Department for Business, Innovation & Skills

RESEARCH

Paul Dolan

Daniel Fujiwara

Robert Metcalfe

Contents

Executive summary

Background

Good quality adult learning generates a range of possible benefits. Some of these will be captured but many of the benefits show up in better lives. They show up in improved health and wellbeing, potentially in civic participation and family interactions, and possibly even in other behaviours and attitudes.

This research project seeks to develop the knowledge base of the wider impacts of adult learning through reviews of the literature since 2008 and new data analysis.

Questions of validity

The substantive challenge is the extent to which causality can be attributed i.e. whether we can say anything meaningful about the causal impact of adult learning on these outcomes of interest.

We could find a positive relationship between adult learning and wellbeing but inferences about the internal validity of the relationship could be biased by many problems, such as selection (people who would have higher levels of wellbeing anyway may undertake adult learning courses) and reverse causality (an increase in wellbeing may have actually driven people to undertake more adult learning).

We must therefore classify the existing literature, and our own work, according to the rigour of the analysis in establishing causality. We use an established five point scale, which ranges from 1 (studies with no comparison group) to 5 (randomised trails), with the addition of 5* (randomised trials where people do not know they are being watched).

For policy purposes, it is also important to understand the degree to which any result can be generalised to the larger population of interest. This is the external validity of the study. It depends greatly on the samples and methodologies used. In general, the more robust studies are those whose results are representative of the whole population or of the sub-population of interest.

Review results

All of the qualitative studies we reviewed are level 1. These studies generally ask learners about the outcomes that they perceived. They provide important insights that would be missing if those studies did not exist but they cannot be used to say anything meaningful about the causal effects of adult learning.

This is because these types of study can be biased by the ways in which we tend to recall our past experiences and because they do not explicitly take the counterfactual into account. Further these types of study tend to rely on small sample sizes, which make the results hard to generalise.

There have only been a handful of quantitative studies since 2008. These studies have looked at the effect of adult learning on a range of outcomes including, wellbeing, and child test scores (for parents undertaking adult learning) and labour market outcomes. These studies are predominantly level 3 study designs as they are able to account for the counterfactual to some extent and hence provide better measures of causality. We also found two level 5 designs.

In general, after controlling for some of the observed differences between learners and non-learners, adult learning shows some important impacts but we do find that some associations become insignificant. This confirms the importance of controlling for other factors when trying to understand causality.

There is clearly the need for more quantitative research, especially into the impact of adult learning (rather than school-based learning) on non-economic benefits (rather than the economic returns to education, where there is a vast literature).

This project relies on observational data and so we explore some of these relationships and a number of new areas in more detail using level 3 study designs to control for observed differences across learners and non-learners. This is the optimal technique given the nature of the data.

We also provide recommendations for future research in this area using level 5* study designs which would allow us to get more robust estimates of the causal effects of adult learning.

New analyses

The analysis uses the British Household Panel Survey (BHPS), which is a nationally representative sample of over 10,000 adult individuals conducted between September and December of each year from 1991. Respondents are interviewed in successive waves, and all adult members and children aged 11-16 in a household are interviewed.

The BHPS contains a wide range of questions on adult learning and its panel format (that tracks people over a number of years) make it a very useful data source for examining questions on adult learning.

We focus on two types of adult learning: 1) formal learning that leads to a qualification; and 2) informal learning that does not lead to a qualification and we assess outcomes that have not been looked at in the previous literature.

Our study uses a level three design: match treated and non-treated groups based on observable characteristics. Our method exploits the longitudinal nature of the BHPS panel by using ordinary least squares (OLS) regression with fixed effects (FE).

Results

Here we summarise the main findings according the outcomes of interest; the wider benefits of adult learning. We report those factors that are significantly associated with

6

adult learning at the 5% level of significance and, where possible, provide information on the magnitude of the effects.

In a nutshell, adult learning has its greatest impacts in the domain of health and wellbeing; the impacts on civic participation and attitudes are less pronounced.

Mental health and wellbeing

- Improvements in reported life satisfaction and happiness

- Improvements in self-confidence (especially for formal learning)

    o This is more than twice the impact of being employed

- Improvements in own perceptions of self-worth

- Reductions in self-reported depression

- Increases in satisfaction with social life

- Increases in satisfaction with use of one's leisure time

Physical health

- Reductions in the number of visits to a GP

    o this is about one-seventh of the impact of being employed

- Improvements in self-reported overall health satisfaction

    o This is about half of the impact of being employed

Family and parenting

- Increases in the probability that the children in the household speak more frequently with the mother about serious issues

Civic participation

- Increases in trade union membership (especially for formal adult learning)

- Greater involvement in voluntary work (for formal learning only)

Attitudes and behaviours

- Greater desire to find a better job (especially for informal learning)

- Improved financial expectations (especially for formal learning)

Differences by demographics

Focussing resources on male and older learners may produce the largest impacts on some of the health outcomes, and encouraging parents and low income groups into formal learning would produce relatively large health benefits.

Discussion

We have shown that inferences about the degree to which this is true depend greatly on the quality of the evidence. Our high level 3 study, and those from other studies at a similar or sometimes higher level, do show that adult learning has health and wellbeing benefits – but perhaps not as much as might be concluded from level 1 qualitative studies.

One of the novel findings from this research is that, in addition to health and wellbeing benefits, adult learning appears to have some impact upon attitudes, resulting in increased expectation about work and income. Future research should consider the degree to which these shifts in expectations are found elsewhere.

To increase the robustness of the evidence we need to focus efforts on conducting level 5* studies – *natural field experiments* where subjects are randomised to receive different interventions but without them knowing they are part of a trial. Such studies will be internally valid (they can establish causality) and can be externally valid too (if representative samples are used).

That is all for the future. There are a few important things we can say now. From our analysis, the main wider benefits of adult learning show up in health, mental health and job-related outcomes. The previous literature generally supports this. Both formal and informal types of learning tend to matter, suggesting that participation in learning in itself is important.

The impact on self-confidence is worth a special and final mention. Adult learning has more than twice the impact on self-confidence than does being employed. This is an especially large effect and there are potential positive spillovers for a range of market and non-market outcomes from feeling better about oneself.

1. Introduction

This research project seeks to develop the knowledge base of the wider impacts of adult learning through reviews of the recent literature and new data analysis.

We aim to:

1. significantly update the knowledge on the wider benefits of adult learning; and

2. quantify these impacts with using statistical techniques

Our objectives are to:

1. assess the wider impacts, including health, wellbeing, civic participation and family interactions;

2. assess the impact on behaviours and attitudes (this has not received significant attention in the past, and it will provide a fuller picture of the wider benefits of learning);

3. use existing datasets to explore areas of analysis that have not been undertaken in the adult learning literature to date;

4. provide robust and careful assessment of the extent to which the methods used and those that employ provide a causal interpretation;

5. provide recommendations for new data collection and novel suggestions for employing experimental techniques to assess causality e.g. the use of field experiments is growing in economics and the social sciences and it has the potential to add great value in this area.

2. Telling the chicken from the egg

The main question, and substantive challenge, for this project is the extent to which causality can be attributed i.e. whether we can say anything meaningful about the causal impact of adult learning on a range of outcomes of interest, such as health, civic participation, and subjective wellbeing (SWB).

We seek to classify the existing literature according to the rigour of the analysis in establishing causality. Some of the problems surrounding causality in the impact of adult learning have been discussed previously in the literature (eg, Preston and Feinstein, 2004). Here, we extend that discussion to include a number of other potential factors.

We consider four problems that may lead to biased estimates of the causal effect of adult learning on other domains in life. We discuss them in the context of adult learning and wellbeing (broadly defined), but they apply to any study seeking to show how the change in one variable causes a change in another. So, say we find a positive relationship between adult learning and wellbeing. Inferences about the **internal validity** of the relationship could be biased by:

i. **Selection bias**. People who would have higher levels of wellbeing anyway may undertake adult learning courses. In this context, we could find that it was some third (confounding) variable that was driving the positive relationship between wellbeing and learning.

ii. **Reverse causality**. An increase in wellbeing may have actually driven people to undertake more adult learning.

iii. **Measurement error**. In statistical techniques that are often used to estimate the effect of learning, if adult learning were measured with error (say hours of learning was mis-reported) it would result in the coefficient on adult learning being biased.

iv. **Indirect effects**. Adult learning may impact on wellbeing directly and indirectly, say through improved social relationships that also in turn impact on wellbeing. The statistical techniques that are often used to estimate the effect of learning control for a large number of other factors (including social relationships) mean that these indirect effects are not acknowledged in the results and we do not derive a true effect of adult learning on wellbeing.

We cannot say with any certainty the overall likely direction of bias due to these four issues since they potentially work in different directions. The literature review and the methodologies applied in this study will, so far as is possible, account for these estimation problems.

For policy purposes, it is also important to understand the degree to which any result can be generalised to the larger population of interest. This is the **external validity** of the

study. It depends greatly on the samples and methodologies used. In general, the more robust studies are those whose results are representative of the whole population or of the sub-population of interest.

Internal validity is the first order concern, though, since without it we cannot start to make judgements about the generalisability of the results; external validity is inoperable without internal validity.

To account for the differences in the quality of research designs, we use a 5-point scale as a way to adjust the reported results. This is based on the five-point scale developed by researchers at the University of Maryland (Sherman et al., 1998).

We have added a 'five star' category to distinguish between randomised trials where participants do and do not know they are in a study. The latter is preferred because there is good evidence that people behave differently when they know they are being watched (most often referred to as the Hawthorne effect).

Table 1: Modified Maryland five-point scale

Level	Design	Statistical method
5*	Natural field experiments	*As per randomised trials, but where the participants do not know that they are part of a study*
5	Randomised trials	*Evaluations with well implemented random assignment of adult learning to subjects in treatment and control groups.*
4	Quasi-experiments	*Evaluations that use a naturally occurring event (that makes the adult learning assignment as good as random)*
3	Matching techniques; Regression analysis	*Non-experimental evaluations where treatment and comparison groups are matched on observable characteristics*
2	Simple comparisons	*Studies with a treated and comparison group, but with no attempt made to control for differences among the groups.*
1	Pre- and post-analysis	*Studies where no comparison group is used. Outcomes are measured pre and post-treatment.*

This is a ranking system of the internal validity of the study's results. As we move down from a five star rating, less and less confidence can be placed in any reported effects (or lack of effects) of the programme or policy intervention.

Level 5 and 4 designs provide solutions to the problems of selection bias, reverse causality and measurement error since the treatment (here adult learning) is derived exogenously to the system of equations. Treatment effects therefore will have a robust causal interpretation. Level 5 and some level 4 methods can also account for the indirect effects of the treatment.

Level 3 analyses are able to control for many of the differences between treatment and control groups, leading to a better understanding of causality that level 2 and 1 designs. Only under strict assumptions (selection on observables and no measurement error), however, can we confidentially conclude that we have an unbiased estimate of causality and the statistical design often masks many of the important indirect effects of the treatment.

Level 2 and 1 studies provide poor estimates of causality. In line with the Washington State Institute for Public Policy (WSIPP) (2006) we reject as problematic and non-robust any study below level 3; only studies of level 3 and above should be considered.

We shall rank the literature since 2008 according to these metrics and discuss the external validity of the findings separately.

3. Literature review

In this section, we provide a review of the literature on adult learning since 2008. We review papers that look at the impacts of adult learning on wellbeing defined very broadly. We searched the literature using the following regularly employed academic search applications: *EconLit, Google Scholar, JSTOR* and *Elsevier*.

A variety of different techniques have been employed in the literature since 2008, including both quantitative and qualitative methods. Both types of methodology can be assessed against the Modified Maryland scale. Overall, however, we conclude that the results and findings are not as robust as they are often claimed to be.

3.1. Qualitative studies

A fairly common technique has been to use surveys with case studies of people who have undertaken adult learning. The typical methodology is to ask a sample of adult learners whether they have experienced any benefits as a result of their learning.

The Mental Health Foundation (MHF, 2011) interviewed 256 adult learners before attending their course and at intervals up to 12 months after completion of the course. They report participants experiencing better wellbeing and less severe symptoms of depression and anxiety due to undertaking the course up to 12 months afterwards.

Estyn (2012) sampled 125 older learners in Wales and asked them to identify key outcomes of their learning experiences. They found that older learners reported having improved social contacts, skills, confidence and self-esteem, better health and greater involvement in the community (including volunteering) because of their learning. They also report that learning leads to better employment opportunities as evidenced by an increase in the level of qualification attainment among older learners.

The National Audit Learning Organisation (AONTAS) (2009) use a similar survey technique to assess the impacts of adult learning on improved job outcomes using only a sample of three and report significant benefits.

The National Institute of Adult Continuing Education (NIACE) (2009) use a number of case studies and report that adult and family learning leads to gains in confidence, better communication, new skills, changed behaviours and improved relationships with family and community.

The National Research and Development Centre for Adult Literacy and Numeracy (NRDC) (2009) assess the impact of 74 family literacy programmes by surveying a total of 1,100 children and adults involved in the courses. Looking at literacy before and after the courses, NRDC found that both parents and children made progress in reading and that a majority of parents were able to attain qualifications. Qualitative surveys found that parents reported (i) being more involved with children's school work; (ii) having great self-confidence; and (iii) being more willing to enrol in further courses and to seek employment.

OFSTED (2009) undertook an assessment of 36 family learning classes at Sure Start centres and local libraries. It is not clear what the overall sample size was. OFSTED used end-of-course surveys that asked participants to assess how their circumstances had changed during the courses. A range of benefits for parents were reported, including, increased self-confidence and communication skills, better parenting skills, better numeracy and literacy skills often leading to level 1 and 2 qualifications. Teachers reported that most children improved their literacy and numerical skills and their concentration and behaviour in class.

All of these studies fall into the level 1 category in the table above. They post course trends, sometimes in comparison with pre-course, but without a suitable counterfactual. They provide some useful insight into the perceived benefits of adult learning. But they are problematic for a number of reasons. First, there is cognitive dissonance. When our beliefs and our behaviour come into conflict, we often realign the beliefs to fit the behaviour (Festinger, 1957). So, we have just done a course. Well, I spent all that time on it, so it must have been good right? Even if it wasn't, I will tell myself it was. As with most of the biases in life, they operate largely below conscious awareness, so we cannot be expected to know when they are present (Wilson, 2002).

Second, and related, we tend to behave in a way that supports the impression of a positive and consistent self-image. As one consequence of this, people who have benefitted the most from the education will self-select into the ex post survey and will have a greater change of attitudes and ratings of the course (Dolan et al., 2012).

Third, we are prone to focusing effects. So by asking questions about the educational course, you are drawing attention to the course, and people will think it is important because they are being asked about it. And they think it will continue to be important into the future because they are thinking about how important it will be (Kahneman et al., 2006; Schkade and Kahneman, 1998).

Fourth, and related, we cannot be expected to know what the counterfactual would have looked like. So we make it easier for ourselves by attributing any effect to what we are being asked about. Implicitly, then the control or counterfactual outcomes are created by the survey respondent. There is no way to assess whether their predictions and statements are correct and there is a large amount of evidence that people are unable to forecast counterfactual scenarios accurately or sometimes at all (Hastie and Dawes, 2010; Khemlani and Oppenheimer, 2010).

Fifth, our memories do not recall past utilities and their duration particularly well. It appears that we use a 'Peak-End Rule', which ignores the full set of experiences and how long these experiences last and instead we remember the most salient and the last experience. So we will remember the best and last moment of the education course, but not necessary how we feel throughout the whole course (Kahneman et al., 1993).

Sixth, the sample sizes can be very small, meaning that the results are also unlikely to have much external validity.

This is not to say that the results from such studies are worthless; they do provide important insights that would be missing if those studies did not exist. Specifically, they

help us develop a better understanding of the routes and mechanisms through which adult learning may impact upon the various wider benefits of adult learning, which is something that large quantitative datasets are sometimes less able to do.

It is also worth noting that quantitative studies are not without problems. Even randomised trials (level 5) suffer from Hawthorne effects (people behave differently when they know they are being watched), and so the results may not be generalisable to wider populations. Wherever possible, we would recommend a level 5* but sometimes this is clearly not possible.

So we must judge the evidence according to what would be known in its absence and, to restate, qualitative data is better than nothing. As Erasmus, the famous Dutch philosopher said "In the land of the blind, the one eyed man is king".

3.2. Quantitative studies

Since 2008 very few quantitative studies of adult learning have been undertaken (prior to 2008 a number of quantitative studies were published by the Centre for Research on the Wider Benefits of Learning (WBL), including some level 2 evaluations).

Sabates and Duckworth (2009) use ordinary least squares (OLS) regression on data from the 2004 Avon Longitudinal Study of Parents and Children (ALSPAC) to estimate the impacts of mother's learning on their children's academic performance. This survey was completed by about 5,400 parents and carers. They find, after controlling for a number of confounding factors, that mother's learning did not have any impact on children's maths and English scores. This study uses a level 3 design as it controls for other determinants of children's attainment.

This study shows the importance of controlling for differences among treatment and control groups as when these differences are controlled for the impacts of adult learning changes dramatically. It should be noted, however, that this study only controls for observable characteristics and variables and therefore results do not have a full causal interpretation if there are also unobservable variables on which people self-select into adult learning.

The same estimation methodology was employed by Jenkins (2011) using the English Longitudinal Study of Ageing (ELSA) to look at the impacts of adult learning on wellbeing (GHQ-12, life satisfaction and quality of life (CASP-19) of people aged 50 and over. Jenkins found that after controlling for a number of determinants of wellbeing, music, arts and evening classes had a significant positive effect on all three measures of wellbeing, whilst formal courses did not have any impact. On whole the whole sample size was too small to derive meaningful results broken down by demographic characteristics. Although the previously discussed issues surrounding self-selection may bias the results in this study it is interesting to note that formal learning has little impact for older generations.

Matrix (2009) also use a level 3 design in its study of the impacts of adult learning on four measures of wellbeing: the General Health Questionnaire (GHQ 12 and 36), happiness as measured in the GHQ and life satisfaction from the British Household Panel Survey (BHPS). They used eight years (waves) of the BHPS with a sample size of about 22,000

which was taken from the nationally representative data (it is unclear whether they set any restrictions on the sample and so we assume that all individuals with the relevant data would have been included in this sample). They found that undertaking formal learning (defined as part-time learning that led to a qualification) had positive impacts on all four measures of wellbeing, whilst informal learning (defined as part-time learning that did not lead to a qualification) had positive impacts on all measures except GHQ12.

The study controls for the main determinants of wellbeing and also uses a fixed effects estimator that can also control for unobservable time-invariant variables, such as ability, preferences for adult learning and possibly motivation. Although time-variant factors cannot be included in the model Matrix's OLS with fixed effects model design will provide better estimates of causality than standard OLS as the model can control for more factors.

We can be more confident in the findings of these level 3 quantitative studies than the level 1 qualitative studies discussed above. It should be acknowledged, however, that the validity of these studies depends greatly on the extent to which all of the main determining factors have been captured in the model. Even if we can make the assumption that selection into adult learning is exogenous conditional on observable characteristics, the study still needs to have controlled for the all relevant characteristics. OLS with fixed effects can provide part of the solution as it controls for time-invariant factors.

These studies cannot, however, account for time-variant unobservable factors. For example, say that motivation varies over time and that highly motivated people are more likely to undertake adult learning and also be happy or satisfied with life anyway. Then even under a fixed effects design we will have an overestimate of the effect of adult learning on wellbeing. Furthermore, issues surrounding measurement error and indirect effects still cause problems in the OLS approach and in fact fixed effects estimators can exacerbate measurement error problems by increasing the ratio of measurement error in a variables variance (Deaton, 1993).

External validity is maximised when the results tell us the *average treatment effect* (ATE). Taking adult learning to be the "treatment" here, the ATE would be optimal because it would tell us the causal effect of adult learning (on the outcome of interest) for anyone plucked out at random from the population being sampled. With a sample that is nationally representative, such as the BHPS, we could interpret this result as being the expected causal effect of adult learning for anyone in the UK. If there is selection on unobservable factors, ATE can only be estimated through randomisation.

Sabates and Duckworth (2009) and Matrix (2009) use observational data with regression analysis. Assuming that in their studies the assumption of selection on observables holds (we have discussed above that this is unlikely to be the case) so that adult learning is exogenous, we do have a causal interpretation but it is not the ATE.

In fact, what is estimated is suggested to fall between the average treatment effect for the treated (ATT) and the average treatment effect for the non-treated (ATNT) (Humphreys, 2009). The ATT is the causal effect for those that did take the treatment and the ATNT is the effect that would have happened if the untreated would have taken the treatment. These effects refer to specific sub-groups in the population and are therefore narrower than the interpretation for ATE. The difficulty is that it is unclear to which groups in the

population we can generalise the results. We discuss this issue in more depth in our results section.

We are aware of two level 5 study designs in the area of adult learning. Wolter and Messer (2009) use a large-scale field Swiss experiment, where in 2006 education vouchers were randomly distributed to a sample of 2,400 people. They compared this treatment group's behaviour with that of a control group of 10,000. Since vouchers were randomly assigned we can be confident that on average all attributes across the treatment and control groups were equal pre-treatment and that any differences in outcomes observed between the two groups were due to adult education only. They found that the vouchers lead to causally higher level of participation in adult education. Using an instrumental variable approach, where receipt of the voucher instruments for undertaking learning, they also found that there was no causal effect of adult learning on employability, wages and job security over a one-year period.

This is a much more robust study design: (i) all (observable and unobservable) characteristics/factors are controlled for through randomisation; (ii) since adult learning is exogenous under randomisation, there is no correlation between adult learning and error terms and hence issues regarding measurement error are solved for; and (iii) indirect effects are accounted for because the study only looks at labour market outcomes (without controlling for any other factors).

Wolter and Messer's study finds there to be a significant amount of selection into adult learning. It is clear that even after controlling for observable factors, there is likely to be a source of bias due to selection on observables and measurement error, making the case stronger for focusing on higher ranking study designs.

Finally, Trivette et al. (2009) look at an interesting aspect of adult learning. They focus on the effects of learning and teaching methods, rather than just adult learning per se. Their meta-analysis of large number of randomised trials (level 5 study designs) finds that all main adult learning styles under consideration had a causal effect on knowledge, attitudes and self-efficacy beliefs.

Under random allocation of the treatment (adult learning via vouchers in Wolter and Messer (2009) and adult learning styles in Trivette et al. (2009)), we can estimate the ATE (provided that randomisation worked properly and there is no need to control for any other variables). Wolter and Messer's random assignment of educational vouchers in Switzerland used a large sample that was randomly generated from the whole Swiss population. Their finding of the positive causal effect of vouchers on adult learning take-up is the ATE. In other words, if we were to give out an educational voucher, we can expect a higher level of participation in adult learning for anyone in Switzerland (regardless of their characteristics).

The non-existent causal effect of adult learning on employment factors has a slightly different interpretation. Since there was some non-compliance in Wolter and Messer's sample (i.e. some people that received vouchers did not enrol into a course), the effect on employment factors is only relevant for some of the sample population and these are those people who did undertake a course due to the voucher.

We therefore have a more localised effect than the ATE, known as the *local average treatment effect* (LATE). The interpretation turns out to be that for people in Switzerland, who would undertake a course if provided an educational voucher, there is no effect for them of adult learning on employment factors.

We cannot say from this study whether there would be an effect on employment factors for the types of people that *do not* take-up adult learning after being given a voucher. This may seem like a rather minor and pedantic point, but it is important to the extent that it shows that the (non-existent) effect of adult learning on employment factors is not one that should be generalised to the *whole* population without further research.

4. Scoping and analysis of existing datasets

4.1. Theoretical background

Human capital theory suggests that individuals and society derive economic benefits from investments in people. Although types of human capital investment generally include health and nutrition (Schultz, 1981) education consistently emerges as the prime human capital investment for empirical analysis. One main reason for this is that education is perceived to contribute to health and nutritional improvements (Schultz, 1963)

In addition, there is theory and evidence that education tends to impact on health, attitudes and behaviours, civic participation and family life. Education may impact on population growth and increase overall quality of life (Becker, 1993). There are large differences in life expectancy by education across every demographic group (Meara et al., 1981). Education gradients in health behaviours are large; controlling for age, gender, and parental background, better educated people are less likely to smoke, less likely to be obese, less likely to be heavy drinkers, more likely to drive safely and live in a safe house, and more likely to use preventive care (Cutler and Lleras-Muney, 2007).

Education is also theorized to provide the means to "an enlightened citizenry", meaning that people are able to participate in the democratic and legal system, and to pursue values such as equality, fraternity, and liberty at both private and social levels (Swanson and King, 1991)

While these qualitative benefits may represent the most important contributions made by education, each is difficult to measure quantitatively. Perhaps this explains why wage outcomes and economic growth have become the benefit of choice for empirical analysis (Woodhall, 1987).

The aim of the present paper is to try to shed light on some of these wider non-economic benefits. The studies quoted above focus mainly on school-level education but the theory of human capital can be applied to adult learning too and we would expect to see similar types of benefits to adult learning. So as suggested by human capital theory we test for the impacts of adult learning on:

- Health and mental health;

- Family and parenting;

- Civic participation

- Attitudes and behaviours.

4.2. Data

The analysis uses the British Household Panel Survey (BHPS), which is a nationally representative sample of over 10,000 adult individuals conducted between September and December of each year from 1991. Respondents are interviewed in successive waves, and all adult members and children aged 11-16 in a household are interviewed. The BHPS contains a wide range of questions on adult learning and its panel format (that tracks people over a number of years) make it a very useful data source for examining questions on adult learning.

We focus on people aged 18 and over and exclude adults in full-time higher education from the analysis. Data on adult learning have been taken in the BHPS since 1997 (Wave H).

We focus on two types of adult learning:

i. **Formal learning** that leads to a qualification; and

ii. **Informal learning** that does not lead to a qualification.

In the BHPS, people are asked questions about up to three courses that they have taken over the previous year. We do not look at the impacts of a qualification on its own since the focus of the report is on informal learning.

It should be noted that these adult learning variables are self-reported and hence will invariably be measured with some error. As noted above this leads to a downward-bias in the coefficient on adult learning in equation (1) and so our results to some extent may be conservative (if we can assume that measurement error is the only problem with the model in (1)).

Where possible (where models are robust and sample sizes are large enough to give meaningful results), we also differentiate the impacts of both types of learning across different demographic groups broken down by:

- Age (18-40 year olds and 40+ year olds);

- Gender;

- Household income (£0-£25,000/£25,000+)

- Parental status (children or no children).

We use only two groups in each case in order to maintain sufficiently large samples in each group. Other groups we could have included are ethnicity, prior educational status/qualifications and disability. We found that data on ethnic group is poor in the BHPS, with a large number of non-responses which led to a lot of insignificant results. Prior qualifications and disability (or health status) map closely with income (i.e. disabled people and people with few qualifications generally have low income) and so we focus on

income in the analysis here. Income groups were determined based on the distribution of household income in the sample we used in the BHPS (0-50th/50th+ percentiles).

We assessed the full BHPS dataset for all variables that are associated with the four outcome domains of interest:

- Health and mental health;

- Family and parenting;

- Civic participation;

- Attitudes and behaviours.

In total we found 36 outcome variables across these four domains. There were a number of other variables that we could have used as potential outcomes but these have been excluded either because (i) they were only collected before 1997 (the first of data on adult learning), or (ii) they were collected after 1997 but only for one or two years which would not create enough of a sample size.

4.3. Methods

With data like the BHPS, we would ideally use a level 4 design according to the Modified Maryland scale (level 5 requires experiments to specifically test for the impact of adult learning). We are essentially reliant on the data observed containing some events that can be assumed to have been randomly assigned. Very often these types of events can be found where Government introduces a new policy to only a sub-sample of the population, or accessibility or eligibility to a treatment or policy intervention is based on some form of random assignment, such as a lottery.

Here we are essentially looking for events that have randomly assigned participation in adult learning across a sub-population. This approach has been used regularly in the US literature but has not been used very much in the UK due to lack of data and suitable policy settings.

A rare exception is Feinstein and Sabates (2005), who use the Educational Maintenance Allowance created by the then Department for Education and Skills, which distributed study allowances to encourage people between 16-18 to take up further education. They assessed the impacts on juvenile crime levels. This study does not technically address the issue of adult learning, showing the difficulty and lack of suitable data to implement this type of quasi-experimental study with adult learning. Indeed, we were not able to find a robust quasi-experimental setting in the BHPS data.

Alternatively, level 4 designs can use randomisation in a factor or variable that impacts on adult learning, rather than randomisation in adult learning itself. Unbiased estimates of the causal effect of adult learning can be derived in this approach by using an instrumental variable (IV) technique. In intuitive terms, the IV technique finds an event that is essentially random and which directly affects the treatment of interest (here adult learning). The

technique can then feed this randomisation through so that the treatment (adult learning) also becomes a random event and we can then attribute causality to it.

The study by Wolter and Messer (2009) is an example of using IVs. Here educational vouchers were randomly assigned. For some people, this led to an increase in adult learning take-up. The IV technique uses the changes in adult learning take-up that are due to the voucher scheme only, which was random. Apart from Wolter and Messer (2009), we are unaware of this kind of IV technique being used before in the adult learning literature.

We made a systematic assessment of the BHPS data to see if an IV technique could be viable. We looked at possible factors that could influence adult learning uptake by making changes in people's financial and time resources. We hypothesised that events or factors that decreased people's time or finances would have a negative impact on the likelihood of them undertaking learning. For events that lead to time constraints we used birth of a child (controlling for factors that are associated with fertility), but found the instrument to be weak. For finance constraints we used both involuntary redundancy and lottery wins as instruments.

Sample sizes were small, however, and finances and time are likely to directly impact on the outcomes of interest (such as mental health and family interaction) even after controlling for other determinants, which rules them out in terms of being valid instruments. We conclude, therefore, that we cannot use a level 4 study design with the adult learning data in the BHPS. This is quite a common finding in the empirical literature since finding suitable IVs is a notoriously difficult challenge.

Level 3 designs match treated and non-treated groups based on observable characteristics. The preferred method here is to exploit the longitudinal nature of the BHPS panel by using ordinary least squares (OLS) regression with fixed effects (FE). This is similar to the estimation methodology used by Matrix (2009), as described in the literature review. The basic technique will be to estimate the following function:

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \qquad (1)$$

where Y_{it} is the outcome of interest and X_{it} is a vector of explanatory variables that are determinants of Y_{it}. Adult learning variables will enter this vector. The subscripts i and t refer to the individual and time period respectively. ε_{it} is the error term which can be broken down into:

$$\varepsilon_{it} = \mu_i + \omega_{it} \qquad (2)$$

where μ_i is the time-invariant unobservable factors that affect Y_{it} and ω_{it} is the time-variant unobservable factors that affect Y_{it}. Through OLS FE we can control for μ_i and this is the benefit that we can derive from panel data using FE relative to OLS on cross-sectional data. In essence, in the error term we are left with ω_{it} only. If ω_{it} is uncorrelated with the other determinants of Y_{it} (i.e. X_{it}), then the effect of adult learning on Y_{it} will have a causal interpretation. This proposition, however, cannot be tested and hence here we have to assume that there is no correlation between X_{it} and ω_{it} with sensible caveats.

A caveat to this approach is that by using FE we lose all between-individual variation and hence are reliant on variation of the variables within individuals over time. If theirs is no or little variation over time within individuals then variables like adult learning will not have significant effects on the outcome of interest and this is a trade-off we make in order to be able to control for unobservable time-invariant factors using FE.

To further increase the validity of the causal interpretation a possibility would be to use lagged adult learning variables in equation (1). This uses the notion of Granger causality, where causality could be more strongly inferred in circumstances where the determinants (explanatory variables) occur in a previous timeframe. This helps to reduce the bias caused by reverse causality but it is not a panacea since there may still be confounding unobservable factors at work.

Lagging in panel data for the current exercise requires that we have a large number of individuals with year-on-year observations, but this is often not the case as there is attrition in the sample or some people only respond in odd years. This is what we found to some extent in the BHPS as when using lags for adult learning we lost a lot of observations making sample sizes small with largely insignificant results. Even if we did have large samples after lagging we would still need to question whether using lagged adult learning variables would provide plausible results.

This is because, for example, an individual may have taken a course in October 2005 (a month after completing the BHPS survey). He would then take the survey in September 2006 and 2007. For this individual we would be estimating the effect of a course taken in October 2005 on say health in September 2007 (23 months later). The impact is unlikely to show up especially if it was a short course. We therefore do not use lagged adult learning variables in the final results.

As described in equations (1) and (2) we use FE models. For outcomes measured on continuous scales or interval scales with a clear ranking we use ordinary least squares with FE. For binary variables we use a conditional logit model with FE. In general, therefore, in our results models with binary outcomes tend to have fewer observations since many observations tend to be dropped due to no within-individual variation of the outcome variable. This is not a problem we see with continuous variable outcomes since variation in these variables is much greater.

For regression models in each domain, the other explanatory variables (apart from adult learning) were determined from reviews of the relevant literature on the topic (references can be found in the notes at the end of each table below).

In general, sample sizes are large and hence we put greatest weight on results that are significant up to the 5% level. If neither adult learning variable (formal and informal learning) is significant even at the 10% level under FE estimation, we estimate OLS or logit models instead. These models are more likely to find significant effects as they will include between-individual variation in the variables of interest, but this is at the expense of not being able to control for unobservable time-invariant factors which may lead to bias in the coefficients on the adult learning variables.

For logit models without FE it is possible to estimate the marginal impacts of adult learning. These marginal percentage impacts are presented in brackets in Table 2 where

23

FE has not been used. When FE is used in logit models marginal percentage term effects cannot be estimated because we cannot make any assumptions about the distribution of the unobserved FE. In logit models with FE the coefficients purely represent changes in the log odds ratio and hence only have intuitive meaning in relative terms compared to other coefficients.

In order to provide more context to the FE results where appropriate (e.g. for the health models), we have compared the size of the impacts for adult learning to the impact of being employed (where employment is significant) since employment status was included as a key determinant in most models. It was not appropriate to do this for all FE models since, on occasions, employment had the opposite effect and hence a comparison was not meaningful.

4.4. Results

4.4.1 Overall results

We present the results in Table 2. As noted above, for logit models without FE it is possible to estimate the marginal impacts of adult learning. These marginal percentage impacts are presented in brackets in Table 2 where FE has not been used.

A clear message from the results in Table 2 is that the main effects are seen in the domain of health and wellbeing. Impacts of adult learning on all bar one health and wellbeing outcome are significant at the 5% level under FE estimation. In the other three domains, results tend to mainly be significant when estimated without using FE and in the family and parenting domain many outcomes are insignificant even under non-FE estimation.

These findings together with Matrix's (2009) study, which showed significant positive impacts of adult learning on life satisfaction, GHQ happiness and GHQ 36, using the same study design (OLS with FE), suggest that adult learning has its greatest impacts in the domain of mental and physical health.

4.4.1.1 Health and mental health
After controlling for other determinants and fixed effects, participation in formal and informal adult learning courses is associated with:

- Improvements in self-confidence

- Improvements in own perceptions of self-worth

- Reductions in self-reported depression

Impacts on self-confidence and perceptions of self-worth are greater for formal learning.

After controlling for other determinants and fixed effects, participation in formal adult learning courses is associated with:

- A reduction in the number of visits to a GP.

- Improvements in self-reported overall health satisfaction.

After controlling for other determinants and fixed effects, participation in informal adult learning courses is associated with:

- Increases in satisfaction with social life

- Increases in satisfaction with use of one's leisure time.

Adult learning and employment both have positive effects on health. They both reduce the number of GP visits per year and we find that the impact of formal learning on the number of GP visits has about one-seventh of the impact of employment on GP visits. The positive impact of formal learning on health satisfaction is about half of the impact that being employed has on health satisfaction. The impact of both types of learning (formal and informal) on self-confidence is very large; more than twice the impact of employment on self-confidence.

There is also some evidence that formal learning may reduce self-reported drug and alcohol problems but these results are not significant in the more robust FE models. Here a marginal effect can be attributed, and we see that participation in formal learning is associated with a 0.1% decrease in the likelihood of reporting drug and alcohol problems, which is not especially large.

4.4.1.2 Family and parenting

After controlling for other determinants and fixed effects, participation by one or both parents in informal adult learning courses is associated with:

- Increases in the probability that the children in the household (11 - 16 year olds) speak more frequently with the mother about serious issues.

For all other family and parenting, we find no effect of formal and informal adult learning when using FE models. Under less robust non-FE models there is some evidence that formal learning is associated with a 15% increase in the probability that the learner will help their adult children with small errands and chores (such as paying bills and writing formal letters). There is also a small impact (about 1%) on the probability of separation, and a negative impact on how children (11 - 16 year olds) feel about their own family life (children feel less happy about family life if parents undertake adult learning), but these results are only significant at the 10% level under non-FE models.

4.4.1.3 Civic participation

After controlling for other determinants and fixed effects, participation in formal and informal adult learning courses is associated with:

- An increase in the likelihood of trade union membership. Formal adult learning has a greater impact.

After controlling for other determinants and fixed effects, participation in formal adult learning courses is associated with:

- Greater involvement in voluntary work.

There is also some evidence in non-FE models that formal and informal learning may increase the likelihood of becoming a member of:

- a political party by 0.4%;

- an environmental group by about 1%;

- a parents association by about 1%;

Furthermore, in these models we find that formal and informal learning may increase the likelihood of:

- frequently attending local groups by about 2%

- doing sport by about 4%.

It should be noted, however, that these results are not significant in the more robust FE models.

4.4.1.4 Attitudes and behaviours

After controlling for other determinants and fixed effects, participation in formal and informal adult learning courses is associated with:

- A desire to find a better job. The impact of informal learning is greater.

- An increased likelihood of reporting improved financial expectations for the following year. Formal learning has a slightly larger effect

There is also some evidence that formal and informal learning may reduce the perception of how important money is in life. On a scale of 1 to 10, on average, people report being 0.2 index points less likely to perceive money as important which is about a 3% fall from initial average levels. There is a similar *increase* in the perception of the importance of having a good job. Informal learning may also be associated with a reduction in people agreeing with traditional roles for men and women; with informal learning people are about 3% less likely to agree with the statement that *'husband should earn while wife stays at home'*. It should be noted, however, that these results are not significant in the more robust FE models.

4.4.2 Differences by demographics

As discussed above we favour FE estimation. For this reason we undertake analysis of demographic breakdowns on models that are only significant under FE estimation (this is similar to the method and process undertaken by Jenkins (2011)). This leaves us with 11 outcomes. It should be noted that when breaking down samples by demographic groups there are a few rare occasions when the estimates by demographic group do not lie in the range around the overall estimate as would be expected. This happens if there are missing

values for the demographic variables on which we split the samples or it can happen in binary logit FE models as splitting down samples loses some observations when using these models.

Table 3 shows how these main impacts vary across different demographic groups. On a number of occasions we find a significant overall effect for learning but insignificant effects once broken down by demographic group and this was usually due to wider standard errors due to small sample sizes. Also, results tend to be varied, but there are a few overall messages that we can discern.

For health and mental health, where comparisons are possible (for self-confidence, depression, self-worth) we find that the impacts on health tend to generally be larger for men and those aged 40 and over. An interesting finding is that impacts for parents and different income groups differ by learning type. With respect to these three health outcomes parents and lower income groups derive relatively more health benefits from formal learning. Non-parents and higher income groups derive more health benefits from informal learning.

For civic participation, the positive impact of learning on volunteering is greater for women, younger age groups (18-40 year olds) and parents. The impact of learning on trade union membership is greater for older age groups.

In terms of attitudes and behaviours, the positive impacts of learning on the desire to find a better job and future financial expectations are generally larger for men, younger age groups, parents and lower income groups.

5. Discussion

5.1. Summary

Using human capital theory as our theoretical starting point, we explored the impact of adult learning on health and wellbeing, family life, civic participation and attitudes and behaviours. We focussed on outcomes that have not been looked at in a robust manner in the previous literature. Using the BHPS dataset we found that the best statistical methods we could employ to maximise causal interpretation was regression analysis with fixed effects.

These methods equate to the top end of level 3 study designs according to our Maryland Scale of evaluation techniques. Under the assumption that selection into adult learning is based on observable characteristics and unobservable time-invariant factors, our fixed effects results would have a causal interpretation.

Of course, there may be some important factors that we have not managed to pick up which would jeopardise the interpretation somewhat. We therefore prefer the language of 'statistical associations' after controlling for other determinants in describing our results. We deem that fixed effects methods provide the best estimates of causality in our framework and have focused the discussion on those models here.

It is quite clear that **adult learning has its greatest impacts in the domain of health and mental health**. Both types of learning (formal and informal) have significant impacts in this domain.

Our analysis also highlights that **non-health related impacts of adult learning tend to cluster around job-related or labour market outcomes**. There are significant impacts on trade union membership, job preferences, volunteering and future financial expectations. Here formal learning tends to have a greater impact although both types of learning are important. These outcomes are different to more traditional labour market outcomes that have been analysed in the past literature, such as wages and employment probability, and hence provide new results and directions for future research.

The analysis here would suggest that **impacts of civic participation and attitudes are less pronounced** and less robust. This may be a general observation or outcome or it may just be due to the types of outcomes that the data have allowed us to look at in this domain.

We are able at various times to consider the magnitude of the effect of adult learning compared to being employed. In the domain of health, we find that the negative effect of formal learning on the number of GP visits per year is about one-seventh of the impact of being employed on GP visits. There is a positive effect on health satisfaction from formal learning, which is about half of the corresponding impact of being employed. The impact of adult learning (both formal and informal) on self-confidence is very large; more than twice the effect that being employed has on self-confidence.

As discussed, it has been difficult to attribute magnitudes of effects to these impacts since we attempted to use FE models wherever possible in order to maximise the robustness of results. The trade-off is that FE model outputs then do not have easily interpretable marginal effects. Where marginal effects are discernible (because we have not used FE modelling), the impacts are sometimes quite large e.g. formal learning is associated with a 15% increases in the probability that the learner will help their adult children with small errands and chores, but at most other times quite small e.g. a 4% increase in the likelihood of participating in sport due to adult learning and a 1% increase in the likelihood of joining a school parents association.

If the methodologies and results presented here are deemed robust enough (i.e. the methodological assumptions hold to make the modelling work valid) this would suggest that we conclude that adult learning does not have significant benefits in this area and therefore that further research on the domain on family outcomes should not be pursued. On the other hand, if the assumptions underlying the methods presented here can be contested, then this would set a call for further research in this area. As we shall discuss below it is hard to make a definitive conclusion without more robust (level 4 or 5) evidence.

The analysis in this paper suggests that adult learning is likely to impact most on health and labour market related outcomes. A good starting point for future research would therefore be to focus on these two outcome domains and to try to derive more definitive and robust results. As we shall discuss, this can only be done by introducing experimental design studies.

5.2 Internal and external validity revisited

As discussed earlier, if the assumption of selection on observable and unobservable time-invariant characteristics holds then our estimates in Table 2 have a causal interpretation. There may, however, be issues related to measurement errors in the adult learning variables and we cannot solve for these issues with a FE model. Thus the coefficient on the adult learning variables may be biased downward.

Furthermore, we have controlled for a number of variables in the regressions analyses and some of these are likely to be impacted on by adult learning, which would dilute some of the effects of adult learning. For example, undertaking adult learning may impact on income (through higher earnings) and since income is a determinant of mental health (say depression) and is controlled for in the depression regression the effect of adult learning through income on depression would not be picked up. But we cannot drop depression entirely form the regression as it would result in omitted variable bias and impact on all coefficients in the model.

Given some likelihood of measurement error in self-reported variables like adult learning and the parametric restrictions imposed by the modelling framework, we can tentatively conclude that the estimates of the effects of adult learning are likely to be biased downwards to some degree and hence are conservative (provided that we can assume selection on observable and unobservable time-invariant characteristics and hence rule out any bias due to endogeneity). If there is also selection on unobservable time-varying characteristics and reverse causality then we can make no claim about the direction of bias since endogeneity bias may work in the opposite direction.

29

Turning to the issue of external validity, it is helpful to know for which groups in the population the results are relevant. Here we assume that selection on observable and unobservable time-invariant characteristics holds (and hence no endogeneity bias) so that our results have internal validity.

Take, for example, the positive impact on self-confidence found in Table 2 under FE modelling. Let us assume that the coefficient on informal learning (0.295) equates to a 10 per cent increase in self-reported confidence (as explained above we cannot estimate actual percentage changes in FE models). This is an averaged impact of adult learning calculated from our designated sample (people aged 18 and over who are not in full-time higher education). Does this mean that we can state that we would expect *anyone* fitting the characteristics of our sample to expect this kind of effect on average? This is what is known as the average treatment effect for the sample (ATE).

If 'treatment' (undertaking adult learning) was a truly random event across our sample, then we could infer that our estimates represent the ATE. But treatment is not random in our sample and hence we have to control for other factors in the regression analyses. In short this takes us away from the ATE. Therefore, it would not be correct to interpret the coefficients in Table 2 as the effects of adult learning we would expect any adult in the UK to experience on average. The effect that we actually measure is somewhere between the treatment effect for the treated (ATT) and the treatment effect for the non-treated (ATNT).

On average in our sample in the BHPS, about 17% of people undertake adult learning. OLS estimation with non-common treatment (here adult learning) can be shown to lie closer to the ATT than the ATNT (Humphreys, 2009). Thus, we take it that our estimates approximate the ATT. The ATT provides information on what the effects of the treatment were for the treated group. In strict technical terms they should not be translated as the effects we could expect to see for someone who did not undertake adult learning (this is the ATNT).

In a nutshell, our results are likely to represent the effects we would see on average for the type of people who are likely to undertake adult learning anyway. They may not accurately reflect the effects we would expect to see if we encouraged more people in to learning (who would not have done it of their own doing). This is true for any previous study on adult learning that has used regression analysis. It is an unavoidable outcome of using OLS with selection on covariate models and is a further argument to direct future research towards experimental designs that can provide more general estimates such as the ATE, which are more meaningful for policy.

5.3 Implications for future research

We have reviewed the qualitative and quantitative literature since 2008 and we have also undertaken a large modelling exercise using one of the most comprehensive panel datasets in the UK in terms of adult learning variables. A clear conclusion from the report is that for the most part we are restricted to level 3 study designs. These studies rely on observational datasets to tease out causality by controlling for differences in other factors that may impact on the outcomes of interest.

Because of the panel nature of the data we were able to control for time-invariant unobservable characteristics. This, however, was at the cost of a loss in variation in adult learning variables since we could only use within-individual variation in the FE model. If the assumptions implicit in FE regression modelling holds then we can assume that the effects we have derived have a causal interpretation. We feel that this is probably the best one can do given the datasets available to us currently in the UK.

This is a strict assumption, of course, and future research in this area should start to move away from level 3 study designs so that causality can be more robustly attributed – so that we can address issues of reverse causality, measurement error and indirect effects as well as selection effects.

We need to seriously consider the possibility of conducting natural field experiments (NFEs) at level 5*. NFEs maximise internal and external validity which will provide the most robust evidence for input into policy-making.

To causality with internal validity requires that we have knowledge of the counterfactual. In essence, causality can only be understood if we know for individual i the outcome that occurred with treatment (here an adult learning course) and the outcome that would have occurred if *i* had not been treated. For a given individual both of these outcomes of course cannot be observed and this is the fundamental problem in causal inference {Holland, 1986 #7}. All study designs in Table 1 attempt to solve for this problem and are ranked according to how well this can be achieved.

The main aim is work at aggregate levels – treated and control groups (since we cannot ever view the treatment and counterfactual for a given individual simultaneously) – and to control for factors that confound the relationship between learning and the outcome of interest (including the outcome itself as there may be reverse causality). Randomisation of treatment provides the most robust way of doing this. If treatment is randomised correctly, all characteristics (observed and unobserved) between the treatment and control groups will on average be equal. Since treatment has been randomly assigned it cannot be based on the outcome itself either (i.e. for example happier people would not be assigned more treatment), and hence endogeneity and reverse causality will both be solved for.

In sum, through randomisation treatment and control group will be identical on all characteristics except the treatment intake and so any differences in outcomes can be attributed *solely* to the treatment. Given this, we can compare means in outcomes across the treatment and control groups to determine the causal effect. This is akin to running the following regression:

$$Y_i = \alpha + \beta D_i + \varepsilon_i \qquad (3)$$

where Y_i is the outcome of interest for individual *i* and D_i is treatment status, such that $D_i = 1$ if *i* has taken a course and $D_i = 0$ if not. Since the randomisation allows us to run this regression without controlling for any other characteristics, we do not end up controlling for any indirect effects of adult learning. Further since treatment status is observed fully by the researcher measurement error is minimised.

Overall, therefore, randomised treatment in experimental settings can solve for the threats to internal validity. It should be noted that this does not nullify the need for qualitative

evidence since such data can be used as outcomes in the NFE. In other words, we can look at the causal impact of randomised adult learning assignment on self-reported outcomes such as confidence or parenting.

NFEs differ from traditional randomised controlled trials (RCTs) that are common in the medical field in that typically treatment is randomised across different groups in real-life settings. Crucially, those involved in the trial do not know they are part of a trial. Behaviour is known to be different when people know they are being watched. This is known as the Hawthorne effect - a term first coined by Landsberger (1958). This is one reason why the results of RCTs rarely translate to natural settings.

An important issue we have discussed is how far we can extrapolate the results to other populations – external validity. Ideally, we would like to know the average treatment effect of adult learning for the whole population and populations of interest. As discussed these are average treatment effects (ATE) for the population or the sample of interest. We noted that the regression analyses to date have measured an effect that lies somewhere between the ATT and ATNT. We were able to provide a more concrete estimate in our analysis by looking at the proportion of people who undertake learning in the BHPS data and concluded that our estimates were closer to the ATT. The ATT provides historical information on what the effects of the treatment were for the treated group.

In strict technical terms, they should not be translated as the effects we could expect to see for someone who did not undertake adult learning (this is the ATNT), or for some person randomly chosen from the wider UK population. The ATT is useful information for cost-benefit analyses of previously implemented policies but the evidence could not be used to infer cost-benefit ratios for future adult learning interventions with different target groups.

In experimental settings, where the researcher can assign treatment and determine the appropriate samples, it is possible to derive the treatment effect of interest for the target group of the policy. We therefore can get much more focussed estimates of the treatment effect, which is harder to achieve with observational data. In fact NFEs will provide the ATE and ATT for the population of interest which is probably more useful for policy, since the ATE will tell us the effect of adult learning we can expect on average for the sample we have used.

Nothing is ever perfect, of course, and NFEs are no exception, especially if they are not designed properly. There are, in fact, four concerns with them. First, NFEs can fail to provide internally valid results if the randomisation protocol fails in ways such that characteristics are not balanced between the treatment and control groups. Outcomes will then not be solely due to treatment intake. Second, there may be non-compliance with the experimental protocol such that some of the controls manage to get treated or vice versa leading to similar problems. Controls may also get treated with alternative programmes – a problem known as substitution bias. Third, a related problem is that of attrition where people drop out after learning more about the benefits of the programme. This causes problems for internal validity if drop outs are a certain type of individual. Fourth, external validity is threatened if non-representative samples are used or if different processes and protocols are in place when the policy is scaled up to a larger population.

Some people would add a fifth concerns – that NFEs may be unethical. For example, it would be unfair to knowingly prevent treatment for some groups. This is only true if we have a good idea that the treatment is likely to have a beneficial effect in the first place. The whole point of an experiment – of any kind of evidence – is to find out what works and so this is a spurious argument. In fact, it is unethical not to find out what works and to roll out policies which may be of no benefit, or even cause harm.

The great advantage of using NFEs is that the outcomes of interest can be designed into the study. We would begin, however, by focussing on the 'big three' – health, wellbeing, and labour market outcomes.

Although labour market outcomes have been the emphasis of a lot of the adult learning literature to date, it would be interesting to focus on this topic once again for two reasons. First, our analysis shows that previously not studied areas of employment, such as trade union membership, volunteering and job preferences, may be important outcomes of adult learning and it would be good to assess these relationships using level 5 study designs. Second, Wolter and Messer's (2009) study which generally uses a more robust study design than the previous literature seems to contradict much of the previous literature in that it finds no causal effect of adult learning on employability and wages. As discussed, these results are specific to the Swiss population and thus it is important to test the labour-market – learning hypothesis further in a UK setting using robust experimental study designs.

One possible way to randomise the intervention would be to randomise vouchers for different types of adult learning (including a control group that received 'standard practice'). Because the NFE can focus on what and whom it likes, and over any time frame, we would need to work with BIS to develop and design this idea further.

Field experiments can test a number of competing hypotheses by including more than two arms in the experiment design. For example participants could be assigned to formal learning, informal learning and no learning through the vouchers. The larger the sample size the more confidence we can have in accepting or rejecting a null hypothesis. Academic papers in psychology run experiments on as few as 50 participants in each arm. Ideally a few hundred participants would be assigned to each arm at the very least. Wolter and Messer (2009) use much larger sample sizes (2,400 in the treatment group and 10,000 in the control group). Larger samples also help to deal with sample problems like attrition and non-compliance. It is also important to run experiments with different sample populations so that we can account for heterogeneity in treatment effects.

In theory only post-treatment data need be taken but it is common for experiments to also take some pre-treatment data so that the effectiveness of randomisation can be tested by looking at how well the two groups match pre-treatment. Through virtue of randomisation we can assume that the only difference between the treatment and controls groups is the receipt of treatment (eg, a voucher for adult learning). Outcomes can be tracked for any period of time and as long as attrition and non-compliance do not become a large problem, this can be for many years. In theory we could observe the effect of learning on an outcome like health for months or years after the field experiment and still attribute causality. These outcomes can be objective, such as GP visits or self-reported subjective outcomes, like self-confidence, so field experiments are very flexible.

By shifting future funding and resource to these areas of research we would be able to explore in greater detail the relationships and associations between learning and its outcomes demonstrated in the current paper and the previous literature. The strength of the study design would allow us to confirm, adjust or reject previous findings, in the process consolidating our knowledge of the impacts of adult learning in the UK.

5.4 Implications for policy

That is all for the future. There are a few important things we can say now. From our analysis, the main wider benefits of adult learning show up in health, mental health and job-related outcomes. The previous literature generally supports this. Both formal and informal types of learning tend to matter, suggesting that participation in learning in itself is important. We also found evidence to suggest that there may be some impacts in the domains of civic participation and family and parenting, although the results here are less robust.

Overall, these results suggest that the wider impacts of adult learning need to be considered in any policy decisions related to this area. Also, it would be useful to assess budgetary impacts of these outcomes e.g. NHS expenditure.

Exploring the impacts by different demographic groups brought up some interesting results. Focussing resources on male and older learners may produce the largest impacts on some of the health outcomes, and encouraging parents and low income groups into formal learning would produce relatively large health benefits.

Given the increasingly important role attached to voluntary work by government, it would potentially be useful to further explore why the impacts of adult learning on volunteering are larger for women, younger age groups and parents.

Table 2: The impacts of adult learning

	Health and mental health					
Dependent variable	**Type of variable**	**Formal learning**	**Informal learning**	**Model**	*No. of Obs*	*Fixed effects*
Number of visits to GP	Continuous variable	**-0.039****	*Insignificant*	OLS	*62,057*	Y
Self-confidence (GHQ)	Binary outcome (1 = Not losing confidence; 0 = Losing confidence)	**0.324*****	**0.295*****	Logit	*37,996*	Y
Not depressed or unhappy (GHQ)	Binary outcome (1 = Not feeling depressed or unhappy; 0 = otherwise)	**0.276*****	**0.276*****	Logit	*35,133*	Y
Self-worth (GHQ)	Binary outcome (1 = Not feeling worthless; 0 = otherwise)	**0.558*****	**0.446*****	Logit	*39,818*	Y
Satisfaction with health	Scale: 1='not satisfied'; 7 = 'completely important'	**0.044****	*Insignificant*	OLS	*54,877*	Y
Satisfaction with social life	Scale: 1='not satisfied'; 7 = 'completely important'	*Insignificant*	**0.033****	OLS	*60,481*	Y
Satisfaction with use of own leisure time	Scale: 1='not satisfied'; 7 = 'completely important'	*Insignificant*	**0.034****	OLS	*60,481*	Y
Self-reported problems with drugs or alcohol	Binary outcome (1 = Yes; 0 = No)	-1.19** (-0.08%)	*Insignificant*	Logit	*56,300*	N

Notes: marginal probability effects brackets. * significant at 10%, **5%, ***1%. Explanatory variables used in the health and wellbeing regressions were employment status; marital status; spouse employment status; age; education; income; quality of the neighbourhood (pollution and crime factors); carer status; frequency of meeting with friends; house ownership; household size; debt problems and region (Dolan et al., 2008).

Family and parenting						
Dependent variable	**Type of variable**	**Formal learning**	**Informal learning**	**Model**	*No. of Obs*	*Fixed effects*
Probability of divorce	Binary outcome (1 = Yes; 0 = No)	*Insignificant*	*Insignificant*	Logit	*8,902*	N
Probability of separation	Binary outcome (1 = Yes; 0 = No)	*Insignificant*	0.423* (+0.8%)	Logit	*8,902*	N
Satisfaction with spouse/partner	Scale: 1='not satisfied'; 10 = 'completely important'	*Insignificant*	*Insignificant*	OLS	*3,703*	N
Frequency of contacting adult children	Scale: 1='daily'; 6 = 'never'	*Insignificant*	*Insignificant*	OLS	*958*	N
Frequency of seeing parents	Scale: 1='daily'; 6 = 'never'	*Insignificant*	*Insignificant*	OLS	*1,900*	N
Probability of looking after grandchildren	Binary outcome (1 = Yes; 0 = No)	*Insignificant*	*Insignificant*	Logit	*958*	N
Probability of helping adult children with chores (e.g., paying bills, writing letters)	Binary outcome (1 = Yes; 0 = No)	1.95** (+14.5%)	*Insignificant*	Logit	*959*	N
Child happiness (self-reported by 11-16 year olds)	Scale: 1='completely unhappy'; 7 = 'completely happy'	*Insignificant*	*Insignificant*	OLS	*9,203*	N
How child feels about own family (reported by 11-16 year olds)	Scale: 1='completely unhappy'; 7 = 'completely happy'	-0.057*	*Insignificant*	OLS	*8,212*	N
Frequency of arguing with parents (reported by 11-16 year olds)	Scale: 1='most days'; 4 = 'hardly ever'	*Insignificant*	*Insignificant*	OLS	*9,135*	N
Probability of talking to mother about things that matter (reported by 11-16 year olds)	Binary outcome (1 = at least once per week; 0 = otherwise)	*Insignificant*	**0.233****	Logit	*4,721*	Y

Family and parenting						
Dependent variable	**Type of variable**	**Formal learning**	**Informal learning**	**Model**	*No. of Obs*	*Fixed effects*
Probability of talking to father about things that matter (reported by 11-16 year olds)	Binary outcome (1 = at least once per week; 0 = otherwise)	*Insignificant*	*Insignificant*	Logit	*8,327*	N
Number of evening meals with family in past week (reported by 11-16 year olds)	Bounded response (1 - 7 times)	*Insignificant*	*Insignificant*	OLS	*8,213*	N

Notes: marginal probability effects brackets. * significant at 10%, **5%, ***1%. Explanatory variables used in the family and parenting regressions were employment status; marital status; number of children; age; work hours; education; income; number of previous marriages; quality of neighbourhood (crime factors); self-reported depression; self-reported self-worth; caring responsibilities. For regressions for the sample of 11-16 year olds we also included household size, number of friends, whether they are bullied, marital status of parents, how they feel about their appearance (Bradbury et al., 2000; Fomby and Cherlin, 2007; South and Spitze, 1986).

Civic Participation						
Dependent variable	**Type of variable**	**Formal learning**	**Informal learning**	**Model**	*No. of Obs*	*Fixed effects*
Member of a political party	Binary outcome (1 = Yes; 0 = No)	*Insignificant*	0.308*** (+0.4%)	Logit	*43,262*	**N**
Member of an environmental group	Binary outcome (1 = Yes; 0 = No)	0.269** (+0.5%)	0.488*** (+1%)	Logit	*43,262*	**N**
Member of a trade union	Binary outcome (1 = Yes; 0 = No)	**0.313***	**0.203**	Logit	*5,044*	**Y**
Member of a parents association	Binary outcome (1 = Yes; 0 = No)	0.571*** (+1.2%)	0.483*** (+1%)	Logit	*43,262*	**N**
Member of a social group	Binary outcome (1 = Yes; 0 = No)	*Insignificant*	-0.104* (-0.7%)	Logit	*43,262*	**N**
Frequently attends local groups	Binary outcome (1 = Yes; 0 = No)	0.417*** (+2.5%)	0.295*** (+1.7%)	Logit	*54,001*	**N**
Frequently does voluntary work	Binary outcome (1 = Yes; 0 = No)	**0.398***	*Insignificant*	Logit	*5,982*	**Y**
Frequently does sports	Binary outcome (1 = Yes; 0 = No)	0.156*** (+3.8%)	0.151*** (+3.7%)	Logit	*54,001*	**N**

37

Notes: marginal probability effects brackets. * significant at 10%, **5%, ***1%. Explanatory variables used in the civic participation regressions were employment status; marital status; health status; age; education; income; frequency of meeting with friends and family; number of children (Dee, 2004; Gottlieb and Gillespie, 2008).

Attitudes and behaviours						
Dependent variable	**Type of variable**	**Formal learning**	**Informal learning**	**Model**	*No. of Obs*	*Fixed effects*
Likelihood of wanting a better job	Binary outcome (1 = Yes; 0 = No)	**0.12****	**0.219****	Logit	*22,400*	**Y**
Likelihood of wanting to start own business	Binary outcome (1 = Yes; 0 = No)	*Insignificant*	*Insignificant*	Logit	*10,751*	**N**
Likelihood of stating that improved financial expectations	Binary outcome (1 = Yes; 0 = No)	**0.32*****	**0.311*****	Logit	*27,749*	**Y**
How important is having a lot of money	Scale: 1='not important'; 10 = 'very important'	-0.153**	-0.196***	OLS	*12,105*	N
How important is having a good job	Scale: 1='not important'; 10 = 'very important'	0.288***	0.15**	OLS	*12,105*	N
How important is being independent	Scale: 1='not important'; 10 = 'very important'	*Insignificant*	*Insignificant*	OLS	*12,302*	N
How important is having good friends	Scale: 1='not important'; 10 = 'very important'	*Insignificant*	*Insignificant*	OLS	*12,128*	N
Probability of agreeing with statement that 'husband should earn and wife stay at home'	Binary outcome (1 = Yes; 0 = No)	*Insignificant*	-0.339*** (-2.5%)	Logit	*11,483*	N

Notes: marginal probability effects brackets. * significant at 10%, **5%, ***1%. Explanatory variables used in the attitudes and behaviours regressions were employment status; marital status; health status; job satisfaction; satisfaction with leisure time; job hours; commuting time; trade union membership; spouse employment status; age; education; number of children; income; caring responsibilities; house ownership; housing quality; debt problems.

Table 3: Demographic breakdown of adult impacts

Health and mental health					
Dependent variable	**Type of variable**	**Formal learning**		**Informal learning**	
Number of visits to GP	Continuous variable	**-0.037****		*Insignificant*	
Gender (Male | Female)		*Insignificant*	-0.056**		
Age (18-14 | 40+)		*Insignificant*	*Insignificant*		
Parent status (Children | No children)		*Insignificant*	*Insignificant*		
Income group (£0-£25K | £25K+)		*Insignificant*	*Insignificant*		
Self-confidence (GHQ)	Binary outcome (1 = Not losing confidence; 0 = Losing confidence)	**0.324****		**0.295****	
Gender (Male | Female)		0.375***	0.276***	0.349***	0.252***
Age (18-14 | 40+)		0.27***	0.376***	0.219***	0.351***
Parent status (Children | No children)		0.337***	0.272***	0.254***	0.284***
Income group (£0-£25K | £25K+)		0.404***	0.26***	0.209***	0.309***
Satisfaction with health	Scale: 1='not satisfied'; 7 = 'completely important'	**0.044****		*Insignificant*	
Gender (Male | Female)		*Insignificant*	*Insignificant*		
Age (18-14 | 40+)		*Insignificant*	0.069**		
Parent status (Children | No children)		0.11***	*Insignificant*		
Income group (£0-£25K | £25K+)		0.101***	*Insignificant*		

Health and mental health						
Dependent variable	**Type of variable**	**Formal learning**		**Informal learning**		
Satisfaction with social life	Scale: 1='not satisfied'; 7 = 'completely important'	*Insignificant*		**0.033****		
Gender (Male | Female)				*Insignificant*	0.053**	
Age (18-14 | 40+)				*Insignificant*	*Insignificant*	
Parent status (Children | No children)				*Insignificant*	*Insignificant*	
Income group (£0-£25K | £25K+)				0.076**	*Insignificant*	
Satisfaction with use of own leisure time	Scale: 1='not satisfied'; 7 = 'completely important'	*Insignificant*		**0.034****		
Gender (Male | Female)				*Insignificant*	0.054**	
Age (18-14 | 40+)				0.048*	*Insignificant*	
Parent status (Children | No children)				0.071***	*Insignificant*	
Income group (£0-£25K | £25K+)				0.1***	*Insignificant*	
Not depressed or unhappy (GHQ)	Binary outcome (1 = Not feeling depressed or unhappy; 0 = otherwise)	**0.276*****		**0.276*****		
Gender (Male | Female)		0.392***	0.177**	0.283***	0.271***	
Age (18-14 | 40+)		0.278***	0.284***	0.181***	0.341***	
Parent status (Children | No children)		0.312***	0.212***	0.248***	0.276***	
Income group (£0-£25K | £25K+)		0.277***	0.218***	0.211***	0.256***	
Self-worth (GHQ)	Binary outcome (1 = Not feeling worthless; 0 = otherwise)	**0.558*****		**0.446*****		
Gender (Male | Female)		0.612***	0.509***	0.598***	0.354***	

Health and mental health					
Dependent variable	**Type of variable**	**Formal learning**		**Informal learning**	
Age (18-14 | 40+)		0.536***	0.542***	0.358***	0.494***
Parent status (Children | No children)		0.634***	0.427***	0.314***	0.496***
Income group (£0-£25K | £25K+)		0.616***	0.467***	0.328***	0.448***

Civic Participation					
Dependent variable	**Type of variable**	**Formal learning**		**Informal learning**	
Frequently does voluntary work	Binary outcome (1 = Yes; 0 = No)	**0.398***		*Insignificant*	
Gender (Male | Female)		0.36*	0.411***		
Age (18-14 | 40+)		0.422**	0.347**		
Parent status (Children | No children)		0.411**	0.345**		
Income group (£0-£25K | £25K+)		0.785***	*Insignificant*		
Member of a trade union	Binary outcome (1 = Yes; 0 = No)	**0.313***		**0.203**	
Gender (Male | Female)		*Insignificant*	0.324**	*Insignificant*	*Insignificant*
Age (18-14 | 40+)		0.269*	0.365**	*Insignificant*	0.346**
Parent status (Children | No children)		*Insignificant*	0.535***	*Insignificant*	0.313**
Income group (£0-£25K | £25K+)		*Insignificant*	0.341**	0.447***	*Insignificant*

Attitudes and behaviours					
Dependent variable	**Type of variable**	**Formal learning**		**Informal learning**	
Likelihood of wanting a better job	Binary outcome (1 = Yes; 0 = No)	**0.116****		**0.219*****	
Gender (Male | Female)		0.133*	*Insignificant*	0.268***	0.178***
Age (18-14 | 40+)		*Insignificant*	0.156*	0.240***	0.171**
Parent status (Children | No children)		*Insignificant*	0.18**	0.191***	0.196***
Income group (£0-£25K | £25K+)		*Insignificant*	*Insignificant*	0.239***	0.161***
Likelihood of stating that improved financial expectations	Binary outcome (1 = Yes; 0 = No)	**0.321*****		**0.311*****	
Gender (Male | Female)		0.365***	0.276***	0.366***	0.267***
Age (18-14 | 40+)		0.379***	0.21***	0.327***	0.24***
Parent status (Children | No children)		0.335***	*Insignificant*	0.324***	0.19**
Income group (£0-£25K | £25K+)		0.368***	0.168*	0.411***	0.193***

References

AONTAS, 2009. The Costs and Benefits of Undertaking Adult Education Courses from the Perspective of the Individual, in: Organisation, T.N.A.L. (Ed.).

Becker, G., 1993. The Economic Value of Education. Columbia University Press, New York.

Bradbury, T.N., Fincham, F.D., Beach, S.R.H., 2000. Research on the Nature and Determinants of Marital Satisfaction: A Decade in Review. Journal of Marriage and Family 62, 964-980.

Cutler, D., Lleras-Muney, A., 2007. Understanding differences in health behaviors by education. Unpublished manuscript.

Deaton, A., 1993. Data and Econometric Tools for Development Analysis, in:

Chenery, H., Srinivasan, T.N. (Eds.), Handbook of Development Economics. Elsevier.

Dee, T.S., 2004. Are there civic returns to education? Journal of Public Economics 88, 1697-1720.

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., Vlaev, I., 2012. Influencing behaviour: The mindspace way. Journal of Economic Psychology 33, 264-277.

Dolan, P., Peasgood, T., White, M., 2008. Do we really know what makes us happy A review of the economic literature on the factors associated with subjective well-being. Journal of Economic Psychology 29, 94-122.

Estyn, 2012. Skills for Older Learners.

Feinstein, L., Sabates, R., 2005. Education and youth crime: effects of introducing the Education Maintenance Allowance programme. Centre for Research on the Wider Benefits of Learning Research Report 14.

Festinger, L., 1957. A theory of cognitive dissonance. Stanford University Press, Stanford, CA.

Fomby, P., Cherlin, A., 2007. Family Instability and Child Well-Being. American Sociological Review.

Gottlieb, B., Gillespie, A., 2008. Volunteerism, health, and civic engagement among older adults. . Canadian Journal on Aging 399-406.

Hastie, R., Dawes, R., 2010. Rational Choice in an Uncertain World. Sage, London.

Humphreys, M., 2009. Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities. Columbia University.

Jenkins, A., 2011. Participation in learning and wellbeing among older adults. International Journal of Lifelong Education 30, 403-420.

Kahneman, D., Fredrickson, B.L., Schreiber, C.A., Redelmeier, D.A., 1993. When More Pain Is Preferred to Less: Adding a Better End. Psychological Science 4, 401-405.

Kahneman, D., Krueger, A., Schkade, D., Schwarz, N., Stone, A., 2006. Would You Be Happier If You Were Richer? A Focusing Illusion. CEPS Working Paper 125.

Khemlani, S.S., Oppenheimer, D.M., 2010. When One Model Casts Doubt on Another: A Levels-of-Analysis Approach to Causal Discounting. Psychological Bulletin 13.

Landsberger, H. 1958 *Hawthorne Revisited*, Ithaca.

Mallows, D., 2010. The impact of basic skills education. National Research and Development Centre for Adult Literacy and Numeracy.

Matrix, 2009. Lifelong Learning and Well-being: An Analysis of the Relationship Between Adult Learning and Subjective Well-being. IFLL Public Value Paper 3.

Meara, E., Richards, S., Cutler, D., 1981. The gap gets bigger: Changes in mortality and life expectancy, by education, 1981-2000. Health Affairs 27.

MHF, 2011. Learning for Life. Mental Health Foundation.

National Research and Development Centre for Adult Literacy and Numeracy (NRDC) 2009. Impact and effectiveness of Literacy on Parents, Children, Families and Schools.

NIACE, 2009. Providing the evidence: the impact of wider family learning. National Institute of Adult Continuing Education.

OFSTED 2009. Family learning: an evaluation of the benefits of family learning for participants, their families and the wider community.

Preston, J., Feinstein, L., 2004. Adult education and attitude change. Centre for Research on the Wider Benefits of Learning.

Sabates, R., Duckworthm, K., 2009. The impact of mothers' learning on their children's academic performance at Key Stage 3: evidence from ALSPAC. Centre for Research on the Wider Benefits of Learning.

Schkade, D.A., Kahneman, D., 1998. Does Living in California Make People Happy? A Focusing Illusion in Judgments of Life Satisfaction. Psychological Science 9, 340-346.

Schultz, T., 1963. The Economic Value of Education. Columbia University Press, New York.

Schultz, T., 1981. Investing in people. University of Chicago Press.

Sherman, L., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., Bushway, S., 1998. Preventing Crime: What Works, What Doesn't, What's Promising. U.S. Department of Justice.

South, S.J., Spitze, G., 1986. Determinants of Divorce over the Marital Life Course. American Sociological Review 51, 583-590.

Swanson, A.D., King, R.A., 1991. School finance: Its economics and politics. Longman Publishing Group, White Plains, NY.

Trivette, C., Dunst, C., Hamby, D., O'Herin, C., 2009. Characteristics and Consequences of Adult Learning Models. Research Brief 3.

Wilson, T., 2002. Strangers to Ourselves: Discovering the Adaptive Unconscious. Harvard University Press, USA.

Wolter, S., Messer, D., 2009. Continuing Education and Education Vouchers, in: Affairs, S.F.D.o.E. (Ed.).

Woodhall, M., 1987. Lending for learning: designing a student loan programme for developing countries., in: Secretariat., C. (Ed.), London.

WSIPP, 2006. Evidence-Based Adult Corrections Programs: What Works and What Does Not, in: document, W.S.I.f.P.P.p. (Ed.).