RESEARCH

BIS Department for Business Innovation & Skills

BIS RESEARCH PAPER NUMBER 72

Estimating the Impact of Training on Productivity using Firm-level Data

MAY 2012

Prepared by London Economics on behalf of the Department for Business, Innovation and Skills

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Business, Innovation and Skills.

Department for Business, Innovation and Skills

1 Victoria Street

London SW1H 0ET

www.bis.gov.uk

Research paper number 72

May 2012

About London Economics

London Economics is one of Europe's leading specialist economics and policy consultancies and has its head office in London. We also have offices in Brussels, Dublin, Cardiff and Budapest, and associated offices in Paris and Valletta.

We advise clients in both the public and private sectors on economic and financial analysis, policy development and evaluation, business strategy, and regulatory and competition policy. Our consultants are highly-qualified economists with experience in applying a wide variety of analytical techniques to assist our work, including cost-benefit analysis, multi-criteria analysis, policy simulation, scenario building, statistical analysis and mathematical modelling. We are also experienced in using a wide range of data collection techniques including literature reviews, survey questionnaires, interviews and focus groups.

Authors: Mr Pietro Patrignani and Dr Gavan Conlon

Head Office: 11-15 Betterton Street, London, WC2H 9BP, United Kingdom.
w: www.londecon.co.uk
e: info@londecon.co.uk

t: +44 (0)20 7866 8185 f: +44 (0)20 7866 8186

Acknowledgements

We would like to acknowledge the useful guidance and feedback provided by Matt Bursnall and Simon McKee from BIS. We would also like to thank Dr. Steven McIntosh for very helpful comments on the study and the MAUS team at the ONS for support with the ONS data sets. Responsibility for the contents of this report remains with London Economics.

Disclaimer

Contents

Glossary

Terminology abbreviations

ILR	Individualised Learner Record
EDS	Employer Data Service
BSD	Business Structure Database
IDBR	Inter Departmental Business Register
ABI	Annual Business Inquiry
ARD	Annual Respondent Database
TTG	Train to Gain
SFA	Skills Funding Agency
LSPA	Young People's Learning Agency
NESS	National Employer Skills Survey for England
ESF	European Social Fund
SIC	Standard Industrial Classification
NESS	National Employer Skills Survey
entref	Enterprise reference number
luref	Local unit reference number
ruref	Reporting unit reference number
LU	Local Unit
RU	Reporting Unit

Abstract

London Economics were commissioned by the Department for Business Innovation and Skills to undertake an assessment of the impact of publicly funded training on firm-level productivity. Firm-level analysis has traditionally been hampered by a lack of reliable and comprehensive data. For the first time in the UK a matched employer-employee dataset with information on government funded training provided through the *Train to Gain* programme was made available thanks to an employer identifier contained in the ILR. Information on training was then matched to firm level data on productivity and other firm characteristics, meaning that a dataset spanning more than one year and containing information on productivity and employee level engagement in government funded training was available for analysis. In analysing this data, the short time frame available and the likely presence of unobserved factors affecting training decisions and productivity meant it was not possible to draw robust conclusions about the relationship between publicly funded training and productivity. The fact that some patterns did emerge suggests that the lack of robust conclusions is because of the limited amount of data currently available, rather than the nature of the data itself. Econometric techniques controlling for time and firm specific effects and for the likely endogeneity of training decisions typically require a longitudinal data set of at least 4 years, preferably more.

We recommend the analysis of TTG should be repeated when the data set can be extended to 5 calendar years (2007 to 2011) when the ABI for 2011 becomes available (probably in late 2013 or 2014)[1]. Firms typically use a combination of three types of government funded training: apprenticeships; *Train to Gain* and work place training. Given that employer details for all three programmes were first recorded in 09/10, future research will have access to information on training undertaken through the different programmes and on how the picture has evolved over time (especially after *Train to Gain* support was withdrawn in 2011). Data for 4 calendar years covering all 3 programmes first becomes available in 13/14 (i.e. 2010 to 2013) and we recommend further analysis for all three programmes be conducted when ABI 2013 becomes available (probably in late 2015 or 2016). More generally we recommend that matching the ILR data to the IDBR (and the ABI) become an on-going exercise in order to improve the match rate[2] and to provide a dataset for on-going analysis of firms' engagement in publicly funded training.

[1] Starting from the ABI 2009, there will be some significant changes to the structure of the ABI survey: employment data will be collected at the local unit level (BRES survey) and finance data at reporting unit level (ABS survey), using 2 separate surveys with slightly different reference points. While this might potentially add another level of complexity to the analysis we would expect to be still possible to derive the total employment data for the reporting unit (and possibly even to use better imputation methods to allocate financial information to local units).

[2] The ONS typically match to a live version of the IDBR. Therefore, if matching were delayed until 2015 there would be as loss of firms who had changed their structure and thus their address.

Executive summary

London Economics were commissioned by the Department for Business Innovation and Skills to undertake an assessment of publicly funded training on firm-level productivity. The report is organised in two different stages: Stage 1 is an in-depth assessment of available data sources, sample composition and matching issues in order to understand the scope for econometric analysis; Stage 2 provides the basis for the suggested econometric analysis developed from the findings presented in Stage 1.

The impact of training undertaken at firm level on productivity is of key interest for both firms and policy makers. However, firm-level analysis has traditionally been hampered by a lack of reliable and comprehensive data. Moreover, various methodological issues, including the presence of firm-level and time specific effects, and the probable endogeneity of training decisions and productivity performance, typically require the availability of a panel data set that follows the same firm over several years.

For the first time at UK level firm level data is available spanning more than one year and containing information on productivity and employee level engagement in government funded training. However, for the bulk of the training we only have productivity data one year after it was completed so measurement potential is limited to the short-term impact of training.

Due to the exploratory nature of the data we investigate a variety of model specifications and the parameters estimating the impact of training on productivity vary to a high degree between them. Nevertheless some patterns begin to emerge and these would suggest that there is a positive relationship between training and productivity for small firms (less than 50 employees) and a negative relationship between training and productivity for large firms (250 employees or more). There are however two major drawbacks with this evidence.

(i) Firstly it relates to 2008 only with no evidence found of a relationship in 2007. This may be because the programme only became fully established in 2008 but equally it may indicate that the relationship is explained by other factors, including external economic shocks, such as the downturn, rather than training.

(ii) Secondly it does not show a causal relationship between training and productivity. Firm's engagement may be determined by their productivity, or factors relating to it, and the above estimates may be capturing these effects.

When we try a number of approaches to control for these, there is is still a tendency to observe positive significant coefficients for small firms and negative significant coefficients for large firms but there values are less stable and sometimes non-significant. This instability to changes in the model specification means it is not possible to draw robust conclusions about the relationship between training and productivity given current data availability.

The fact that patterns did emerge suggest that the lack of robust conclusions is because of the limited amount of data currently available, rather than the nature of the data itself, and that as the quality and volume of the data set increases in future years there is every

chance the coefficients will coalesce sufficiently to draw robust conclusions. One year of additional data will enhance the analysis to some extent but for reasons outlined in the summary of empirical results, a step change in the reliability of the analysis is not likely to occur until two further year's data is available. Further improvements to data collection and processing should also enhance the reliability of future analysis and this is discussed under recommendations for future analysis.

Brief review of recent related literature

Two studies using UK data fully reflect these limitations: in fact, due to the lack of firm-level data, Dearden, Reed and Van Reenen (2005) use industry-level data on training and productivity level for the period 1983-1996. Haskel and Galindo-Rueda (2005) use firm-level data from the Annual Business Inquiry and match firms with data on skills from the 2001 Employer Skills Survey; however, are forced to restrict their analysis to a single time period, due to the cross-sectional nature of the NESS.

Haskel and Galindo-Rueda (2005) find that higher level qualifications have a positive effect on firms' productivity, with the results varying by sector and being robust only for full-time male workers. On the other hand, low-level qualifications have a negligible effect on productivity. The authors also estimate the impact of qualifications on wages and compare this with the effect on productivity. They find that, for higher level qualifications, the effect is higher for services and lower, but not statistically significant, for manufacturing[3].

Three studies have investigated the impact of training on both wages and productivity. They use a variety of methods and are based on different countries but all show an impact on productivity which is greater than that on wages - with the productivity effect being of the order of double the wage effect.

Dearden et al. (2005) combine individual-level data on training from the Labour Force Survey with industry level data from the Annual Census of Production. Their findings suggest that the overall effect of training on productivity at industry level is positive and robust, around twice as high as the wage effect and consistent across different specification. In fact, the coefficient associated to training varies from around 0.7 in the random effect and fixed effect specifications to 0.6 in the system GMM specification (which accounts for endogeneity). For wages, the coefficient is around 0.35, roughly similar across different specification. Based on these results, the authors report that an increase in training by one percentage point at industry level is associated with an increase in productivity (value added) of about 0.6%, and in wages of about 0.3%.

Colombo and Stanca (2008) use a panel of Italian firms and find that a one percentage point increase in training intensity boosts firms' productivity by about 0.07 per cent. They also find that not controlling for unobserved firm characteristics leads to over-estimate the

[3] The authors also explore the presence of skills externalities and estimate production and wage regressions controlling for local skills characteristics (proxied by the share of both local workers and residents with higher levels of qualifications). Their findings suggest the presence of area-based skills externalities

effect of training on productivity, while ignoring endogeneity leads to an underestimate of the effect (the coefficient on training varies from 0.045 in the OLS regression to 0.028 in the fixed effects specification to 0.074 in the system GMM specification[4]). Moreover, the effect is even larger when they control for training duration (available only in a subsample of cases). Training has also a positive effect on wages in the firms undertaking the training, but this is found to be significantly smaller than the effect on productivity (coefficient around 0.02 and 0.044 in the fixed effect and system GMM specification respectively). The impact of training by occupational groups is varied, with high returns found for blue-collar workers (0.13%) and negligible returns for executives and clerks.

In a more recent paper, Konings and Vanormelingen (2010) use longitudinal data on Belgian firms and find that the productivity effect of training in the aggregate equation (controlling for the endogeneity of training and inputs) is around 0.24%, while the wage effect is around 0.17%[5]. When estimating the production function by industry the unweighted average for the training coefficient is around 0.18 in the productivity equation and around 0.12 in the wage equation (indicating that a one percentage point increase in training raises productivity by approximately 0.18% and wages by 0.12%).

Differences in findings across studies are probably explained by a combination of the following factors: different levels of analysis (industry level for the UK, which captures both within firm and within industry spillovers and firm level for the other countries); different definition of training; differences in data sources, relative measurement errors and internal consistency of datasets; differences in sector coverage, (with Dearden et al. focusing in the production sectors only); different labour market characteristics in different countries (including the role of unions); different time periods. But in summary, all suggest an impact on productivity which is substantially higher than (of the order of double) the impact on wages

In Table 47 in the Annex we present a detailed comparison of the data, methodology and main findings of the three studies.

The *Train to Gain* programme

The training data used in this study refer almost exclusively to training undertaken through the *Train to Gain* programme. This is because it was the first Further Education programme for which training providers were required to submit an employer identifier on their administrative data return – the Individualised Learner Record, and thus requires the necessary information to be matched to the firm-level data. The TTG programme was introduced in April 2006 and provided funding towards training costs for NVQs undertaken

[4] The authors find a significant effect on training on productivity in the system GMM specification when treating training as pre-determined (responding to past, but not current productivity shocks) and therefore instrumented using lagged values from t-1. When training is treated as strictly exogenous coefficient turns large and not statistically significant.

[5] The main estimates referred to the specification controlling for the endogeneity of training and inputs. OLS estimates in the aggregate equation are around 0.31% for productivity and 0.2% for wages.

through the employer. More specifically, firms were normally cold contacted, either by a skills broker or by a provider. The skills broker or provider identified skills gaps and training needs and when training was needed, and identified learners eligible for training. Courses could be fully funded (basic skills and NVQs at level 2 and NVQ3 for learners aged between 19 and 24) or partially funded (higher level NVQs). The vast majority of training undertaken took place at level 2 (around 78%) or level 3 (around 16%). The programme was wound down from the start of July 2011.

Overview of the matching process undertaken in the current study

In this study, we combine administrative data on publicly funded training undertaken through the *Train to Gain* programme between 2007 and 2010 with firm-level data from the Inter Departmental Business Register [IDBR] and the Annual Respondents Database [ARD] (containing the response to the Annual Business Inquiry [ABI]). The ABI is an annual survey of around 50,000 firms, taking a census of larger companies (250 or more employees) and a stratified sample of smaller firms that change between years.

For firms undertaking training through the *Train to Gain* (TTG) programme, information on training activities was collected through the Individualised Learner Record that also contained a firm identifier (A44) and firm-level details from the Employer Data Service.

Below we briefly review and summarise the matching process and the available sample size, explaining how individual level information from the ILR was aggregated to firm level and how the information was then convoyed through to the IDBR and the ABI. More detailed information is contained in section 3 below.

The following steps were carried out in the reshaping and matching process:

- The ILR is organised by academic year. To align the timing of training information to the timing of financial and other variables collected in the ABI, each learner was assigned to the relevant calendar year (with a rule assigning learners to a calendar year when at least 75% of training activities had been completed by the end of that year).

- Training information was then aggregated to firm level using the A44 identifier (224,000 firms), keeping track of various characteristics of training undertaken (such as level etc.),.

- The ONS were able to identify around half of the original sample of firms in the IDBR at enterprise level. The subsequent match at local level was less successful (around 78,000 observations were identified at local level compared with 114,000 at enterprise level).

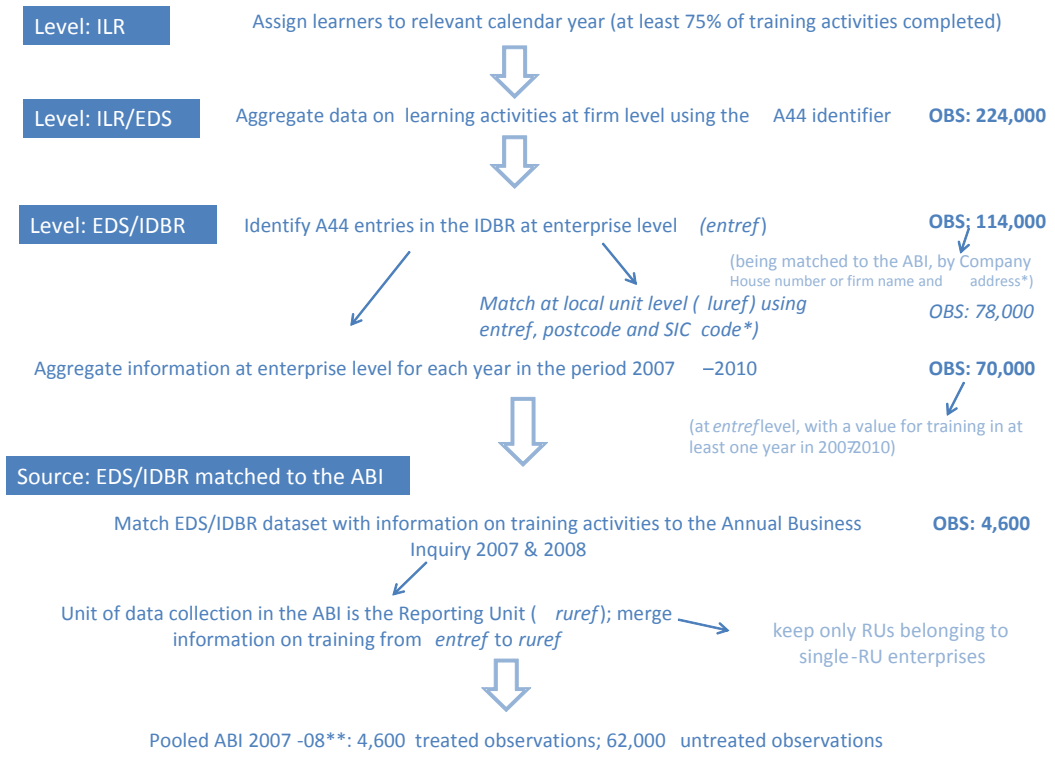
- Training information was then aggregated at enterprise and local unit level, with around 70,000 enterprises having participated in TTG activities in at least one year between 2007 and 2010.

Data on productivity and other detailed firm-level characteristics is asked in the Annual Business Inquiry. Therefore, training information was matched to firms surveyed in the ABI

in 2007 and 2008. Information from the 2006 ABI was also retained to look at firms' characteristics before engaging in TTG activities. Overall the pooled 2007-08 ABI dataset with training information from the ILR contains around 4,300 firms undertaking TTG activities and 61,000 non-treated firms (or firms for which identification was not possible).

Figure 1 below summarise the matching process described.

Figure 1: Overview of the matching process and sample size availability



Note: * Done in-house by ONS; ** These figures refer to the pooled dataset, with a number of firms appearing in both 2007 and 2008
Source: London Economics

Data issues and main limitations of the analysis

While the matched dataset is a novelty at UK level, containing comprehensive information on training at firm-level (from the ILR/EDS database) and information on firm-level characteristics over more than one year (from the BSD and the ABI), the final matched dataset and the empirical approach used are affected by a series of key limitations that should be considered when analysing the results (also see section 4.2.2 for more details):

- We only observe a value for training for firms identified in the IDBR and we restrict that further when we match with ABI information. As a result sample size composition is shifted towards larger companies.

- The ILR and the ABI have different timings. Even if we realign information from the two sources to the relevant calendar year, residual mismatches may still occur.

Also, it should be taken into account that the nature of the data sources is different and that they might be subject to different biases

- We cannot control for other training being undertaken at firm-level and we cannot characterise the selection rule determining how firms are selected and decide to engage in TTG activities.

- Due to data availability (the latest edition of the ABI currently available is the ABI 2008), we have to restrict our attention to the period 2007-2008. This means that we cannot control for any deferred impact of training and more crucially, we cannot control for time- or firm-specific fixed effects and the potential endogeneity of training decisions. As such, we cannot infer on the causal impact of training on productivity. Also the period was characterised by external shocks at macro level (the economic downturn) potentially affecting productivity and training decisions.

- The capital stock series currently available in the Virtual Microdata Laboratory and used in the analysis is not completely up to date. Therefore, data contained in the Volume Index of Capital Services on deflators by asset type and industry was projected forward using 2002 data. Moreover, even applying different tolerance imputation ratios (i.e. the ratio of missing to observed values), controlling for the capital stock will omit a substantial number of observations. Therefore we also used a specification controlling for capital expenditure (which is collected in the ABI) rather than stock.

Overview of empirical results

For all the reasons presented above, it is clear that no strong conclusions can be drawn from the analysis of results described in section 4.3; especially no causal inference can be drawn from the observed results. However, we present the key findings below, also highlighting patterns and unusual values for the training coefficients. Due to the exploratory nature of the data we investigate a variety of model specifications and the parameters estimating the impact of training on productivity vary to a high degree between them.

One key difference between specifications is whether our (physical) capital covariate is based on capital stock or in-year capital expenditure. It is vital that we control for capital because firm's production strategies will typically vary in the extent to which they favour physical or human capital inputs. Theoretically we should be controlling for capital stock rather than capital expenditure but this leads to a greatly reduced sample size (around 25,000 firms in the capital stock specification compared to 63,000 firms in the specification controlling for capital expenditure). We therefore fit each model twice, once with capital stock and once using capital expenditure, and compare the results.

At the aggregate level the effect of training on productivity is negative in the specification using capital stock (around -0.16) and turns positive in the specification controlling for capital expenditure (around 0.14). However the coefficient is never statistically significant. This result may be explained by changes in the sample composition as looking at capital expenditure allows us to include a much larger number of small and micro firms in the analysis (with less than 50 employees), which seem to benefit from training while this does

not appear to be the case for large firms (250 employees or more), where the relevant coefficient indicating the impact of training on productivity is negative.

Results by sector of industrial activity show strong evidence of a positive association between training and productivity for the construction sector (significant at the 1% level). This is observed using either capital stock or capital expenditure indicating that, for the construction sector, increasing the proportion of employees receiving training by one percentage point is associated with an increase in productivity between 1% (capital expenditure) and 1.6% (capital stock). There is also weak evidence (limited to one specification only) of a positive impact for the motor trades and the retail sectors (significant only at the 10% level): indicating a one percentage point increase in training is associated with an increase in productivity of 0.8% in retail (capital stock) and 1.2% in motor trades (capital stock). So the coefficients are again quite large even at sectoral level (but again, we cannot rule out the presence of unobserved effects we are unable to control for given the short time frame available). It should also be noticed that the available sample sizes for the treatment group (those receiving TTG training) might be limited for the analysis at sectoral level.

Turning in more detail to the analysis by firm size, some patterns begin to emerge for small and large firms across all specifications; suggesting that even in the short-term training may have a significant positive impact on the productivity of small firms (less than 50 employees) but the relevant coefficient appears to be significant negative for large firms (250 or more employees). However the pooled estimates over the period 2007-08 are driven by the coefficient for 2008 in both cases, while no significant effect is observed for 2007.

The parameters are significantly positive for small firms and significantly negative for large firms using either stock or expenditure as the capital covariate: the results would indicate that raising the proportion of employees trained by one percentage point would be associated with an increase in productivity of around 0.5-0.9% for small firms and a decrease in productivity of between 1.5 and 1.8% percentage points for large firms. In both cases, the magnitude and significance of the coefficient seems to be drawn by 2008 values. It is worth restating that no causal effect can be inferred with the current dataset and that the nature of the data could yield potentially large biases. While it is feasible to conceive that in the short term training could have a small negative impact on productivity, it is infeasible that training alone could have a negative impact of this size. A coefficient if this size is more likely to be explaining the impact on the training decision on the factors relating to productivity.

Given the short time frame available, we are unable to control for either the potential endogeneity of training decisions or for the possible presence of time-invariant individual components or individual-invariant time effects. Endogeneity could affect the relationship between training decision and productivity (and thus the estimated results) in two different ways:

1) **Firms may adjust their training decisions in response to past or current productivity shocks. In other words firms with higher (or lower) productivity may be more or less likely to engage in (publicly-funded) training activities.** Although our analysis found no evidence that productivity at time t influences the incidence of training at time t+1 at the aggregate level, and may only have an

(significant) impact on the incidence of training in three of the eight industrial sectors considered (motor trades, production and property), this does not mean there is evidence that there is no endogeneity, particularly because there may be unmeasured factors related to productivity and simultaneous shocks that do influence the training decision.

2) **There may also be unmeasured factors that influence productivity.** Even though we have controlled for a range of factors affecting productivity (industrial sector, capital stock, expenditure on software and advertising, proportion of part time workers etc.) the parameter estimates for training may be capturing the impact of other determinants of productivity that we were unable to control for (such as firms existing human capital, product-market strategy). Furthermore, the importance of such factors in explaining productivity may vary according to firm size, potentially explaining the different results for small and large firms discussed above.

Best econometric practice suggests using a full Generalised Method of Moments (GMM) system with at least 3 to 4 years lagged values to control for potential endogeneity and time or firm-specific effects. This was not possible due to the limited longitudinal nature of the data. Given the experimental nature of the work, we attempted to further explore the relationship between training and productivity over time using the limited data in three ways with varying results:

(i) A first very limited attempt to control for endogeneity was performed using the one year's lagged values for training proportion and the capital variable as instruments: the estimated coefficient for training is around 0.4-0.5 in the aggregate regression, but never statistically significant (for the aggregate regression or when disaggregating by size).

(ii) Controlling for both the current and lagged values of the trained proportion and the lagged value of productivity we observed some evidence that also past training may be associated with productivity levels for the group of small firms.

(iii) Using the firm level change in productivity between years, we observed a significant positive relationship for small firms using capital expenditure but no significant result for large firms. However at the aggregate level we observed (weakly significant) positive effects using capital expenditure or capital stock (10% level): the results would indicate that raising the proportion of employees trained by one percentage point would be associated with an increase in productivity growth of around 0.7-0.8%.

Overall Conclusions

Measuring the causal impact of training on firm-level productivity is a demanding exercise in terms of data availability (we typically require a longitudinal dataset) and comprehensiveness of the dataset (we need to control for a series of key variables other than training). Moreover when modelling real data it is never possible to measure a sufficient number of covariates to be absolutely certain that findings are entirely robust. Due to current limitations of the dataset available (only two years of data available and various issues with the matching process), we have not reached the position of establishing a causal link between training and productivity here. The main findings of the

analysis showed some evidence that government funded training in small firms may be associated with higher productivity, while the reverse is true for large firms (where government funded training was associated with lower productivity on average).

Further analysis is needed to assess whether these estimates reflect any causal effects, given that results were statistically significant for 2008 only and that external factors such as the downturn affected firm-level productivity and probably also the decision to engage in TTG training. Crucially, it should also be noticed that we cannot currently control for other training undertaken at firm-level and for the dynamics of training decisions over time[6].

Recommendations for future analysis

Although the aim of this work was to investigate the impact of any government supported training on firm level productivity we focussed on the TTG programme because firm level information on training for other programmes does not currently overlap with available information on productivity variables from the ABI. Even so the analysis was limited because the TTG programme was launched in 2006/07, while the latest available edition of the ABI is currently the 2008 edition, meaning that we only had two years (2007 and 2008) with information available on both TTG training activities (from the ILR) and on productivity and other firm-level characteristics (from the ABI). As a consequence, the matched ILR/EDS/ABI data with information on TTG training covered an insufficient period to tackle the likely presence of endogeneity and time and firm-specific effects (typically tackled in the literature through the application of a GMM estimation technique). An additional limitation was a lack of data on private training undertaken by firms at the same time.

The most immediately obvious alternative data set is the National Employer Skills Survey (NESS). NESS currently surveys around 80,000 English firms, but does not follow firms over time, it is only collected every two years and also the match with the IDBR (and hence, the ABI) is limited. Therefore the NESS in its current form has a very limited potential for panel techniques but it does contain information on training undertaken independent of government funding[7]. Clearly, if the level of privately funded training is relatively stable over time, it should be accounted for when controlling for the presence of time-invariant effects and have limited influence on the analysis. However, it is likely that access to publicly funded training may have an effect on the provision of privately funded training and that should be taken into account. Nonetheless, weighing up all of these factors we believe that in its current format NESS has less potential for robust analysis than the ILR/EDS.

---

[6] For example it might be possible that while small firms involved in the TTG programme increased their overall level of training, other firms (possibly larger firms already engaging in training activities) experiencing financial difficulties replaced privately-funded training with publicly-funded training. However lack of data does not currently allow any inference on training dynamics.

[7] However the NESS 2009 did not report detailed information on the number of employees trained through Train-to-Gain, but only participation in the programme. Detailed numbers were available for Apprenticeships.

The first part of this section looks at the potential for future analysis of the ILR. At the end of the section we look at options using other current data collections to build a firm-level panel dataset.

Potential for further analysis using the ILR

The *Train to Gain* programme ran from 2006 to 2011. The dataset used in this analysis had data on training undertaken available up to the academic year 2010, while data on productivity and other firm-level characteristics are currently available in the Annual Respondents Database up to 2008. Extending the analysis to 2011, adding information on training for the academic year 2010-11 (already available in the ILR collection) and ABI survey data up to 2011 would yield more robust estimates and allow for further disaggregation. Furthermore, a significant proportion of firms engaging in *Train to Gain* activities appear to train a very small proportion of their employees. While this might be explained by a variety of factors[8], it might also undermine the possibility to detect a significant impact of training on productivity in the presence of "noise" correlated with training activities and productivity. Looking at firms over time should allow controlling for unobserved time-invariant firm characteristics[9]. Also, while in the short term, firms may adjust their training strategy in response to the *Train to Gain* initiative (in a way we cannot observe due to the lack of information on other training undertaken), using the 5-year dataset allows us to observe changes over time after the initial adjustment. Finally, a 5-year panel dataset (covering the period 2007-2011) would also allow the use, at least to some extent, of panel-data econometric techniques, controlling for the presence of time invariant effects and endogeneity. It will be possible to construct this data set when the ABI for 2011 becomes available (probably in late 2013 or 2014)[10].

An additional consideration when looking at training over time is that firms will typically use a combination of three types of government funded training: apprenticeships; *Train to Gain* and other work place training. Firm identifiers were collected for apprenticeships from 2009/10 and for other workplace training (elsewhere referred to as non-TTG) from 08/09 (see Table 1). As such, employer details for all three programmes were first recorded in 09/10 so data for 4 calendar years first becomes available in 13/14, i.e. the calendar years 2010 to 2013. Basing the analysis on ILR data between 09/10 and 13/14 should also improve the match rate because stricter quality assurance of employer identifiers was introduced in 09/10. Prior to 09/10 providers has the option of submitting employer

[8] For example the low matching ratio to the IDBR, the fact that we are forced to use the reporting unit level for productivity analysis and the fact that the Train to Gain may have not fully reached all recipients by the end of 2008

[9] Obviously it is possible that the effect of training on productivity is negligible and the "true" value of the coefficient is close to zero.

[10] Starting from the ABI 2009, there will be some significant changes to the structure of the ABI survey: employment data will be collected at the local unit level (BRES survey) and finance data at reporting unit level (ABS survey), using 2 separate surveys with slightly different reference points. While this might potentially add another level of complexity to the analysis we would expect to be still possible to derive the total employment data for the reporting unit (and possibly even to use better imputation methods to allocate financial information to local units).

information using free text, this was often incomplete and difficult to match to the IDBR. From 09/10 they could only use the blue sheep employer identification number which linked back to the blue sheep data base provides full employer information which is much easier to match to the IDBR. We therefore recommend further analysis of all three programmes when ABI 2013 becomes available (probably in late 2015 or 2016).

Table 1: History of inclusion and quality of the Employer Identifier (A44) in the ILR

	06/07	07/08	08/09	09/10	10/11
TTG	Required	Required	Required	Required + some validation	Required + strict Validation
Other workplace Training			Required	Required + strict Validation	Required + strict Validation
Apprenticeships				Required + some validation	Required + strict Validation

Source: Review of ILR Specification 08/09-10/11, http://www.theia.org.uk/ilr/ilrdocuments/

While collecting ILR data at employer level is a useful exercise to identify training patterns, particular care should be paid to ensure that the quality and coverage of the information gathered is maximised. In particular, key improvements to the data collection are possible in the following areas:

- Ensuring that the A44 field is correctly filled by learning providers with a valid entry, the FE data service introduced more rigorous validity testing of this variable for all funding streams from 10/11;

- An increased reliability of the A44 identifier should in turn boost the probability of identifying employers in the IDBR: in fact, the match rate for matching conducted by the ONS between ILR/EDS firm information and the IDBR was 50%, whereas other projects matching firm data to the IDBR typically have match rates between 65 and 70%. There is therefore considerable potential for improvement;

- In the longer term it might be possible to extend the coverage, requiring learning providers (or awarding bodies) to record a firm identifier for all courses towards a recognised qualification undertaken through the employer (whether publicly funded, co-funded or employer funded). While this would increase the reporting burden for learning providers or awarding bodies, it would also provide the policy maker and the researcher with invaluable information on training dynamics, with far-reaching applications in the analysis.

- More generally we recommend that matching the ILR data to the IDBR (and the ABI) become an on-going exercise. Firstly it will improve the match rate because the ONS typically match to a live version of the IDBR – timely matching to the IDBR when TTG participation data becomes available reduces the risk of losing firms by the time ABI data is available if they change their structure and thus their address. Secondly it means that matching ratios, sample characteristics and trends could be monitored on a yearly basis and the matched LR/IDBR dataset could be used for regular statistical analysis.

An improved dataset should then be used for further analysis of the impact of training on productivity We believe that taken together these changes will reduce the compositional bias and provide a larger sample for analysis and a clearer picture on different forms of training undertaken. Clearly there might still be other factors we are unable to control for, such as firm's product market strategy, skills structure and internal firm organisation. Finally, it should be noted that linking multiple years of the ABI together would result in a loss of all micro firms and a substantial loss of small firms, due to the survey design.

Options using other data collections

While we believe enhancing the quality and coverage of training data collected through the ILR may be a viable route, alternative or complementary approaches could be used with other current data collections to build a firm-level panel dataset covering training undertaken and including data on productivity and other firm-level characteristics. These include the following:

- Introduce a panel component in the National Employer Skills Survey. Following a representative sample of firms over time would provide firm-level evidence on both privately and publicly funded training undertaken over several time periods. Also, given that the NESS questionnaire is currently administered to surveyed firms every two years, it would be useful to send annual follow ups, related to training activities only, to firms forming the panel database. Finally, the current match rate between the NESS and the IDBR is around 50%: improving the match between the NESS and the IDBR (and hence the ABI) is another key enhancement that could be undertaken.

- Include some questions on training activities undertaken and skills structure, at least for a random sample, in the ABI questionnaire enquiring on employment and staff.

- Potentially consider also using the FAME database as an alternative source for productivity data. While FAME database has inherent limitations and would not solve the lack of information on training, it has been used for productivity analysis in the past. Datasets similar to FAME have also been used for analysis in other EU countries. An exploratory analysis of quality availability might provide a clearer picture on whether it is worth using the FAME dataset (matched with the ILR) for the analysis of training and productivity.

1 Introduction

London Economics were commissioned by the Department for Business Innovation and Skills to undertake an assessment of publicly funded training on firm-level productivity. The report is organised in two different stages: Stage 1 is an in-depth assessment of available data sources, sample composition and matching issues in order to understand the scope for econometric analysis; Stage 2 provides the basis for the suggested econometric analysis developed from the findings presented in Stage 1.

The remainder of the report is organised as follows: Section 2 provides a description of data sources, while section 3 describes in detail the matching process and presents a series of descriptive statistics on training and firms' characteristics. Section 4 contains a descriptive analysis of key financial and other variables for treated and untreated firms, the model specification and the econometric results. Finally, Section 5 contains the conclusions and recommendations for future analysis and data collection.

2 Data sources

In this section, we provide some information on the various data sets that were used as part of this data analysis.

2.1 Individualised Learner Record

The Individualised Learner Record contains detailed information on course and personal characteristics for Further Education courses and is supplied by learning providers throughout the Further Education (FE) system. The ILR is organised by academic year (1st August – 31st July) and the data specification may vary to some extent from year to year. The ILR data is collected from providers that are in receipt of funding from the Skills Funding Agency (SFA), the Young People's Learning Agency (LPSA) and from providers funded by co-financed European Social Funds (ESF).

The current study focuses on the three academic years from 2007-08 to 2009-10, which are the years when information on employers was added to the ILR dataset. The ILR dataset we received made use of information on learner and course characteristics for the relevant learning aims, together with the A44 employer identifier (when available).

2.2 Employer Data Service (EDS)

The Employer Data Service (or Blue Sheep) dataset contains firm-level identifiers of those firms which engaged in publicly provided training. The EDS also contains information on firm characteristics, the so-called “Blue Sheep” data. “Blue Sheep” data contains details of the firm’s postcode, number of employees and turnover at site and group level, sector of activity as defined by SIC code, year of foundation, companies house number where available and a range of other firm level characteristics. The EDS dataset we received contains both the A44 employer identifier, provided by the EDS, and also, when available, the ‘*entref*’ and ‘*luref*’ identifiers (enterprise and local unit reference number respectively), supplied by the Office for National Statistics (ONS). The A44 identifier makes it possible to match the EDS with ILR information, while the ‘*entref*’ and ‘*luref*’ identifiers allow subsequent linkage to the Business Structure Database (BSD) and the other ONS surveys at firm level (such as the Annual Business Inquiry (ABI)) .

2.3 Inter Departmental Business Register (IDBR) and Business Structure Database (BSD)

The Business Structure Database is the version of the Inter Departmental Business Register (IDBR), which is available for research purposes. The BSD contains details on around four million firms in the United Kingdom (of which slightly more than two million are active, covering nearly 99% of UK economic activity) and contains information on company’s postcode, SIC code, employment profile and employees, turnover, legal status (company, sole proprietor, partnership, public corporation/nationalised body, Local Authority or non-profit body), enterprise group links, and country of ownership. Data in the IDBR (and the BSD) are updated from the following sources: HM Revenue and Customs (VAT and PAYE), ONS surveys, Companies House (company registrations) and Dun and

Bradstreet (group structures). Data contained in the BSD represent a snapshot of the company profile taken every year in March; however, employment and turnover figures will correspond to a point in a previous year (or earlier). Employment and turnover data suffer from some weaknesses because of the use of several sources (VAT, PAYE, etc.) with different definitions and time periods, along with data imputation.

The information is provided at 'enterprise group', 'enterprise level' (*entref* identifier) and 'local unit (LU) level' (*luref* identifier) and using the *entref* and *luref* identifiers it is possible to link the BSD with the ONS surveys conducted at firm level.

2.4 Annual Business Inquiry (ABI)

The Annual Business Inquiry contains detailed information on company characteristics. The ABI is intended to be a census of larger companies (250 or more employees) and a stratified sample of smaller firms. Data in the ABI is collected at the reporting unit (RU) level, which may be a subset of an enterprise (meaning that an enterprise can have one or more reporting units) and a superset of local units (a reporting unit can be formed by one or more local units). Responses to different editions of the ABI are available through the Annual Respondent Database (ARD). The ARD contains three different files for each year and sector (Catering, Construction, Motor Trades, Production, Property, Retail, Service Trades and Wholesale): the DAT file, which contains responses for all reporting units selected and providing responses in a particular year; the NUL file, containing non selected reporting units; and the SNUL file, with details at both RU and LU level for all units (selected and non-selected). Details on financial and other variables are only available through the DAT (selected files).

The ABI has two parts, ABI/1 which covers employment and ABI/2, which covers financial data (including gross value added). ABI/1 has wider coverage than ABI/2 and asks for employment on a set date in December. ABI/2 is only available for selected reporting units, a subset of those for which ABI/1 is available split by sector, contains detailed financial information, and is available for approximately 50,000 companies. Businesses are expected to report on the year from 1st January to 31st December or, if their business year is different from the calendar year, to any 12 month period ending within the financial year (from 6th April to 5th of April of the following year). Some sectors of the economy are not covered by the ABI (mainly agriculture, financial intermediation, public administration and defence).

Information in the ABI on employment and turnover may differ from information reported in the BSD. When this happens, the ABI is considered to be the more accurate source of information on employment and turnover[11].

[11] In fact as reported by the ONS (http://www.neighbourhood.statistics.gov.uk/dissemination/Info.do?page=analysisandguidance/analysisarticles/idbr-analysis-to-support-local-authorities.htm): "The ABI uses a specific methodology, modelling and grossing up survey results into estimates for a specific point in time, although at the more detailed level these become subject to a (potentially) lesser degree of accuracy. IDBR employment data for local units could come from the Business Register Survey, ABI or PAYE data from HMRC. It too is subject to suppression but the variety of sources means data are usually available at lower levels than for the ABI. IDBR employment does not have the usual sampling error inherent in the ABI and any other survey.

2.5 National Employer Skills Survey (NESS)

The National Employer Skills Survey for England (NESS) is an employer survey that asks a representative sample of English employers about recruitment issues and problems, workforce and skills and their approach to providing training, including the participation in publicly funded schemes. The latest available editions of the survey were undertaken in 2007 and 2009.

However, it will have other, potentially larger errors, not least because its sources have different definitions, and dates at which employment is taken. IDBR employment is shown for the enterprises and local units that were active on the IDBR in March, but the employment figure will correspond to a point in a previous year or earlier; as will turnover."

3 Reshaping and matching process

3.1 Step 1: Allocating learners to a calendar year and initial cleaning

The ILR dataset contains detailed information at course and learner level. This study focuses on the academic years 2007/08 to 2009/10. Since 2007/08, an employer identifier is available in the ILR for publicly funded courses undertaken through the employer. The ILR dataset is organised at learner level, while the main elements of the analysis are to be undertaken at firm level. Therefore, in order to convoy significant information at firm level, we needed to generate summary information for relevant variables using the employer identifier variable A44. ILR data was cleaned and aggregate at firm level using the procedure described below. Summary information on training was then attached to the EDS dataset containing variables on firm characteristics.

An initial preliminary step related to assessing the availability of the A44 identifier in the dataset. When the A44 variable is unavailable any linking at firm level for further analysis is impossible and the information is unusable and records were removed from the datasets. The availability of the A44 identifier depended on the funding streams for the learning aim. The funding streams under consideration were *Train to Gain* (TTG) or non–TTG, with the latter including Further Education aims delivered in part or fully at an employers premises in 2007/08. FE and *Train to Gain* were then merged into one funding stream from 2008/09. It should be noticed that for non-TTG aims it was not compulsory to record the employer identifier and therefore the A44 variables is seldom available for those records.

Table 2: Number of ILR records (learners)

	Calendar year								
	2007/08			2008/09			2009/10		
	ILR Records	A44 Available	A44 Missing	ILR Records	A44 Available	A44 Missing	ILR Records	A44 Available	A44 Missing
Non-TTG	436,566	3,785	99.1%	56,553	10,704	81.1%	226,181	10,671	95.3%
TTG	493,937	492,210	0.3%	1,179,330	1,076,205	8.7%	981,426	976,276	0.5%
Total	930,503			1,235,883			1,207,607		

Note:"Non-TTG" is used here as shorthand for FE courses delivered in part or fully at employer's premises and this group excludes other Skills Funding agency provision for employers such as apprenticeships.
Source: London Economics analysis of the ILR

For the purposes of the current analysis we decided to consider together the records available for the academic years 2007/08 to 2009/10 and then to allocate them to different calendar years. This step was motivated by the different timings of the Individualised Learner Record (which covers the academic year, from August to July), the Business Structure Database (which is a "snapshot" taken in March) and the Annual Business Inquiry, where the information provided by businesses normally refers to a period close to the calendar year (employment figures refers to the situation in December, while financial and other information should normally refer to the calendar year).

In order to do that, we needed to set a rule to assign learners to a calendar year, according to the proportion of the course already undertaken. Below we detail the preliminary steps undertaken to clean the dataset from unnecessary or unusable information and the rules used to assign a learner to the relevant calendar year (2007, 2008, 2009 and 2010):

a) keep latest available record when the learner is present in more than one academic year for the same course and keep highest learning aim if a learner attended multiple learning aims in the same academic year;

b) remove record when course length is zero and learner flagged as non-completer;

c) remove record when learner has withdrawn from the course and less than 75% of the average length of the course has been completed (average length was computed using non-missing records for courses with similar characteristics);

d) for those learners flagged as continuers but not available in subsequent academic years consider the end of the academic year as end date (consider the records as having a missing end date);

e) for learners ending their course after July 2010 the end date is currently unknown; use 31st December 2010 as the end date and keep the record if at least 75% of the average course has been completed by then (see point f) below);

f) in case of unknown end dates, assign as normal (using the rule detailed in point g) below) for 2008 and 2009; for 2010 see if they have completed at least 75% than the average length of the course, otherwise discard;

g) for records covering more than one academic year, consider the proportion of training completed by the end of the calendar year; then define the year up to December and assign the learner to that calendar year if at least 75% of the length of the course has been completed by December 31st.

In Table 3, we summarise the different cleaning steps, and the assumptions used to allocate learners (and learning aims) to a specific calendar year.

Table 3: Cleaning steps and allocation rule	
CASE	**ACTION**
CLEANING	
Same learner and course available in more than one academic year	Keep latest available record
The learner is flagged as non-completer and the course length is zero	Remove from dataset
Learner has withdrawn from course and less than 75% of the course has been completed	Remove from dataset
MISSING END	
Continuing learners not available in subsequent academic years	Set the end of the academic year (31st July) as the end date
Learners still enrolled in a learning course at the end of the 2009/10 academic year	Set 31st December 2010 as the end date
Learners with end of course unknown	Set the end of the academic year (31st July) as the end date
ASSIGN LEARNER TO A CALENDAR YEAR	
If record only covers one calendar year	Assign to the relevant calendar year
If record covers more than one calendar year	Assign learner to a calendar year if at least 75% of the length of the course has been completed by December 31st

Source: London Economics analysis of the ILR

In the table below we describe data availability at learner level after data cleaning, which included removing all records without an A44 identifier and allocating learners to calendar year. In the table we also include the disaggregation by funding stream (*Train to Gain* and non-TTG) to confirm that it is not feasible to disaggregate information by funding stream, given that the number of non-TTG cases is very limited and unlikely to be representative. We therefore only retained records with an A44 identifier, with no further disaggregation by funding stream and refer to funding received as TTG funding.

Table 4: Number of ILR records (learners)				
	Calendar year			
	2007	**2008**	**2009**	**2010**
Non-TTG	524	4,932	5,141	5,348
TTG	90,588	359,339	647,505	497,856
Total	91,112	364,271	652,646	503,204

Source: London Economics analysis of the ILR

3.2 Step 2: Aggregating summary training information at firm level

Further to this initial cleaning, a number of variables were appended at firm level in the EDS dataset in order to identify the main features of the training undertaken during the academic year. Summary information was retained and aggregated at firm level using the A44 identifier; more specifically we identified the following variables at firm level:

- Number of employees receiving training in each year;

- Proportion of employees receiving training at different levels (among those trained) to capture the intensity of training.

- A series of summary demographic characteristics on employees receiving training (gender, ethnicity and age).

- Proportion of higher or lower level TTG funding received (lower level funding is generally reserved for cases where it is largely certification of existing skills).

In Table 5, we present a snapshot of the dataset for three fictitious firms in one academic year. Numbers presented in the table are entirely fictitious and the table should only be seen as an aid for a better understanding of the dataset.

Table 5: An example of training information retained

A44	Employees trained	% L2	% L3	% female	% non white	Average age	% Higher level
X	32	81.3%	12.5%	46.9%	37.5%	24.5	68.8%
Y	3	33.3%	66.6%	33.3%	0.0%	28.7	100.0%
Z	121	69.4%	25.6%	40.5%	31.4%	26.2	71.9%

Note: Numbers presented in the table are fictitious and do not represent any specific firm in the dataset
Source: London Economics analysis of the ILR

3.3 Step 3: The Employer Data Service (EDS) dataset

3.3.1 Matching the EDS with the IDBR

The original EDS dataset contained the A44 identifier together with financial and other firm level information collected by the Employer Data Service using different data sources (this information is commonly known as “Blue Sheep” data). In total there were around 224,000 firms identified by the A44 identifier.

The Office for National Statistics carried out a matching exercise, trying to identify firms in the EDS dataset in the Inter Departmental Business Register (IDBR). The matching exercise was performed at enterprise level. A subsequent match at local unit level was carried out using *entref* number, postcode and SIC code. In total, the ONS were able to add an *entref* number to around 114,000 observations[12], while around 78,000 observations were also matched at local unit level and therefore have also a *luref* number. The *entref* and *luref* reference numbers are necessary for subsequent linking with all the variables available in the BSD and the ABI.

All remaining observations do not have an *entref* number, which implies it was not possible to identify them in the BSD. The main problem is associated to the reduction in sample size and the possible distortion of sample characteristics when we match the information

[12] However, more than 3,000 *entref* numbers were unusable and they were discarded from subsequent analysis.

to the BSD. In fact, if there is any characteristic that predicts the probability of a company being identified in the BSD, the BSD sample could be different to some extent from the original sample. For example, if larger firms are more likely to be identified and have an *entref* number, they will be over-represented in the BSD sample.

Unfortunately it is difficult to assess if and to what extent these biases occur. We decided to compare matched and unmatched companies using Blue Sheep data. The advantage of this approach consists in having information on key variables (region, sector, size) for both groups of companies; however, the approach relies on the Blue Sheep data being accurate.

The first approach has the advantage of having internal consistency, in the sense that data are homogenous across the group of matched and unmatched companies. It obviously relies on the data to be accurate, something that we are unable to control for at this stage. In Table 6 below we present information on matched and unmatched firms (firm definition follows the variable A44) using company size (defined by number of employees at both site and group levels), region (generated using information on postcodes), sector of activity (generated using SIC code) and company status.

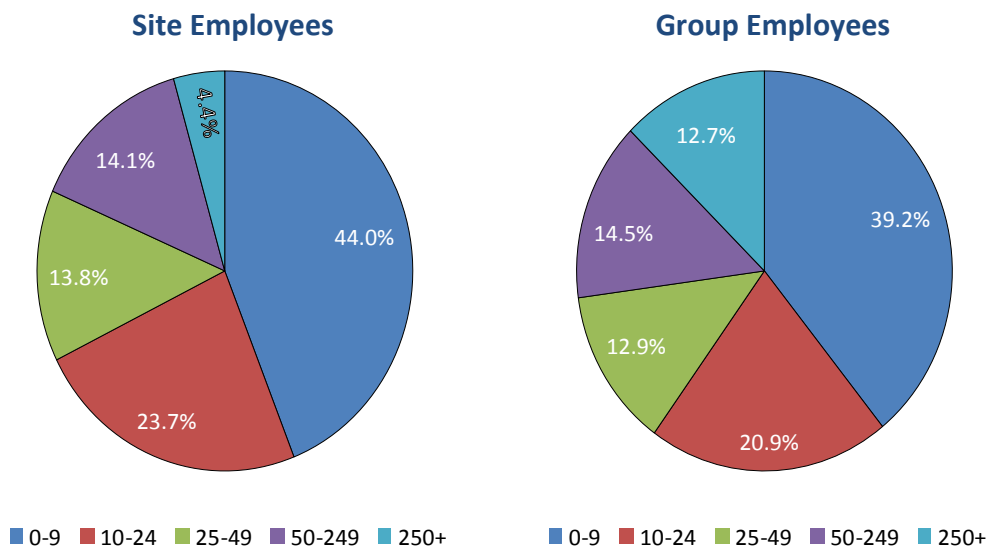
In the remainder of the section we first describe firm characteristics using information contained in the EDS dataset on all firms (as identified by the variable A44) engaging in Train to Gain, then we highlight the differences between the group of firms identified in the IDBR, for which subsequent analysis is possible and the group of firms without an IDBR identifier, and therefore not linkable to the Business Structure Database or the Annual Business Inquiry.

3.3.2 Firms' characteristics using the EDS

The Employer Data Service attached firm-level information (the "Blue Sheep" data) to the dataset containing the A44 firm identifier. We will present a brief description of the characteristics of the overall group of firms engaging with *Train to Gain* activities between 2006/07 and 2009/10 using the Blue Sheep data. We will present details on firm size (at site and group level), region of location, sector of activity and company status. The EDS collected the data using a variety of data sources (mainly Thomson Directories, Companies House, Dun and Bradstreet and Equifax). In particular, information on the number of employees at group level was sourced from Companies House.

In Figure 2 and Figure 3 we present firm characteristics for the overall sample available in the EDS dataset (the same information is presented in Table 6 and Table 7, column Total or T). Using data at site level (left panel of Figure 2) we can see that the sample of firms engaging in *Train to Gain* activities is made up of 44% of micro firms (less than 10 employees), while another quarter are small firms (10 to 24 employees), and around 14% are firms with more than 25 and less than 50 employees. Medium to large and very large companies (50 to 250 employees and more than 250 employees) account for 14% and 4.4% of the sample respectively. At group level the proportion for the three middle categories (10-24, 25-49 and 50-249 employees) are similar to those reported for site level, while less than 40% are micro firms and almost 13% are very large enterprises.

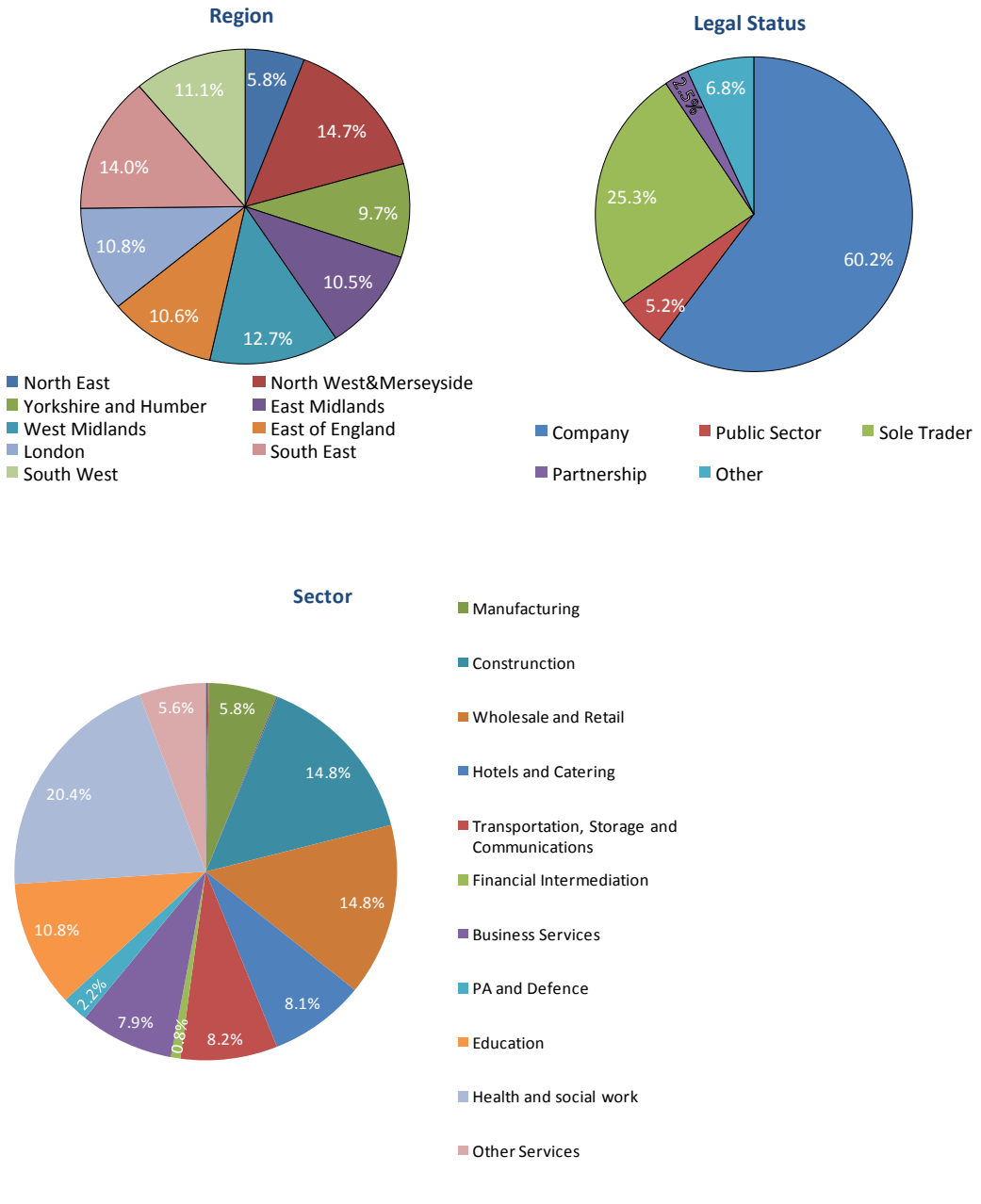
Figure 2: Firm size using Blue Sheep data



Source: London Economics analysis of the EDS using Blue Sheep data

Figure 3 describes firms' characteristics using region of residence, sector of activity and legal status: the different English regions account for a proportion of the total sample varying from 6% in the North East to almost 15% in the North West and Merseyside (all other regions account for between 10% and 14%). The sectors more represented among firms in the EDS dataset are Health and Social Work (more than one in five), Wholesale and Retail and Construction (both around 15%). Finally, more than 60% of firms are described as private companies (LTD or PLC), while more than a quarter are Sole Traders and around 5% Public Sector Organisations.

Figure 1: Region, sector and Legal Status using Blue Sheep data

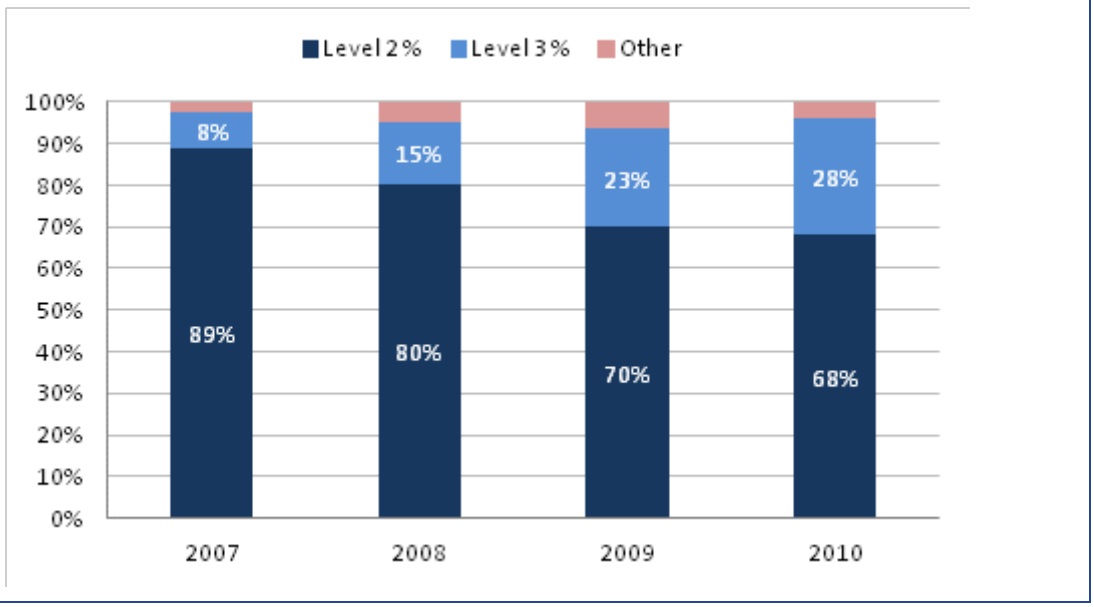


Note: For "Sector" only categories accounting for at least 0.5% of the sample are labelled
Source: London Economics analysis of the EDS using Blue Sheep data

In Figure 4 we describe the characteristics of training undertaken by firms in the EDS dataset over the period 2007-2010. Data refers to the entire sample and firms are defined by the A44 identifier. We report the proportion of employees trained at level 2 or level 3 and 'other' level (level 1 or level 4) as a residual category. The proportion of employees (among those trained) receiving training at Level 2 accounted for almost 90% in 2007 and then gradually declined to less than 70% in 2010. Conversely, the proportion of employees

receiving training at Level 3 increased from less than 10% in 2007 to almost 30% in 2010[13].

Figure 4: Level of funded TTG training over time



Note: 'Other': residual training undertaken at level 1 or level 4
Source: London Economics analysis of the EDS using Blue Sheep data

3.3.3 Comparing the groups of firms matched and unmatched to the IDBR

However, as already mentioned, not all observations were identified in the IDBR. Below, we compare the characteristics of the group of firms matched and unmatched to the IDBR to assess to what extent the two group differ and therefore to what extent the group of firms retained for further analysis is different from the original sample.

As shown in Table 6, according to Blue Sheep data, matched firms are on average bigger than unmatched firms at site level (more than 53% of unmatched firms are micro firms with less than 10 employees, compared to 35% for matched firms). At group level the proportion of firms falling in each band is very similar (compared to site level) for matched firms, while the group of unmatched firms has an unexpectedly high proportion of very large firms (information on group employees in the Blue Sheep dataset is sourced from Companies House, and would normally refer to information at the holding company level, rather than site level).

[13] Also, the proportion of funding received at higher level was around 70% and fairly stable between 2007 and 2009 (with a slight increase in 2010, where information on TTG funding level was missing for a significant number of records). Lower level funding normally reflects simple skills accreditation, while higher level funding indicates that training activities were undertaken.

Table 6: Matched and unmatched companies: comparison by size of firm						
	Site Employees			Group employees		
No of employees	Unmatched	Matched	Total	Unmatched	Matched	Total
0-9	53.4%	34.9%	44.0%	46.5%	32.0%	39.2%
10-24	19.5%	27.8%	23.7%	14.7%	26.8%	20.9%
25-49	11.9%	15.6%	13.8%	10.0%	15.6%	12.9%
50-249	11.1%	16.9%	14.1%	10.4%	18.4%	14.5%
250+	4.1%	4.7%	4.4%	18.4%	7.1%	12.7%
Total	106,104	109,538	215,642	106,874	110,924	217,798

Source: London Economics analysis of the EDS using Blue Sheep data

In Table 7 we present a comparison between the groups of matched and unmatched companies using government region, sector of activity and company status.

All variables used present some variation across groups: the group of matched firms presents a significantly lower proportion of companies based in London or West Midlands, while all other regions (apart from the East of England where the proportions are very similar) have a higher proportion of firms in the matched group compared to the unmatched group of companies.

Differences are even more marked when we take into account sector of activity: in fact the group of matched firms contains around 9% of companies from the Manufacturing sector, compared to 2.4% for the unmatched group. Similarly there is a higher proportion of companies operating in Business services and in the Health and Social Work sector in the matched group compared to the unmatched (5% compared to 10% and 21% compared to 19% respectively). Conversely, the Construction, Transportation and Storage and Education sectors seem to be significantly under-represented in the matched group of firms (17% vs. 13%, 12% vs. 5% and 14% vs. 8% respectively).

The variables identifying company status follow two slightly different classifications in the Blue Sheep and BSD dataset. For subsequent comparison with the BSD, we decided to use a common classification across the Blue Sheep and BSD datasets. Consequently the variables identifying company status in the two datasets have been re-coded using the following classification: “Company” (including both LTD and PLC), “Public Sector Organisation” (which includes Public Corporation, Central Government Body and Local Authority), “Sole Trader”, “Partnership”, and “Other” (category which also covers Non-profit making body). As expected, the category identifying “Company” seems to be heavily over-represented in the matched group (79% compared to 39%), while Sole Traders are heavily under-represented (they account for only 8% of matched observations, but around 45% in the unmatched group). Public Sector organisations and “Other” organisations are slightly under-represented (4% compared to almost 7% and 6% compared to 8% respectively), while “Partnerships” are slightly over-represented (3% compared to 2%).

Table 7: Matched and unmatched companies: comparison by region, sector of activity and company status

Region				Sector				Legal status			
	U	M	T		U	M	T		U	M	T
North East	5.5%	6.1%	5.8%	Agriculture	0.1%	0.2%	0.2%	Company	38.7%	79.1%	60.2%
North West & Merseyside	14.3%	15.1%	14.7%	Mining and Quarrying	0.0%	0.1%	0.1%	Public Sector	6.7%	3.9%	5.2%
Yorkshire & Humber	9.1%	10.2%	9.7%	Manufacturing	2.4%	8.9%	5.8%	Sole Trader	45.0%	7.8%	25.3%
East Midlands	9.8%	11.1%	10.5%	Electricity/Gas/ Water	0.1%	0.1%	0.1%	Partnership	1.9%	3.0%	2.5%
West Midlands	13.8%	11.7%	12.7%	Construction	16.8%	13.0%	14.8%	Other	7.8%	6.0%	6.8%
East of England	10.6%	10.5%	10.6%	Wholesale/ Retail	15.1%	14.4%	14.8%				
London	12.3%	9.5%	10.8%	Hotels / restaurants	7.3%	8.9%	8.1%				
South East	13.8%	14.2%	14.0%	Transport/ Storage	11.7%	4.9%	8.2%				
South West	10.7%	11.4%	11.1%	Financial/ Insurance	0.7%	0.9%	0.8%				
				Business services	5.2%	10.5%	7.9%				
				PA and Defence	2.2%	2.3%	2.2%				
				Education	13.9%	8.0%	10.8%				
				Health/ Social Work	19.4%	21.3%	20.4%				
				Other services	5.0%	6.2%	5.6%				
Total	82,361	100,067	182,428		102,768	111,384	214,152		71,731	81,360	153,091

Note: Sector of activity is generated using the SIC92 classification; U=Unmatched, M=Matched, T=Total
Source: London Economics analysis of the EDS using Blue Sheep data

We also present, in Table 8 below, the various training indicators for the matched (M) and unmatched (U) group of companies and the value for the entire sample (T). Matched companies seem to train on average a higher number of employees, probably reflecting the higher proportion of larger companies in the matched group. Matched companies also seem to train a marginally higher proportion of employees at level 3 (and conversely a slightly lower proportion at level 2), but the difference is minimal, probably suggesting that there is no significant difference in the structure of training between matched and unmatched firms[14].

[14] The proportion of TTG funding received at higher level is around 70% for both groups in the first three years and increase to around 77-79% in 2010. The difference across the two groups is between one and two percentage points.

Table 8: Training in matched and unmatched firms												
	Calendar year											
	2007			2008			2009			2010		
	U	M	T	U	M	T	U	M	T	U	M	T
Total trained	3.56	4.27	3.98	4.67	5.94	5.38	4.54	6.37	5.50	4.47	5.79	5.17
Proportion Level 2	88.9%	89.3%	89.1%	80.9%	79.5%	80.1%	71.4%	69.0%	70.1%	68.7%	67.5%	68.1%
Proportion Level 3	8.5%	8.3%	8.4%	14.5%	15.7%	15.2%	22.9%	23.9%	23.4%	27.6%	28.0%	27.8%

Source: London Economics analysis of the ILR and EDS

To summarise, the analysis suggests that, relative to unmatched firms, the group of firms matched in the BSD seem to be formed by a higher proportion of larger firms, a smaller proportion of firms based in London and West Midlands, a higher proportion of firms in Manufacturing, Business services and Health and Social Work and a smaller proportion of firms in the Construction, Education and Transportation and Storage sectors. Also firms with a "company" status seem to be significantly over-represented while firms with a "sole proprietor" status seem to be significantly under-represented in the group of matched firms. On average, the group of matched firms trains a higher proportion of employees, probably reflecting the larger firm size, but reported training characteristics do not seem to be very dissimilar between matched and unmatched firms, probably reflecting that training type does not depend on firm size and number of employees trained[15].

In conclusion, matched and unmatched firms that have received funding appear to have a relatively similar training profile, but there are some differences in company characteristics.

3.4 Step 4: Business Structure Database (BSD)

3.4.1 Comparing data sources

In the previous paragraph we presented evidence relative to the difference between firms matched and firms not matched in the BSD. The data presented relied on information provided by Blue Sheep. In the next steps we use information from the BSD and ABI, although we are forced to restrict our attention to only the group of companies identified in the BSD. As mentioned in the previous section, there are around 114,000 (out of 224,000) companies that were matched in the BSD. Matching with the BSD is performed at enterprise level (*entref*). The ONS also tried to identify the local unit reference number (*luref*) using *entref*, postcode and SIC code (around 78,000 observations were identified at local unit level).

Before analysing the characteristics of matched companies using BSD data, it is useful to compare information from Blue Sheep and BSD data on the group of matched firms. In other words, while in Step 2 we compared matched and unmatched companies using the same data source (Blue Sheep), here we compare information from two data sources (Blue Sheep and BSD) using the same group of firms (those matched in the BSD).

The data sources and the units of analysis used in the two datasets are of different nature: the BSD is a snapshot of the IDBR taken in March and the units of analysis in the BSD are

[15] On this issue also see Table 18

the statistical units at enterprise and the local unit level[16]. As previously discussed, Blue Sheep data is gathered using different sources, such as Thomson Directories, Companies House, Dun and Bradstreet and Equifax. Information is likely to differ especially for the number of employees at enterprise level (BSD) and the number of group employees (from Companies House). In particular, the definition of "Group" in the Blue Sheep data (which corresponds to company registered at Companies House) and enterprise (in the BSD) seem to be not directly comparable: firms with the same *entref* (and therefore the same number of employees in the BSD) may have completely different numbers of employees at group level (as reported by Blue Sheep, sourced from Companies House). It is therefore practically impossible to readily compare *entref* level and group level and we focus on the comparison between *luref* level and site level. In all comparisons, we use data from the 2009 Business Structure Database and compare with the information provided in the EDS dataset (Blue Sheep data).

In Table 9 we present summary statistics for company size and region at group and *entref* level, where group is used by EDS to refer to enterprise level. As said, the number of employees at group and *entref* level is not directly comparable and the table only provides some insight on the two distributions. In terms of firm size, the two distributions are very similar in the proportion of micro enterprises identified, but considerably different in other categories, especially very large enterprises. Conversely, the two distributions by region of location and sector of activity are generally highly similar.

Table 9: Blue Sheep and BSD data: distribution by size and region (group vs *entref*)

Firm size (employees)			Region			Sector		
	Blue Sheep (group)	BSD (*entref*)		Blue Sheep (group)	BSD (*entref*)		Blue Sheep (group)	BSD (*entref*)
0-9	32.7%	33.8%	North East	6.1%	5.8%	Agriculture	0.2%	0.6%
10-24	28.6%	17.5%	North West/ Mersey	15.0%	14.3%	Mining and Quarrying	0.1%	0.1%
25-49	16.2%	11.0%	Yorkshire & Humber	10.2%	9.8%	Manufacturing	9.2%	9.4%
50-249	17.6%	13.9%	East Midlands	11.2%	10.5%	Electricity/Gas/ Water	0.1%	0.1%
250+	4.8%	23.8%	West Midlands	11.7%	11.7%	Construction	13.1%	12.8%
			East of England	10.5%	10.6%	Wholesale/Retail	14.6%	14.5%
			London	9.4%	11.3%	Hotels /restaurants	8.6%	8.9%
			South East	14.3%	14.2%	Transport/Storage	4.8%	4.4%
			South West	11.4%	10.7%	Financial/Insurance	0.9%	0.9%
						Business Services	10.5%	12.1%
						PA and Defence	2.3%	0.7%
						Education	8.0%	8.1%
						Health/ Social Work	21.2%	21.4%
						Other services	6.1%	5.8%
Total	100,565	100,565		94,111	94,111	Total	105,089	105,089

Source: London Economics analysis of the EDS using Blue Sheep and BSD data

[16] For more information on the definition of statistical units see the Business Structure Database User Guide available at http://www.esds.ac.uk/findingData/snDescription.asp?sn=6697

Local unit and site level are likely to be more homogenous and in the tables below we present the characteristics and the cross tabulation of the two distributions.

In Table 10 and Table 11 we present information on the number of employees, geographical region, sector of activity and company status focusing on site level (Blue Sheep) versus local unit (BSD) , while Table 12 shows the cross tabulation of the Blue Sheep and BSD definition of company size. The analysis of these two tables suggest that the differences in terms of company size are less marked than when using enterprise and group level, but significant mismatch may still occur, either because they are measured at different points in time or because there are differences in the definition of site and local unit (also there may be measurement error). Conversely, both datasets are nearly identical in relation to region of location[17] and sector of activity, and also fairly similar in terms of legal status, suggesting that the firm is correctly identified, but there might be differences in the measurement of relevant characteristics (such as employees) due to the different data sources used.

Table 10: Blue Sheep and BSD data: distribution by size and region (site vs *luref*)

	Firm size		Region		
No of employees	**Blue Sheep (site)**	**BSD (*luref*)**	**Region**	**Blue Sheep**	**BSD (*luref*)**
0-9	32.4%	40.3%	North East	6.1%	6.1%
10-24	28.4%	24.3%	North West/ Mersey	15.1%	15.1%
25-49	16.5%	15.9%	Yorkshire & Humber	10.2%	10.2%
50-249	18.0%	16.2%	East Midlands	11.2%	11.2%
250+	4.7%	3.3%	West Midlands	11.7%	11.7%
			East of England	10.5%	10.5%
			London	9.3%	9.3%
			South East	14.3%	14.3%
			South West	11.5%	11.5%
Total	72,613	72,613		69,718	69,718

Source: London Economics analysis of the EDS using Blue Sheep and BSD data

[17] A cross-tab of region showed that nearly 100% of cases are on the main diagonal (i.e. Blue Sheep and BSD dataset record the same region of location)

Table 11: Blue Sheep and BSD data: distribution by sector and legal status (site vs *luref*)

	Sector of activity		Company Status		
Sector of activity	**Blue Sheep (site)**	**BSD (*luref*)**		**Blue Sheep**	**BSD (*luref*)**
Agriculture	0.3%	0.7%	Company	78.2%	74.7%
Mining and Quarrying	0.1%	0.1%	Sole proprietor	8.0%	6.5%
Manufacturing	9.4%	9.3%	Partnership	3.4%	5.1%
Electricity, gas and Water	0.1%	0.1%	Public	4.2%	4.1%
Construction	12.9%	12.4%	Other	6.2%	9.6%
Wholesale/Retail	13.9%	13.9%			
Hotels /restaurants	8.7%	9.2%			
Transport/Storage	4.7%	4.5%			
Financial/Insurance	0.8%	0.7%			
Business Services	10.2%	11.5%			
PA and Defence	2.4%	1.3%			
Education	8.2%	7.4%			
Health/ Social Work	22.1%	22.7%			
Other services	6.1%	6.1%			
Total	75,848	75,848	Total	54,751	54,751

Source: London Economics analysis of the EDS using Blue Sheep and BSD data

Table 12a: Blue Sheep and BSD data: cross-tab by size (site vs *luref*) – number of firms

	No of employees					
BSD Blue Sheep	0-9	10-24	25-49	50-249	250+	Total
0-9	**18,576**	3,068	887	803	162	23,496
10-24	6,850	**10,263**	2,371	1,025	143	20,652
25-49	1,764	2,856	**5,705**	1,536	127	11,988
50-249	1,582	1,171	2,291	**7,381**	619	13,044
250+	457	279	298	1,036	**1,363**	3,433
Total	29,229	17,637	11,552	11,781	2,414	72,613

Source: London Economics analysis of the EDS using Blue Sheep and BSD data

Table 12b: Blue Sheep and BSD data: cross-tab by size (site vs *luref*) – percentage of firms

	No of employees					
BSD Blue Sheep	0-9	10-24	25-49	50-249	250+	Total
0-9	**26**	4	1	1	0	32
Oct-24	9	**14**	3	1	0	28
25-49	2	4	**8**	2	0	17
50-249	2	2	3	**10**	1	18
250+	1	0	0	1	**2**	5
Total	40	24	16	16	3	100

In this section we have compared two different data sources ("Blue Sheep" and BSD) on the common sample of companies matched in the BSD (i.e. having an *entref* and, when available, a *luref* identifier). The exercise is useful to assess the reliability of the different data sources and comparison was mainly carried out at site ("Blue Sheep" definition) and local unit (BSD definition) level. The close correspondence between the two datasets regarding region of location, sector of activity and also legal status seems to suggest that they both identify the same firm. However, data on number of employees may be dissimilar across the two sources, given the different data sources used. Moreover differences can also arise due to different definitions of the unit of analysis (site vs. local unit) and/or because the timing of data coverage and collection is not entirely comparable.

3.4.2 Descriptive analysis of matched firms using the BSD

In this paragraph we describe in detail the characteristics of matched firms using data contained in the Business Structure database. The analysis is carried out at *entref* level, given the better match rate at enterprise level. There are a total of around 70,000 enterprises which used publicly funded training between 2007 and 2010 and were identified in the BSD. Some of these enterprises appear in one or more years, making up a total of slightly less than 110,000 observations over the four-year period (approximately 9,000 in 2007, 25,000 in 2008 40,000 in 2009 and 34,000 in 2010). In Table 13 we present the pattern of data availability over time. In around 12% of cases the same enterprise is available in the dataset in at least three of the four years, 21% in two of the four years and around 65% of enterprises are available in one year only.

Table 13: Pattern of availability over time

	Frequency	%	Pattern
	18,055	25.9%	••1•
	14,466	20.7%	•••1
	8,614	12.3%	••11
	8,001	11.5%	•1••
	5,157	7.4%	•111
	4,694	6.7%	•11•
	3,237	4.6%	1111
	2,111	3.0%	1•••
	1,351	1.9%	•1•1
	4,144	5.9%	(other)
Total	69,830	100.0%	

Source: London Economics analysis of the EDS using BSD data

Below we present a brief description of the characteristics statistics on company size, company age, region of location, sector of activity and company status for firms engaging in TTG activities that we were able to identify on the IDBR so this should not be considered as a fully representative analysis. The full tables containing year by year details are presented in Annex 1.

From the information presented in Table 14 we can see that around 55% of enterprises that engaged in publicly funded training and were identified in the BSD have less than 25 employees, while about 15% have between 25 and 50 employees, 18% have more than 50 and up to 250 employees and around 11% are very large enterprises (more than 250 employees). Moreover, around 37% of enterprises engaging in the programme are relatively young companies (established for less than 10 years), while another 36% are

aged between 10 and 20 years, and a further 27% of enterprises were founded more than 20 years prior to participation in the programme. Turning to the analysis of firms by legal status, the vast majority of enterprises are recorded as "Company" (around 70%), while Sole Proprietors, Partnerships and Non-for-profit organisations all account for approximately around 10% of firms.

As for region of location, approximately 15% of the enterprises are located in the North West, and a slightly smaller proportion is located in the South East. Apart from the North East (where approximately 5% of the companies are located), the proportion of enterprises located in each of the other regions is between 10-12%. Finally, the sectors with the highest proportion of enterprises is Health and Social care activities (at 24%), followed by Construction, Wholesale and Retail Trade, Real Estate and Manufacturing (all between 10% and 15%).

Table 14: Enterprise characteristics using the BSD

BSD 2007-10									
Size		**Age**		**Legal status**		**Region of location**		**Sector of activity**	
0-9	33.5%	0-5	18.4%	Company	69.0%	North East	5.5%	Agriculture	1.0%
10-24	21.7%	6-10	18.5%	Sole proprietor	8.9%	North West & Merseyside	15.5%	Mining and Quarrying	0.1%
25-49	15.1%	11-15	18.6%	Partnership	8.7%	Yorkshire & Humber	10.1%	Manufacturing	11.0%
50-249	18.3%	16-20	17.4%	Public corporation	0.2%	East Midlands	10.4%	Electricity/Gas/ Water	0.1%
250+	11.4%	20+	27.1%	Central government	1.0%	West Midlands	11.8%	Construction	14.7%
				Local authority	1.4%	East of England	10.1%	Wholesale/Retail	12.4%
				Non-profit	10.8%	London	11.3%	Hotels /restaurants	6.7%
						South East	14.5%	Transport/Storage	4.6%
						South West	10.8%	Financial/Insurance	0.7%
						North East	5.5%	Business Services	12.1%
								PA and Defence	0.6%
								Education	6.0%
								Health/Social Work	24.3%
								Other services	5.7%

Source: London Economics analysis of the EDS using BSD data

3.4.3 Characteristics of enterprises engaging and not engaging in TTG using the BSD

In Table 14 we have described in detail the characteristics of enterprises engaging in *Train to Gain* and identified in the BSD. Below we compare the characteristics of enterprises engaging in TTG (around 40,000) to the rest of the population (around 2 million enterprises). Data for the comparison is drawn from the BSD 2009. As expected, the distribution by company size for the general population is heavily skewed towards micro firms (0 to 9 employees), while only 35% of firms engaging in TTG are micro firms, compared to the vast majority (90%) in the rest of the population. Clearly, the reverse is true for all other size bands, which are over-represented in the TTG sample compared to the rest of the population. This is also confirmed by the distribution of firms using company age (which is likely to be correlated with size): in fact TTG enterprises are on average much older compared to the rest of the population: only 19% of TTG enterprises are less than 5 year old (compared to 40% in the overall population) while around 45% belong to the 16+ age bracket (compared to only 27% in the rest of the population).

The distribution by region of location is also slightly different compared to enterprises not engaging in TTG activities: some regions are over-represented in the TTG sample

compared to the rest of the population (North East, North West, Yorkshire and the Humber, East Midlands, West Midlands and the South West), while other regions are under-represented in the TTG sample (London, the South East and the East of England).

The last three columns of Table 14 show the composition of the TTG group and the rest of the population by sector of industrial activity. TTG companies seem to over-represent the Manufacturing, Construction, Education and especially, the Health and Social Work sectors compared to the rest of the population. Conversely, the Agriculture, Wholesale and Retail and, quite significantly, the Business Services sectors seem to be under-represented in the TTG group.

There are a series of factors explaining this substantial mismatch: the matching exercise was less successful for smaller firms, micro firms with zero or one employee may have little or no incentive in engaging in training and also the data may not be completely up to date for the category of micro firms[18]. Moreover, even if the propensity to training was constant across firms of different size, the higher number of employees in larger companies implies that they are more likely to train at least one employee. Characteristics of enterprises involved and not involved in TTG.

Table 15: Characteristics of enterprises involved and not involved in TTG

BSD 2009											
Size	Non-TTG	TTG	Company Age	Non-TTG	TTG	Region	Non-TTG	TTG	Sector	Non-TTG	TTG
0-9	90.0%	35.3%	0-5 years	40.2%	18.8%	North East	3.1%	5.5%	Agriculture	6.8%	1.2%
10-24	6.5%	22.7%	6-10	20.8%	19.1%	North West & Merseyside	12.1%	14.9%	Manufacturing	6.8%	10.9%
25-49	2.0%	14.8%	11-15	12.1%	16.8%	Yorkshire & Humber	8.2%	9.9%	Electricity/Gas/ Water	0.0%	0.1%
50-249	1.2%	17.1%	16-20	9.5%	18.5%	East Midlands	7.9%	10.6%	Construction	11.2%	14.3%
250+	0.2%	10.0%	20+	17.5%	26.8%	West Midlands	9.4%	11.5%	Wholesale/Retail	17.0%	12.8%
						East of England	11.6%	10.1%	Hotels /restaurants	6.2%	6.9%
						London	18.7%	11.1%	Transport/Storage	3.8%	4.5%
						South East	18.4%	14.5%	Financial/Insurance	1.8%	0.8%
						South West	3.1%	5.5%	Business Services	33.6%	12.4%
									PA and Defence	0.1%	0.6%
									Education	1.2%	6.2%
									Health/Social Work	3.3%	23.4%
									Other services	8.1%	5.8%

Source: London Economics analysis using BSD data matched with information from the EDS/ILR

We further explore on this in Table 16, where we present the average proportion of employees receiving TTG training, computed on the entire IDBR population. In total, the proportion of employees receiving TTG training in 2009 is around 0.5% of the workforce (however this is an under-estimate, given that only around 50% of original firms were identified in the IDBR). Also BSD data on the number of employees may not be entirely accurate and/or up to date. Data in Table 16 show that the overall proportion of employees receiving training is lowest for micro firms (0.4%) and highest for medium and medium/large firms (1.7%). Similarly, firms in the age band from 0 to 5 years train the lowest proportion of employees on average (0.4%) compared to 0.6% for both medium age and older firms.

[18] Even after removing all non-live units from the dataset.

Firms in the North East trained on average 1.2% of the workforce through TTG in 2009, followed by the East Midlands. At the other end of the spectrum the average proportion of employees trained through TTG training was around 0.3% in the London region and 0.4% in the South East.

The proportion of the workforce receiving TTG training was particularly high for the Health and Social Work and the Education sectors (2.3% and 1.7% respectively) and low for the Financial and the Real Estate sectors (both around 0.2%).

Table 16: Proportion of employees receiving training – entire IDBR population

BSD 2009							
Size	**%**	**Company Age**	**%**	**Region**	**%**	**Sector**	**%**
0-9	0.4%	0-5 years	0.4%	North East	1.2%	Manufacturing	0.6%
10-24	1.3%	6-10	0.5%	North West & Merseyside	0.7%	Electricity/Gas/Water	0.7%
25-49	1.7%	11-15	0.6%	Yorkshire & Humber	0.7%	Construction	0.9%
50-249	1.7%	16-20	0.6%	East Midlands	0.8%	Wholesale/Retail	0.4%
250+	1.4%	20+	0.5%	West Midlands	0.7%	Hotels /restaurants	0.5%
				East of England	0.5%	Transport/Storage	0.8%
				London	0.3%	Financial/Insurance	0.2%
				South East	0.4%	Business Services	0.2%
				South West	0.5%	PA and Defence	0.4%
						Education	1.7%
						Health/Social Work	2.3%
						Other services	0.4%
Total	0.5%		0.5%		0.5%		0.5%

Source: London Economics analysis using BSD data matched with information from the EDS/ILR

In Table 49 in Annex 1 we also present the proportion of firms involved in TTG activities by sector comparing information from BSD (matched with the EDS/ILR) and information publicly available in the NESS. While the proportions across the two data sources are substantially different, given the nature of the two databases, the relative ordering is quite similar, with both sources indicating that the top four sectors in terms of involvement in TTG activities are Education, Health and Social Work, Electricity, Gas and Water and PA and Defence, followed by Manufacturing, Construction and Transport Storage & Communications. The bottom five sectors for TTG involvement are, according to both data sources, Other Services, Hotels and Catering, Business Services, Retail and Wholesale and Financial Intermediation.

3.4.4 Analysis of training using the BSD

In this section we present a series of descriptive statistics on training activities for those firms participating in *Train to Gain* and identified in the IDBR. Table 17 shows, for each year, the average number of employees trained, the proportion of trained employees compared to the total number of employees and the percentage trained at level 2 and level 3[19]. The average number of employees receiving training increased from 4.8 in 2008 to

[19] In some cases (around 8% of the total) the number of employees receiving training exceeds the overall number of employees (proportion trained is greater than one). This might be due to the different nature of the two data sources (ILR/EDS dataset and the BSD) and also to the fact that the numbers in the BSD might be imputed and not be entirely up to date. When the discrepancy was small, we considered these observations as training their entire workforce (proportion trained=100%). Other cases when number of trained employees was greater than the overall number of employees were removed from the analysis. For more details see Table 55 in Annex 1.

around 7.5 in 2009 and 2010. The average proportion of trained employees out of total company employees is also increasing over time, from 19% in 2007 to around 25% in the period 2008-2010. The average proportion of trained employees is likely to be boosted by a number of companies training a relatively high proportion of employees (especially small and micro companies); in fact the median value proportion is around 12.5% and it is also increasing over time (from 7% in 2007 to 14.5% in 2009). The proportion of employees trained at level 2 is decreasing over time (from 90% in 2008 to 68% in 2010), while the proportion of employees trained at level 3 is increasing from 8% in 2008 to 28% in 2010[20].

Table 17: TTG training by year, for firms identified in the IDBR

	Year				
	BSD 2007	BSD 2008	BSD 2009	BSD 2010	Total
	Mean	Mean	Mean	Mean	Mean
Proportion trained	18.5%	25.4%	27.2%	24.5%	25.2%
Total trained	4.8	6.8	7.6	7.3	7.1
% trained at level 2	89.5%	79.5%	69.6%	67.6%	72.9%
% trained at level 3	8.3%	15.9%	23.8%	28.2%	22.1%

Source: London Economics analysis of the EDS using ILR data

As we would expect, the total number of employees trained increases considerably with company size (from 2 for micro firms to 31 for large enterprises), while the proportion of total employees trained declines sharply as company size increases (from 49% for micro companies down to less than 3% for large enterprises). The other indicators show a slightly higher proportion of employees trained at level 3 for micro and small firms (and conversely a slightly lower proportion trained at level 2) compared to medium to large firms.

Table 18: TTG training by enterprise size, for firms identified in the IDBR

	Training indicators			
Size	Proportion trained	Total trained	% trained at level 2	% trained at level 3
0-9	49.4%	2.0	72.4%	23.3%
10-24	20.6%	3.2	71.3%	23.5%
25-49	13.2%	4.6	73.1%	21.9%
50-249	8.1%	7.8	73.9%	20.6%
250+	2.6%	31.0	73.7%	20.0%
Total	24.3%	7.2	72.7%	22.2%

Source: London Economics analysis of the EDS using BSD and ILR data

Table 19 presents summary statistics on training indicators by sector of activity. In general, the distribution by sector would reflect the relative proportion of smaller and larger companies in each sector. In fact, sectors with a low number of employees trained have a

[20] Also, the proportion of TTG funding received at higher level is between 69-71% in the first three years and rises to 76% in 2010 (but in the ILR 2009/10 information on TTG funding level was missing for a significant number of records).

high proportion of trained employees (as a share of total employees), reflecting smaller size on average (see for example Agriculture and Construction), while companies in sectors with a high number of trained employees have a low proportion of trained employees, probably reflecting the larger size on average (see Public Administration and Defence and Electricity, Gas and Water). As for the structure of training, the Education and Health and Social Work sector seem to train a relatively high proportion of employees to level 3 qualifications: the average proportion trained at level 3 (as a proportion of total employees receiving training through TTG) exceeds 30% for both sectors, compared to an overall average of 22%.

Table 19: TTG training by sector of activity, for firms identified in the IDBR

	Training Indicators			
Sector	Proportion trained	Total trained	% trained at level 2	% trained at level 3
Agriculture	37.5%	3.7	81.9%	14.7%
Mining and Quarrying	16.7%	10.8	81.5%	13.0%
Manufacturing	18.5%	10.3	80.7%	15.0%
Electricity/Gas/Water	7.2%	18.1	75.4%	18.8%
Construction	39.2%	5.1	80.4%	17.0%
Wholesale/Retail	25.8%	4.9	76.0%	20.6%
Hotels /restaurants	25.3%	4.5	79.8%	13.9%
Transport/Storage	35.2%	12.1	88.9%	8.4%
Financial/Insurance	19.6%	7.5	71.8%	21.0%
Business Services	28.5%	6.9	71.8%	19.4%
PA and Defence	4.7%	26.9	76.3%	18.0%
Education	17.5%	12.8	57.4%	36.5%
Health/ Social Work	18.1%	6.4	63.3%	31.0%
Other services	28.2%	5.6	69.4%	25.9%
Total	25.2%	7.1	72.9%	22.1%

Source: London Economics analysis of the EDS using BSD and ILR data

Summary statistics on training by region of location are presented in Table 20. The average number of employees trained per firm varies from 5.6 in the South West to 9.9 in the North East. Similarly, the proportion of employees trained ranges from 23% in the South East to 28% in the North East. There is also some variation in the proportion of employees trained at different levels, with companies based in London, the South East, North West and West Midlands training, on average, a higher proportion of employees at level 3.

Table 20: TTG training by region of location, for firms identified in the IDBR				
	Training Indicators			
Region	Proportion trained	Total trained	% trained at level 2	% trained at level 3
North East	28.2%	9.9	76.6%	17.9%
North West & Merseyside	25.4%	6.5	72.0%	23.1%
Yorkshire & Humber	26.5%	7.2	75.9%	18.6%
East Midlands	26.3%	7.2	73.7%	21.5%
West Midlands	26.9%	7.5	72.7%	23.1%
East of England	24.7%	6.8	75.3%	20.5%
London	25.2%	7.8	66.8%	26.8%
South East	23.0%	6.5	72.2%	23.3%
South West	23.7%	5.6	72.9%	21.4%
Total	25.2%	7.1	72.9%	22.1%

Source: London Economics analysis of the EDS using BSD and ILR data

3.5 Step 5: Matching to the ABI

3.5.1 Structure of the Annual Respondents Database

The Annual Business Inquiry (ABI) collects detailed input and output data on an annual basis. The information is then available to researchers through the Annual Respondents Database (ARD). The latest available edition is currently the ABI 2008, which implies we have a maximum of two years of training data when we merge information from the EDS/IDBR dataset to the ABI.

The unit of data collection in the ABI is the reporting unit (RU)[21]. RUs may be a subset of an enterprise (the same enterprise might have multiple reporting units) and a superset of local units (the same RU may report on multiple local units). In general the reporting unit is the same as the enterprise, but in some cases, especially for very large enterprises, the same enterprise has more than one RU.

In terms of database structure, ARD files are divided into sectors (catering, construction, motor trade, property, retail trade, other services and wholesale) and types of contributors:

- The main ARD file for analysis on firms' productivity is the DAT file, which contains detailed information on contributors who were selected and returned data. The dataset contains around 47,000 observations.

- The NUL file reports indicative information for contributors who were selected but did not return data and contributors who were not selected.

- While both the datasets above contain information at reporting unit level, the SNUL dataset contains indicative information for both contributors who were selected and the rest of the population at local unit level. In other words, the SNUL dataset contains indicative information on all local units and the *luref* and *ruref* identifiers. It is therefore possible to map between local units and reporting units.

3.5.2 Matching training information to the ABI

The EDS/IDBR data containing information on training activities undertaken through TTG is organised at enterprise (*entref* identifier) or local unit (*luref* identifier) level, with the original match at local unit level being less successful compared to the match at enterprise level. However, the unit of data collection and analysis for ONS surveys (including the ABI) is the reporting unit (*ruref* identifier), which does not correspond to enterprise or local units (in general it is the same as the enterprise, but there are some enterprises with multiple

[21] Definition of reporting unit from the Annual Respondents Database User Guide: "The reporting unit holds the mailing address for the business and is the unit for which businesses report their survey data to ONS. In general, the reporting unit is the same as the enterprise. In some of the more complex cases, enterprises are subdivided into reporting units that correspond to KAUs, and are defined by specifying the appropriate local units from within an enterprise. Note that unlike on the CSO business register, local units and reporting units are distinct on the IDBR. In particular reporting unit is not also a local unit. A reporting unit and a local unit may be co-located but have distinct identities. The local units that form a reporting unit have employment that sums to the reporting unit, there is no residual employment accounted for by the reporting unit itself."

RUs)[22]. In order to conduct analysis on the impact of training on productivity we need to match the information on training from the EDS/IDBR with the ABI to use a series of detailed variables on inputs and outputs contained in the ABI.

In turn, this implies a considerable loss of treated observations (those engaging in TTG and identified in the EDS/IDBR dataset) and a dilemma on the level of aggregation to use in the analysis. In fact neither the IDBR definition of statistical units used in the BSD (enterprise and local unit) has a one-to-one match with the definition of reporting unit. We considered three different approaches:

- Performing the analysis at **enterprise level**, which implies aggregating the financial information available at RU level; the main weakness of this approach is not all reporting units forming the same enterprise are surveyed (or return information) each year. Consequently, we cannot generate a total value for financial variables at enterprise level aggregating values for those RUs included in the ABI-DAT files variables (RUs returning information do not account for 100% of the enterprise).

- Performing the analysis at **reporting unit level**; no transformation of ABI variables is needed since data is collected at RU level. However, we need to generate a value for the amount of training undertaken at reporting unit level, either aggregating the value for all local units covered by a reporting unit or apportioning the value at enterprise level across RUs. We will return to this issue in more detail below.

- Performing the analysis at **local unit level**; there are two main disadvantages associated to this option: firstly the original match to the IDBR at local unit level was less successful than at enterprise level; secondly and more crucially, no detailed data on inputs and outputs is collected at LU level and it is not clear how to allocate values for inputs and outputs from RUs to LUs. The only feasible method would rely on using information on LU employment to apportion data on inputs and outputs. However, this methodology relies on strong assumptions on the relationship between employment and other inputs and outputs.

The approach based on reporting unit requires weaker data assumptions and we therefore decided to conduct the data and econometric analysis using the ABI at the reporting unit level.

However, data on training activities through the EDS/IDBR dataset are available at either the enterprise or local unit level and we need to identify how to generate a value at RU level. For single-RU enterprises, the match is straightforward: the reporting unit corresponds to the enterprise and data on training can be merged with information contained in the ABI. For enterprises comprising more than one reporting unit it is less

[22] For details on matching the BSD to the ABI and the structure of businesses see Criscuolo, C., Haskel, J. and Martin, R. (2003). “Building the evidence base for productivity policy using business data linking”, Economic Trends, 600, pp. 39-61. and Galindo-Rueda, F. & Haskel, J. (2005)."Skills, Workforce Characteristics and Firm-Level Productivity: Evidence from the Matched ABI/Employer Skills Survey,"IZA Discussion Papers 1542, Institute for the Study of Labor (IZA).

clear how to impute the value for training at RU level. If the match at enterprise and LU level had been equally successful, the amount of training identified at enterprise level would correspond to the sum of the training variables allocated across different LUs of the same enterprise. As a consequence we would be able to identify values for the training variables at RU level, aggregating information available at LU. However, the original match at LU was less successful than the match at enterprise level and as a result the value for training at enterprise level may be more than the sum of the values at LU level. Therefore, for multi-RU enterprises, we may be unable to allocate all training undertaken and identified at enterprise level, to the different reporting units. In other words, for many multi-RU enterprises, we have that some proportion of training being undertaken at enterprise level that is not assigned to any RUs, which might lead us to underestimate the training activities at RU level.

Below we try to summarise the different steps in the matching process using a fictitious case. In the example presented in Figure 5, four sites at A44 level (ILR/EDS dataset) are matched to the IDBR with the same *entref* number (they all belong to the same enterprise, E1). In total E1 has trained 50 employees in a given year. However, only 25 of these employees are assigned to a local unit, while 25 remained unassigned to any of the local units forming E1. This implies that in the ABI, at RU level, we are unable to assign 25 employees receiving training. In theory they could belong to RU1 or RU2 (surveyed in the ABI) or RU3 (not surveyed in the ABI). This could lead us to underestimate (in the matched EDS/IDBR/ABI file) the proportion of training being undertaken by RUs if the employees receiving training belonged to either RU1 or RU2.

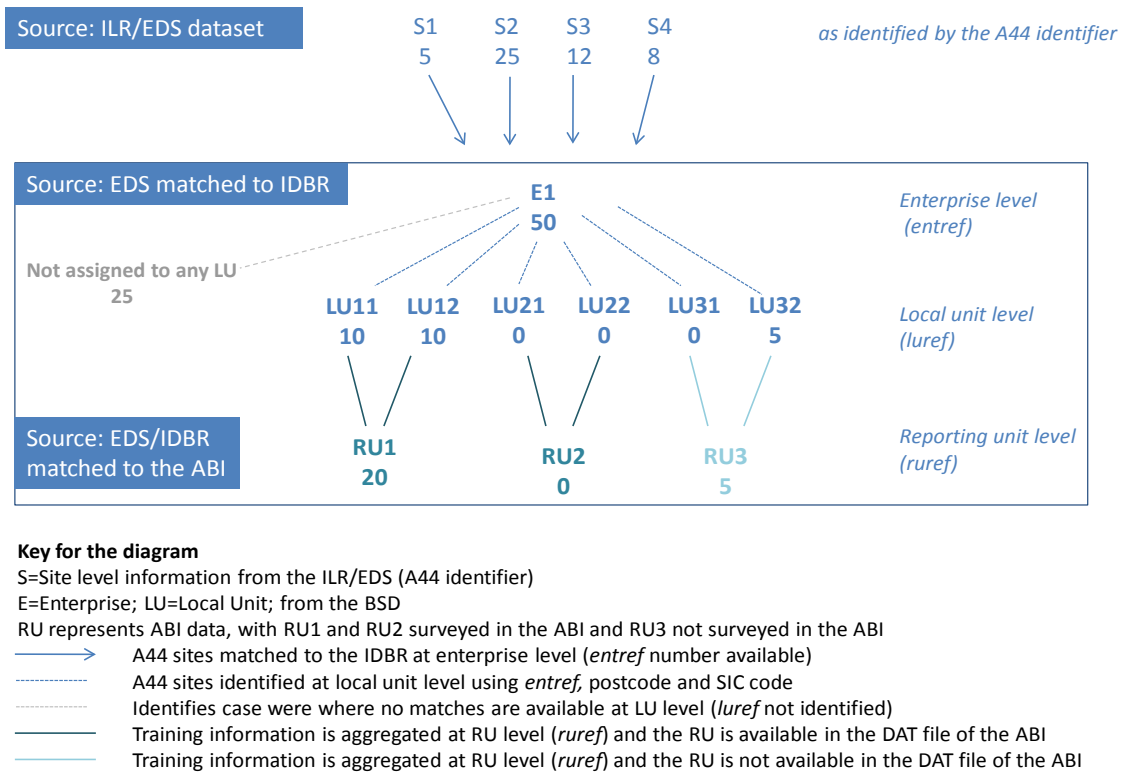
We can make explicit the degree of information available at RU level by identifying the overall proportion identified at RU level (from LUs) and explained by each reporting RU:

Table 21: Matching at RU level

Enterprise		Reporting Units				
Code	Employees trained	Code	Employees trained (identified)	Proportion Identified	Proportion Explained	Available in the ABI
E1	50	RU1	20	*50%* (25/50)	*80%* (20/25)	Yes
E1	50	RU2	0	*50%* (25/50)	0% (0/25)	Yes
E1	50	RU3	5	*50%* (25/50)	*20%* (5/25)	No

Source: London Economics

Figure 5: Matching training information



Key for the diagram
S=Site level information from the ILR/EDS (A44 identifier)
E=Enterprise; LU=Local Unit; from the BSD
RU represents ABI data, with RU1 and RU2 surveyed in the ABI and RU3 not surveyed in the ABI
A44 sites matched to the IDBR at enterprise level (*entref* number available)
A44 sites identified at local unit level using *entref,* postcode and SIC code
Identifies case were where no matches are available at LU level (*luref* not identified)
Training information is aggregated at RU level (*ruref*) and the RU is available in the DAT file of the ABI
Training information is aggregated at RU level (*ruref*) and the RU is not available in the DAT file of the ABI

Note: the match took place at Enterprise level and then at LU level (using *entref*, postcode and SIC code). RU data was then aggregated from LU information – using information in the ABI (SNUL) dataset, which maps LUs to RUs, but does not contain any financial information. Financial information is only available for RU1 & RU2 in our example, since they have been selected for the ABI(DAT) and returned the questionnaire. We are still able to identify the number of employees trained by RU3 and exclude them from the analysis (since RU3 did not take part in the ABI (DAT)). But we are unable to allocate the other 25 employees to any LUs and, consequently, RUs.
Source: London Economics

Clearly, unless the proportion identified at RU level (from local units) is 100% of the training undertaken at enterprise level, we will have a degree of uncertainty on how training should be allocated among RUs. While different approaches are possible (for example using the proportion identified to gross up the amount of training undertaken at RU level, or considering only cases where the proportion identified is above a certain threshold, e.g. 50%), none of them is likely to provide a good approximation on training effectively undertaken at RU levels, unless the proportion identified is close to 100%. Consequently, and considering that the vast majority of RUs in the ABI corresponds to the enterprise, we have decided to discard all RUs belonging to multi-RU enterprises and to restrict our attention to cases where the RU is reporting on the entire enterprise (single-RU enterprises)[23].

[23] In Table 62 we also provide the estimated results of a regression including all RUs: a first specification with all RUs, non controlling for the proportion of total training undertaken at enterprise level that was assigned to RUs and a second specification including only RUs belonging to multi RU-enterprise for which at least 50% of training undertaken was identified.

Below we show the impact this would have on the sample size for the analysis: more than 97% of RUs surveyed in the ABI between 2006 and 2008 report for the whole enterprise, while less than 3% of RUs belong to multi-RU enterprises. Discarding the latter group of RUs would lead to a loss of around 306 observations over the period 2007-2008 (around 6% of the total)[24].

Table 22: Number of reporting units per enterprise

Number of reporting units in the ABI 2006-08	Not engaged in TTG		Engaged in TTG		Total	
1	121,150	97.2%	4,917	94.1%	126,067	97.1%
2	814	0.7%	91	1.7%	905	0.7%
3	477	0.4%	52	1.0%	529	0.4%
4	323	0.3%	30	0.6%	353	0.3%
5	300	0.2%	20	0.4%	320	0.2%
6-10	854	0.7%	65	1.2%	919	0.7%
11+	745	0.6%	48	0.9%	793	0.6%
Total	124,663		5,223		129,886	

Note: Training activities were matched in the ABI 2007 and ABI 2008 only
Source: London Economics analysis of the ABI using matched EDS/IDBR data

[24] The figures refer to RUs belonging to enterprises undertaking training in the period considered, which does not necessarily imply that some form of training was undertaken by the specific RU. As a consequence, the loss of "treated" observations is less than the number presented in Table 22

4 Stage 2: Data and econometric analysis

Stage 1 of the analysis focused on assessing the potential loss of representativeness moving from the original ILR/EDS sample (with Blue Sheep data) to the sample formed of companies matched in the BSD at enterprise (and when possible local unit level) and, subsequently, to the sample of companies matched in the ABI. While the move from the original sample to the BSD matched sample seems to generate some composition bias, with larger companies being over-represented, the final ABI matched sample is likely to contain a disproportionate percentage of large and very large companies compared to the original sample. We start with describing the characteristics of the matched ABI sample and comparing them with the characteristics of other firms available in the ABI. We will also briefly refer to the characteristics of all enterprises identified in the Business Structure Database (section 3.4.2).

The last part of the chapter will present the emerging econometric findings and all the data issues associated to the datasets in use. It is important to notice that the group of firms undertaking training through *Train to Gain* and identified in the ABI is considerably different from the initial group of firms available in the EDS dataset. However, we have little chance to control for the bias induced by the different matching steps (EDS to IDBR and EDS/IDBR to the ABI): firstly since information is derived from different data sources not directly comparable (EDS and IDBR) and secondly since the main factor explaining the loss of observations (the probability of being observed in the EDS/IDBR and final ABI sample) is company size, which is a relevant control variable in the econometric model. All results are therefore subject to observing the treated firm in the final sample. We will try to attenuate this bias stratifying the analysis by company size and sector when possible.

In the remainder of this section we will use equivalently the terms TTG group or treatment group to identify firms undertaking training through the TTG programme and matched to the ABI. Similarly the rest of the population is defined as non-TTG or comparison group.

4.1 Firms' characteristics using the ABI

In this section we present a series of descriptive statistics for the group of firms available in the ABI and engaging in TTG activities (information merged from the EDS/IDBR database) and the rest of the population. Data refer to 2007 and 2008, when training activities took place. The classification used in the ABI is different from the classification used in the BSD with respect to sector of activity (and also to some extent for region of location). We have decided to keep and use the ABI classification for the analysis undertaken in this section. The analysis should also be seen as a first indication of the available sample size for the subsequent econometric analysis. However, in the econometric analysis we restrict the sample size further by controlling for a series of variables which might not be available in 100% of cases.

Overall, we have around 4,400 TTG observations (6.1% of the total) over the period 2007-08, compared to almost 68,000 (93.9%) for the untreated group.

The *Train to Gain* group of firms identified in the ABI seems to be heavily skewed towards larger firms compared to the group of firms not belonging to the TTG group: almost half of firms in the non-TTG group are micro enterprises compared to just 2% in the TTG group; conversely, medium/large and large companies (50-250 and 250+) account for 28% of firms in the non-TTG group, but almost 85% in the TTG group.

If we compare the distribution by company size of the TTG group identified in the ABI with the general TTG group matched to the IDBR (at *entref* level) presented in Table 14, we can see how the proportion of micro and small enterprises (less than 25 employees) has dropped dramatically from slightly more than 50% to less than 8%, while the proportion of larger firms (more than 50) has risen sharply from less than 30% to more than 85%. The huge disproportion of medium/large and large enterprises in the group of TTG firms available in the ABI is due to the ABI survey structure, with smaller firms having a much lower probability of being surveyed.

Table 23: Company Size using the ABI

Size	Firms identified as not engaging in TTG		Firms identified as engaging in TTG		Total	
	No. of firms	%	No. of firms	%	No. of firms	%
0-9	31,548	46.4%	100	2.1%	31,648	43.5%
10-24	10,600	15.6%	262	5.6%	10,862	14.9%
25-49	6,597	9.7%	345	7.4%	6,942	9.5%
50-249	12,547	18.4%	1,465	31.3%	14,012	19.3%
250+	6,738	9.9%	2,510	53.6%	9,248	12.7%
Total	68,030		4,682		72,712	

Note: Pooled ABI 2007-2008, England only
Source: London Economics analysis of the ABI using matched EDS/IDBR data

In Table 24 we present the distribution by region of location: the two main differences across the TTG and non-TTG group of companies relate to the proportion of enterprises based in London and the South East (43% in the non-TTG sample but only 32% in the TTG sample) and the North West (around 10% in the non-TTG sample and 15% in the TTG sample).

Table 24: Region of location using the ABI

Size	Firms not engaging in TTG		Firms engaging in TTG		Total	
	No. of firms	%	No. of firms	%	No. of firms	%
South East & London	29,236	43.0%	1,525	32.6%	30,761	42.3%
East Anglia	3,293	4.8%	205	4.4%	3,498	4.8%
South West	6,903	10.1%	442	9.4%	7,345	10.1%
West Midlands	6,907	10.2%	590	12.6%	7,497	10.3%
East Midlands	5,576	8.2%	448	9.6%	6,024	8.3%
Yorkshire & Humberside	6,134	9.0%	488	10.4%	6,622	9.1%
North West	7,144	10.5%	715	15.3%	7,859	10.8%
North East	2,845	4.2%	269	5.7%	3,114	4.3%
Total	68,038		4,682		72,720	

Note: Pooled ABI 2007-2008, England only
Source: London Economics analysis of the ABI using matched EDS/IDBR data

Table 25 presents the sector of activity (using the ABI classification) for the two groups: firms active in the Construction, Production sectors are over-represented in the group of TTG firms compared to the rest of the population; all other sectors are under-represented to some extent in the TTG group of firms compared to the rest of the population.

Table 25: Sector of activity using the ABI

Size	Firms not engaging in TTG		Firms engaging in TTG		Total	
	No. of firms	%	No. of firms	%	No. of firms	%
Catering	1,881	2.8%	231	4.9%	2,112	2.9%
Construction	5,503	8.1%	536	11.4%	6,039	8.3%
Motor Trades	2,608	3.8%	121	2.6%	2,729	3.8%
Production	11,959	17.6%	1,076	23.0%	13,035	17.9%
Property	3,142	4.6%	122	2.6%	3,264	4.5%
Retail	5,653	8.3%	359	7.7%	6,012	8.3%
Service Trades	28,002	41.2%	1,801	38.5%	29,803	41.0%
Wholesale	9,290	13.7%	436	9.3%	9,726	13.4%
Total	68,038		4,682		72,720	

Note: Pooled ABI 2007-2008, England only

Source: London Economics analysis of the ABI using matched EDS/IDBR data

Ideally, we would like to stratify the econometric analysis by company size and sector (and also possibly the cross-tab of sector and company size). However the descriptive statistics presented in this section suggest that sample size availability might constrain the scope of our analysis. In Table 26 we present a cross-tab by company size and sector to assess the scope for clustering the analysis by size and sector. Data refer to the treatment group only and do not condition on the availability of the dependent variable and the different variables controlled for in the analysis, which are likely to reduce the sample size available (see section 4.3.1 for more details); it should be seen as an upper bound and indicative only of data availability. Even at this stage, we can see that for most cells no robust analysis can be undertaken combining size and sector: most cells are below 50 (indicated in italics) and a high number between 50 and 100 (underlined). Only eight cells, crossing the four larger sectors (Construction, Production, Service trades and Wholesale) and the two larger size bands have more than 100 observations. We therefore decided to conduct the econometric analysis by company size and sector, without further stratification by size and sector combined.

Table 26: Cross-tab of company size and sector of activity using the ABI – TTG group only					
Company size / **Sector**	0-9	10-24	25-49	50-249	250+
Catering	*<10*	*<20*	*<10*	*46*	**162**
Construction	*20*	*50*	69	223	**174**
Motor Trades	*<10*	*<10*	*<20*	*50*	*50*
Production	*<10*	>50&<60	89	**445**	**481**
Property	*<10*	*<10*	*<10*	*<20*	92
Retail	*<20*	*<20*	*<20*	61	**263**
Service Trades	*40*	84	97	**453**	**1,128**
Wholesale	*<20*	*35*	57	**171**	**160**

Note: Pooled ABI 2007-2008, England only, Data refer to the TTG group only. Italics: sample size up to 50; Underlined: sample size between 51 and 100; Bold: sample size greater than 100.
Source: London Economics analysis of the ABI using matched EDS/IDBR data

In Table 27 and Table 28 we present summary information on the different steps undertaken and the differences across the original ILR/EDS sample, the BSD matched sample and the EDS sample identified in the ABI. The final column presents characteristics for the entire ABI sample (treated and non-treated observations). As already mentioned, small firms are heavily under-represented and large firms heavily over-represented in the final EDS-ABI sample compared to the original sample or the sample identified in the IDBR. Manufacturing firms and firms operating in Wholesale/Retail seem to be overly represented in the final EDS/ABI sample, while firms in the Health and Social Work and the Education sector seem to be under-represented compared to the original sample or the sample of firms identified in the IDBR.

Table 27: Compositional changes through matching by enterprise size				
	Employees in Enterprise/Group			
No of employees	**ILR/EDS[1]**	**BSD[2]**	**EDS-ABI matches[3]**	**Full ABI[3]**
0-9	39.2%	33.8%	2.1%	43.5%
10-24	20.9%	17.5%	5.6%	14.9%
25-49	12.9%	11.0%	7.4%	9.5%
50-249	14.5%	13.9%	31.3%	19.3%
250+	12.7%	23.8%	53.6%	12.7%
Number of firms	217,798	100,565	4,682	72,712

Note: 1 Taken from Table 6 (group level); 2 Taken from Table 9 (enterprise level); 3 Taken from Table 25 (reporting unit level).
Source: London Economics analysis of the EDS, IDBR and ABI data

Table 28: Compositional changes through matching by sector				
	Employees in Enterprise/Group			
Sector	ILR/EDS[1]	BSD[2]	EDS-ABI matches[3]	Full ABI[3]
Agriculture	0.20%	0.60%	na	na
Construction	14.80%	12.80%	11.2%	7.3%
Education	10.80%	8.10%	5.7%	3.1%
Electricity/Gas/Water	0.10%	0.10%	0.4%	0.2%
Financial/Insurance	0.80%	0.90%	0.3%	0.3%
Health/ Social Work	20.40%	21.40%	11.6%	2.4%
Hotels /restaurants	8.10%	8.90%	5.1%	3.0%
Manufacturing	5.80%	9.40%	22.3%	17.3%
Mining and Quarrying	0.10%	0.10%	0.3%	0.3%
Other services	5.60%	5.80%	4.4%	7.2%
PA and Defence	2.20%	0.70%	na	na
Business Services	7.90%	12.10%	12.4%	28.0%
Transport/Storage	8.20%	4.40%	6.5%	5.4%
Wholesale/Retail	14.80%	14.50%	19.7%	25.5%

Note: 1 Taken from Table 6 (group level); 2 Taken from Table 9 (enterprise level); 3 Reporting unit level.
Source: London Economics analysis of the EDS, IDBR and ABI data

In the next section we introduce the econometric model used in the analysis and discuss various data issues associated to the variables used in the model.

4.1.1 Training characteristics

Below we present a series of characteristics on TTG activities undertaken in the period 2007-08, by company size and sector. We report the proportion of employees receiving TTG training and also the proportion of firms with different training intensity (divided in three bands, 0-5%, 5-25% and 25% or above). Statistics are reported both relatively to the entire sample (including non-TTG firms) and to the TTG group of firms only. Fuller information by both sector and size band is provided in table 56 in annex 1.

When we restrict the attention to TTG firms only, smaller firms have clearly a much higher proportion of employees receiving TTG training than larger firms. Looking at sectors of activity, the Construction sector has the highest proportion of TTG employees, followed by firms in the Production and Wholesale sectors.

The tables show a significant proportion of firms engaging in *Train to Gain* activities appear to train a very small proportion of their employees. While this might be explained by a variety of factors[25], it might also undermine the possibility to detect a significant impact of training on productivity in the presence of "noise" correlated with training activities and productivity.

[25] For example the low matching ratio to the IDBR, the fact that we are forced to use the reporting unit level for productivity analysis and the fact that the Train to Gain may have not fully reached all recipients by the end of 2008

Table 29:	TTG Training characteristics by firm size– ABI sample										
	%trained *All*		%trained *TTG only*		Training band *All*				Training band *TTG only*		
Firm size	*Avg*	*SD*	*Avg*	*SD*	*Nil*	*0-5%*	*5-25%*	*> 25%*	*0-5%*	*5-25%*	*> 25%*
1-9	0.1%	0.03	47.8%	0.33	99.7%	Na	0.1%	0.2%	na	40.0%	60.0%
10-24	0.5%	0.05	20.4%	0.22	97.6%	0.3%	1.6%	0.5%	10.8%	66.4%	22.8%
25-49	0.6%	0.04	11.9%	0.14	95.0%	2.0%	2.3%	0.6%	40.1%	47.1%	12.8%
50-99	0.7%	0.04	8.3%	0.11	91.9%	4.5%	3.0%	0.6%	55.2%	37.7%	7.1%
100-249	0.7%	0.04	5.6%	0.09	87.3%	9.0%	3.1%	0.6%	70.9%	24.5%	4.6%
250-499	0.7%	0.03	3.4%	0.06	80.5%	16.1 %	3.2%	0.3%	82.2%	16.2%	1.6%
500-5000	0.5%	0.02	1.7%	0.04	68.7%	29.2 %	2.1%	na	93.2%	6.8%	na
5000+	0.2%	0.01	0.4%	0.01	36.0%	64.0 %	na	na	100.0%	na	na
Total	0.4%	0.04	6.2%	0.13	93.6%	4.6%	1.4%	0.3%	72.2%	22.4%	5.5%

Note: When cell size was smaller than 10 the value was removed (na) and the total rescaled to the value of the remaining "valid" cells.
Avg= average proportion of employees receiving TTG training SD= Standard deviation
Source: London Economics analysis of the EDS and ABI data

Table 30:	TTG Training characteristics by sector of activity – ABI sample										
	%trained *All*		%trained *TTG only*		Training band *All*				Training band *TTG only*		
Firm size	*Avg*	*SD*	*Avg*	*SD*	*Nil*	*0-5%*	*5-25%*	*> 25%*	*0-5%*	*5-25%*	*> 25%*
Catering	0.3%	0.02	2.6%	0.07	89.2%	9.5%	1.3%	na	87.7%	12.3%	na
Construction	1.1%	0.07	12.1%	0.19	91.2%	4.6%	3.0%	1.2%	52.5%	33.8%	13.8%
Motor Trades	0.2%	0.02	4.2%	0.08	95.7%	3.5%	0.8%	na	80.3%	19.7%	na
Production	0.5%	0.04	6.6%	0.12	91.9%	5.4%	2.2%	0.5%	66.9%	27.5%	5.7%
Property	0.2%	0.03	4.8%	0.16	96.4%	3.1%	0.5%	na	85.6%	14.4%	na
Retail	0.2%	0.02	3.2%	0.10	94.1%	5.2%	0.5%	0.2%	87.9%	9.0%	3.1%
Service Trades	0.3%	0.03	5.5%	0.12	94.0%	4.5%	1.2%	0.3%	75.4%	19.9%	4.7%
Wholesale	0.3%	0.03	6.2%	0.10	95.5%	3.1%	1.2%	0.2%	68.9%	25.8%	5.3%
Total	0.4%	0.04	6.2%	0.13	93.6%	4.6%	1.4%	0.3%	72.2%	22.4%	5.5%

Note: When cell size was smaller than 10 the value was removed (na) and the total rescaled to the value of the remaining "valid" cells.
Avg= average proportion of employees receiving TTG training SD= Standard deviation
Source: London Economics analysis of the EDS and ABI data

4.2 Econometric Approach

4.2.1 Measures of productivity, data sources and sample composition issues

Labour productivity has normally been measured in the literature by gross value added (GVA) per employee. An alternative measure used in the literature is the average wage bill, which only captures the return to the individual, rather than the company. Both variables are available in the ABI, but not in the BSD, which implies that if we wish to estimate the impact of education and training on firm level productivity, then we are restricted to the ABI matched group of companies. Alternative routes, such as measuring productivity using BSD information on turnover per employee, are not viable for methodological reasons (the measure does not take into account the cost structure) and for lack of necessary data quality (data on employees and turnover are likely to be imputed and to be subject to measurement error).

4.2.2 Datasets and methodological issues

We have at maximum two years of training data overlapping with inputs and outputs information from the ABI (2007 and 2008). ABI datasets from 2006 to 2008 were appended into one dataset covering the period 2006-2008. While the main analysis focuses on 2007 and 2008, we have also used 2006 data to control for the characteristics of firms engaging in *Train to Gain* activities.

Ideally we would like to have access to a longitudinal dataset (following the same observation over a period of time). The current dataset gives us at maximum two years of data.

Estimates of the impact of training on productivity using cross-sectional data (the unit of analysis is observed at one point in time) can produce biased results due to the presence of endogeneity in training decisions, mainly identifiable in the two following sources:

- Unobserved heterogeneity: there are time-invariant firm-specific characteristics that affect both the training and productivity choice (for example organisational and managerial quality) that we are unable to control for; and

- Training can be a choice variable, i.e. training decision and productivity can be jointly determined (for example, a productivity shock might affect training decisions).

In the presence of endogeneity, we need some exogenous instrument. The absence of suitable instruments in cross-sections implies that we are unable to tackle endogeneity and estimates are therefore likely to be biased; however, with longitudinal data we can exploit the longitudinal structure of the dataset and deal with both problems (using lagged values of training and productivity as instruments). However, we need at the very least three or four years of data to apply the appropriate panel data techniques that control for endogeneity and time-invariant firm-specific characteristics (normally System and Difference GMM). Depending on the nature of the error even three to four years may not be sufficient, we are restricted to one lagged value, and as such, panel data techniques are clearly not applicable.

Also, linking across ABI surveys will exacerbate the loss of small and medium companies in the sample, given that they are not kept in the sample every year. Below we introduce the model specification and the proposed estimation approach that we have adopted given that data limitations do not allow full GMM panel data techniques.

4.2.3 Model specification

In the cross section (at time t) we can specify the following model:

$$y_i = \beta' X_i + \varepsilon_i \qquad (1)$$

Where y is the (natural logarithm) of productivity (measured by value added per worker[26]), X identifies a series of (potentially endogenous) regressors including training and ε is the error term (and *i* identifies the company).

The explanatory variable of primary focus included in X is the proportion of training undertaken at firm level (number of employees trained through TTG divided by the total employment).

The other explanatory variables included in X are:

- Capital per worker;
- Company size;
- Age of firm;
- Expenditure on software and advertising (as a proportion of total purchases);
- Sector;
- Region;
- Legal status;
- Domestic/foreign ownership;
- Proportion of part-time and full-time workers[27];

[26] Nominal values were deflated using the MM22 Producer Price Series for the manufacturing sector and the Services Producer Price Series when available. For all other sectors information from the Structural Trade Analysis Database from the OECD was used.

[27] Information on employment structure is only asked in the long-form of the ABI/1 questionnaire (the one on employment). These variables are therefore missing for a part of the sample. We decided to impute missing variables controlling for a series of relevant variables identifying firms' characteristics. The other alternative would be to exclude employment structure from the analysis.

- Yearly/seasonal dummies.

4.2.4 Estimation approach

- Firstly, we decided to run a probit model highlighting the association between firms' characteristics (including gross value added) on the decision to participate in the TTG programme. Participation in the TTG programme is defined as a binary variable (1 identifying firms engaging in TTG activities, 0 otherwise) and we used lagged data (2006 for firms participating in TTG in 2007 and 2007 data for firms participating in TTG in 2008) to explore the pre-training characteristics of firms taking part in TTG activities.

- We then estimated equation (1) in the cross-section using OLS (pooled over the two years 2007 and 2008 and separately).

- When sample size permitted equation (1) was estimated disaggregating by company size and sector of activity.

- We also estimated a differenced equation model (1) using the training variable as a flow; the dependent variable is therefore the change in gross value added between 2007 and 2008. Similarly, we estimated equation (1) controlling for the lagged values of gross value added and training. This is also a first attempt to control for any deferred impact of training undertaken on productivity.

- Finally, we tried to exploit the (limited) longitudinal component, using lagged values of the training and capital variables to instrument current values.

- In 0 we present results of equation (1) estimated carrying out various robustness checks: namely, including only those firms falling in a similar company size band according to the EDS and BSD database; controlling for the sequential year of *Train to Gain* training (i.e. whether the firm was in the first or second year of TTG training); controlling for the 3 digit sic code; a model controlling for training as a binary variable.

As already mentioned, linking two or more years of data together exacerbates the sample composition bias toward larger firms (which are surveyed every year), but it is a necessary condition to exploit the longitudinal framework of the dataset.

4.2.5 Data and methodological issues

There are a series of relevant issues with the data which should be considered when interpreting the estimation results:

1) **Sample composition bias**: the sample available in the ABI of TTG firms is considerably different from the original sample of firms participating in TTG. Sample selection primarily occurs due to company size and therefore there is little we can do to control for that. Results are therefore conditional on being in the matched dataset.

2) **Different data source and timing of information on training and ABI variables**: as described in section 3.1, training undertaken is recorded following the academic year and a course may extend over more than one academic year. Data on gross value added and other financial variables in the ABI typically refer to the calendar year, but firm can report in relation to any 12-month-period during the financial year. As a consequence, even if we have aligned the ILR dataset to the calendar year, residual differences may occur and training may have not been finished when we observe productivity data. Moreover, we are combining administrative data from the ILR and survey data from the ABI and the two data sources may be subject to different biases.

3) As a result of the matching process, we have that just 6% of the sample has a value for TTG training different from zero and for the rest of the sample we have a mass point at zero. While this could potentially affect the estimates, we have no way to control for training undertaken privately at firm level or other forms of publicly funded training (such as Apprenticeships). Similarly, we cannot characterise the selection rule determining how firms are selected and decide to engage in TTG activities.

4) Given that data on training covers the period from 2007 to 2010, but data on productivity are only available for the first two years, in most cases we can only observe the direct impact of training on productivity (also subject to the limitation on timings discussed in point 2) above). However, we also estimated one specification controlling for both the lagged and present value of training (see section 4.3.5 below).

5) **Endogeneity of results**: with only two years of linked data available, we cannot apply panel data techniques to control for time-invariant components and possible endogeneity of training (and other) variables. The Instrumental Variable approach presented above is a first attempt to control for the possible endogeneity. However we acknowledge that we cannot fully control for endogeneity with such a short time frame available.

6) **Capital stock series**: information on the capital stock is not asked in the ABI. A version of the Capital Stock series is available if the value for the capital stock is generated using a Perpetual Inventory Method (PIM). As explained in the VML technical guide on the estimation of capital stock[28], the PIM allocates shares of the aggregate capital stock series to firms for the first year that they appear in the ABI, then depreciates this value and adds the net investment (collected in the ABI) the following year and repeats these steps for all years. The aggregate capital stock series used is the Volume in Capital Services (VICS); however, the current version of the VICS available in the VML and compatible for matching with the Annual Respondents Database uses 2002 data. While there is currently an ongoing project at the ONS to update the VICS files, no updated information was going to be available before the end of the project and we therefore decided to project 2002 data forward.

[28] See the ONS “Technical Guide: Estimating capital stock“ by Robert Gilhooly

7) **Missing values and tolerance level in the capital stock series**: not all firms are surveyed every year in the Annual Business Inquiry and therefore the investment series in the Annual Respondents Database is affected by repeated missing values over time. The Microdata Analysis & User Support (MAUS) at the ONS uses employment figures to update missing values for the investment series. The tolerance level (i.e. the ratio of missing to observed values) can be altered to pick-up more small and medium sized firms at the cost of a diminished reliability of the series. In the analysis we decided to generate three values for the capital stock series, with the ratio moving from 1:1 (one missing per each observed value) to 2:1 and 3:1 (three missing per each observed value). However, given that even the 3:1 tolerance ratio was excluding the majority of small and medium enterprises, we decided to also run a specification using capital expenditure instead of capital stock. All tables clearly state the capital measure used in the analysis.

4.3 Descriptive statistics and estimation results

In this section we first look at how using different measures of the capital stock affects the available sample size for the analysis and then present a series of summary statistics for relevant variables for the TTG and non-TTG group. After that we describe firms' characteristics of both groups in the year prior to participating in training and run a probit model with training decisions as the dependent variable. Finally, we present all the estimates of model (1) using the different approaches highlighted in section 4.2.4.

4.3.1 Effect of different capital measures on the sample size

As already mentioned, we generated three different versions of the capital stock series, using different tolerance levels for the ratio of imputed to observed values. More specifically, versions of the capital stock series were generated altering the ratio from 1:1 (allowing one imputed value for each observed value) to 2:1 (allowing two imputed values for each observed value) and to 3:1 (allowing three imputed values for each observed value). In Table 31 we present the different sample sizes: the first column (GVA only) only conditions on the availability of gross value added[29] (the dependent variable), while columns two to five also condition on different versions of the capital stock series. Finally, in the rightmost column, the sample size shows the effect of conditioning on gross value added and capital expenditure. All sample sizes are presented in aggregate and divided by company size. When we condition on availability of gross value added only, we have more than 4,600 observations in the treatment group (pooled 2007-2008 data) and around 62,000 observations in the comparison group. When we condition on different versions of the capital stock series, we significantly restrict the sample size available and lose a vast proportion of micro and small enterprises: the treatment group drops to around 2,400 using the 1:1 tolerance level, 3,100 using the 2:1 tolerance level and to around 3,500 using the 3:1 tolerance level. However, the aggregate number of firms in the treatment group with

[29] The measure of gross value added is derived from the values of turnover, total purchases, variation in stocks and other minor components. We decided to recode the value of GVA to missing when all components were equal to zero or when either the value for turnover or total purchases was equal to zero, possibly reflecting that data were not available (turnover and purchases data should not be zero for a company operating under normal conditions). In total we lose around 7,000 observations in the period 2007-2008, but only 80 in the treatment group (since it's formed by larger companies on average)

less than 50 employees drops to 55, 138 and 229 respectively (with the latter definition capturing a significant higher number of small to medium enterprises, but only a limited extra number of micro firms). Conversely, the capital expenditure (capex) series is collected directly in the ABI and conditioning on capex, rather than capital stock, would imply only a small reduction in the number of observations compared to the original sample size (presented in the GVA only column).

Table 31: Sample size availability using different versions of the capital measure

	GVA only		GVA+CAP STOCK 1:1		GVA+CAP STOCK 2:1		GVA+CAP STOCK 3:1		GVA+CAPEX	
Company size	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG
0-9	26,561	91	3,867	11	5,935	19	7,907	27	24,712	82
10-24	10,139	247	866	19	1,852	41	3,010	74	9,907	239
25-49	6,399	332	663	25	1,671	78	2,657	128	6,254	319
50-249	12,236	1,452	3,896	445	7,296	849	8,979	1,020	11,952	1,371
250+	6,592	2,474	5,310	1,917	6,021	2,156	6,217	2,229	6,413	2,341
Total	61,927	4,596	14,602	2,417	22,775	3,143	28,770	3,478	59,238	4,352

Note: Pooled ABI 2007-2008, England only
Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

In the main econometric specification we perform some sensitivity analysis, showing how results differ as we condition on different values of the capital stock series or use capex (full results are presented in Annex 1). When we disaggregate by size and sector we decided to run only two different specifications: one including the value of the capital stock series generated using the 2:1 imputation ratio and a second specification using capex. While capex measures the capital flow in a given year and it is therefore a highly imperfect proxy for the value of the capital stock, it allows us to conduct the analysis on a much larger sample and to include a much higher number of small and micro enterprises.

4.3.2 Analysis of key variables across the treatment and comparison group

In section 4.1 we have already described the main characteristics of the treatment and comparison group according to company size, region of location and sector of activity. Here we present summary statistics on gross value added and capital (stock and expenditure) across the treatment and comparison group at aggregate level and stratifying by company size and sector. The analysis is based on ABI pooled 2007-2008 and therefore measures shown are being determined simultaneously with the occurrence of training activities. In the next section we will also explore the characteristics of the treatment and comparison group before training occurred.

On aggregate, firms in the treatment and comparison groups seem to have similar values for gross value added, while the average value of both capital stock and expenditure is higher in the group of treated companies. However, looking at the different size bands, we can see that the main differences arise in the group of micro firms: treated firms with less than nine employees have a higher value of GVA on average compared to untreated firms in the same size band and a significantly higher value for both capital stock and capital expenditure. Considering the small number of observations available for the treatment group in this size band, it is likely that the average value is influenced by the presence of outliers. For the class of companies with more than 10 employees and less than 25, values

of GVA and capital are slightly higher in the treatment group, while in all other size bands the value for the comparison group exceeds the corresponding value for the treatment group. The two groups are remarkably similar in terms of capital expenditure once we exclude micro firms.

Table 32: ABI summary statistics by company size

	ln GVA per capita (mean)		ln GVA per capita (median)		ln CAP STOCK per capita 2:1 (mean)		ln CAPEX per capita (mean)	
Company size	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG
0-9	2.86	3.19	3.22	3.12	2.3	3.56	0.36	0.98
10-24	3.21	3.27	3.41	3.41	3.38	3.51	0.52	0.54
25-49	3.32	3.24	3.49	3.43	3.54	3.48	0.63	0.68
50-249	3.32	3.27	3.5	3.42	3.67	3.5	0.74	0.79
250+	3	2.93	3.47	3.26	3.61	3.48	0.96	0.97
Total	3.07	3.09	3.39	3.33	3.21	3.48	0.55	0.87

Note: Pooled ABI 2007-2008, England only; all measures are divided by total employment and expressed in logarithmic terms.
Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

In Table 33, we present a similar analysis disaggregating by sector of activity: firms in the treatment group seem to have a higher value of GVA (both referring to the average and median value) in all sectors apart from service trades. Moreover, the value of capital expenditure is consistently higher for treated firms across all sectors, while the value for capital stock (which clearly refers to a smaller sample) is lower in the treatment group for the catering sector and very similar for the retail and motor trades sectors.

Table 33: ABI summary statistics by sector of activity

	ln GVA per capita (mean)		ln GVA per capita (median)		ln CAP STOCK per capita 2:1 (mean)		ln CAPEX per capita (mean)	
Sector	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG
Catering	2.43	2.91	2.62	2.91	3.78	3.56	0.56	1.08
Construction	3.23	3.68	3.43	3.7	2.22	2.91	0.45	0.67
Motor Trades	2.84	3.27	3.17	3.37	3.2	3.18	0.48	0.68
Production	3.37	3.5	3.5	3.58	4.09	4.2	0.77	1.14
Property	3.01	3.39	3.41	3.45	3.59	4.01	0.59	0.82
Retail	2.53	2.81	2.83	2.92	3.15	3.17	0.39	0.83
Service Trades	3.06	2.62	3.43	3.01	2.87	3.23	0.5	0.76
Wholesale	3.17	3.44	3.57	3.63	3.13	3.23	0.53	0.83
Total	3.07	3.09	3.39	3.33	3.21	3.48	0.55	0.87

Note: Pooled ABI 2007-2008, England only; all measures are divided by total employment and expressed in logarithmic terms.
Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

In Figure 6 we present a scatter plot of log of GVA (vertical axis) and trained proportion (horizontal axis) by year and combined, while Figure 7 describes the same variables disaggregating by company size (the three lower size bands are aggregated into one band including all firms with less than 50 employees due to the limited size of each cell for the treatment group). As mentioned, there is a high concentration of training proportion around zero, given that 94% of firms did not undertake training through TTG[30] and we are not able to control for any other form of training undertaken. In It should also be noticed that in a limited number of cases (33), the values for the total number of employees receiving TTG training was greater than the reported number of employees (this may be explained by discrepancies in the timing and nature of the sources for training data, the ILR, and employment data, the ABI). Cases where the discrepancy was minimal (14 cases) were coded as training 100% of employees, while the other cases with number of employees receiving TTG training greater than total employment were discarded from the analysis (see Table 56 in Annex 1 for further details[31]).

In Table 34 we also provide the correlation coefficient between the log of GVA and the proportion of employees receiving training. Overall the correlation coefficient is very small and positive (larger in 2008) and it appears to be positive and for small firms and negative for large firms.

Table 34: Correlation coefficient between GVA and trained proportion

Sample	All	2007	2008	1-49	50-249	250+
Coeff	0.002**	0.008	0.015***	0.016***	0.004	-0.024**
p-value	0.012	0.130	0.007	0.001	0.605	0.025
Obs	0.002	0.008	0.015	0.016	0.004	-0.024

Note: Pooled ABI 2007-2008, England only; ln(GVA) per capita;. *** p<0.01, ** p<0.05, * p<0.1.
Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

[30] It is also possible that some firms have engaged in TTG but have not been identified in the IDBR and therefore have a null value for the training variable.

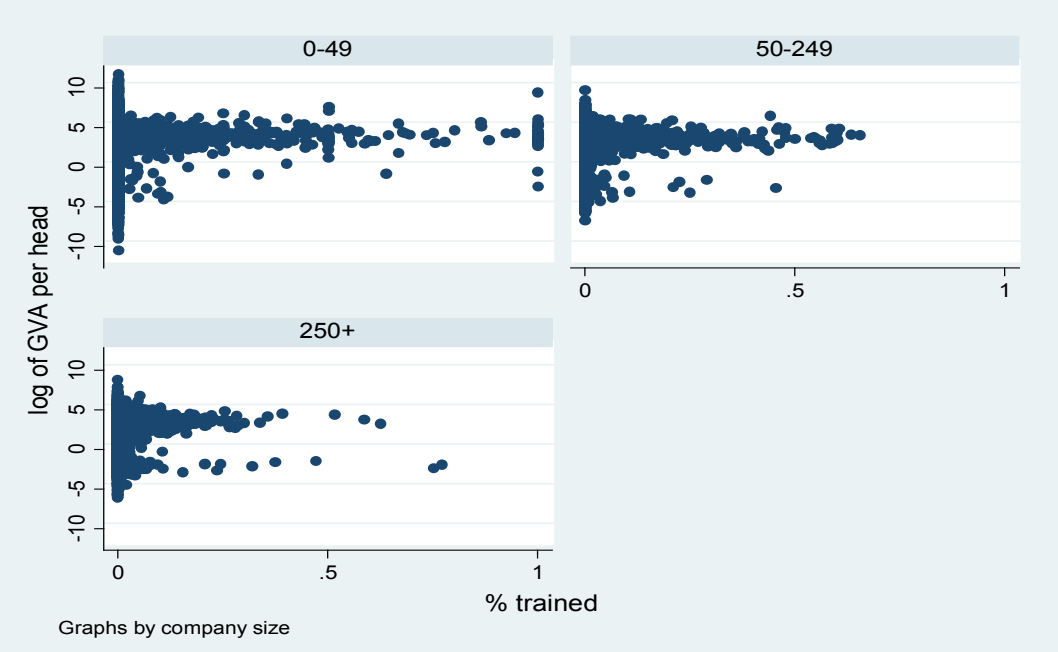
[31] Table 56 also contains the same scatter plots, removing both the 0% and 100% values for training.

Figure 6: GVA and trained proportion by year

Note: Pooled ABI 2007-2008, England only;.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Figure 7: GVA and trained proportion by company size



Note: Pooled ABI 2007-2008, England only.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

4.3.3 Training decisions and firms' characteristics pre-training

The longitudinal structure of the Annual Respondent Database allows us to follow firms over time. It is interesting to look at firms' characteristics before they engage in publicly funded training, also to calculate changes in firm level productivity variables and compare these for firms engaging in TTG with firms that did not. However, linking one or more years of the ABI together implies losing almost all small and micro firms. Below we present summary statistics for the treatment and comparison groups the year before training took place (using 2006 ABI values for 2007 and 2007 ABI values for 2008). Clearly firms will be available only if they are surveyed in at least two consecutive years.

When data is split by company size (Table 35), the first size band groups all firms with less than 50 employees, given that we lose the majority of observations for small and micro firms in the treatment group. Non-treated firms generally have higher values for GVA and capital stock, but lower values for capital expenditure. However differences in GVA only reflect a sample composition bias: the group of treated firms has a much higher proportion of large enterprises, which reported lower average values compared to other size groups. The average values of GVA across the two groups in each size band are remarkably similar. Treated firms with less than 50 employees have higher values for the capital stock and expenditure.

Table 35: Firms' characteristics pre-training: company size

	No. of firms		ln GVA per capita (mean)		ln GVA per capita (median)		ln CAP STOCK per capita 2:1 (mean)		ln CAPEX per capita (mean)	
Company Size	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG
0-49	6,356	216	3.32	3.29	3.46	3.4	3.73	3.87	0.63	0.82
50-249	6,425	781	3.36	3.32	3.52	3.42	3.73	3.57	0.79	0.83
250+	5,416	2,063	2.98	2.96	3.49	3.33	3.66	3.49	0.97	0.97
Total	18,197	3,060	3.24	3.09	3.49	3.37	3.7	3.52	0.79	0.92

Note: ABI data available in at least two consecutive years between 2006 and 2008; England only.
Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

In Table 36 we present a similar analysis disaggregating by sector of activity: average values of GVA are higher for the treatment group in the Construction, Motor Trades and Property sectors, lower in Service Trades and Wholesale and substantially similar in the remaining sectors.

Table 36:	Firms' characteristics pre-training: sector of activity									
	No. of firms		ln GVA per capita (mean)		ln GVA per capita (median)		ln CAP STOCK per capita 2:1 (mean)		ln CAPEX per capita (mean)	
Sector	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG
Catering	503	162	2.64	2.91	2.75	2.94	3.55	3.36	0.68	1.04
Construction	612	251	3.47	3.68	3.55	3.73	2.93	3.04	0.51	0.58
Motor Trades	695	70	3.19	3.41	3.39	3.41	3.4	3.13	0.63	0.67
Production	5,562	803	3.55	3.54	3.57	3.6	4.12	4.06	0.9	1.12
Property	427	91	3.26	3.53	3.7	3.53	4.24	3.92	1.13	0.96
Retail	1,058	268	2.85	2.82	2.93	2.91	3.07	3.21	0.54	0.86
Service Trades	6,697	1,140	2.92	2.57	3.39	3.03	3.59	3.37	0.8	0.87
Wholesale	2,643	275	3.61	3.46	3.75	3.66	3.44	3.25	0.68	0.84
Total	18,197	3,060	3.24	3.09	3.49	3.37	3.7	3.52	0.79	0.92

Note: ABI data available in at least two consecutive years between 2006 and 2008; England only.
Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Probit model on training decisions

We used the linked data to explore the association between a set of variables (the same presented in section 4.2.3, with GVA among the set of regressors) observed at time t and the decision to engage in TTG activities at time t+1 (defined as a binary variable). The results of the estimation should be seen as complementing the information presented in this section and no causal effect should be attributed to the probit model, given that little (or no) information available on other training activities undertaken at time t, how firms were contacted and decided to take part in the TTG programme, and all the sample composition bias due to matching issues. The variable identifying the log of GVA at time t (with training undertaken at time t+1) is small and never statistically significant at the aggregate level, although it seems to have a positive effect on the probability of engaging in training for firms in the Motor Trades and Property sectors (see Table 38). Total employment seems to have a positive effect on the decision to engage with TTG after controlling for size band, while the effect of capital stock seems to be positive for smaller firms and negative for larger firms. When we control for capital expenditure, instead of capital stock, we see a positive effect of capex on the probability of undertaking training at time t+1.

Table 37:	Probit model: training decisions, all and by company size							
	Dependent Variable: training at t+1							
Sample	All	All	Size 0-49	Size 0-49	Size 50-250	Size 50-250	Size 250+	Size 250+
ln GVA	0.000	-0.001	-0.003	0.002	0.003	-0.001	-0.001	-0.000
ln CapStock 2	-0.006***		0.003	-0.009***	-0.016***			
ln Capex		0.006***				0.006***	0.005*	0.010**
Size: 10-24	-0.071*	0.004						
Size: 25-49	-0.050	0.046						
Size: 50-249	0.036	0.123***						
Size: 250+	0.132***	0.237***						
Catering	-0.015	-0.028**	-0.014	-0.016	-0.029	0.018	-0.043**	-0.063*
Motor Trades	-0.095***	-0.069***		-0.084***	-0.201***	0.005	-0.087***	-0.203***
Production	-0.074***	-0.061***	-0.017	-0.078***	-0.143***	0.013	-0.082***	-0.168***
Property	-0.062***	-0.053***	-0.011	-0.064**	-0.122***	-0.016	-0.080***	-0.128***
Retail	-0.055***	-0.048***	-0.021	-0.075***	-0.080**	0.001	-0.073***	-0.103***
Service Trades	-0.094***	-0.074***	-0.017	-0.090***	-0.179***	0.010	-0.089***	-0.200***
Wholesale	-0.078***	-0.061***	-0.020	-0.086***	-0.122***	0.002	-0.083***	-0.134***
Foreign owner	-0.016**	-0.016***	-0.008	-0.013	-0.032***	-0.007	-0.019**	-0.036***
Multi-LUs	0.034***	0.025***	0.018	0.011	0.102***	0.010*	0.015**	0.101***
Software int	-0.175***	-0.140***	-0.066	-0.105	-0.469***	-0.074**	-0.124	-0.467***
Female FT (%)	-0.031	-0.007	-0.004	-0.062**	-0.014	0.013	-0.031	0.005
Female PT (%)	0.115***	0.102***	0.002	0.081**	0.244***	0.007	0.127***	0.291***
Male PT (%)	0.002	0.012	0.010	-0.116	0.060	0.024	-0.094	0.057
Obs	13,645	19,894	1,882	5,351	7,205	6,382	7,103	7,388
Capital measure	*CapStock*	*Capex*	*CapStock*	*Capex*	*CapStock*	*Capex*	*CapStock*	*Capex*

Note: Reporting marginal effects. ABI data available in at least two consecutive years between 2006 and 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies. Capital Stock generated using the 2:1 tolerance level on missing values. *** p<0.01, ** p<0.05, * p<0.1. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 38: Probit model: training decisions, by sector								
	Dependent Variable: training at t+1							
Sample	Catering	Construction	Motor trades	Production	Property	Retail	Service Trade	Wholesale
ln GVA	0.041	0.012	0.023**	-0.010*	0.041***	-0.003	0.003	-0.004
ln CapStock 2	0.003	-0.002	-0.010	-0.006	-0.007	0.033*	-0.014***	-0.008
Size: 10-24	-0.252	0.036		0.918***		0.911	-0.119	0.938
Size: 25-49				0.930***		0.903	-0.096	0.978***
Size: 50-249	-0.137	0.089	0.630***	0.914***	0.104	0.963	-0.048	0.900***
Size: 250+	-0.045	0.223	0.633***	0.983***	0.227	0.839	0.063	0.995***
Foreign owner	0.173***	-0.152***	-0.008	-0.009	-0.019	-0.028	-0.041***	0.018
Multi-LUs	0.189***	0.103*	0.021	0.015	-0.047	0.093**	0.062***	0.042***
Software int	-0.491	-4.326	1.957***	0.176	-0.747**	-0.157	-0.295**	-0.139
Female FT (%)	0.790**	-0.479*	-0.493*	-0.017	-0.053	-0.179	-0.019	-0.103**
Female PT (%)	0.650***	-0.436	0.506	0.007	0.699***	0.042	0.215***	-0.033
Male PT (%)	0.121	0.075	0.398	0.137	-1.030*	0.085	-0.049	0.035
Obs	417	560	430	4404	349	811	5622	1751

Note: Reporting marginal effects. ABI data available in at least two consecutive years between 2006 and 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies. Capital Stock generated using the 2:1 tolerance level on missing values. *** p<0.01, ** p<0.05, * p<0.1.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

4.3.4 Impact of training on firm-level productivity using ARD 2007-08

In this section we present the estimated results of equation (1) at the aggregate level and disaggregating by year, firm size and sector. We also use different definitions of capital stocks (varying the tolerance level for missing values from 1:1 to 3:1) and also estimate a specification using capital expenditure instead of capital stock. All other control variables are those outlined in section 4.2.3.

For clarity, in the rest of the analysis we will present the results from two specifications: one controlling for capital stock using the 2:1 tolerance ratio and another specification using capital expenditure. The main regressions controlling for the capital stock measures constructed using the 1:1 and 3:1 tolerance ratios are presented in 0.

For the period 2007-08, we have a total sample available of around 24,000 using the definition of capital stock (2:1), while controlling for capital expenditure (rather than stock), allows us to retain almost all the observations available in the pooled 2007-08 dataset (more than 61,000) and a much higher number of micro and small firms (particularly in the treatment group).

The pooled coefficient associated to the training variable is negative in the capital stock (2:1) specification (-0.16), which would imply that a one percentage point increase in the volume of training undertaken reduces productivity by approximately 0.16%. However, the effect is not statistically significant at any conventional level. The estimate turns positive and it stands at around 0.14 in the specification using capital expenditure, although the effect is still not statistically significant.

When we disaggregate by year we can see a large and positive coefficient in 2007 (although again never statistically significant) when using the capital stock measure and a large and negative coefficient associated with training in 2008[32]. The trend by year is opposite when using capital expenditure, negative in 2007 and positive in 2008.

The magnitude of the coefficient associated to the capital variable is around 0.08 in 2007, nil and insignificant in the second year and around 0.05 on aggregate), but double when using capital expenditure (around 0.11 on aggregate and positive and significant in all years). The variables identifying company size probably reflect the change in sample composition across specifications: the coefficient associated to company size (the baseline category is the group of micro firms) is positive and significant (apart from large enterprises, where it is not significant) and diminishing as we move across different capital stock measures. However the coefficient turns negative for larger companies in the specification using capital expenditure, which includes a much higher number of small and micro firms.

It is very difficult to sensibly interpret the results, given the difference in sample composition associated to moving from the specification with capital stock to the specification controlling for capital expenditure. Moreover, given the short time series available we cannot rule out the presence of transitory external shocks affecting productivity (notably the economic downturn), but not directly related to any inputs of the production function. It is difficult to know whether these external factors affected homogenously firms in the treatment and comparison groups, especially remembering the differences in sample composition across the two groups (with larger firms over-represented in the treatment group). This is one potential explanation; however, it is more likely that the various methodological issues that have been highlighted throughout the report are more important in explaining the reason for the ambiguous results.

Disaggregated results: Firm size and industrial sector

Table 40 presents results of the analysis split by company size (using three bands: 0-49 employees, 50-250 employees and 250+ employees). In the rightmost part of the table we present the results using capital expenditure. The results confirm that sample composition and time can have a significant effect on the estimated coefficients: the effect of training on productivity is positive and significant for small firms, with the pooled coefficient indicating that a one percentage point increase in the proportion of employees trained (through TTG) is associated with an increase of between 0.5% and 0.9% on firm productivity. The effect is never statistically significant for firms in the size band 50-249 employees (apart from 2007 for the specification using capital stock), while it is implausibly large and negative, ranging from -1.5 to -1.8 (significant on aggregate and for 2008) for large enterprises. In all cases the sign and significance of the coefficient seems to be driven by values observed in 2008.

Looking at results by sector for the pooled 2007-08 dataset (Table 41), the coefficient of the training variable shows significant variation across different sectors, and it is only

[32] Looking at the full results using all measures of capital stock presented in table 38, it can be observed that the magnitude of the coefficients gets smaller (and more plausible) as we move to specifications with larger sample size (explained by the higher tolerance level used in the capital stock series).

statistically significant (and with a positive sign) for the Construction sector (both specifications) the Retail sector (specification controlling for capital stock only) and the Motor Trades sectors (specification controlling for capital expenditure only). However it is difficult to assess the robustness and reliability of the results.

Table 39: OLS estimates: Impact of training on productivity, by year

	Dependent Variable: log of GVA					
Year	2007-08	2007	2008	2007-08	2007	2008
% Trained	-0.158	1.147	-0.496	0.138	-0.079	0.165
ln CapStock 2	0.045***	0.082***	0.009			
ln Capex				0.109***	0.124***	0.084***
Size: 10-24	0.164***	0.186**	0.190**	0.153***	0.134***	0.172***
Size: 25-49	0.235***	0.221**	0.323***	0.104***	0.064**	0.148***
Size: 50-249	0.199***	0.246***	0.201***	0.049**	0.067**	0.010
Size: 250+	0.075	0.094	0.144*	-0.090***	-0.082**	-0.083*
Catering	-0.291***	-0.483***	-0.107	-0.188***	-0.362***	0.017
Motor Trades	-0.634***	-0.557***	-0.676***	-0.455***	-0.434***	-0.476***
Production	-0.282***	-0.406***	-0.163**	-0.162***	-0.204***	-0.117***
Property	0.248*	0.092	-1.746	0.087	-0.014	-0.172
Retail	-0.389***	-0.554***	-0.217**	-0.360***	-0.450***	-0.294***
Service Trades	0.077	-0.182***	0.308***	0.167***	-0.025	0.365***
Wholesale	-0.352***	-0.355***	-0.335***	-0.212***	-0.195***	-0.201***
Foreign owner	0.147***	0.135***	0.158***	0.162***	0.155***	0.147***
Multi-LUs	-0.004	0.017	-0.028	-0.003	0.013	-0.022
Software int	0.254**	0.330*	0.216	0.154**	0.104	0.239**
Female FT (%)	-0.019	-0.060	0.061	-0.100**	-0.184***	0.115
Female PT (%)	-2.084***	-1.742***	-2.549***	-1.868***	-1.531***	-2.408***
Male PT (%)	-2.399***	-2.055***	-2.855***	-2.382***	-1.852***	-3.514***
Obs	25740	13230	12510	63332	33446	29886
R-squared	0.208	0.221	0.206	0.165	0.173	0.169
Capital measure	*CapStock2*	*CapStock2*	*CapStock2*	*Capex*	*Capex*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 40: OLS estimates: Impact of training on productivity, by company size and year																		
	Dependent Variable: log of GVA																	
Sample	Size 0-49			Size 50-249			Size 250+			Size 0-49			Size 50-249			Size 250+		
Year	2007-08	2007	2008	2007-08	2007	2008	2007-08	2007	2008	2007-08	2007	2008	2007-08	2007	2008	2007-08	2007	2008
% Trained	0.943**	1.639	0.726***	-0.157	1.216***	-0.590	-1.820**	-0.386	-1.976**	0.504***	0.089	0.567***	0.007	-0.114	-0.011	-1.528**	-0.789	-1.634**
ln CapStock 2	-0.021	0.023	-0.076***	0.136***	0.151***	0.117***	0.099***	0.118***	0.077***									
ln Capex										0.112***	0.130***	0.077***	0.115***	0.113***	0.121***	0.106***	0.136***	0.071***
Catering	-0.260**	-0.643***	-0.010	-0.405***	-0.411***	-0.377	-0.149*	-0.139	-0.155	-0.210***	-0.417***	-0.072	-0.240***	-0.234**	-0.222	-0.124	-0.134	-0.105
Motor Trades	-0.749***	-0.756***	-0.793***	-0.579***	-0.297**	-0.758***	-0.460***	-0.439***	-0.478**	-0.456***	-0.483***	-0.473***	-0.428***	-0.136	-0.648***	-0.468***	-0.460***	-0.465**
Production	-0.297***	-0.607***	0.036	-0.294***	-0.260***	-0.331***	-0.170**	-0.141*	-0.203*	-0.149***	-0.229***	-0.051	-0.118***	-0.023	-0.273***	-0.115*	-0.095	-0.130
Property	-0.031	-0.214	-5.170***	-0.223	-0.168	-2.232	0.642***	0.622***	0.335**	-0.120	-0.249**	-0.020	0.112	0.211	-1.287	0.729***	0.715***	0.303**
Retail	-0.587***	-0.894***	-0.374**	-0.296***	-0.289**	-0.292*	-0.104	-0.086	-0.118	-0.418***	-0.519***	-0.398***	-0.210***	-0.197**	-0.234**	-0.121	-0.108	-0.122
Service Trades	0.283***	-0.136	0.599***	-0.083	-0.123	-0.022	-0.049	-0.085	-0.009	0.236***	0.005	0.478***	0.083*	0.086	0.059	-0.009	-0.062	0.048
Wholesale	-0.461***	-0.484***	-0.412***	-0.166*	-0.129	-0.195	-0.255***	-0.234*	-0.281*	-0.228***	-0.223***	-0.202***	-0.038	0.074	-0.183*	-0.257***	-0.246**	-0.266*
Foreign owner	-0.032	-0.027	-0.048	0.133***	0.142***	0.128*	0.125***	0.079*	0.184***	0.067	0.084	-0.017	0.167***	0.148***	0.197***	0.163***	0.106**	0.232***
Multi-LUs	0.006	0.058	0.003	0.001	0.022	-0.031	0.005	0.029	-0.020	-0.008	0.006	-0.016	-0.006	0.005	-0.027	0.031	0.042	0.020
Software int	-0.115	-0.237	-0.044	0.989***	0.820***	1.166***	0.681*	0.598	0.897	-0.012	-0.075	0.092	0.579***	0.446***	0.737***	0.185	0.068	0.599
Female FT (%)	0.011	-0.188	0.575**	-0.082	-0.066	-0.106	-0.225**	-0.177	-0.285*	-0.109*	-0.216***	0.292***	-0.112	-0.212**	0.076	-0.170	-0.130	-0.234
Female PT (%)	-2.908***	-1.989***	-4.279***	-1.709***	-1.763***	-1.536***	-1.608***	-1.546***	-1.681***	-2.012***	-1.467***	-3.012***	-1.892***	-1.898***	-1.794***	-1.551***	-1.570***	-1.541***
Male PT (%)	-2.901***	-2.030***	-4.422***	-2.175***	-2.049***	-2.378***	-2.116***	-2.076***	-2.163***	-2.495***	-1.704***	-4.594***	-2.216***	-2.145***	-2.370***	-2.205***	-2.091***	-2.317***
Obs	9550	4097	5453	8089	4818	3271	8101	4315	3786	41401	20976	20425	13254	7926	5328	8677	4544	4133
R-squared	0.129	0.126	0.161	0.183	0.221	0.143	0.364	0.355	0.376	0.122	0.124	0.144	0.179	0.209	0.149	0.353	0.354	0.354
Capital measure	*CapStock2*			*CapStock2*			*CapStock2*			*Capex*			*Capex*			*Capex*		

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 41: OLS estimates: Impact of training on productivity, by sector

	Dependent Variable: log of GVA													
Sample	Catering	Construction	Motor Trades	Production	Retail	Service Trades	Wholesale	Catering	Construction	Motor Trades	Production	Retail	Service Trades	Wholesale
Year	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08
% Trained	-0.231	1.600*	1.383	-0.002	0.802*	-0.725	-0.774	0.989	0.968***	1.243*	0.111	0.459	-0.335	0.395
ln CapStock 2	0.064	0.003	0.166***	0.108***	0.079	0.013	0.061							
ln Capex								0.127***	0.077**	0.166***	0.132***	0.157***	0.070***	0.104***
Size: 10-24	0.259*	0.139	-0.091	0.372**	0.587***	-0.078	0.590***	-0.023	0.207***	0.173*	0.206***	0.202***	0.069**	0.262***
Size: 25-49	0.702***	0.214	0.084	0.383**	0.866***	0.065	0.505**	0.195*	0.107	0.043	0.131**	0.255***	0.037	0.143**
Size: 50-249	0.843***	0.160	0.351	0.425***	0.849***	0.048	0.479**	0.228**	0.193**	0.255**	0.120**	0.261***	-0.034	0.095
Size: 250+	1.190***	0.163	0.543*	0.493***	0.906***	-0.249***	0.379	0.451***	0.292**	0.353**	0.187***	0.307***	-0.325***	-0.028
Foreign owner	-0.127	-0.168	-0.087	0.023	0.140	0.202***	0.099	0.030	-0.069	-0.008	0.050	0.187**	0.210***	0.062
Multi-LUs	-0.320**	0.130	-0.248*	0.026	-0.082	-0.065	-0.070	-0.117	0.013	-0.309***	-0.008	-0.146**	-0.032	-0.006
R&D intensity	-2.859	0.604	-1.325***	1.642***	0.197	0.370***	0.448	-0.859	0.508	-0.794*	0.284	-0.381	0.251***	0.433
Female FT (%)	-0.755	1.456***	0.855	-0.272**	-0.098	-0.159	0.505*	0.076	1.595***	-0.021	-0.169**	-0.208	-0.300***	0.198
Female PT (%)	-1.546***	-1.682***	-3.422***	-1.537***	-1.045***	-2.483***	-3.597***	-0.900***	-1.485***	-1.474***	-1.693***	-0.901***	-2.250***	-3.199***
Male PT (%)	-0.773**	-1.048**	-1.128**	-2.739***	-1.425***	-2.697***	-3.087***	-0.679***	-2.622***	-0.967***	-2.216***	-1.241***	-2.934***	-4.048***
Obs	912	1363	1066	6077	1782	10949	3242	2021	4715	2634	12445	5896	25459	9306
R-squared	0.191	0.096	0.112	0.098	0.123	0.295	0.122	0.139	0.078	0.066	0.106	0.084	0.243	0.106
Capital measure	*Cap stock2*	*Cap stock2*	*Cap stock2*	*Cap stock2*	*Cap stock2*	*Cap stock2*	*Cap stock2*	*Capex*	*Capex*	*Capex*	*Capex*	*Capex*	*Capex*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1. No robust estimates were available for the Property sector, due to the small sample size.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

4.3.5 Controlling for the lagged value of GVA and training

In Table 42 we present results (for 2008 only given data availability) controlling for the lagged values of the trained proportion and productivity. Linking two years of the ABI together implies a loss of observations especially in the band of micro and small firms, due to the survey design. Results show that current training has a positive impact on productivity for the group of small firms and some evidence that training has also a deferred impact on training for the same group of firms. However both coefficients seem to be implausibly large.

Table 42: Impact of training on productivity, controlling for lags – all and by company size (2008)

	Dependent Variable: log of GVA							
Sample	All	All	Size 0-49	Size 50-249	Size 250+	Size 0-49	Size 50-249	Size 250+
% Trained	0.290	0.441	2.400**	-0.367	-0.239	1.303***	-0.116	-0.057
L. % Trained	-0.133	-0.434	1.605*	-0.418	0.894	0.488	-0.464	-0.314
L. GVA	0.635***	0.608***	0.270***	0.495***	0.798***	0.372***	0.517***	0.796***
ln CapStock 2	0.015		-0.067	0.055*	0.045**			
ln Capex		0.041**				0.085**	0.054*	-0.000
Size: 50-249	-0.067	-0.062						
Size: 250+	-0.018	-0.052						
Catering	-0.199	-0.209*	-0.603	-0.387	-0.150	-0.360	-0.299	-0.036
Motor Trades	-0.490***	-0.455***	-0.885**	-0.688***	-0.259	-0.314	-0.633***	-0.231
Production	-0.167	-0.166*	-0.275	-0.284	-0.156	0.029	-0.294**	-0.038
Property	0.483***	0.441***		0.150			0.177	
Retail	-0.126	-0.174	-0.480	-0.245	-0.122	-0.202	-0.221	-0.002
Service Trades	-0.079	-0.036	-0.340	-0.093	-0.079	0.101	-0.033	0.042
Wholesale	-0.165	-0.172	-0.252	-0.156	-0.212	0.045	-0.223	-0.115
Foreign owner	0.065	0.084**	-0.229	0.048	0.087*	-0.068	0.081	0.115**
Multi-LUs	-0.005	-0.015	-0.196	-0.006	0.006	-0.114	-0.013	0.009
Software int	1.039***	0.732***	0.479	1.531***	1.041**	0.667	0.526***	1.100***
Female FT (%)	-0.088	-0.012	0.396	-0.082	-0.105	0.153	0.054	-0.141
Female PT (%)	-0.711***	-0.649***	-2.303***	-0.709**	-0.286***	-1.106***	-0.728***	-0.326***
Male PT (%)	-0.625***	-0.810***	-0.347	-1.578***	-0.410***	-1.084**	-1.460***	-0.516***
Obs	6237	8810	845	2137	3255	2595	2845	3370
R-Squared	0.533	0.493	0.147	0.315	0.770	0.214	0.333	0.760
Capital measure	*CapStock2*	*Capex*	*CapStock2*	*CapStock2*	*CapStock2*	*Capex*	*Capex*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status and company age

*** p<0.01, ** p<0.05, * p<0.1. L. indicates the first time lag of a variable. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

4.3.6 Instrumental Variable Approach

While the implementation of a full GMM system is not feasible, given the short longitudinal component available, we employed an instrumental variable approach, using the lagged

values of training and capital stock or capital expenditure (from 2007) to instrument the 2008 value. Clearly, the sample will only be formed by observations available in both 2007 and 2008. We only present data at aggregate level and disaggregating by company size[33]. The results using the entire sample (first two columns) are positive (magnitude between 0.40 and 0.49) but not statistically significant. When we disaggregate by company size the results assume implausible values (especially for smaller firms given the limited number of observations available in two consecutive years and are never statistically significant).

Table 43: IV approach: Impact of training on productivity, all and by company size (2008)

	Dependent Variable: log of GVA							
Sample	All	All	Size 0-49	Size 50-249	Size 250+	Size 0-49	Size 50-249	Size 250+
% Trained	0.402	0.486	6.981	-0.438	-0.696	3.411	0.169	-0.696
ln CapStock 2	0.085***		-0.019	0.133***	0.085***			
ln Capex		0.216***				0.296**	0.318***	0.124**
Size: 50-249	0.026	-0.074*						
Size: 250+	0.023	-0.151***						
Catering	0.293*	0.267**	1.130*	0.561*	0.083	0.092	0.528**	0.013
Motor Trades	-0.406***	-0.326***	-0.311	-0.274	-0.402**	0.080	-0.292	-0.477**
Production	0.013	0.058	0.351	0.186	-0.107	0.492***	0.118	-0.130
Property	1.540	1.349		0.964			0.771	
Retail	0.101	0.041	0.085	0.231	0.059	0.220	0.111	0.038
Service Trades	0.231**	0.347***	0.449	0.450*	0.106	0.700***	0.495***	0.118
Wholesale	0.114	0.152	0.496	0.406	-0.139	0.599***	0.270	-0.160
Foreign owner	0.293*	0.267**	1.130*	0.561*	0.083	0.092	0.528**	0.013
Multi-LUs	-0.036	-0.050	-0.171	-0.028	-0.017	-0.159*	-0.039	-0.013
Software int	1.259***	0.689***	0.291	2.412***	1.136**	-0.020	0.627**	1.030**
Female FT (%)	-0.258	-0.064	0.082	-0.240	-0.381*	0.138	0.114	-0.327
Female PT (%)	-1.736***	-1.650***	-2.457***	-1.521***	-1.645***	-1.585***	-1.454***	-1.665***
Male PT (%)	-2.111***	-2.049***	-1.349	-2.723***	-2.203***	-1.829***	-2.386***	-2.237***
Obs	6282	8878	863	2149	3270	2630	2863	3385
R-Squared	0.279	0.241	0.068	0.166	0.386	0.079	0.134	0.380
Capital measure	*CapStock2*	*Capex*	*CapStock2*	*CapStock2*	*CapStock2*	*Capex*	*Capex*	*Capex*

Note: 2SLS instrumenting present values of training and capital measure with lagged values. ABI data on observations available in 2007 and 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies. Capital Stock generated using the 2:1 tolerance level on missing values. *** p<0.01, ** p<0.05, * p<0.1. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

[33] Estimation at sector level resulted in implausible coefficients, probably explained by the restricted sample size and the short time frame available, causing significant "noise" when applying IV

4.3.7 Change in GVA over time

In the final section dedicated to the empirical analysis, we have looked at the change in GVA over time controlling for training undertaken during the year and the change in other relevant characteristics. All time-invariant characteristics would therefore disappear from the equation. Below we present summary statistics on the average change over time in GVA and the results of equation (1) estimated in differences. The first two columns of Table 44 presents the average and the median change over the period 2006-08, while the other columns give the detail by year (i.e. summary statistics for the change in GVA between 2006 and 2007 and 2007 and 2008 respectively). Changes are expressed as log differences and can therefore be interpreted as growth rates of GVA over time.

TTG firms experienced a positive productivity growth between 2006 and 2007, while the average and median change between 2007 and 2008 and on aggregate were negative. The pattern for non-TTG firms is quite similar at aggregate level. Looking at disaggregation by size, it seems that the change in GVA for large and medium firms in the TTG group was positive (especially for large firms) between 2006 and 2007, but negative between 2007 and 2008 (particularly for medium-size firms). For small firms in the TTG group, the average and median changes have opposite sign in both periods: over the period 2006-07, the average change is negative (-0.18), while the median change is positive (0.04); the reverse is true for the period 2007-08, with a positive average change (0.06) and a negative median change (-0.05). Compared to the non-TTG group, TTG firms experienced a lower average and growth rate in almost all cases, with the exception of large firms between 2006 and 2007 and small firms between 2007 and 2008.

Table 45 presents the average and median change in GVA by sector of industrial activity: values are notably different across the TTG and non-TTG firms for the Property sector (positive change for non-TTG and negative for TTG firms). The Motor Trades sector seems to have experienced the largest fall in productivity between 2007 and 2008 and on aggregate (stronger fall for TTG firms) while firms in the Catering and Service Trades sectors seem to have experienced a positive growth rate in productivity (with TTG firms outperforming non-TTG firms in both sectors in 2008).

Table 44: Change in GVA over time:summary statistics by year and company size

Company size	Change between 2006 and 2007 & 2007 and 20081				Change between 2006 and 2007				Change between 2007 and 2008			
	Mean		Median		Mean		Median		Mean		Median	
	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG
0-49	-0.06	-0.05	-0.01	-0.02	-0.03	-0.18	0.01	0.04	-0.1	0.06	-0.04	-0.05
50-249	-0.02	-0.07	0.01	-0.01	0.03	0	0.02	0.01	-0.09	-0.15	-0.02	-0.04
250+	0.01	0	0.02	0.01	0	0.04	0.04	0.03	0.02	-0.05	0.01	-0.01
Total	-0.02	-0.03	0.01	0	0	0.01	0.02	0.03	-0.06	-0.06	-0.02	-0.02

Note: Average difference in GVA over time.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 45:	Change in GVA over time:summary statistics by year and sector											
Sector	Change between 2006 and 2007 & 2007 and 2008				Change between 2006 and 2007				Change between 2007 and 2008			
	Mean		Median		Mean		Median		Mean		Median	
	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG	Non-TTG	TTG
Catering	0.02	0.03	0.01	-0.03	0.06	0	0.05	-0.01	-0.03	0.05	-0.02	-0.05
Construction	-0.01	0.01	0.02	0.06	0.05	0.13	0.02	0.06	-0.13	-0.12	0.02	0.04
Motor Trades	-0.16	-0.23	-0.06	-0.15	0.01	-0.01	-0.02	0.04	-0.33	-0.33	-0.14	-0.18
Production	-0.03	-0.04	0.01	-0.02	0	-0.03	0.02	0.02	-0.07	-0.05	-0.03	-0.07
Property	0.05	-0.13	-0.02	-0.05	-0.03	-0.07	0.01	-0.05	0.15	-0.16	-0.06	-0.05
Retail	-0.01	-0.04	0	-0.01	0	0.02	0.03	0.02	-0.02	-0.11	-0.03	-0.07
Service Trades	0.01	0.01	0.02	0.02	0.01	-0.02	0.03	0.03	0	0.03	0	0.01
Wholesale	-0.07	-0.09	0	-0.02	-0.03	0.2	0.02	0.05	-0.12	-0.38	-0.02	-0.08
Total	-0.02	-0.03	0.01	0	0	0.01	0.02	0.03	-0.06	-0.06	-0.02	-0.02

Note: Average difference in GVA over time.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

In Table 46 we present the results of a regression looking at GVA growth over time as the dependent variable and. All time invariant characteristics cancel out in the first-difference model and we control for the change over time for time-varying variables. The effect on training undertaken during the year on productivity growth seems to be large, positive and statistically significant in 2008 and in the pooled regression. The effect seems to be driven by a positive and large change for small companies. However results should be taken with extreme care and do not reflect any causal impact of training on productivity growth. The analysis suggests that an increase in the volume of training is associated with a positive growth rate in firm level productivity, at least for small firms.

Table 46: OLS estimates: change in GVA over time									
	Dependent Variable: change in log GVA								
	All	06/07	07/08	All	06/07	07/08	Size:1-49	Size:50-249	Size:250+
% Trained	0.754*	1.152	0.835*	0.533*	0.250	0.734**	1.515**	-0.189	0.266
D.lnempl	-0.514***	-0.523***	-0.508***	-0.635***	-0.652***	-0.615***	-0.503***	-0.804***	-0.623***
D.lncap	0.122**	0.127**	0.115						
D.ln capex				0.022**	0.023*	0.020	0.072***	-0.002	0.003
D.R&D intensity	0.518***	0.587***	0.446**	0.402***	0.428***	0.373***	0.515**	0.314***	0.371**
L.Female FT (%)	-0.159	-0.034	-0.264**	-0.053	0.043	-0.155	0.044	-0.056	-0.249
L.Female PT (%)	-0.218*	-0.119	-0.324*	-0.114	-0.072	-0.146	-0.070	-0.344***	0.099
L.Male PT (%)	-0.189	-0.176	-0.186	-0.089	0.095	-0.352	0.068	-0.491**	0.056
Obs	13742	7505	6237	19838	11028	8810	6172	6829	6837
R-Squared	0.024	0.031	0.019	0.025	0.032	0.019	0.019	0.043	0.024

Note: ABI data available in at least two consecutive years between 2006 and 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies. Capital Stock generated using the 2:1 tolerance level on missing values. *** p<0.01, ** p<0.05, * p<0.1. All: change between 2006 and 2007 and 2007 and 2008; 06/07: change between 2006 and 2007; 07/08:change between 2007 and 2008.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

5 Conclusions

The absence of a firm-level dataset containing information on firm characteristics as well as information on training undertaken has historically hindered the possibility of performing firm-level analysis on the impact of training on productivity for UK firms. In fact, the most comprehensive source reporting training undertaken at firm level (the National Employer Skills Survey) does not follow firms over time, while the Annual Respondents Database, which contains information on financial variables and other firm-level characteristics collected in the Annual Business Inquiry does not include any variable related to training or skills.

In this study we have used a novel dataset containing administrative data from the Individualised Learner Record, aggregated at firm level thanks to an employer identifier and matched with the Inter-Departmental Business Register. Training data refer to training undertaken through the *Train to Gain* programme between 2007 and 2010. Training information was then matched to financial and other data contained in the Annual Respondents Database 2007-2008 to perform firm-level analysis of the impact of training on productivity.

However, a series of data and methodological issues (including a shift in sample composition and the inability to control for other training undertaken) may affect the empirical results and therefore the evidence presented should be considered with extreme caution (see section 4.2.5 for more details). In particular no causal interpretation may be given to the results given the presence of time and firm specific fixed effects and the potential endogeneity of training decisions.

In general, the reported results seem to vary substantially across specifications used and are normally of a magnitude in line with previous studies only when using the pooled dataset, although the coefficients are never statistically significant. Disaggregation by size seems to suggest that for small firms an increase in the proportion of employees receiving training is associated with an increase in productivity, while a negative coefficient was observed for large firms. Results by sector of industrial activity show some evidence of a positive association between training and productivity for the construction sector and the motor trades and the retail sectors (however the magnitude of the effects seems to be quite large compared with previous estimates at firm-level).

The *Train to Gain* programme ran from 2006 to 2011. Currently data on training undertaken are available up to the academic year 2010, while data on productivity and other firm-level characteristics are only available in the Annual Respondents Database up to 2008. Extending the analysis to 2011, adding information on training for the academic year 2010-11 and ABI survey data up to 2011 would yield more robust estimates and allow for further disaggregation. A 5-year panel dataset (covering the period 2007-2011) would also allow the use, at least to some extent, of panel-data econometric techniques, controlling for the presence of time or firm-specific components and endogeneity. Furthermore, a significant proportion of firms engaging in *Train to Gain* activities appear to train a very small proportion of their employees. While this might be explained by a variety

of factors[34], it might also undermine the possibility to detect a significant impact of training on productivity in the presence of "noise" correlated with training activities and productivity. Looking at firms over time should allow controlling for unobserved time-invariant firm characteristics[35]. Also, while in the short term, firms may adjust their training strategy in response to the *Train to Gain* initiative (in a way we cannot observe due to the lack of information on other training undertaken), using the 5-year dataset allows us to observe changes over time after the initial adjustment. Finally, a 5-year panel dataset (covering the period 2007-2011) would also allow the use, at least to some extent, of panel-data econometric techniques, controlling for the presence of time invariant effects and endogeneity. It will be possible to construct this data set when the ABI for 2011 becomes available (probably in late 2013 or 2014)[36].

An additional consideration when looking at training over time is that firms will typically use a combination of three types of government funded training: apprenticeships; *Train to Gain* and other work place training. Firm identifiers were collected for apprenticeships from 2009/10 and for other workplace training (elsewhere referred to as non-TTG) from 08/09 (see Table 1). As such, employer details for all three programmes were first recorded in 09/10 so data for 4 calendar years first becomes available in 13/14, i.e. the calendar years 2010 to 2013. Basing the analysis on ILR data between 09/10 and 13/14 should also improve the match rate because stricter quality assurance of employer identifiers was introduced in 09/10. Prior to 09/10 providers has the option of submitting employer information using free text, this was often incomplete and difficult to match to the IDBR. From 09/10 they could only use the blue sheep employer identification number which linked back to the blue sheep data base provides full employer information which is much easier to match to the IDBR. We therefore recommend further analysis of all three programmes when ABI 2013 becomes available (probably in late 2015 or 2016).

[34] For example the low matching ratio to the IDBR, the fact that we are forced to use the reporting unit level for productivity analysis and the fact that the Train to Gain may have not fully reached all recipients by the end of 2008

[35] Obviously it is possible that the effect of training on productivity is negligible and the "true" value of the coefficient is close to zero.

[36] Starting from the ABI 2009, there will be some significant changes to the structure of the ABI survey: employment data will be collected at the local unit level (BRES survey) and finance data at reporting unit level (ABS survey), using 2 separate surveys with slightly different reference points. While this might potentially add another level of complexity to the analysis we would expect to be still possible to derive the total employment data for the reporting unit (and possibly even to use better imputation methods to allocate financial information to local units).

Table 47: History of inclusion and quality of the Employer Identifier (A44) in the ILR					
	06/07	07/08	08/09	09/10	10/11
TTG	Required	Required	Required	Required + some validation	Required + strict Validation
Other workplace Training			Required	Required + strict Validation	Required + strict Validation
Apprenticeships				Required + some validation	Required + strict Validation

Source: Review of ILR Specification 08/09-10/11, http://www.theia.org.uk/ilr/ilrdocuments/

While collecting ILR data at employer level is a useful exercise to identify training patterns, particular care should be paid to ensure that the quality and coverage of the information gathered is maximised.

Key improvements to the ILR data collected at employer level are possible in the following areas:

- Ensuring that the A44 field is correctly filled by learning providers with a valid entry;
- An increased reliability of the A44 identifier should, in turn, lead to a more successful match of entries to the IDBR: in fact, the matching conducted by the ONS between ILR/EDS firm information and the IDBR was around 50%, considerably lower than other projects matching firm data to the IDBR (with match rates between 65 and 70%).
- Extending the coverage, requiring learning providers (or awarding bodies) to record a firm identifier for all courses undertaken through the employer (whether publicly funded, co-funded or employer funded). The coverage has already been extended over the years for publicly funded training, with firm identifiers being collected for apprenticeships from 2009/10 and for other workplace training (formerly referred to as non-TTG) from 2008/09.

Training data could then be matched to the IDBR and the ABI for analysis. Revisiting the matching in future work would therefore be advisable. Taken together we believe these changes will reduce the compositional bias and provide a larger sample for analysis and a clearer picture on different forms of training undertaken. Clearly there might still be other factors we are unable to control for, such as firm's product market strategy, skills structure and internal firm organisation. Finally, it should be noted that linking multiple years of the ABI together would result in a loss of all micro firms and a substantial loss of small firms, due to the survey design.

While we believe enhancing the quality and coverage of training data collected through the ILR may be a viable route, alternative or complementary approaches could be used to build a firm-level panel dataset covering training undertaken and including data on productivity and other firm-level characteristics. These include the following:

- Introduce a panel component in the National Employer Skills Survey: currently the NESS surveys around 80,000 English firms, but does not follow firms over time so in its current form its potential for panel techniques is more limited than the ILR/EDS

but it does contain information on training undertaken independent of government funding. Following a representative sample of firms over time would provide firm-level evidence on both privately and publicly funded training undertaken over several time periods. Also, given that the NESS questionnaire is currently administered to surveyed firms every two years, it would be useful to send annual follow ups, related to training activities only, to firms forming the panel database. Finally, the current match rate between the NESS and the IDBR is around 50%: improving the match between the NESS and the IDBR (and hence the ABI) is another key enhancement that could be undertaken.

- Include some questions on training activities undertaken and skills structure, at least for a random sample, in the ABI questionnaire enquiring on employment and staff.

- Potentially consider also using the FAME database as an alternative source for productivity data. While FAME database has inherent limitations and would not solve the lack of information on training, it has been used for productivity analysis in the past.

6 References

Colombo, E, & Stanca, L. (2008). ‘The Impact of Training on Productivity: Evidence from a Large Panel of Firms’, Working Papers 134, University of Milano-Bicocca, Department of Economics.

Dearden, L., Reed, H. & Reenen, J.V. (2005). 'The Impact of Training on Productivity and Wages: Evidence from British Panel Data', IFS Working Papers W05/16,

Galindo-Rueda, F. & Haskel, J. (2005). 'Skills, Workforce Characteristics and Firm-Level Productivity: Evidence from the Matched ABI/Employer Skills Survey', SSRN eLibrary.

Konings, J. & Vanormelingen, S., (2010). ‘The Impact of Training on Productivity and Wages: Firm Level Evidence’, IZA Discussion Papers 4731, Institute for the Study of Labor (IZA).

Annex 1

A1.1 Summary of recent literature

Table 48: Comparison of recent related literature

	Dearden et al. (2005)	Colombo & Stanca (2008)	Konings & Vanormelingen (201)
Main Findings: Impact of a 1ppt increase in the percentage of employees trained on productivity	Value added 0.6% Wages 0.35%	Value added 0.07% Wages 0.04%	Value added 0.23% Wages 0.11%
Further results		- Looking at type of worker, effect for blue collars is strong and significant (0.13%), small and not significant for white collars - Controlling for effective training (considering training length) rather than training intensity only indicate a slightly stronger effect of training	- Estimates drop to 0.17% (productivity) and 0.11% (wages) when controlling for worker heterogeneity - Effect of training on productivity slightly higher for the non-manufacturing sector, the reverse is true for wages. - Each additional hour of training increases productivity by 0.004% and wages by 0.003%.
Country	UK	Italy	Belgium
Period considered	1983-1996	2002-2005	1997-2006
Unit of analysis	Industry level (94 industries)	Firm-level (ca 11,000 Italian firms)	Firm-level (ca 170,000 Belgian firms)
Spillovers in-scope	Intra - firm and intra - industry	Intra - firm	Intra - firm
Industry Coverage	Production sectors only	Excludes financial sectors	All
Training Coverage	Training or education in last 4 weeks relating to job or job might be able to do in future (not necessarily employer provided].	Number of employees undertaking some form of training.	Number of employees undertaking some kind of formal training. Excludes training undertaken at the work floor and self-study.
% of firms providing training		71.5%	9.1%
Trained proportion (avg)	0.14 (post-'90)	0.28	0.53*
Length of training (avg)	14 days (median)	8.85 days	39.8 hours*
Econometric approach and estimators used	RE, FE, GMM (using t-2 and t-3 lags as instruments)	OLS, RE, FE, GMM (using t-2 lags as instruments)	OLS and non-parametric estimation. (after Ackerberg et al (2006))
Consistency of data sources	Productivity and wage from ACOP (ABI forerunner), training and hours worked from LFS	Training data from a survey of Italian firms (Excelsior), merged with annual company account data provided by Bureau Van Dijck (AIDA)	Annual income statement available through the Belfirst database (commercialised by Bureau Van Dijck). Includes training information

Note: * relative to firms providing training only
Source: London Economics based on cited authors

A1.2 Matching process and related analysis

Table 49: Involvement in TTG by sector using the IIR/EDS/BSD dataset and the NESS			
Source			
ILR/EDS matched to BSD 2009		NESS 2009	
Sector	*%*	*Sector*	*%*
Health and social work	10.9%	Education	32.0%
Education	8.4%	Health and Social Work	28.0%
PA and Defence	6.7%	Electricity, Gas and Water	17.0%
Electricity, gas and Water	5.3%	PA and Defence	17.0%
Manufacturing	2.7%	Manufacturing	11.0%
Construction	2.2%	Construction	11.0%
Transport Storage & Communications	2.1%	Transport Storage & Communications	10.0%
Hotels and Catering	1.9%	Other Services	10.0%
Wholesale and Retail	1.3%	Hotels and Catering	9.0%
Other Services	1.2%	Business Services	9.0%
Financial Intermediation	0.8%	Retail and Wholesale	7.0%
Business Services	0.6%	Financial Intermediation	6.0%
Total	1.7%	Total	11.0%

Source: London Economics analysis of the EDS using BSD data and NESS2009 report

Detailed characteristics of firms matched in the BSD by year

Enterprise size

Table 50 presents summary information on company size (measured by total number of employees), using the BSD 2008 to 2010. Around 55% of enterprises that engaged in publicly funded training and were identified in the BSD have less than 25 employees, while about 15% have between 25 and 50 employees, 18% have more than 50 and up to 250 employees and around 11% are very large enterprises (more than 250 employees).

As shown in Table 50, there is variation in the distribution of enterprises by size for the bottom and top categories: in fact, there is a lower proportion of micro companies (less than 10 employees) in 2007 compared with the following years (25% vs. 35% in 2009 and 2010) and a higher proportion of large companies in 2007 compared with the following years (more than 16% in 2007, declining to 10% in 2009 and 11% in 2010).

Table 50: Enterprise size					
	Year				
	BSD 2007	BSD 2008	BSD 2009	BSD 2010	Total
0-9	24.8%	32.0%	35.3%	34.7%	33.5%
10-24	20.7%	21.7%	22.7%	20.7%	21.7%
25-49	17.0%	15.8%	14.8%	14.4%	15.1%
50-249	21.2%	18.5%	17.1%	18.9%	18.3%
250+	16.4%	12.0%	10.0%	11.4%	11.4%
Total	8,875	23,822	39,605	33,383	105,685

Source: London Economics analysis of the EDS using BSD data

Company age

In Table 51 we show the distribution by company age: around 37% of enterprises engaging in the programme are relatively young companies (established for less than 10 years), while another 36% are aged between 10 and 20 years, and a further 27% of enterprises were founded more than 20 years prior to participation in the programme. There is some variation between 2008 and 2009, with an higher proportion of companies in the 11 to 15 year age bracket in 2008 (23% vs. 17% in 2009) and a lower proportion of companies in the 16 to 20 year age bracket in 2008 (14% vs. 18% in 2009)

Table 51: Enterprise age

	Year				
	BSD 2007	BSD 2008	BSD 2009	BSD 2010	Total
0-5 years	17.2%	18.8%	18.8%	18.0%	18.4%
6-10	17.2%	17.6%	19.1%	18.8%	18.5%
11-15	24.8%	22.8%	16.8%	16.2%	18.6%
16-20	12.8%	13.9%	18.5%	19.7%	17.4%
20+	28.1%	26.9%	26.8%	27.3%	27.1%
Total	9,010	24,284	39,706	33,956	106,956

Source: London Economics analysis of the EDS using BSD data

Region of location

Table 52 presents details on region of location. Around 15% of the enterprises are located in the North West, and a slightly smaller proportion is located in the South East. Apart from the North East (where approximately 5% of the companies are located), the proportion of enterprises located in each of the other regions is between 10-12%. Furthermore, data presented in Table 52 do not show any notable variation in the relative proportions across years.

Table 52: Enterprise region of location

	Year				
	BSD 2007	BSD 2008	BSD 2009	BSD 2010	Total
North East	5.5%	5.0%	5.6%	5.6%	5.5%
North West & Merseyside	16.4%	15.6%	15.1%	15.7%	15.5%
Yorkshire & Humber	10.2%	9.9%	10.0%	10.2%	10.1%
East Midlands	9.9%	9.7%	10.7%	10.7%	10.4%
West Midlands	11.8%	12.0%	11.6%	11.7%	11.8%
East of England	10.4%	10.3%	10.2%	9.9%	10.1%
London	10.0%	12.3%	11.3%	10.9%	11.3%
South East	14.7%	14.4%	14.6%	14.4%	14.5%
South West	11.1%	10.8%	10.8%	10.9%	10.8%
Total	9,010	24,284	39,706	33,956	106,956

Source: London Economics analysis of the EDS using BSD data

Industrial Classification

Summary data on sector of activity (based on SIC codes) are shown in Table 53. The sectors with the highest proportion of enterprises is Health and Social care activities (at 24%), followed by Construction, Wholesale and Retail Trade, Business Services and Manufacturing (all between 10% and 15%). There is some variation over time, with the

Manufacturing, Construction and Health and Social care activities sector representations declining between 2-3.5 percentage points in the period 2007 - 2010.

Table 53: Enterprise sector of activity

	Year				
	BSD 2007	BSD 2008	BSD 2009	BSD 2010	Total
Agriculture	0.8%	0.9%	1.2%	1.1%	1.0%
Mining and Quarrying	n.a.	0.1%	0.1%	0.1%	0.1%
Manufacturing	12.3%	11.8%	10.8%	10.2%	11.0%
Electricity/Gas/Water	n.a.	0.1%	0.1%	0.1%	0.1%
Construction	15.3%	17.4%	14.4%	13.1%	14.7%
Wholesale/Retail	12.6%	12.3%	12.7%	12.2%	12.4%
Hotels /restaurants	6.2%	6.2%	6.8%	7.0%	6.7%
Transport/Storage	3.5%	4.2%	5.0%	4.6%	4.6%
Financial/Insurance	0.8%	0.6%	0.8%	0.8%	0.7%
Business Services	11.4%	11.6%	12.7%	11.9%	12.1%
PA and Defence	0.7%	0.6%	0.5%	0.6%	0.6%
Education	4.0%	4.9%	6.2%	7.0%	6.0%
Health/ Social Work	28.3%	24.3%	22.9%	24.8%	24.3%
Other services	4.0%	4.9%	5.8%	6.5%	5.7%

Source: London Economics analysis of the EDS using BSD data

Legal Status

Turning to the analysis of legal status, the vast majority of enterprises are recorded as "Company" (around 70%), while Sole Proprietors, Partnerships and Non-for-profit organisations all account for approximately around 10% of firms.

Table 54: Enterprise status

	Year				
	BSD 2007	BSD 2008	BSD 2009	BSD 2010	Total
Company	71.5%	70.9%	68.2%	67.8%	69.0%
Sole proprietor	7.4%	8.4%	9.6%	8.9%	8.9%
Partnership	10.1%	9.0%	8.7%	8.1%	8.7%
Public corporation	0.2%	0.2%	0.2%	0.3%	0.2%
Central government body	1.3%	1.0%	0.9%	1.1%	1.0%
Local authority	1.8%	1.4%	1.2%	1.5%	1.4%
Non-profit making body	7.7%	9.1%	11.3%	12.2%	10.8%
Total	9,010	24,284	39,706	33,956	106,956

Source: London Economics analysis of the EDS using BSD data

A1.2.1 Further analysis related to the training variable

Mismatch of training proportions occurring in the BSD and the ABI

Table 55: Comparing total number of employees receiving training (ILR\EDS) and total employment (BSD)

Case	Occurrence	%
1) Total employment>trained employees	94720	87.6%
2) Total employment=trained employees	5247	4.9%
3) Total employment<trained employees	8165	7.6%
Average mismatch in 3)	*11.6*	
Median mismatch in 3)	*3.0*	
Observations kept in 3) accounting for minor mismatches	*4033*	

Source: London Economics analysis of the EDS/ILR dataset matched with the BSD

Table 56: Comparing total number of employees receiving training (ILR\EDS) and total employment (ABI)

Case	Occurrence	%
4) Total employment>trained employees	4869	99.0%
5) Total employment=trained employees	16	0.3%
6) Total employment<trained employees	33	0.7%
Observations kept in 3) accounting for minor mismatches	*14*	

Source: London Economics analysis of the EDS/ILR dataset matched with the BSD and the ABI

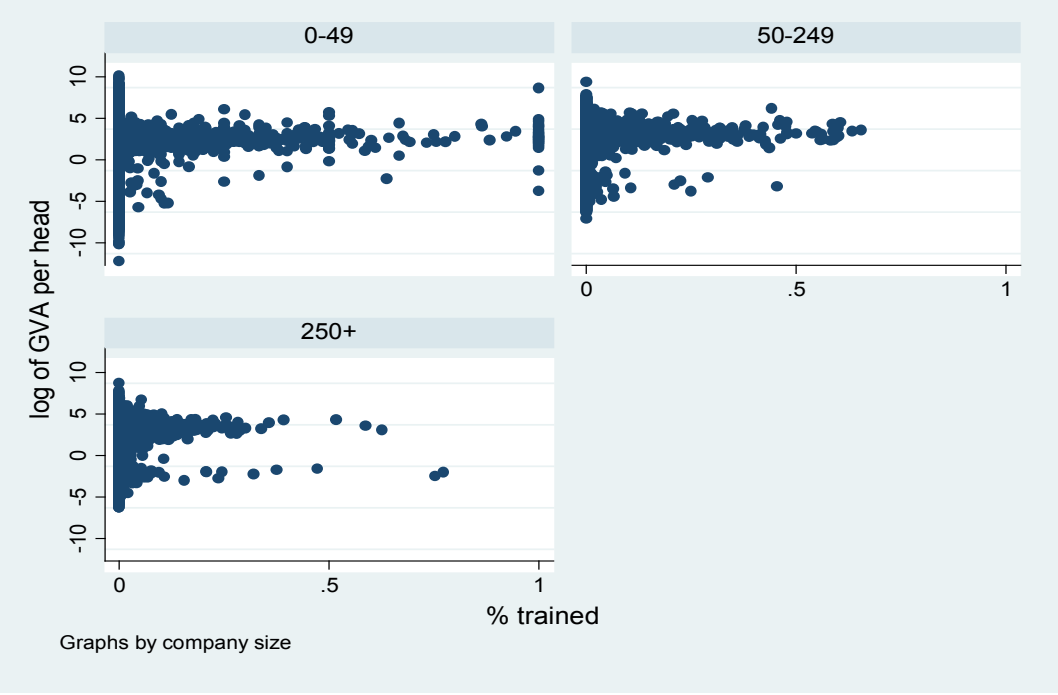
GVA and trained proportion

Figure 8: GVA and trained proportion by year, 0%<proportion trained<100%

Note: Pooled ABI 2007-2008, England only; all measures are divided by total employment and expressed in logarithmic terms.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Figure 9: GVA and trained proportion by company size, 0%<proportion trained<100%



Note: Pooled ABI 2007-2008, England only; all measures are divided by total employment and expressed in logarithmic terms.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 57: TTG Training characteristics by firm size and sector of activity– ABI sample																
	CA	CN	MT	PD	PR	RE	ST	WH	CA	CN	MT	PD	PR	RE	ST	WH
	All								Trained only							
Firm Size	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
1-49	0.2%	0.8%	0.3%	0.2%	0.1%	0.1%	0.2%	0.2%	12.3%	28.0%	11.3%	16.2%	32.0%	18.9%	20.8%	14.0%
50-249	0.4%	2.1%	0.8%	0.2%	0.6%	0.6%	0.5%	0.5%	3.4%	8.9%	4.4%	7.2%	2.1%	5.9%	6.2%	5.4%
250+	0.4%	1.5%	0.8%	0.4%	0.2%	0.2%	0.5%	0.5%	1.0%	3.6%	0.9%	3.0%	1.5%	0.4%	2.3%	2.0%
Total	0.3%	1.1%	0.5%	0.2%	0.2%	0.2%	0.3%	0.3%	2.6%	12.1%	4.2%	6.6%	4.8%	3.2%	5.5%	6.2%

Note: When cell size was smaller than 10 the value was removed (na) and the total rescaled to the value of the remaining "valid" cells. Avg= average proportion of employees receiving TTG training

CA Construction; CN Catering; MT Motor Trades; PD Production; PR Property; RE Retail; ST Service Trades; WH Wholesale .

Source: London Economics analysis of the EDS and ABI data

Annex 2 Further econometric results

Table 58: OLS estimates: Impact of training on productivity,different measures of capital stock												
	Dependent Variable: log of GVA											
Year	2007-08	2007	2008	2007-08	2007	2008	2007-08	2007	2008	2007-08	2007	2008
% Trained	-0.252	1.753*	-0.844*	-0.158	1.147	-0.496	0.018	0.745	-0.165	0.138	-0.079	0.165
ln CapStock 1	0.042***	0.091***	0.007									
ln CapStock 2				0.045***	0.082***	0.009						
ln CapStock 3							0.041***	0.078***	0.001			
ln Capex										0.109***	0.124***	0.084***
Size: 10-24	0.263***	0.337***	0.265**	0.164***	0.186**	0.190**	0.129***	0.147**	0.141**	0.153***	0.134***	0.172***
Size: 25-49	0.311***	0.315**	0.401***	0.235***	0.221**	0.323***	0.164***	0.139*	0.249***	0.104***	0.064**	0.148***
Size: 50-249	0.306***	0.394***	0.305***	0.199***	0.246***	0.201***	0.122***	0.157**	0.112*	0.049**	0.067**	0.010
Size: 250+	0.180**	0.231**	0.259***	0.075	0.094	0.144*	0.003	0.018	0.047	-0.090***	-0.082**	-0.083*
Catering	-0.194**	-0.335***	-0.096	-0.291***	-0.483***	-0.107	-0.241***	-0.458***	-0.019	-0.188***	-0.362***	0.017
Motor Trades	-0.652***	-0.560***	-0.724***	-0.634***	-0.557***	-0.676***	-0.606***	-0.565***	-0.617***	-0.455***	-0.434***	-0.476***
Production	-0.275***	-0.412***	-0.181*	-0.282***	-0.406***	-0.163**	-0.262***	-0.406***	-0.110	-0.162***	-0.204***	-0.117***
Property	0.341*	0.173	-0.976	0.248*	0.092	-1.746	0.228*	0.059	-1.663	0.087	-0.014	-0.172
Retail	-0.320***	-0.437***	-0.231**	-0.389***	-0.554***	-0.217**	-0.406***	-0.585***	-0.207**	-0.360***	-0.450***	-0.294***
Service Trades	0.117*	-0.155**	0.307***	0.077	-0.182***	0.308***	0.109**	-0.153***	0.358***	0.167***	-0.025	0.365***
Wholesale	-0.387***	-0.351***	-0.426***	-0.352***	-0.355***	-0.335***	-0.291***	-0.327***	-0.241***	-0.212***	-0.195***	-0.201***
Foreign owner	0.139***	0.094**	0.189***	0.147***	0.135***	0.158***	0.146***	0.118***	0.174***	0.162***	0.155***	0.147***
Multi-LUs	-0.046	-0.017	-0.075	-0.004	0.017	-0.028	0.005	0.018	-0.008	-0.003	0.013	-0.022
R&D intensity	0.167	0.378	0.081	0.254**	0.330*	0.216	0.263***	0.288*	0.263**	0.154**	0.104	0.239**
Female FT (%)	-0.013	-0.054	0.062	-0.019	-0.060	0.061	-0.106*	-0.175**	0.054	-0.100**	-0.184***	0.115
Female PT (%)	-2.083***	-1.700***	-2.551***	-2.084***	-1.742***	-2.549***	-1.997***	-1.661***	-2.496***	-1.868***	-1.531***	-2.408***
Male PT (%)	-2.651***	-2.346***	-2.930***	-2.399***	-2.055***	-2.855***	-2.489***	-2.071***	-3.128***	-2.382***	-1.852***	-3.514***
Obs	16863	8282	8581	25740	13230	12510	32048	16801	15247	63332	33446	29886
R-squared	0.240	0.258	0.234	0.208	0.221	0.206	0.194	0.205	0.193	0.165	0.173	0.169
Capital measure	*CapStock 1*	*CapStock 1*	*CapStock 1*	*CapStock 2*	*CapStock 2*	*CapStock 2*	*CapStock 3*	*CapStock 3*	*CapStock 3*	*Capex*	*Capex*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 59:	OLS estimates: Impact of training on productivity - correctly identified					
	Dependent Variable: log of GVA					
Year	2007-08	2007	2008	2007-08	2007	2008
% Trained	-0.039	1.372*	-0.510	0.126	0.510	-0.015
ln CapStock 2	0.098***	0.112***	0.078***			
ln Capex						
Size: 10-24	-0.535**	-0.582*	-0.413	-0.209**	-0.267**	-0.134
Size: 25-49	-0.345*	-0.453	-0.123	-0.259***	-0.384***	-0.068
Size: 50-249	-0.402**	-0.426	-0.323	-0.357***	-0.420***	-0.271*
Size: 250+	-0.489**	-0.487*	-0.422	-0.485***	-0.499***	-0.451***
Catering	-0.268***	-0.301**	-0.223	-0.234***	-0.288***	-0.148
Motor Trades	-0.354***	-0.258	-0.447***	-0.322***	-0.205*	-0.450***
Production	-0.259***	-0.206**	-0.313***	-0.174***	-0.116**	-0.243***
Property	0.438	0.401		0.455*	0.415	
Retail	-0.171*	-0.160	-0.171	-0.320***	-0.280***	-0.366***
Service Trades	-0.167**	-0.182*	-0.143	-0.095*	-0.129*	-0.045
Wholesale	-0.178**	-0.137	-0.227*	-0.118**	-0.098	-0.144*
Foreign owner	0.142***	0.135***	0.149**	0.158***	0.138***	0.179***
Multi-LUs	-0.018	-0.031	-0.006	-0.001	-0.010	0.012
R&D intensity	2.172***	2.251***	2.102***	1.129***	1.165***	1.132***
Female FT (%)	-0.344***	-0.331**	-0.372**	-0.212**	-0.215*	-0.193
Female PT (%)	-1.586***	-1.577***	-1.595***	-1.544***	-1.563***	-1.472***
Male PT (%)	-2.334***	-2.264***	-2.405***	-2.138***	-1.943***	-2.487***
Obs	6559	3651	2908	10143	5811	4332
R-squared	0.315	0.308	0.329	0.271	0.262	0.286
Capital measure	*CapStock2*	*CapStock2*	*CapStock2*	*Capex*	*Capex*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1. Correctly identified: observations having a similar value for company size according to the EDS and BSD. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 60:	OLS estimates: Impact of training on productivity – sequential year	
	Dependent Variable: log of GVA	
Year	2007-08	2007-08
% Trained	-0.557	-0.063
ln CapStock 2	0.132***	
ln Capex		0.149***
Size: 10-24	-1.110**	-0.134
Size: 25-49	-1.126**	-0.278
Size: 50-249	-1.127**	-0.282
Size: 250+	-1.222***	-0.373*
Catering	-0.363***	-0.246***
Motor Trades	-0.402***	-0.324***
Production	-0.431***	-0.254***
Property	1.198***	1.410***
Retail	-0.405***	-0.277***
Service Trades	-0.277***	-0.159**
Wholesale	-0.425***	-0.231***
Foreign owner	0.113**	0.164***
Multi-LUs	-0.071	-0.060
R&D intensity	1.387***	0.418
Female FT (%)	-0.348**	-0.353***
Female PT (%)	-1.126***	-1.207***
Male PT (%)	-2.083***	-2.038***
Training year 2	0.175***	0.117**
Obs	3098	4288
R-squared	0.329	0.296
Capital measure	*CapStock2*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1. Sequential year: the sequential year of a firm's participation in TTG activities. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 61:	OLS estimates: Impact of training on productivity – controoling for 3 digit SIC codes					
	Dependent Variable: log of GVA					
Year	2007-08	2007	2008	2007-08	2007	2008
% Trained	0.027	1.205	-0.296	0.245	0.042	0.259*
ln CapStock 2	0.062***					
ln Capex				0.118***	0.129***	0.096***
Size: 10-24	0.172***	0.182**	0.183**	0.164***	0.143***	0.179***
Size: 25-49	0.287***	0.259***	0.361***	0.131***	0.097***	0.162***
Size: 50-249	0.293***	0.315***	0.294***	0.116***	0.133***	0.059*
Size: 250+	0.233***	0.225**	0.300***	0.072**	0.059	0.087*
Foreign owner	0.093***	0.078**	0.114**	0.118***	0.107***	0.110***
Multi-LUs	-0.061**	-0.036	-0.082**	-0.043**	-0.024	-0.064**
R&D intensity	0.180*	0.269	0.158	0.089	0.017	0.192*
Female FT (%)	-0.023	0.015	-0.045	-0.031	-0.125**	0.187**
Female PT (%)	-1.869***	-1.478***	-2.398***	-1.704***	-1.328***	-2.320***
Male PT (%)	-2.002***	-1.613***	-2.536***	-2.094***	-1.546***	-3.256***
Obs	25740	13230	12510	63332	33446	29886
R-squared	0.250	0.269	0.247	0.193	0.205	0.197
Capital measure	*CapStock2*	*CapStock2*	*CapStock2*	*Capex*	*Capex*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 62:	OLS estimates: Impact of training on productivity – all RUs			
	Dependent Variable: log of GVA			
Assumption on training at RU level	Use the trained proportion for all RUs		Use the trained proportion only for single RU enterprises or for multi-RU enterprises where at least 50% of training was assigned to RUs	
Year	2007-08		2007-08	
% Trained	-0.064	0.155	-0.141	0.133
ln CapStock 2	0.062***		0.062***	
ln Capex		0.116***		0.116***
Size: 10-24	0.107*	0.146***	0.111**	0.147***
Size: 25-49	0.177***	0.100***	0.180***	0.100***
Size: 50-249	0.152***	0.043**	0.160***	0.045**
Size: 250+	0.050	-0.084***	0.055	-0.084***
Catering	0.338***	0.203***	0.323***	0.199***
Motor Trades	-0.388***	-0.291***	-0.384***	-0.288***
Production	0.034	0.037	0.036	0.039
Property	0.639***	0.333***	0.620***	0.320***
Retail	-0.086	-0.169***	-0.087	-0.168***
Service Trades	0.344***	0.342***	0.354***	0.347***
Wholesale	-0.035	-0.018	-0.028	-0.014
Foreign owner	0.092***	0.135***	0.092***	0.135***
Multi-RUs	-0.044	0.029	-0.067	0.010
R&D intensity	0.252**	0.147**	0.279**	0.160**
Female FT (%)	-0.021	-0.096**	-0.022	-0.097**
Female PT (%)	-2.014***	-1.832***	-2.032***	-1.838***
Male PT (%)	-2.286***	-2.319***	-2.328***	-2.336***
Obs	28,698	66,419	28,310	66,018
R-squared	0.194	0.160	0.195	0.161
Capital measure	*CapStock2*	*Capex*	*CapStock2*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1. RUs: using all RUs, including RUs belonging to multi-RU enterprises. Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

Table 63:	OLS estimates: Impact of training as a binary variable on productivity, by year and size											
	Dependent Variable: log of GVA											
Year	2007-08	2007	2008	2007-08	2007	2008	2007-08	2007-08	2007-08	2007-08	2007-08	2007-08
Sample	All	All	All	All	All	All	0-49	50-249	250+	0-49	50-249	250+
Trained (Y/N)	0.032	0.100**	-0.001	0.033	0.036	0.045	0.143	0.010	-0.002	0.107**	-0.011	-0.004
ln CapStock 2	0.045***	0.082***	0.009				-0.020	0.136***	0.100***			
ln Capex	(4.048)	(5.189)	(0.592)				(-1.013)	(8.786)	(6.634)			
Size: 10-24	0.163***	0.186**	0.187**	0.153***	0.135***	0.171***						
Size: 25-49	0.233***	0.221**	0.318***	0.103***	0.064**	0.146***						
Size: 50-249	0.195***	0.244***	0.195***	0.047**	0.066**	0.005						
Size: 250+	0.066	0.081	0.138*	-0.097***	-0.087**	-0.098**						
Catering	-0.290***	-0.491***	-0.102	-0.189***	-0.362***	0.015	-0.265**	-0.402***	-0.132	-0.212***	-0.240***	-0.108
Motor Trades	-0.628***	-0.556***	-0.669***	-0.455***	-0.433***	-0.476***	-0.753***	-0.574***	-0.439***	-0.459***	-0.429***	-0.451***
Production	-0.279***	-0.407***	-0.159**	-0.163***	-0.203***	-0.117***	-0.301***	-0.291***	-0.161**	-0.151***	-0.119***	-0.106*
Property	0.252*	0.090	-1.737	0.087	-0.013	-0.169	-0.034	-0.219	0.658***	-0.122	0.111	0.744***
Retail	-0.386***	-0.561***	-0.213**	-0.361***	-0.449***	-0.294***	-0.590***	-0.293***	-0.081	-0.419***	-0.210***	-0.100
Service Trades	0.081*	-0.182***	0.312***	0.166***	-0.025	0.366***	0.280***	-0.079	-0.034	0.234***	0.083	0.004
Wholesale	-0.348***	-0.356***	-0.329***	-0.213***	-0.194***	-0.201***	-0.466***	-0.162*	-0.240**	-0.230***	-0.039	-0.245***
Foreign owner	0.148***	0.135***	0.159***	0.161***	0.153***	0.148***	-0.031	0.134***	0.127***	0.063	0.167***	0.165***
Multi-LUs	-0.005	0.013	-0.026	-0.004	0.013	-0.024	0.002	0.001	0.005	-0.010	-0.005	0.031
Software int	0.256**	0.334*	0.219	0.155**	0.106	0.240**	-0.117	0.990***	0.706**	-0.013	0.578***	0.198
Female FT (%)	-0.018	-0.059	0.062	-0.098**	-0.184***	0.118	0.009	-0.082	-0.237**	-0.106*	-0.112	-0.180*
Female PT (%)	-2.087***	-1.749***	-2.548***	-1.871***	-1.534***	-2.413***	-2.907***	-1.710***	-1.613***	-2.013***	-1.891***	-1.559***
Male PT (%)	-2.400***	-2.054***	-2.852***	-2.384***	-1.854***	-3.517***	-2.903***	-2.174***	-2.115***	-2.497***	-2.217***	-2.201***
Obs	25744	13230	12514	63348	33450	29898	9552	8089	8103	41413	13256	8679
R-squared	0.208	0.221	0.206	0.165	0.173	0.169	0.129	0.183	0.363	0.122	0.179	0.352
Capital measure	*CapStock 2*	*CapStock 2*	*CapStock 2*	*CapStock 2*	*CapStock 2*	*CapStock 2*	*CapStock 3*	*CapStock 3*	*CapStock 3*	*Capex*	*Capex*	*Capex*

Note: Reporting coefficients. ABI 2007- 2008; England only. All regressions include a full set of regional, legal status, company age and yearly dummies (pooled data only). *** p<0.01, ** p<0.05, * p<0.1. Employment: controlling for the (log of) employment. Training is inserted as a binary variable (whether undertaken any TTG activity in the period considered). Missing category for sector of activity: Construction.

Source: London Economics analysis of the ABI and Capital Stock Series using matched EDS/IDBR data

This publication is also available on our website at www.bis.gov.uk

Any enquiries regarding this publication should be sent to:

Department for Business, Innovation and Skills
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000

If you require this publication in an alternative format, email enquiries@bis.gsi.gov.uk, or call 020 7215 5000.

URN 12/766