

Discriminating between alternative measures of overeducation

Verhaest, Dieter; Omev, Eddy

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:

Verhaest, D., & Omev, E. (2006). Discriminating between alternative measures of overeducation. *Applied Economics*, 38(18), 2113-2120. <https://doi.org/10.1080/00036840500427387>

Nutzungsbedingungen:

Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: <http://www.peerproject.eu> Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under the "PEER Licence Agreement". For more information regarding the PEER-project see: <http://www.peerproject.eu> This document is solely intended for your personal, non-commercial use. All of the copies of this document must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der

Leibniz-Gemeinschaft



Discriminating between alternative measures of overeducation

Journal:	<i>Applied Economics</i>
Manuscript ID:	APE-04-0241.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	25-Aug-2005
JEL Code:	I21 - Analysis of Education < , J24 - Human Capital Skills Occupational Choice Labor Productivity <
Keywords:	overeducation, underemployment, mismatch, measurement, encompassing

powered by ScholarOne
Manuscript Central™

Discriminating between alternative measures of overeducation

Abstract

Five overeducation measures are evaluated empirically on the basis of encompassing tests. The measures are based on job analysis (JA), worker-assessment of the required level to do the job (WAd), worker-assessment of the required level to get the job (WAg), the mean educational level of realized matches (RMmn), and the modal level of realized matches (RMml). Over- and undereducation are linked to wages, job satisfaction, mobility and training participation. For none of the outcome variables, the JA model is encompassed by another model. Given the risk on systematic errors, this is a sufficient condition to prefer a carefully conducted JA to any other measure. The most reliable solution is to use the JA measure as an instrument for the WAd measure.

Keywords: overeducation, underemployment, mismatch, measurement, encompassing

I. Introduction

A large amount of literature on the economics of over- and undereducation has emerged over the past years. There is substantial evidence that several outcome variables such as wages, job satisfaction, mobility and training participation are related to overeducation. Despite this growing interest, no uniform way of measurement exists. Hartog (2000) classified these measures into three categories. Firstly, subjective Worker self-Assessment (WA) measures are derived from survey questions to the level required to do (WAd) or to get (WAg) the job. Over- and undereducation is measured by comparing this level with the actual educational level. Secondly, Job Analysis (JA) bases the required level on the occupational classification of job experts. Thirdly, Realised Matches (RM) measures look at the difference between the actual educational level and the mean (RMmn) or modal (RMml) level within a person's occupation¹. Comparative research shows that results depend on the applied measure (Rubb, 2003; Verhaest and Omey, 2005b). JA is clearly the most attractive measurement (Hartog, 2000). Conceptually, it reflects the core idea of overeducation in the literature: overeducation as underutilization of skills. Furthermore, there are no reasons to expect systematic bias in measurement. However, a lot of researchers still prefer to use subjective measures since JA is a very costly exercise and requires frequent updates to adapt to changing technology. Additionally, explanatory variables such as job satisfaction or mobility may be influenced by the perception of being overeducated instead of objective overeducation.

The aim of this paper is to examine the reliability of the different measures on the

¹ Sometimes, worker self-assessment measures are based on direct survey questions to their over- or undereducation status. For surveys of the literature and extensive discussions on the applied measures, we refer to Halaby (1994), Green et al. (1999), Hartog (2000), Groot and Maassen van den Brink (2000), Verhaest and Omey (2004, 2005a), and Sloane (2004).

1
2
3
4
5
6
7 basis of encompassing tests (see Mizon and Richard, 1984). In a measurement context, the
8 application of encompassing or related approaches for non-nested hypotheses is rather
9 seldom (see, e.g., Oczkowski and Farrell, 1998; Poor et al., 2001). The idea is simple: one
10 measure should be preferred to another if its outcomes encompass the results of its
11 counterpart. We link five overeducation measures to four outcomes: wages, job satisfaction,
12 mobility and training participation. These outcomes represent the bulk of the literature on
13 the impact of overeducation (see, e.g., Hersch, 1991; Allen and van der Velden, 2001;
14 Verhaest and Omey, 2005b). So, there is a strong case for applying JA if this measurement
15 would prove to be reliable within these different contexts.
16
17
18
19
20
21
22
23

24 The paper is structured as follows. In section II, we give an overview of potential
25 errors and biases when analysing the impact of overeducation. The test procedure is
26 explained in section III. In section IV, we deal with the data. In section V, we outline our
27 empirical model. The results are overviewed in section VI. Finally, in section VII, we
28 discuss the outcomes of the analysis and make some conclusions. Our results strengthen the
29 use of JA. However, the most reliable solution is to instrument the WAd measure on the JA
30 measure.
31
32
33
34
35
36
37
38
39
40

41 II. Potential biases and errors

42
43

44 A first potential bias lies in differences between the concept to be measured and the concept
45 actually measured, i.e. the validity of the instrument. In the literature, overeducation is
46 generally conceptualised in terms of an educational level that exceeds the minimal required
47 level to do the job (Green et al., 1999). Starting from the technology of the job and the
48 activities to be performed, JA clearly reflects this concept (Hartog, 2000). Also subjective
49 measures on the basis of the required level to do the job are in line with this interpretation.
50 Although closely related to this concept, the other indicators clearly measure something
51 different. The required level to get the job, or the mean or modal educational level within
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7 an occupation are not necessarily the same as the level to do the job. The minimal required
8
9 level to get the job reflects a reservation educational level below which employers won't
10
11 hire job applicants (Verhaest and Omey, 2004). The gap between these two concepts is
12
13 often termed qualification inflation and deflation (see, e.g., Green et al., 1999; Dolton and
14
15 Silles, 2001). Defining the required level by the mean or median educational level may
16
17 reflect educational mismatches resulting from searching and matching (Borghans and de
18
19 Grip, 2000). However, it only measures frictional mismatches and neglects structural
20
21 sources of over- and undereducation (Verhaest and Omey, 2005a).

22
23 The second source of biases and errors results from the data collection and
24
25 registration process, i.e. the reliability of the measures. The main critique on JA is about
26
27 random measurement error. It is often argued that heterogeneity within occupations makes
28
29 JA (and RM) measures more prone to this type of error (see, e.g., Halaby, 1994). This is
30
31 true if the starting point of the occupational classification is the job title, and not the
32
33 requirements. Apart from heterogeneity error, also classification error may cause these
34
35 types of error. Contrary to job experts, respondents can rely on all relevant information.
36
37 However, they lack uniform coding instructions. They may report the required level to get
38
39 the job, current hiring standards, the median educational level for similar workers, or their
40
41 own educational level. Verhaest and Omey (2004) indeed found that random error explains
42
43 a significant but small part of the variation of every measure. Social desirability may
44
45 systematically bias subjective levels upwardly. This bias may be enforced if individuals are
46
47 more likely to report their actual educational level. Oppositely, JA delivers an unbiased
48
49 measure of the required level (Rumberger, 1987; Hartog, 2000). It is sometimes argued that
50
51 technological progress biases JA levels systematically downwardly. However, this can be
52
53 avoided by a regular update of the classification scheme or by using a scheme that starts
54
55 from requirements instead of job titles. In this last case, the scheme should be flexible
56
57 enough to incorporate changes in requirements.
58
59
60

1
2
3
4
5
6
7 Highly problematic is when these errors are related to the variables of interest,
8
9 such as job satisfaction, training participation, mobility or wages. Researchers are generally
10 not merely interested in correlations, but in the causal impact of true over- and
11 undereducation on these variables. Well known from the psychological literature is that
12 answers to survey questions are interrelated. Context effects do not only influence answers
13 on attitude questions, but also on non-attitude questions such as one's race (Martin et al.,
14 1990). Respondents will try to keep a degree of consistency in their answers. Furthermore,
15 temporal mood states strongly influences judgements (Schwarz and Strack, 1999). Finally,
16 also social desirability bias leads to interrelated answers. All this creates a spurious relation
17 between the outcome variables and overeducation if measured in a subjective way. It is
18 clear that these problems are severe in the case of overeducation, a concept that, even
19 among social scientists, has no unique interpretation. So, individuals that answer to be
20 satisfied with their job will also tend to evaluate their match position in a more positive
21 way. Being offered a permanent contract or not being fired may be used as an indication for
22 the quality of their match. They may conclude that they were undereducated for their job
23 since they were offered additional training. Or their personal wage may influence their
24 judgement. After all, also researchers use these outcome variables as an alternative way to
25 assess the degree of over- and undereducation or the quality of the match². This approach
26 may be useful to examine overeducation in a different context if no direct measure is
27 available, but not to assess the impact on the variable on which its measurement is based.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 III. Test procedure

50
51
52 A way of testing which of two non-nested models should be preferred is the application of
53
54

55 ² For wages, see, e.g., Gottschalk and Hansen (2003); for satisfaction, see, e.g., Chevalier
56 (2003); for tenure, see, e.g., Bowlus (1995).
57
58
59
60

the encompassing test-principle³. For outcome variable y^* and respective overeducation measures x^a and x^b , consider the following models:

$$H_a : y^* = \beta_1^a x^a + \varepsilon^a$$

$$H_b : y^* = \beta_1^b x^b + \varepsilon^b$$

The encompassing principle is applied by adding the non-nested part of the second equation (i.e. x^b) to the first equation:

$$H_{ab} : y^* = \beta_2^a x^a + \beta_2^b x^b + \varepsilon^{ab}$$

The test is a simple F- or Likelihood Ratio-test of the new equation H_{ab} against H_a , and, similarly, against H_b . H_a is said to encompass H_b if H_a is not rejected against H_{ab} . The possible outcomes are summarized in figure 1.

“figure 1 here”

The interpretation of the outcomes depends on the type of measures being compared. Suppose that x^a is a valid measure (e.g., JA or WAd), and x^b not (e.g., WAg, RMmn or RMml). Since x^b is not valid, x^a should always be preferred. However, x^b is still useful if both concepts do not differ in reality. This is the case if the results of both models encompass each other (figure 1, cell 1). Then, researchers can still use the non-valid measure if a valid measure is not available. In the other cases, the measures differ substantially, and only one (cell 2 or 3) or both concepts (cell 4) are related to the outcome

³ For an overview, see Davidson and MacKinnon (1993).

variable. Suppose that both x^a and x^b are valid (e.g., JA is tested against WAd). If both measures encompass each other (cell 1), they generate statistically similar outcomes. Hence, both indicators are equally effective. Model H_a may encompass H_b , while H_b does not encompass H_a (cell 2). In this case, x^a should be preferred to x^b . In the opposite case (cell 3), we should prefer x^b . Finally, none of the models may encompass its counterpart (cell 4). Then, neither model is satisfactory, e.g. due to substantial errors in both measures.

These standard encompassing interpretations are only valid under the assumptions that $E(\varepsilon^a | x^a) = 0$ and $E(\varepsilon^b | x^b) = 0$. However, one may expect that the error term of one measure is spuriously correlated with the explanatory variable, e.g. for WAd in the likely case of the job satisfaction analysis. Then, the encompassing test procedure is equivalent to the specification test of Hausman (1978) for endogeneity. Suppose that $E(\varepsilon^a | x^a) = 0$ (e.g. for JA) and $E(\varepsilon^b | x^b) \neq 0$ (e.g. for WAd). A way to correct for the bias in measure x^b is to use the unbiased measure x^a as an instrument ($\hat{x}^b = \eta x^a$). If $E(\varepsilon^b | x^b) = 0$, the impact of the residuals of this estimation on y^* equals the impact of the estimated values \hat{x}^b . So, a test for endogeneity is based on the following equation H_e :

$$H_e : y^* = \beta_3^b (x^b - \hat{x}^b) + \beta_3^b \hat{x}^b + \varepsilon^e \Leftrightarrow y^* = (\beta_3^b - \beta_3^a) \hat{x}^b + \beta_3^b x^b + \varepsilon^e$$

The Hausman test is a test of the null hypothesis that $\beta_3^b = \beta_3^a$ or $(\beta_3^b - \beta_3^a) = 0$. Given

$\hat{x}^b = \eta x^a$, H_e can be rewritten as:

$$H_e : y^* = (\beta_3^b - \beta_3^a) \eta x^a + \beta_3^b x^b + \varepsilon^e$$

1
2
3
4
5
6
7 If $\eta \neq 0$, i.e. x^a is correlated with x^b , a test of the null hypothesis that $(\beta_3^{\hat{b}} - \beta_3^b) = 0$ is
8
9 equivalent to a test of $(\beta_3^{\hat{b}} - \beta_3^b)\eta = 0$, i.e. the encompassing test. In this interpretation, the
10
11 rejection of the model with the potentially biased measure x^b (cell 2 or 4) is a sufficient
12
13 condition to prefer the usage of the alternative measure x^a . Which measure may be
14
15 systematically biased can only be determined on the basis of theoretical considerations.
16
17

18 19 20 IV. Data

21
22
23 The analysis is based on the 1978 and 1980 birth cohorts of the SONAR data about the
24
25 transition from school to work in Flanders. These cohorts consist of random samples of
26
27 respectively 3002 and 2993 Flemish individuals, interviewed face-to-face during the last
28
29 four months of 2001 and 2003⁴. The interview consisted of an extensive questioning of the
30
31 educational and early labour market career. For the first job after leaving school⁵, we
32
33 compute the over- and undereducation status of the respondents, based on the five
34
35 measurement methods. A major advantage of an analysis of first jobs is that over-education
36
37 measurement is not biased by additional human capital accumulation through experience
38
39 and on-the-job training⁶. So, educational mismatches are more likely to reflect skill
40
41 mismatches (see Allen and van der Velden, 2001). The first job is not observed for 22.0%
42
43 of the individuals, mainly because they were still student at the age of 23 (15.4%). The
44
45 analysis is restricted to those with observations on all of the five measures, with a first non-
46
47

48
49 ⁴ For the 1978 cohort, some extra interviews were conducted in the first months of 2002.

50
51 ⁵ The first job is defined to be the first job of at least one hour a week and tenure of at least
52
53 one month. For an extensive description of the data, see SONAR (2003, 2005).

54
55 ⁶ It is generally found in the literature that experience and tenure decrease (increase) the
56
57 probability of being classified as overeducated (undereducated) (see, e.g., Groot, 1996;
58
59 Sloane et al., 1996, 1999).
60

1
2
3
4
5
6
7 self employed job inside Flanders (Brussels including) and who did not return to education
8
9 after their first job. This reduces the potential sample to 3759 respondents. These school
10
11 leavers entered their first job during the period 1996-2003. For the derivation of the over-
12
13 and undereducation status, the following educational levels are defined: less than lower
14
15 secondary (<LS: 3.9% of potential sample), lower secondary (LS: 9.3%), higher secondary
16
17 (HS: 50.4%), lower tertiary (LT: 25.8%) and higher tertiary education (HT: 10.6%).

18
19 Occupations are coded following the 2001 update of the 1992 Standard
20
21 Occupation Classification of Statistics Netherlands (CBS, 2001). An advantage of this
22
23 classification is that it starts from the complexity and content of the tasks to be executed on
24
25 the job, and not from the job title (cf. supra). The classification is based on five functional
26
27 levels that correspond to our five distinguished educational levels. It is a flexible
28
29 classification that enables to classify jobs with a similar title at different functional levels⁷.
30
31 The coding was based on open answer information on the following items: tasks to be
32
33 executed, job title and activity of the firm. Interviewers were urged to ask to all details of
34
35 the tasks to be executed and activities of the firm. Additional information used was the
36
37 number of individuals being managed, self-employment and firm size. The WAd measure
38
39 is based on: 'What is, according to your own opinion, the most appropriate educational
40
41 level to do your job?'. Furthermore, respondents were asked: 'Was a qualification required
42
43 to get your job?'. In the case of a positive answer, they were asked: 'To get your job, what
44
45 educational level were you required to have?'. On the basis of this information, the WAg
46
47 measure is computed. Finally, the mean and modal educational level for each 2-digit CBS
48
49 occupation was computed to derive the RMmn and RMml measures⁸. The mean and modal

50
51
52 ⁷ Clerk occupations, e.g., are found at four of five functional levels. In the 1988 ISCO
53
54 classification (ILO, 1990), all these jobs are grouped under one basic job title and are
55
56 assigned one single ISCED level.

57
58 ⁸ In line with the literature, the RMmn overeducation (undereducation) status is defined by
59
60

1
2
3
4
5
6
7 level are probably biased if derived from the 1978 and 1980 SONAR cohort since the first
8
9 job is not observed for those who were still at school at the age of 23. So, we derive the
10
11 mean and modal educational level within each occupation from the 1976 SONAR birth
12
13 cohort. A follow-up survey was conducted for this cohort at the age of 26, delivering
14
15 observations on the first job for 92.7% of the school leavers (see SONAR, 2004)⁹.

16
17 Table 1 reports the incidence of over- and undereducation for the analysed sample
18
19 on the basis of the different measures. The figures diverge substantially between the
20
21 different measures. The incidence of overeducation ranges from 8.7% for RMmn to 51.0%
22
23 for JA. The incidence of undereducation ranges from only 4.7% for WAg to 21.8% for
24
25 RMml. These differences are in line with the outcomes of other studies that also measured
26
27 over- and undereducation in various ways for the same dataset of individuals¹⁰. They
28
29 further underline the importance of additional research on the reliability of these measures.
30
31

32
33 “table 1 here”
34
35

36 V. Empirical model

37
38
39

40 We estimate the effect of different educational mismatch measures on job satisfaction,
41
42 mobility, training participation and wages. For each of these variables, we apply the
43
44

45 years of education that are more (less) than one standard deviation above (below) the mean
46
47 (see, e.g., Verdugo and Verdugo, 1989; Groot, 1996).
48

49 ⁹ The results of the RM measures derived from the 1978 and 1980 cohort were
50
51 encompassed by those on the basis of the 1976 cohort for all of the explanatory variables.
52
53 We do not include the respondents of the 1976 cohort into the estimations since the survey
54
55 did not include questions that enable to compute a WAd measure.
56

57 ¹⁰ For an overview of these studies, see Verhaest and Omey (2005a).
58
59
60

following basic model for measure i :

$$Y = \alpha_i + \mathbf{X}\boldsymbol{\beta}_i + \gamma_i YEDUC + \delta_i YOVER_i + \varphi_i YUNDER_i + \varepsilon_i$$

With Y = dependent variable (job satisfaction, mobility, training participation and wages), \mathbf{X} = a vector of control variables, $YEDUC$ = years of education, $YOVER$ = years of overeducation, $YUNDER$ = years of undereducation. Years of education are based on the minimal study length of the educational levels: <LS = 6 years, LS = 10 years, HS = 12 years, LT = 15 years and HT = 16 years. Since $YEDUC \equiv YREQ_i + YOVER_i - YUNDER_i$, this model is equivalent to the generally applied ORU-specification (see Hartog, 2000), but enables to separate the effect of measurement from the effect of years of education. The non-nested part consists of $YOVER$ and $YUNDER$.

For the wage analysis, net monthly starting wages are deflated and converted into hourly earnings and OLS is applied for $\text{LN}(\text{WAGE})$ ¹¹. Job satisfaction is based on a scale from 1 (very unsatisfied) to 5 (very satisfied) and is estimated on the basis of an ordered probit specification. Mobility out of the first job is measured by a change in function and/or firm and estimated through a Cox proportional hazard model. For the training participation analysis, we apply a binary probit regression whereby the training variable is coded with value 1 if the school leaver participated in formal job training inside or outside the firm during the first job.

Two specifications of models are estimated, depending on the control variables (\mathbf{X}) that are included. Specification A includes a fixed number of individual and job characteristic variables¹². Job satisfaction and mobility may depend on the wage level and

¹¹ Outliers (hourly wages of < 2.5 EURO and > 25 EURO) are excluded from the analysis.

¹² The following control variables are included: inactivity duration between school leaving and job start, father's and mother's years of education, dummies for having a father (1

1
2
3
4
5
6
7 the offered training opportunities in the job. Training may also exert an impact on wages,
8
9 while training participation may depend on the observed job length. So, specification B
10
11 differs by outcome and includes these extra control variables. Job duration may be
12
13 endogenously related to training participation. So, we apply a two step approach for the
14
15 training participation analysis: potential observed job duration, i.e. duration between job
16
17 start and the survey date, is used as an instrument for observed job duration. LN(WAGE)
18
19 and a dummy for formal training participation are included in the mobility and job
20
21 satisfaction analysis. The formal training participation dummy is also included in the wage
22
23 analysis¹³.

24
25 The first job is not observed for a substantial part of the sample and this sample
26
27 selection may bias the results (cf. Heckman, 1979). Some additional tests were executed to
28
29 assess whether this may affect our conclusions. We estimated maximum likelihood
30
31 heckman selection models for the wage analysis and maximum likelihood probit models
32
33 with selection for the training analysis (see STATA, 2003)¹⁴. In a first selection model, we

34
35
36 dummy) and mother (1) being regularly unemployed during the phase of secondary
37
38 education, gender (1), non-European descent (1), cohabiting at job start (1), having a child
39
40 at job start (1), firm size (3), industry (11), professions (3), shift-work (1), night-work (1),
41
42 temporary (1) and part-time contract (1), region of employment (4) and quarter of job start
43
44 (31).

45
46 ¹³ Training participation may be delayed, what is problematic for the wage analysis since
47
48 starting wages are analysed. Training is also endogenously related to mobility in this case.
49
50 Yet, in line with the human capital prediction that training is provided at the job start, we
51
52 did not note a significant impact of observed job duration on the probability of formal
53
54 training participation.

55
56 ¹⁴ The selection equation included the same individual characteristic variables as the
57
58 outcome equation (cf. supra), region of residence (4 dummies), and duration between the
59
60

1
2
3
4
5
6
7 only included school leavers in the selection equation. In a second model, we also included
8 those who were still at school at the age of 23. Yet, we did not find any evidence on sample
9 selection effects since the outcome equation was always found to be independent from the
10 selection equation. Moreover, the encompassing test outcomes were identical to those on
11 the basis of the standard estimations. The sample selection problem is particularly relevant
12 for the HT educated¹⁵. We re-estimated the models for every outcome variable, but
13 excluded those with a HT degree. Also this did not change our conclusions. So, we only
14 report the results on the basis of the full sample.
15
16
17
18
19
20
21
22
23
24

25 VI. Results

26
27
28 Adding the educational mismatch variables (*YOVER* and *YUNDER*) to the equation always
29 leads to a significant improvement of the model (see table 2). Which measure delivers the
30 largest contribution to the model depends on the outcome variable. While *JA* delivers the
31 largest contribution to the explanation of $\text{LN}(\text{wages})$, it is the *WAd* measure that explains
32 the largest part of the variation of job satisfaction and job mobility. With respect to training
33 participation, *WAg* delivers the largest contribution. The results are in line with previous
34 findings in the literature: overeducated workers earn less, are less satisfied, are more mobile
35 and participate less in training in comparison with adequately educated workers with a
36 similar educational level¹⁶.
37
38
39
40
41
42
43
44
45
46

47 date of the first interview in the survey and the date of the individual interview. This last
48 variable is independent from the outcomes but has a significant impact on selection.
49

50
51 ¹⁵ Most of the scholars at the age of 23 were student at the HT level (74.3%). Moreover,
52 those with a HT degree were also overrepresented among the school leavers without a job
53 at the age of 23: 35.6% of this group had a HT degree.
54
55

56
57 ¹⁶ An overview of the literature on earnings is given by Hartog (2000), and for the other
58
59
60

1
2
3
4
5
6
7
8
9 “table 2 here”
10
11

12 We concentrate on the encompassing test results (see table 3). There are few differences in
13 test outcomes between the two model specifications. Moreover, these differences do not
14 affect the general conclusions. Only for the RM measures in the case of the training
15 participation analysis, mutual encompassing between two measures is noted. This confirms
16 previous findings that different measures lead to statistically different outcomes. Both the
17 JA and WAd measure results are never encompassed. Furthermore, the WAg measure
18 results are in some cases encompassed by those of the WAd measure. Finally, the RM
19 measures are never able to encompass the results of any of the three other measures, while
20 their results are encompassed at least once by every other measure.
21
22
23
24
25
26
27
28
29
30
31
32

33 “table 3 here”
34
35

36 VII. Discussion 37 38 39

40 The results of the JA and WAd measures are never encompassed by any of the non-valid
41 measures. Thus, JA or WAd should always be preferred to RM or WAg measures of
42 overeducation. Encompassing tests between JA and WAd deliver no preference: neither JA
43 encompasses WAd, nor WAd encompasses JA for any of the outcome variables. At least
44 for the job satisfaction analysis, this result may sound surprising since the relationship
45 between objective conditions of life and subjective well-being is generally found to be
46 weak (Schwarz and Strack, 1999). It illustrates that both aggregation error and
47 classification error should not be greater on the basis of JA if the analysis is carefully
48
49
50
51
52
53
54
55

56
57
58 outcome variables by Verhaest and Omeij (2005b). Our results are available upon request.
59
60

1
2
3
4
5
6
7 conducted and a requirement based classification is used. Of course, WAd is a cheap
8
9 alternative to the relatively costly JA. However, if carefully implemented, JA can also be
10
11 conducted on a survey-based manner and does not necessarily require a visit at the
12
13 workplace.

14
15 The test can also be interpreted as a Hausman specification test for endogeneity.
16
17 Several sources of measurement error may result in spurious correlation between the
18
19 outcome variables and overeducation if measured by WAd. These problems may be
20
21 particularly severe in the case of the job satisfaction analysis. Oppositely, we have no
22
23 reasons to expect systematic bias of the JA measure with any of the outcome variables. On
24
25 the basis of these theoretical considerations, the rejection of the WAd model is a sufficient
26
27 condition to prefer the usage of JA. Since both measures lead to statistically different
28
29 outcomes, JA is a more secure option. Nevertheless, the ideal solution is relying on both
30
31 measures by instrumental variable techniques. Using the WAd measure as an instrument
32
33 variable for the JA measure corrects for the bias caused by random error, but introduces the
34
35 systematic bias of the WAd measure. Oppositely however, using the JA measure as an
36
37 instrument for the WAd measure is a more reliable solution. In this way, error in the WAd
38
39 measure is corrected for without introducing some new systematic bias.
40
41

42 VIII. References

- 43
44
45
46 Allen J. and van der Velden R. (2001), 'Educational mismatches versus skill mismatches:
47
48 effects on wages, job satisfaction, and on-the-job search', *Oxford Economic*
49
50 *Papers*, 53 (3), 434-452.
51
52 Borghans L. and de Grip A. (2000), 'The debate in economics about skill utilization', in
53
54 *The Over-educated Worker?*, L. Borghans and A. de Grip (eds), Edward Elgar:
55
56 Cheltenham, 3-23.
57
58 Bowlus A., 'Matching workers and jobs: cyclical fluctuations in match quality', *Journal of*
59
60

- 1
2
3
4
5
6
7 *Labor Economics*, 13 (2), 335-350.
- 8
- 9 CBS (2001), *Standaard beroepenclassificatie 1992 - editie 2001*.
- 10
- 11 Chevalier (2003), 'Measuring Over-education', *Economica*, 70 (3), 509-531.
- 12
- 13 Davidson R. and MacKinnon J. (1993), *Estimation and Inference in Econometrics*, Oxford
- 14 University Press.
- 15
- 16 Dolton P. and Silles M. (2001), 'Over-education in the graduate labour market: some
- 17 evidence from alumni data', *CEE Discussion Paper*, June 2001, 40p.
- 18
- 19 Gottschalk P. and Hansen M. (2003), 'Is the Proportion of college workers in noncollege
- 20 jobs increasing?', *Journal of Labor Economics*, 21 (2), 449-471.
- 21
- 22 Green F., McIntosh S. and Vignoles A. (1999), 'Overeducation and skills – clarifying the
- 23 concepts', *CEP Discussion Paper*, 435.
- 24
- 25 Groot W. (1996), 'The incidence of, and returns to overeducation in the UK', *Applied*
- 26 *Economics*, 28 (10), 1345-1350.
- 27
- 28 Groot W. and Maassen van den Brink H. (2000), 'Overeducation in the labor market: a
- 29 meta-analysis', *Economics of Education Review*, 19 (2), 149-158.
- 30
- 31 Halaby C. (1994), 'Overeducation and skill mismatch', *Sociology of Education*, 67 (1),
- 32 111-125.
- 33
- 34 Hartog J. (2000), 'Overeducation and earnings: where are we, where should we go?',
- 35 *Economics of Education Review*, 19 (2), 131-147.
- 36
- 37 Hausman J. (1978), 'Specification tests in econometrics', *Econometrica*, 46 (6), 1251-1271.
- 38
- 39 Heckman J. (1979), 'Sample selection bias as a specification error', *Econometrica*, 47 (1),
- 40 153-161.
- 41
- 42 Hersch J. (1991), 'Education match and job match', *Review of Economics and Statistics*, 73
- 43 (1), 140-144.
- 44
- 45 ILO (1990), *International Standard Classification of Occupations: ISCO-88*, 457p.
- 46
- 47 Martin E., Demaio T., and Campanelli P. (1990), 'Context effects for census measures of
- 48 race and Hispanic origin', *Public Opinion Quarterly*, 54 (4), 551-566.
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1
2
3
4
5
6
7 Mizon G. and Richard J. (1986), 'The encompassing principle and its application to testing
8 non-nested hypotheses', *Econometrica*, 54 (3), 657-678.
9
- 10 Oczkowski E. and Farrell M., (1998), 'Discriminating between measurement scales using
11 non-nested tests and two-stage least squares estimators: the case of market
12 orientation', *International Journal of Research in Marketing*, 15 (4), 349-366.
13
14
15
16 Poor P., Boyle K., Taylor L., and Bouchard R., (2001), 'Objective versus subjective
17 measures of water clarity in hedonic property value models', *Land Economics*, 77
18 (4), 482-493.
19
20
21
22 Rubb S. (2003), 'Overeducation in the labor market: A comment and re-analysis of a meta-
23 analysis', *Economics of Education Review*, 22 (6), 621-629.
24
25
26 Rumberger R. (1987), 'The impact of surplus schooling on productivity and earnings',
27 *Journal of Human Resources*, 22 (1), 24-50.
28
29
30 Schwarz N. and Strack F. (1999), 'Reports of subjective well-being: judgmental processes
31 and their methodological implications', in *Well-being: the foundations of hedonic
32 psychology*, Kahneman D., Diener E. and Schwarz N. (eds), 61-84.
33
34
35
36 Sloane P. (2004), 'Much ado about nothing? What does the overeducation literature really
37 tell us?', in *Overeducation in Europe*, F. Büchel, A. de Grip and A. Mertens (eds.),
38 Edward Elgar: Cheltenham, 11-45.
39
40
41
42 Sloane P., Battu H. and Seaman P. (1996), 'Overeducation and the formal education /
43 experience and training trade-off', *Applied Economics Letters*, 3 (8), 511-515.
44
45
46 Sloane P., Battu H. and Seaman P. (1999), 'Overeducation, undereducation and the British
47 labour market', *Applied Economics*, 31 (11), 1437-1453.
48
49
50 SONAR (2003), *Hoe maken Vlaamse jongeren de overgang van school naar werk?*,
51 basisrapportering cohorte 1978 (eerste golf), 147p.
52
53
54 SONAR (2004), *Hoe maken Vlaamse jongeren de overgang van school naar werk?*,
55 basisrapportering cohorte 1976 (tweede golf), 116p.
56
57
58 SONAR (2005), *Hoe maken Vlaamse jongeren de overgang van school naar werk?*,
59
60

1
2
3
4
5
6
7 technisch rapport cohorte 1980 (eerste golf).

8
9 STATA (2003), *STATA reference manual*, release 8.

10
11 Verdugo R. and Verdugo N. (1989), 'The impact of surplus schooling on earnings: some
12 additional findings', *Journal of Human Resources*, 24 (4), 629-643.

13
14 Verhaest D. and Omey E. (2004), 'What determines measured overeducation?', *UGent WP*,
15 04/216, 37p.

16
17
18 Verhaest D. and Omey E. (2005a), 'Measuring the incidence of over- and undereducation',
19
20 *Quality & Quantity*, forthcoming.

21
22
23 Verhaest D. and Omey E. (2005b), 'The impact of overeducation and its measurement',
24
25 *Social Indicators Research*, forthcoming.

Figure 1: Possible encompassing outcomes

		H_a encompasses H_b ?	
		Yes	No
H_b encompasses H_a ?	Yes	1	3
	No	2	4

For Peer Review

Table 1: Incidence of over- and undereducation for the analysed sample (N = 3759)

	JA	WAd	WAg	RMmn	RMml
Overeducated	51.0%	27.2%	44.2%	8.7%	11.6%
Undereducated	7.8%	9.4%	4.7%	15.8%	21.8%

For Peer Review

Table 2: Contribution of *YOVER* and *YUNDER* to the model (\$)

	LN(Wages)	Job Satisfaction	Mobility	Training
<u>Specification A</u>				
JA	58.7**	94.4**	49.5**	37.1**
WAd	25.1**	251.3**	107.8**	49.8**
WAg	21.5**	92.8**	46.1**	55.4**
RMmn	44.6**	48.4**	35.5**	7.8*
RMml	51.5**	36.7**	24.6**	8.9*
<u>Specification B</u>				
JA	55.0**	64.9**	29.0**	36.8**
WAd	22.4**	214.4**	84.0**	47.1**
WAg	18.7**	73.0**	31.0**	54.6**
RMmn	42.7**	34.2**	24.3**	8.3*
RMml	49.4**	25.5**	15.1**	9.4**

Applied estimation procedure: LN(Wages): OLS (A: N=3303; B: N=3300); Job Satisfaction: Ordered Probit (A: 3463; B: N=3300); Mobility: Cox proportional hazard (A: N=3463; B: N=3300), and Training: Binary Probit (A: N=3460; B: N = 3460).

(\$) F statistic for LN(Wages), and Likelihood Ratio Chi² statistic for other analyses.

*: significant at 5% level; **: significant at 1% level.

Table 3: Encompassed models for different outcome variables (\$)

	LN(Wages)	Job Satisfaction	Mobility	Training
<i>Specification A</i>				
JA – WAd	-	-	-	-
JA – WAg	-	-	-	-
JA – RMmn	-	RMmn	-	RMmn
JA – RMml	-	RMml	-	RMml
WAd – WAg	-	WAg	-	-
WAd - RMmn	-	-	-	RMmn
WAd – RMml	-	-	-	RMml
WAg – RMmn	-	-	-	RMmn
WAg – RMml	-	-	-	RMml
RMmn - RMml	RMmn	RMml	RMml	RMmn / RMml
<i>Specification B</i>				
JA – WAd	-	-	-	-
JA – WAg	-	-	-	-
JA – RMmn	-	RMmn	-	RMmn
JA – RMml	-	RMml	RMml	RMml
WAd – WAg	-	WAg	WAg	-
WAd - RMmn	-	-	-	RMmn
WAd – RMml	-	-	-	RMml
WAg – RMmn	-	-	-	RMmn
WAg – RMml	-	-	-	RMml
RMmn - RMml	RMmn	RMml	RMml	RMmn / RMml

(\$)

(\$) For the wage analysis, e.g., the RMmn model is encompassed by the RMml model; tested on the basis of an F-test for LN(Wages) and on the basis of a Likelihood Ratio-test for the other analyses; evaluated at 5% significance level.