Online Appendix Document 1
HILDA Re-education Project: Main Sample

1 Sample Selection

Our analysis sample includes everyone who was 25 or above and not currently studying in 2001, who are observed in both 2001 and 2019 in terms of the outcome and treatment variables.

We delete any individuals who were currently studying in 2001 if:

- They reported currently studying full part or part time for the main survey

- According to the calendar, they have undertaken any full time or part time studies

- They are currently receiving Abstudy/Austudy payment or had received these last financial year

- They have cited study as the reason for not looking for work

1078 individuals were deleted after applying this sample exclusion.

A Variable description

A.1 Outcome Variables

Weekly earnings from main job in 2019 (*w19_wscmei*) records the weekly earnings from the main job for the individual in 2019.

Employed in 2019 (*w19_employed*) records whether the individual is employed in 2019 or not.

Weekly earnings from all jobs in 2019 (*w19_earning*) records the weekly earnings from all jobs for the individual in 2019.

Working hours in 2019 ($w19\_wkhr$) records the total number of hours the individual works in all jobs in a week on average. Working hours are set to 0 for those not working.

A.2   Treatment Variables: Re-education

Re-education based on highest attainment[1] (*reduhl*) records whether the individual has had re-education between 2002 and 2017, based on whether there was a change in the highest education level attained stated in the two years.

Re-education completion based on detailed qualifications (*redudl*) records whether the individual has completed any one of the following qualifications since last interviewed between 2002 and 2017[2]:

- Trade certificate or apprenticeship

- Technicians cert/Advanced certificate

- Teaching qualification

- Nursing qualification

- Associate Degree

- Advance Diploma (3 years full time or equivalent)

- Bachelor degree but not honours

- Certificate I

- Certificate II

- Certificate III

- Certificate IV

- Certificate of unknown level

- Doctorate

- Diploma NFI

- Diploma (2 years full time or equivalent)

---

[1]The HILDA variable on highest attainment was constructed using three components: the age the individual left school, the highest education attainment in the previous wave and the current level of secondary school attained or currently studying for.

[2]Refer to https://en.wikipedia.org/wiki/Australian_Qualifications_Framework for how most of these degrees are situated relative to each other in a hierarchy and the duration of these qualifications.

- Graduate Certificate

- Graduate Diploma

- Honours

- Masters

- Other

Re-education completion based on both highest attainment and detailed qualifications (*redufl*) records whether the individual has completed re-education based on both the variables *reduhl* and *redudl*. When either of these variables has a value of 1, this variable will take on the value of 1.

A.3 Input Variables

For each variable, missing values (if any) have been set to zero and a new binary variable has been generated to indicate the observations that are missing.

Demographics

Female (*p_fem*) records whether the individual is female.

Age group in 2001 records whether in 2001 the individual was:

- Aged 25-34 (*p_age1*)

- Aged 35-44 (*p_age2*)

- Aged 45-54 (*p_age3*)

- Aged 55-64 (*p_age4*)

- Aged 65 and above (*p_age5*)

Country of birth records whether or not an individual was born in:

- Australia and not indigenous (*p_cob1*)

- English speaking countries (*p_cob2*)

- Non-English speaking countries (*p_cob3*)

- Indigenous (*p_cob4*)

Poor English speaking abilities (*p_poeng*) records whether the individual has poor English speaking abilities.

Remoteness records whether the individual lives in:

- A major city (*p_urdg1*)

- An inner region (*p_urdg2*)

- Outer and remote areas or migratory in nature (*p_urdg3*)

Marital status in 2001 records whether in 2001 the individual was:

- Married (*p_mar1*)

- De facto (*p_mar2*)

- Separated (*p_mar3*)

- Divorced (*p_mar4*)

- Widowed (*p_mar5*)

- Single and never been married (*p_mar6*)

Parental Status

Number of dependents in 2001 (*p_noch*) records the number of dependent children the individual had in 2001.

Physical Health

Severity of health conditions in 2001 records whether the individual had:

- No health conditions (*p_ddeg1*)

- A mild condition (*p_ddeg2*)

- A moderate condition (*p_ddeg3*)

- A severe condition (*p_ddeg4*)

Labour Force Variables

Labour market status in 2001 records whether the individual was:

- Employed (*p_lfs1*)

- Unemployed (*p_lfs2*)

- Not in the labour market (*p_lfs3*)

Extent of working hour match with preferences in 2001 records whether the match between the individuals total weekly working hours across all jobs and their preferred number of working hours made them:

- Not working (*p_whp1*)

- Underemployed by at least 4 hours a week (*p_whp2*)

- Roughly Matched: Preferred and Actual Hours Worked differ by less than 4 hours a week (*p_whp3*)

- Overemployed by at least 4 hours a week (*p_whp4*)

Employee type in 2001 records whether the individual was:

- Not working (*p_emp1*)

- An employee (*p_emp2*)

- An employee of own business (*p_emp3*)

- Self Employed (*p_emp4*)

- Unpaid family worker (*p_emp5*)

Contract type in 2001 records whether the individual was:

- Not working (*p_con1*)

- On a fixed term contract (*p_con2*)

- On a casual contract (*p_con3*)

- On a permanent contract (*p_con4*)

- On other types of contracts (*p_con5*)

Occupation in 2001 records whether the individual was working as:

- Not working (*p_occ1*)

- Armed forces (*p_occ2*)

- Legislators, Senior Officials and Managers (*p_occ3*)

- Professionals (*p_occ4*)

- Technicians and Associate Professionals (*p_occ5*)

- Clerks (*p_occ6*)

- Service Workers and Shop and Market Sales Workers (*p_occ7*)

- Skilled Agriculture and Fishery Workers (*p_occ8*)

- Craft and Related Trades Workers (*p_occ9*)

- Plant and Machine Operators and Assemblers (*p_occ10*)

- Elementary Occupations (*p_occ11*)

Household income in 2001 (*p_rehdi*) records the real value of the individuals total household income indexed at 2012 price levels and adjusted for household size.

Partner labour force status in 2001 records whether the individual had:

- No partner or no resident partner (*p_plfs1*)

- A partner who was employed (*p_plfs2*)

- A partner who was unemployed (*p_plfs3*)

- A partner who was not in the labour force (*p_plfs4*)

Parental information

Fathers country of birth records whether or not the individuals father was born in:

- Australia (*p_fcob1*)

- English speaking countries (*p_fcob2*)

- Non-English speaking countries or indigenous (*p_fcob3*)

Mothers country of birth records whether or not the individuals mother was born in:

- Australia (*p_mcob1*)

- English speaking countries (*p_mcob2*)

- Non-English speaking countries or indigenous (*p_mcob3*)

Fathers education records whether the individuals fathers highest education, as reported in 2005, was:

- None (*p_fedu1*)

- Primary (*p_fedu2*)

- Below secondary (*p_fedu3*)

- Secondary (*p_fedu4*)

- Post-secondary, non-university (*p_fedu5*)

- Post-secondary, university (*p_fedu6*)

Mothers education records whether the individuals mothers highest education, as reported in 2005, was:

- None (*p_medu1*)

- Primary (*p_medu2*)

- Below secondary (*p_medu3*)

- Secondary (*p_medu4*)

- Post-secondary, non-university (*p_medu5*)

- Post-secondary, university (*p_medu6*)

Father undertaken post-school qualification through employer or non-tertiary means (*p_fpsm*) records whether the individuals father had undertaken his highest qualification through employers or other channels other than tertiary education, as reported in 2005.

Mother undertaken post-school qualification through employer or non-tertiary means (*p_mpsm*) records whether the individuals mother had undertaken his highest qualification through employers or other channels other than tertiary education, as reported in 2005.

Fathers Employment at age 14 records whether the individuals father was working when they were aged 14, in the following categories:

- Father deceased or not living with respondent (*p_femp1*)

- Father not employed (*p_femp2*)

- Father employed (*p_femp3*)

Mothers Employment at age 14 (*p_memp*) records whether the individuals mother was working when they were aged 14, in the following categories:

- Mother deceased or not living with respondent (*p_memp1*)

- Mother not employed (*p_memp2*)

- Mother employed (*p_memp3*)

Father substantially unemployed growing up records whether the individuals father had been unemployed for 6 months or more when they were growing up, in the following categories:

- Father not living with respondent (*p_fsue1*)

- Father not substantially unemployed (*p_fsue2*)

- Father substantially unemployed (*p_fsue3*)

Fathers Occupation records whether at age 14 the individuals father was last known working as:

- Father not in household (*p_focc1*)

- Armed forces (*p_focc2*)

- Legislators, Senior Officials and Managers (*p_focc3*)

- Professionals (*p_focc4*)

- Technicians and Associate Professionals (*p_focc5*)

- Clerks (*p_focc6*)

- Service Workers and Shop and Market Sales Workers (*p_focc7*)

- Skilled Agriculture and Fishery Workers (*p_focc8*)

- Craft and Related Trades Workers (*p_focc9*)

- Plant and Machine Operators and Assemblers (*p_focc10*)

- Elementary Occupations (*p_focc11*)

Mothers Occupation records whether at age 14 the individuals mother last known working as:

- Moher not in household (*p_focc1*)

- Armed forces (*p_mocc2*)

- Legislators, Senior Officials and Managers (*p_mocc3*)

- Professionals (*p_mocc4*)

- Technicians and Associate Professionals (*p_mocc5*)

- Clerks (*p_mocc6*)

- Service Workers and Shop and Market Sales Workers (*p_mocc7*)

- Skilled Agriculture and Fishery Workers (*p_mocc8*)

- Craft and Related Trades Workers (*p_mocc9*)

- Plant and Machine Operators and Assemblers (*p_mocc10*)

- Elementary Occupations (*p_mocc11*)

Non-cognitive variables

Well-being in 2001 (*p_losat*) records the life satisfaction score, which ranges from 0 to 10, of the individual reported in 2001. A higher score means the individual is more satisfied with his/her life.

Attitude towards having job in 2001 (*p_jbwk*) records the average score of attitude towards having a job reported by the individual in 2001 across two items (*p_jadnm* and *p_jahpj*), in a scale ranging from 1 to 7, with a higher score indicating a more favourable attitude towards having a job.

Enjoy job without needing money in 2001 (*p_jadnm*) records the extent the individual agreed with the statement that the person would enjoy having a job even if they did not need the money in 2001, in a scale ranging from 1 to 7, with a higher score indicating more agreement.

Important to have paying job in 2001 (*p_jahpj*) records the extent the individual agreed with the statement that in order to be happy in life it is important to have a paying job in 2001, in a scale ranging from 1 to 7, with a higher score indicating more agreement.

Prior Year Outcome variables

Mental health in 2001 (*p_mh01*). This is the transformed mental health scores from the aggregation of mental health items of the SF-36 Health Survey, as reported by the individual in 2001. It ranges from 0 to 100, with higher scores indicating better mental health.

Mental health in 2001 below norm (*p_mb01*) records whether the individuals mental health scores for 2001 was below the average of mental health scores across our analytical sample for that year.

Working hours in 2001 (*p_wh01*) records the number of hours the individual works across all jobs in a week on average. Working hours are set to 0 for those not working.

Hourly Wages in 2001 (*p_hrw01*) records the average hourly wage of the individuals main job in 2001. Hourly wages are set to 0 for those not working and set to missing for those reporting working more than 100 hours a week.

A.4 Variables that are not included in the model

The unique person identifier (*xwaveid*).

Completed re-education after 2017 based on highest education (*rehllt*) records whether the individual had only completed their re-education after 2017, comparing their education level in 2017 and 2019.

Completed re-education after 2017 based on detailed qualifications (*redllt*) records whether the individual has completed any one of the following qualifications since last interviewed between 2018 and 2019:

- Trade certificate or apprenticeship
- Technicians cert/Advanced certificate
- Teaching qualification
- Nursing qualification
- Associate Degree
- Advance Diploma (3 years full time or equivalent)
- Bachelor degree but not honours
- Certificate I
- Certificate II
- Certificate III
- Certificate IV
- Certificate of unknown level
- Doctorate
- Diploma NFI
- Diploma (2 years full time or equivalent)
- Graduate Certificate
- Graduate Diploma
- Honours
- Masters

- Other

Completed re-education after 2017 based on both highest attainment and detailed qualifications (*refllt*) records whether the individual has completed re-education after 2017 based on both the variables *rehllt* and *redllt*. When either of these variables has a value of 1, this variable will take on the value of 1.

Timing of Education Completion

Year of first re-education completion records the year of the first reported instance of re-education completion as provided by the detailed qualification variables and include the following categories:

- 2002 (*p_rcom1*)

- 2003 (*p_rcom2*)

- 2004 (*p_rcom3*)

- 2005 (*p_rcom4*)

- 2006 (*p_rcom5*)

- 2007 (*p_rcom6*)

- 2008 (*p_rcom7*)

- 2009 (*p_rcom8*)

- 2010 (*p_rcom9*)

- 2011 (*p_rcom10*)

- 2012 (*p_rcom11*)

- 2013 (*p_rcom12*)

- 2014 (*p_rcom13*)

- 2015 (*p_rcom14*)

- 2016 (*p_rcom15*)

- 2017 (*p_rcom16*)

- 2018 (*p_rcom17*)

- 2019 (*p_rcom18*)

Locus of control in 2003 (*p_cotrl*) records the transformed composite score[3] for locus of control items reported by the individual in 2003, the first year in HILDA for which this information becomes available. The transformation results in a variable that is ranged between 7 and 49. Locus of control measures the degree to which individuals attribute outcomes to internal versus external factors or the extent their welfare are in their own control compared to external circumstances. A higher score indicates having a more external locus of control, which is considered as a favourable personality trait.

Frequency of reading books in 2012 (*p_rdf*) records the frequency the individual reads books in 2012, the first year in HILDA for which this information becomes available. This is a proxy for love of learning[4]. This is a categorical variable encompassing the following frequencies:

- Every day or most days (*p_rdf1*)

- Several times a week (*p_rdf2*)

- About once a week (*p_rdf3*)

- 2 or 3 times a month (*p_rdf4*)

- About once a month (*p_rdf5*)

- Less than once a month (*p_rdf6*)

- Never (*p_rdf7*)

---

[3]See Buddlemeyer and Powdthavee (2015) for details of the transformation.

[4]HILDA contains a question on reading newspapers and magazines but we feel that reflects a care for or understanding of current issues more than a love of learning.