0.231) | (0.136) | (0.0890) |
| **Interactions with qualification assessment** | | | | | | |
| Qual x EU/EEA | 0.0809 | -0.0191 | -0.0139 | -0.0154 | 0.0348 | 0.0484 |
| | (0.184) | (0.0897) | (0.146) | (0.0666) | (0.123) | (0.134) |
| Qual x Russia/ | -0.370** | 0.0851 | -0.286** | 0.0969 | -0.217 | -0.202* |
| Central Europe | (0.185) | (0.125) | (0.131) | (0.0709) | (0.132) | (0.103) |
| Qual x MENA | 0.220 | 0.286 | -0.0563 | 0.0951 | -0.219 | -0.145 |
| | (0.229) | (0.490) | (0.242) | (0.105) | (0.193) | (0.153) |
| Qual x South East | 0.0327 | 0.0515 | 0.0556 | 0.192*** | 0.0533 | 0.0316 |
| Asia | (0.199) | (0.0985) | (0.163) | (0.0683) | (0.141) | (0.0993) |
| Qual x China/E | 0.0705 | -0.144 | 0.0802 | -0.101 | 0.147 | 0.144 |
| Asia | (0.218) | (0.0898) | (0.204) | (0.0707) | (0.149) | (0.122) |
| Qual x South Asia | -0.0437 | 0.346 | -0.0686 | -0.0335 | -0.0008 | 0.0240 |
| | (0.216) | (0.321) | (0.201) | (0.124) | (0.167) | (0.130) |
| Qual x Latin | -0.470 | 0.199 | -0.402 | 0.305*** | -0.438 | -0.264 |
| America | (0.486) | (0.239) | (0.496) | (0.114) | (0.314) | (0.162) |
| **Interactions with time** | | | | | | |
| Time x qual_as | | | -0.0665** | -0.0682** | -0.0511** | -0.0486* |
| | | | (0.0323) | (0.0344) | (0.0248) | (0.0260) |
| Time x EU/EEA | | | -0.0356* | -0.0360 | -0.0077 | -0.0069 |
| | | | (0.0185) | (0.0239) | (0.0155) | (0.0198) |
| Time x Russia/ | | | -0.0237 | -0.0226 | 0.0077 | 0.0084 |
| Central Europe | | | (0.0326) | (0.0302) | (0.0223) | (0.0207) |
| Time x MENA | | | -0.0510 | -0.0505 | 0.0159 | 0.0153 |
| | | | (0.0372) | (0.0360) | (0.0269) | (0.0256) |
| Time x South East | | | -0.0499* | -0.0485* | 0.0052 | 0.00637 |
| Asia | | | (0.0285) | (0.0283) | (0.0229) | (0.0230) |
| Time x China/E | | | -0.0260 | -0.0243 | 0.0004 | 0.0015 |
| Asia | | | (0.0233) | (0.0272) | (0.0174) | (0.0217) |
| Time x South Asia | | | 0.0995*** | 0.0929*** | 0.0447* | 0.0431* |
| | | | (0.0330) | (0.0325) | (0.0229) | (0.0241) |
| Time x Latin | | | -0.0579 | -0.0560 | 0.0171 | 0.0110 |
| America | | | (0.0775) | (0.0454) | (0.0360) | (0.0305) |
| Constant | 0.232 | 0.221 | 0.254** | 0.255** | 0.423*** | 0.280*** |
| | (0.142) | (0.166) | (0.110) | (0.105) | (0.111) | (0.0934) |
| N | 934 | 670 | 1,604 | 1,604 | 2,457 | 2,457 |
| $R^2$ between | | | | .1802 | | .1948 |
| Adj/overall $R^2$ | 0.153 | 0.086 | 0.135 | .1476 | 0.115 | .1363 |

[A]: interactions are instrumented in regressions on pooled data. [B]: negative adjusted $R^2$

Notes: The dependent variable is the logarithm of the gross weekly wage. The symbol * means that the estimate is significantly different from zero at 10% significance level; ** at 5%; *** at 1%. Pooled includes dummies using ESB as reference group. Standard errors are clustered at individual level and reported in parenthesis. Regression performed using the General Methods of Moments (GMM). The F test tests the null hypothesis that the coefficients of the exclusion restrictions in the first stage regression are statistically equal to zero. The The null hypothesis of the endogeneity test is that the variable qual_as is exogenous.

Source: LSIA 1 and LSIA 2, all waves

## Table 10: Tertiary Education: over-education equations without interactions [A], by region of education

| | Same Employer | | | | All Employers | |
|---|---|---|---|---|---|---|
| | OLS IV wave 1 | OLS IV waves 2,3 | OLS IV | Panel IV | OLS IV | Panel IV |
| Qual_as | -0.509*** | -0.478*** | -0.423*** | -0.388*** | -0.531*** | -0.533*** |
| | (0.119) | (0.158) | (0.106) | (0.0894) | (0.104) | (0.0858) |
| EU/EEA | 0.0290 | -0.0166 | 0.0763 | 0.0238 | 0.0207 | 0.0144 |
| | (0.0985) | (0.126) | (0.102) | (0.0793) | (0.0771) | (0.0658) |
| Russia/other Europe | 0.222* | 0.0207 | 0.134 | 0.174 | 0.259** | 0.231** |
| | (0.133) | (0.164) | (0.129) | (0.109) | (0.117) | (0.0921) |
| MENA | 0.298 | 0.548 | 0.439 | 0.105 | 0.202 | 0.150 |
| | (0.237) | (0.992) | (0.321) | (0.180) | (0.187) | (0.146) |
| SE Asia | 0.147** | 0.108 | 0.152** | 0.173** | 0.152** | 0.158** |
| | (0.0732) | (0.117) | (0.0731) | (0.0747) | (0.0619) | (0.0634) |
| China/E Asia | 0.170* | 0.0182 | 0.141 | 0.0799 | 0.158* | 0.139** |
| | (0.101) | (0.138) | (0.0972) | (0.0840) | (0.0852) | (0.0706) |
| South Asia | 0.315*** | 0.351 | 0.365*** | 0.383*** | 0.324*** | 0.295*** |
| | (0.116) | (0.278) | (0.141) | (0.110) | (0.110) | (0.0887) |
| Latin America | 0.483 | 0.602 | 0.591* | 0.589** | 0.579** | 0.535*** |
| | (0.362) | (0.403) | (0.323) | (0.263) | (0.277) | (0.206) |
| **Interactions with qualification assessment** | | | | | | |
| Qual x EU/EEA | 0.0157 | -0.314 | -0.139 | 0.0128 | 0.0150 | 0.0352 |
| | (0.277) | (0.461) | (0.296) | (0.206) | (0.222) | (0.154) |
| Qual x Russia/ | -0.0767 | 0.0948 | 0.0109 | -0.0642 | -0.165 | -0.120 |
| Central Europe | (0.210) | (0.284) | (0.200) | (0.161) | (0.181) | (0.127) |
| Qual x MENA | 0.0764 | -1.033 | -0.229 | 0.283 | 0.132 | 0.223 |
| | (0.383) | (2.208) | (0.509) | (0.291) | (0.310) | (0.227) |
| Qual x South East | 0.528*** | 0.525** | 0.520*** | 0.461*** | 0.525*** | 0.511*** |
| Asia | (0.156) | (0.220) | (0.149) | (0.138) | (0.129) | (0.109) |
| Qual x China/E | 0.00266 | -0.104 | -0.0450 | 0.0744 | -0.124 | -0.0756 |
| Asia | (0.218) | (0.324) | (0.218) | (0.177) | (0.192) | (0.137) |
| Qual x South Asia | 0.0281 | -0.0791 | -0.0248 | -0.0562 | 0.0229 | 0.0685 |
| | (0.190) | (0.403) | (0.215) | (0.160) | (0.169) | (0.123) |
| Qual x Latin | -0.158 | -0.564 | -0.301 | -0.316 | -0.319 | -0.256 |
| America | (0.514) | (0.669) | (0.439) | (0.346) | (0.374) | (0.265) |
| **Interactions with time** | | | | | | |
| Time x qual_as | | | -0.0675* | -0.0252 | -0.0017 | -0.00537 |
| | | | (0.0401) | (0.0276) | (0.0258) | (0.0315) |
| Time x EU/EEA | | | -0.0526** | -0.0325* | -0.0331* | -0.0336 |
| | | | (0.0247) | (0.0181) | (0.0191) | (0.0240) |
| Time x Russia/ | | | -0.0117 | -0.0132 | 0.0034 | 0.00358 |
| Central Europe | | | (0.0360) | (0.0245) | (0.0196) | (0.0240) |
| Time x MENA | | | -0.0695 | -0.0132 | -0.0216 | -0.0220 |
| | | | (0.0604) | (0.0335) | (0.0275) | (0.0304) |
| Time x South East | | | -0.0199 | -0.0327* | -0.0178 | -0.0162 |
| Asia | | | (0.0213) | (0.0185) | (0.0152) | (0.0221) |
| Time x China/E | | | -0.0331 | -0.0077 | -0.0060 | -0.00654 |
| Asia | | | (0.0259) | (0.0190) | (0.0195) | (0.0227) |
| Time x South Asia | | | -0.0286 | -0.0561** | -0.0279 | -0.0281 |
| | | | (0.0399) | (0.0254) | (0.0246) | (0.0282) |
| Time x Latin | | | -0.0724 | -0.0655 | -0.0345 | -0.0325 |
| America | | | (0.0743) | (0.0402) | (0.0401) | (0.0421) |
| Constant | 0.657*** | 0.924*** | 0.697*** | 0.620*** | 0.891*** | 0.785*** |
| | (0.160) | (0.232) | (0.134) | (0.112) | (0.153) | (0.104) |
| N | 1,344 | 1,107 | 2,451 | 2,451 | 3,439 | 3,439 |
| $R^2$ between | | | | .2480 | | .2075 |
| Adj/overall $R^2$ | 0.175 | | 0.101 | .2406 | 0.102 | .1946 |

[A]: interactions are instrumented in regressions on pooled data. [B]: negative adjusted $R^2$

Notes: The dependent variable is the logarithm of the gross weekly wage. The symbol * means that the estimate is significantly different from zero at 10% significance level; ** at 5%; *** at 1%. Pooled includes dummies using ESB as reference group. Standard errors are clustered at individual level and reported in parenthesis. Regression performed using the General Methods of Moments (GMM). The F test tests the null hypothesis that the coefficients of the exclusion restrictions in the first stage regression are statistically equal to zero. The The null hypothesis of the endogeneity test is that the variable qual_as is exogenous.

Source: LSIA 1 and LSIA 2, all waves