

The Promise of Workplace Training for Non-College-Bound Youth: Theory and Evidence from German Apprenticeship*

Damon Clark[†]
IZA, Bonn

René Fahr[‡]
IZA, Bonn and University of Bonn

January 2002

Abstract

This paper assesses the potential of ‘workplace training’ with reference to German Apprenticeship. When job-shopping is important, we examine the ‘optimal’ training package and outline conditions under which firms will provide this. Since the German system broadly meets these conditions, we evaluate the effectiveness of apprenticeship using a large administrative dataset. We find returns to apprenticeship for even the lowest ability school-leavers comparable to standard estimates of the return to school, and show that training is transferable across a wide range of occupations, such as a one-digit occupation group. We conclude that the positive experience with German Apprenticeship Training may guide the design of similar policies in other countries.

JEL-classification: C29, J24, J31, J62

Keywords: German Apprenticeship Training; Human Capital; Occupational Mobility; Wages.

*This paper has greatly benefited from discussions with Steve Nickell, Hilary Steedman and Margaret Stevens. We thank seminar participants at LSE, IZA, and the University of Bonn for helpful comments. Financial support from ESRC and DFG is gratefully acknowledged.

[†]Corresponding author. IZA, P.O. Box 7240, D-53072 Bonn, Germany. Tel: +49 (0)228 3894 306. E-mail: clark@iza.org

[‡]IZA, P.O. Box 7240, D-53072 Bonn, Germany. Tel: +49 (0)228 3894 533. E-mail: fahr@iza.org

1 Introduction

In countries such as the UK and the US, where there exists no well-established path from school to work, policies designed to help non-college-bound school-leavers into the labour market are high up the policy agenda.¹ Although school-to-work policies comprise a diverse mix of ideas and proposals, the common denominator is an emphasis on workplace training as a means of smoothing the transition from school to work.² Proponents of these policies subscribe to the widely-held view that education and training (particularly of non-college bound youth) is central to economic performance, but also draw on recent findings casting doubt on the efficacy of publicly provided training and education programs (see, for example, Heckman (1999)). In both of these respects, the basic thrust of the workplace training idea is very much in step with the conventional policy wisdom.

In other respects however, the idea of workplace training for school-leavers goes against the grain of modern thinking about the labour market. In particular, whilst it was thought in the 1980s that high rates of job mobility amongst school-leavers were in some sense problematic, influential research in the 1990s has emphasised the importance of the role played by the labour market in sorting young workers into the firms and the occupations in which they are most productive (Topel and Ward (1992), Neal (1999)). In this context, workplace training may restrict school-leaver's job shopping opportunities by tying them to firms and occupations that they may not be well suited to. As Neal (1999) puts it, "institutions that limit returns from search may lead to...an inefficient assignment of workers to tasks in the economy" (p. 257).

Is workplace training for school-leavers a good or a bad thing? It is instructive to start from the social planner's perspective. Workplace training is good to the extent that it improves productivity within a particular firm or occupation, but bad in that it is costly to provide, and because it introduces costs to productive job-shopping. That is, the prospect of losing firm- or occupation-specific capital prevents young workers from shopping for their most productive match. Clearly, this depends on the training being 'specific'. The more transferable is the training provided, the less costly it is to shop for new jobs and occupations.

To analyse the problem more formally, we consider workplace training as training within an occupation. However, we consider the transferability of this training to other occupations to be a choice variable, and we assume

¹See Ryan (2001) for a cross-country survey of the school-to-work transition.

²For the US, see Hughes, Bailey, and Mechur (2001) for a positive assessment of existing policies, Bassi and Ludwig (2000) for a detailed account of several school-to-work schemes and Krueger and Rouse (1998) for an analysis of a workplace training scheme aimed at a wider age group. For the UK, Steedman, Green, and Ryan (1998) propose a radical extension of the UK's relatively small 'Modern Apprenticeship' scheme.

that more transferable training is more costly to provide than less transferable training. Starting from the social planner's perspective, we analyse the dependence of both optimal training levels and optimal transferability on returns to training and the importance of matching (and therefore expected turnover). Under the assumption that there are a large number of firms competing to provide training and recruit trained workers within every occupation, we demonstrate that the privately optimal training programme exactly mirrors the socially optimal one. In principle therefore, we would expect the 'optimal' workplace training programme to be offered to school-leavers.

Since our model is concerned with occupational matching, it is related to the literature that considers this issue in more depth. This includes Miller (1984), McCall (1990) and Neal (1999). In contrast to these papers, the occupational mobility decision is very simple in our model: after finishing training, workers have complete information regarding occupational matches and therefore move when match improvements offset human capital losses. We maintain this simple assumption in order that we can investigate the human capital investment decision.

The effect of future turnover on the human capital investment decision has previously been analysed in the context of firm matching but not occupation matching. In particular, Stevens (1994), Chang and Wang (1996) and Scoones (2000) explore the firm's human capital investment decision when there is matching of firms to workers, or at least some turnover. The essential differences between these models and ours are first, that we deal with occupational turnover and secondly, that since we assume 'competitive' occupations, firms earn no rents, there are no externalities and the privately optimal training decision mirrors the socially optimal one.

To bridge the gap between theory and practise, the second part of the paper considers an existing workplace training programme that broadly corresponds to the situation outlined. German Apprenticeship Training (GAT) is a mass workplace training program regularly completed by more than two-thirds of German school-leavers. Training typically lasts for three years, is organised along occupation lines, and competition for potential apprentices and newly apprenticed workers is stiff.

Interestingly, many advocates of workplace training in the UK and US regard German Apprenticeship Training (GAT) as a blueprint or model for school-to-work policies. For example, with reference to GAT, Baily, Burtless, and Litan (1992) advocate the subsidy and regulation (via skill standards and certification) of workplace training for non-college-bound US youth. Others however, are more cautious, advocating only that "work-based learning, as instituted in the Germany style...is an idea worth trying out on a small scale" (Heckman, Roselius, and Smith (1994), p.141).

An exchange of views about GAT between these authors goes to the

heart of the workplace training debate.³ Burtless (1994) argues that GAT is to be admired because, *inter alia*, the returns to training seem high. Heckman, Roselius, and Smith (1994) counter first, that the returns are up for debate and secondly, that even if the returns to training *within the training occupation* are high, the occupation ‘specificity’ of the training may prevent future occupation shopping: “the very narrow technical training and rigid curriculum of the apprenticeship program may contribute to diminished options in later life” (p.99). On the second point, Burtless counters that GAT must be relatively transferable to facilitate certain types of occupational mobility.⁴

In our empirical work, we attempt to resolve these issues by estimating the returns to training within the training occupation and by estimating the transferability of training to other occupations. The first of these exercises entails a comparison of the earnings of apprentices trained inside their occupation with non-apprentices. The second involves an analysis of the loss of earnings experienced upon moving out of the training occupation. Our approach therefore combines an analysis of the returns to apprenticeship training based on an earnings equation written in levels (see Card (1999) for a review) and an analysis of the effects of displacement on earnings (see for example Neal (1995) and Dustmann and Meghir (1999)). The difference between our approach and the one typically taken to estimate the returns to schooling is that we focus only on apprentices still working within their training occupation. The difference between our approach and the displacement literature is that we focus on displacement from an occupation rather than a firm or industry.

To preview our results, we find returns to GAT within the occupation trained in comparable to the returns to a year of schooling typically reported in the literature. Moreover, we find this training to be transferable within a broad occupational group, such as a 1-digit occupation. Consistent with this finding, we show that there exists a high degree of mobility out of the apprenticeship occupation. Also consistent with this finding, we show that trained German workers have a higher degree of attachment to their 1-digit occupation than untrained German workers.

The paper is organised as follows: section 2 sets up a model of workplace training and in section 3 we describe how we plan to estimate the parameters of interest. Section 4 provides an outline of GAT and section 5 describes the

³See the comments of Burtless (1994) on Heckman, Roselius, and Smith (1994) and the reply of the latter.

⁴For example, it is an oft-stated fact within the GAT literature that apprentices trained as bakers often work for motor companies such as Ford upon completing apprenticeship. Heckman, Roselius, and Smith (1994) might argue that this was a product of the fact that the returns to apprenticeship are low, so that there is very little cost to this type of mobility. Burtless (1994) on the other hand would argue that the costs of this kind of mobility are low because returns to apprenticeship are high and apprenticeship training is transferable across occupations.

data used in later sections. Section 6 describes the occupational mobility of German apprentices and we estimate the returns to GAT, and the transferability of GAT in section 7. Section 8 concludes with some remarks on the policy implications of our findings.

2 A Model of Workplace Training

In this section, we present a model of workplace training. The model will enable us to shed light on the issues raised in the Introduction, and will pave the way for our empirical analysis. The basic structure of the model is simple, but flexible enough to allow us to consider both training and matching as sources of productivity and wage growth. The key feature of our model is the assumption that training is occupational, and that the transferability of this training is chosen by firms. Since this concept of training differs from the conventional approach, we discuss these differences in the next section.

2.1 Training Technology

The existing training literature - both in its theoretical and empirical form - starts from the hugely influential distinction made by Becker (1993) between general training (of equal use to every firm in a labour market) and specific training (of use only to the training firm). Whilst this serves as a neat way of distinguishing which party is most likely to bear the costs of training, it does not necessarily provide an accurate representation of what we often think of as training. For example, it is not clear how the ability to repair TVs, to program computers, or to write economics papers can be said to be the combination of general and specific training. Indeed, Becker himself appreciated that the general-specific dichotomy was not a useful classification of all types of skills: “(s)ome training may be useful not in most firms nor in a single firm but in a set of firms defined by product, type of work, or geographical location” (Becker (1993, p.49)).

Since our interest is in workplace training, we focus on training within an occupation. We denote the level of this training by H , where H may be measured in years. Moreover, as we are interested in the transferability of this training to other occupations, we allow occupational training programs to differ according to the transferability of the training provided. In the simplest case of two occupations, the *transferability* of training (denoted γ) is simply the fraction of training in one occupation that can be transferred to another occupation. Hence, increased transferability does not improve the worker’s performance within her current occupation, but does improve her performance in another occupation.

We adopt a reduced form approach to transferability in the sense that we do not consider *how* training becomes more transferable, only that it is

costly for firms to make it more transferable.⁵ In particular, we make the following assumptions regarding the training cost function $C(H, \gamma)$:

$$\begin{aligned} C(0, \gamma) &= 0, C_H(H, \gamma) > 0, C_\gamma(H, \gamma) > 0 \\ \lim_{H \rightarrow 0} C_H(H, \gamma) &= 0; \lim_{\gamma \rightarrow 0} C_\gamma(H, \gamma) = 0 \end{aligned}$$

The first set of assumptions imply that there are no costs if no training is undertaken ($H=0$), but that increased training increases costs. They also ensure that the costs of increased training are rising in the level of transferability. Were this not case, there would be no costs to making training completely transferable. Hence this assumption makes the problem both interesting, and realistic. The final pair of assumptions ensure that in equilibrium, the social planner (and it will turn out, private firms) will provide positive quantities of training, of which a non-zero proportion will be transferable.

2.2 The Model

We assume that a finite number of tasks can be performed in the economy. In the case of GAT, these will refer to occupations, hence we use tasks and occupations interchangeably. We focus attention on the two-occupation case.⁶ Importantly, we assume that there are a large number of identical firms offering jobs within each of these occupations. These firms are price-takers in the output market, and we normalise the output price to unity.

This ‘identical firms’ assumption implies that there is no such thing as firm-specific human capital, and results in an equilibrium in which workers pay all of the costs of their training. Of course this identical firms assumption may not be a reasonable description of certain occupations. However, as we will see in section 4, apprenticeship occupations are defined very narrowly. Moreover, we can provide an informal test of this assumption using the QaC data (described in more detail in section 7 and in the appendix). The survey asks apprentices how many of the apprenticeship skills are used on the current job. Comparing those in the training occupation and the training firm with those in the training occupation but a different training firm, the

⁵An example of a technology that would generate this reduced form structure is one in which firms can invest in occupational training (H) and training that is specifically geared to increasing transferability (X). This training might involve learning about related occupations and acquiring a deeper understanding of the occupation being trained in. Supposing that productivity in the training occupation is H and in the other occupation is $\frac{XH}{1+X}$, transferability would then be $\gamma = \frac{X}{1+X}$. When $X=0$, none of the training is transferred to the other occupation, whilst increases in X increase the transferability of training.

⁶The n -occupation case can be considered if occupations are assumed to be equally spaced around a circle representing occupational ‘distance’. This model has the implication that moves are more common between closely related occupations, but otherwise generates no new insights.

results are almost identical. For the former (latter) group, the responses are very few or none 0.78% (0.95%); few 3.25% (4.31%); some 10.53% (11.04%); many 24.30% (7.34%); and very many 61.14% (56.36%) based on sample sizes of 893 and 951 respectively. Also consistent with this assumption is the fact that whilst firms subsidise training costs in certain occupations, in many occupations, workers do indeed pay for the training themselves.⁷ Acemoglu and Pischke (1998) provide a model designed to explain the phenomenon of cost-sharing based on the superior information that training firms have regarding worker ability.

Match Assumptions

We assume a continuum of workers who supply labour inelastically. Like the firms in this economy, the workers are risk-neutral. The product of worker i in every firm in occupations A and B is described in equations (1) and (2):

$$\Pi_{Ai}^{NOAPP} = m(q_i) + \xi_{Ai} \quad (1)$$

$$\Pi_{Bi}^{NOAPP} = m(q_i) + \xi_{Bi} \quad (2)$$

Equations (1) and (2) say that worker product is an additive function of ‘general ability’ $m(q_i)$ and the worker-occupation match ξ_{ji} , $j \in \{A, B\}$. General ability is a function of worker’s innate ability q_i , where $q_i \in (\underline{q}, \bar{q})$ and the worker-occupation match is assumed to be drawn randomly from a uniform distribution $\xi_{ji} \sim U(\frac{-\bar{\xi}}{2}, \frac{\bar{\xi}}{2})$. The advantage of the uniform distribution is that it allows us to characterise the importance of matching in terms of only one parameter $\bar{\xi}$.

Training

Now consider worker i after she has received training quantity H in occupation A , where a proportion γ of this training can be transferred to occupation B . This worker has the following product when working in occupations A and B :

$$\Pi_{Ai}^{APP} = \alpha(q_i)H + m(q_i) + \xi_{Ai} \quad (3)$$

$$\Pi_{Bi}^{APP} = \alpha(q_i)H\gamma + m(q_i) + \xi_{Bi} \quad (4)$$

Comparing (1) and (3), worker i apprenticed and working in occupation A is more productive by $\alpha(q_i)H$ than untrained worker i working in occupation A . This is the value of training - the annual return to training $\alpha(q_i)$ multiplied by the years of training undertaken (H). We assume that the annual return to training is an increasing function of worker quality: $\alpha'(q_i) > 0$.

⁷Of course apprentices do not pay the costs of training up front. Instead, they earn wages lower than the value of their apprenticeship product during the training period.

Comparing (2) and (4), worker i trained in occupation A and working in occupation B is more productive by $\gamma\alpha(q_i)H$ than untrained worker i working in occupation B .

Timing and Information

As is standard in the training literature, we assume that workers exist in the labour market for two periods: a training period and a working period. Prior to the first period, we assume that workers' optimal matches are unknown to all parties. After the first period, workers' optimal matches in both occupations are revealed to trainees and all firms. We assume that there is no uncertainty relating to any other aspect of the economy, and we assume that training levels are both observable and verifiable. Under these assumptions, at the start of the first period, firms within each occupation will compete for new workers by offering training packages (W, H, γ) , where W is the first-period wage, and these packages may depend on observed worker ability q_i . At the start of the second period, after the occupational matches have been revealed, firms compete for newly trained workers. From the assumption that there are many firms within every occupation, new apprentices will receive wage offers equal to their marginal product. Whether they choose to leave the training occupation will then depend whether the improved match that can be realised by leaving the training occupation will exceed the loss of skills incurred upon making such a move.

2.3 Social Optimum

In this section we consider the economy from the social planner's perspective. Given our assumptions, per capita social welfare (SW_i) is simply expected per capita output less the direct costs of training $C(H, \gamma)$:

$$\begin{aligned}
SW_i &= P_i(stay)[\alpha(q_i)H + E(\xi_{Ai}|stay)] \\
&\quad + [1 - P_i(stay)][\alpha(q_i)H\gamma + E(\xi_{Bi}|leave)] \\
&\quad - C(H, \gamma) \\
&= \alpha(q_i)H + \alpha(q_i)H(\gamma - 1)[1 - P_i(stay)] \\
&\quad + \{P_i(stay)E(\xi_{Ai}|stay) + [1 - P_i(stay)]E(\xi_{Bi}|leave)\} \\
&\quad - C(H, \gamma) \tag{5}
\end{aligned}$$

The right hand side of this equation consists of four components: the direct value of training, the expected loss of skills resulting from turnover, the benefits of occupational matching (which correspond to the second line in equation (5)) and the direct costs of training. We argue in the Appendix that the third term will be increasing in γ . This formalises the notion that an increase in transferability can increase the expected benefits of matching.

The social planner now has three choices: for a given worker i she chooses the optimal quantity of training and the optimal transferability of this training in the event that this worker is trained. She then decides whether this worker should be trained. Consider first the choice of training levels. We state the results without proof here. All proofs can be found in Appendix A.

Proposition 1 *If the social optimum exists, then $H^* > 0, \gamma^* > 0$.*

Proof. See Appendix A

In other words, the equilibrium is characterised by positive levels of training, and a positive degree of transferability. This result comes about because of our assumptions regarding the cost function. This allows us to consider the more interesting question of how changes in the importance of matching (as represented by $\bar{\varepsilon}$) and how changes in the returns to training (as represented by $\alpha(q_i)$) will affect the optimal mix of training.

Regarding the first question, we might expect that an increase in the importance of matching (increased $\bar{\varepsilon}$) would increase optimal transferability and decrease optimal training. Indeed, when $SW_{H\gamma} < 0$, we can show that this is indeed the case. However, we can not be sure that $SW_{H\gamma} < 0$ since this depends *inter alia* on $C_{H\gamma}(H, \gamma)$.⁸ Hence we can not rule out the possibility that increased $\bar{\varepsilon}$ will result in increases in both H and γ or decreases in both H and γ . The important point however is that the social planner takes account of matching when choosing the optimal training programme.

Proposition 2 $\frac{dH^*}{d\alpha} > 0, \frac{dH^*}{dq} > 0$

Proof. See Appendix A.

This Proposition says that an increase in the returns to training increase the optimal level of training. Since we assumed that $\alpha'(q_i) > 0$, more able workers (with higher (q_i)) will receive more training. The effects of increased returns on transferability are ambiguous, and will depend on $SW_{H\gamma}$ and our assumption that γ is not a function of q_i . We do not consider this possibility, not least because we do not have enough data to examine transferability according to ability empirically.

To analyse the optimal number of workers trained, we examine the net per capita social return to training given optimal levels of H and γ and given worker ability q_i .

Proposition 3 *Training will be socially optimal for any worker of ability $q \geq q^+$ where $SW(H^*, \gamma^*; q^+) = 0$.*

⁸See appendix for a more formal discussion

Proof. See Appendix A.

We use the envelope theorem to show that social welfare at the optimal training levels is continuously increasing in ability. This implies that all workers above a cut-off ability level will be trained whilst all those below this cut-off ability level will not.

2.4 Private Optimum

Since we assume that firms are able to credibly commit to training packages (W, H, γ) , firms will maximise profits subject to a constraint that the expected utility of the worker is at least as high as that offered by another firm. Competition among firms ensures that firms make zero profits, and so the problem is equivalent to firms maximising the utility of young school-leavers subject to a zero profit condition. Since utility of a worker is just the sum of expected wages over the two periods, the firm's problem is:

$$\begin{aligned} \max_{H, \gamma} U_i(W, H, \gamma) &= W_i + P_i(\text{stay})E[\Pi_{A_i}^{APP} | \text{stay}] + P_i(\text{leave})E[\Pi_{B_i}^{APP} | \text{leave}] \\ \text{s.t. } W_i &= -C(H, \gamma) \end{aligned}$$

We know that second-period wages must equal second-period productivity in this model, hence this reduces to precisely the problem faced by the social planner in equation (5), we have therefore proved the following proposition:

Proposition 4 *In equilibrium, the training package offered by private firms is identical to that chosen by the social planner, and an identical number of workers are trained.*

2.5 Summary

When firms compete to offer training packages to workers, they will offer socially optimal training packages. As such, the training offered will take account of the expected degree of turnover, and more able workers will be offered more training than less able workers, since their returns will be higher. In our empirical analysis, we will examine the returns to GAT within the training occupation and the transferability of GAT across occupations. The next section outlines how we will identify these parameters.

3 Empirical Framework

In this section, we describe how we will identify the key parameters of interest: the return to GAT and the transferability of GAT. Again, we consider only the two-occupation case (A and B), although our results generalise readily to the n -occupation case. We start by describing the log-earnings of

worker i in occupations A and B . This follows directly from equations (1), (2) (3) and (4):

$$\ln W_{Ai}^{NOAPP} = m(a_i) + \xi_{ai} \quad (1')$$

$$\ln W_{Bi}^{NOAPP} = m(a_i) + \xi_{bi} \quad (2')$$

$$\ln W_{Ai}^{APP} = \alpha(a_i)H + m(a_i) + \xi_{ai} \quad (3')$$

$$\ln W_{Bi}^{APP} = \alpha(a_i)\gamma H + m(a_i) + \xi_{bi} \quad (4')$$

where A is the training occupation and where ‘NOAPP’ (‘APP’) refers to worker i without (with) an apprenticeship qualification. From our theoretical analysis, we know that the worker *leaves* the training occupation when:

$$(\xi_B - \xi_A) > \alpha(q_i)H(1 - \gamma) \quad (6)$$

In other words, when the improved (occupational) match exceeds the cost of transferring apprenticeship training.

3.1 Relation to the Existing Literature

We estimate the return to training in a relatively standard fashion (see for example Lynch (1992)). Whilst we show that the presence of matching has implications for our estimates, we are able to derive an approximate lower bound to this return. Controlling for the presence of matching to identify transferability (a function of the costs of transferring training across occupations) is more difficult. Intuitively, workers leave the apprenticeship occupation to realise a better match (equation (6)). Hence estimates based on a comparison of movers and stayers are likely to overstate the transferability of training.

The approach that we take - we use a sample of displaced workers to generate what are effectively ‘exogenous’ occupation changes - is similar to that taken by other papers that attempt to control for match-driven mobility. For example, Dustmann and Meghir (1999) focus on the returns to tenure and experience in the context of a model in which wages are also driven by worker-firm matching.⁹ As in our framework, the job change decision trades off match improvements with losses of human capital (firm-specific human capital in their case) and a sample of displaced workers is used to generate ‘exogenous’ firm changes.¹⁰

⁹In fact the model is more general, as the returns to tenure are allowed to vary across firms and individuals, and the returns to experience across individuals.

¹⁰Returns to experience are estimated using a sample of workers starting a new job, and these estimates are used to calculate the within-firm wage growth which must be due to a combination of tenure and changes in match quality. Here, the selection problems are that workers observed working for a particular firm chose to join this firm and chose not to leave it. A sample of firm closures is used to control for the first problem and age is used as an instrument for tenure.

The empirical paper closest to ours is Werwatz (1998). Werwatz also addresses the question of how transferable is GAT between occupations, finding that occupational ‘movers’ earn similar wages to occupational ‘stayers’, and concluding that GAT must be fairly transferable across occupations. Since Werwatz has only cross-sectional data, he controls for endogenous mobility by estimating a switching regression model. The fact that the selection terms are rarely found to be statistically significant could indicate that selection biases are not a problem for his results. However, it is more likely that the selection equation has been inadequately specified. In particular, it is not clear that a variety of ‘quality of work’ measures (such as standing up at work) capture what is driving occupational mobility (the search for better occupational matches) or that these do not belong in the earnings equation. Fortunately, the panel nature of the IAB data enables us to improve upon this strategy. The following subsections describe how we do this.

3.2 The Returns to Apprenticeship Training

We estimate the returns to apprenticeship training under two assumptions: first, that school-leavers do not differ in ability and secondly that school-leavers do differ in ability.

Homogenous School-Leaver Quality

The first of these assumptions implies that $q_i = q$ for all i , so that $m(q_i) = m_0$, $\alpha(q_i) = \alpha$ and $H(q_i) = H$, so that the return to one year’s worth of workplace training (α) is a homogenous parameter. Suppose that we wish to estimate this parameter. We do this by estimating the total value of workplace training (αH) and then dividing by the average number of years spent in apprenticeship training H (which we can measure from the data). Consider estimating by ordinary least squares an equation of the following form (ignoring covariates such as experience that have ‘common effects’):

$$\ln W = a_0 + a_1 APPIN + \eta \quad (7)$$

where $APPIN$ refers to an apprentice working inside of the training occupation, η is a random disturbance term and the base group are those without any apprenticeship training. Then from (1’), (3’) and (6), the probability limits of \hat{a}_0 and \hat{a}_1 are given by:

$$\begin{aligned} p \lim \hat{a}_0 &= m_0 + E(\xi_a | \xi_a - \xi_b > 0) \\ p \lim \hat{a}_1 &= \{m_0 + \alpha H + E(\xi_a | \xi_a - \xi_b \geq -\alpha H[1 - \gamma])\} \\ &\quad - \{m_0 + E(\xi_a | \xi_a - \xi_b > 0)\} \\ &= \alpha H + E(\xi_a | \xi_a - \xi_b \geq -\alpha H[1 - \gamma]) - E(\xi_a | \xi_a - \xi_b > 0) \\ &\leq \alpha H \end{aligned}$$

since $\alpha H[1 - \gamma] \geq 0$. Hence, using the estimated value of a_1 from (7), we will estimate a lower bound to αH and therefore a lower bound to α .

Heterogeneous School-Leaver Quality

It is perhaps more plausible to assume that school-leavers differ in ability q_i . Hence as assumed in section 2, α is an increasing functions of q_i , H is an increasing function of q_i , and α is a heterogeneous parameter. Under this assumption, estimates of $\alpha(q_i)$ based on equation (7) represent lower bounds on the returns to intensive apprenticeship training *for those choosing to become apprentices*. Since we have shown in proposition 3 that only those school-leavers with $q > q^+$ will actually be apprenticed, it should be obvious that this is neither the population mean return to apprenticeship nor the return for those on the margins of apprenticeship and non-apprenticeship (i.e. those with $q \simeq q^+$). Since we are often concerned with the effects of apprenticeship-type programs on hard-to-educate workers in other countries, this latter parameter may be of particular interest.

We provide an estimate of this parameter using data disaggregated by training firm size. It is often asserted that there is a clear ranking in both the quality of apprenticeship programs offered and the ability of applicants. Consistent with our model, apprenticeship programs tend to last up to one year longer than those offered in smaller firms (as we will see in section 4). Moreover, as Steedman (1993) notes, “In the public mind in Germany, a definite and complex ranking of apprenticeship places exists linked to expected lifetime returns. As a general rule, *Industrie* apprenticeships are more highly sought-after than *Handwerk* apprenticeships” (p.1285). In fact, *Industrie* and *Handwerk* broadly correspond to large and small firms respectively, and Steedman’s claim is supported by evidence presented in Harhoff and Kane (1993), who find that the proportion of apprentices reporting good mathematics and good German scores in school are strongly increasing functions of apprenticeship firm size.

Since we have data on firm size, we can disaggregate the apprenticeship variable in equation (7) and use estimates of the returns to those trained in the smallest firms as an approximation to $\alpha(q^+)$, the return to apprenticeship training to those workers of lowest ability.¹¹

¹¹In the model, there is no heterogeneity among firms and so we offer no formal explanation as to why the highest quality school-leavers train in the largest firms (as opposed to a situation in which every firm offers a range of programs to cater for different abilities). One possibility is that if training in large firms generates complementary firm-specific skills, the returns to these skills will be shared between trainee and firm, hence the expected lifetime utility of apprentices is higher in larger firms and larger firms earn higher rents on higher quality workers. Of course this assumes a fixed number of large firms.

3.3 The Transferability of Apprenticeship Training

In order to investigate the worst-case scenario, we would like to identify a lower bound to transferability. Since we will estimate transferability by considering the costs of leaving the apprenticeship occupation, this corresponds to an upper bound to the costs of transferring apprenticeship skills.¹² It should be clear that we will not estimate such a bound by simply examining changes in log earnings between periods $t-1$ and t . Since we know from equation (6) that workers will only move out of the training occupation when $\xi_{Bi} - \xi_{Ai} > \alpha H(1 - \gamma)$, simply comparing the wages of ‘movers’ and ‘stayers’ would cause us to *under-estimate* the costs of transferring training.

We can, however, make some progress by basing this comparison on a sample of workers displaced for exogenous reasons (e.g. plant closure). To see this, consider estimating the following equation, where ‘MOVOUT’ is a dummy variable indicating whether the worker has left the training occupation:

$$\Delta \ln W = b_0 + b_1 \text{MOVOUT} + \eta \quad (8)$$

Then we can estimate an upper bound to the costs of transferring training (lower bound to transferability) under the following assumption:

A1 Workers are randomly displaced from their firms. These workers accept the first job that they are offered and may decide to search for a more suitable position on the job.

In a formal model of search with offers arriving exogenously on and off the job, assumption A1 requires that the arrival rate of offers to unemployed workers is no greater than the arrival rate to workers in a job and that search costs for unemployed workers are at least as large as those employed workers.¹³ For workers accepting an offer within their training occupation, the expected change in log earnings is then zero. For those accepting a job outside of the training occupation, the change in log earnings is from equations (3’), (4’) and (6):

¹²An upper bound to transferability (lower bound to the costs of transferring training) is easily derived by augmenting equation (7) to include a dummy for being an apprentice outside of the training occupation. A lower bound to the cost of transferring training then corresponds to the difference between the estimated coefficient on this dummy and $\widehat{\alpha}_1$.

¹³See for example Mortensen (1986).

$$\begin{aligned}
E(\Delta \ln W | MOVOUT_i) &= \alpha H(\gamma - 1) + E(\xi_B | \xi_A - \xi_B \geq \alpha H(\gamma - 1)) \\
&\quad - E(\xi_A | \xi_A - \xi_B \geq \alpha H(\gamma - 1)) \\
&= \alpha H(\gamma - 1) + E(\xi_B | \xi_B - \xi_A < \alpha H(1 - \gamma)) \\
&\quad - E(\xi_A | \xi_A - \xi_B \geq \alpha H(\gamma - 1)) \\
&< \alpha H(\gamma - 1)
\end{aligned}$$

Hence we have that:

$$\begin{aligned}
p \lim \widehat{b}_1 &= \{ \alpha H(\gamma - 1) + E(\xi_B | \xi_A - \xi_B \geq \alpha H(\gamma - 1)) \\
&\quad - E(\xi_A | \xi_A - \xi_B \geq \alpha H(\gamma - 1)) \} - \{0\} \\
&< \alpha H(\gamma - 1)
\end{aligned}$$

This implies that estimates of b_1 derived from equation (8) for a sample of displaced workers are downward-biased under assumption A1 so that we *over-estimate* the costs of transferring skills and so underestimate transferability. This can be explained as follows: the initial occupation was chosen in period $(t-1)$ and the new occupation was not chosen in period $(t-1)$. Hence the expected value of the match in the former is positive, the expected match in the latter negative and so a move outside of the training occupation entails an expected loss of match productivity as well as the cost of transferring skills.

This result may not hold under a less restrictive assumption regarding the job-search behaviour of displaced workers. A more plausible assumption is A2:

A2 There is some cost U to rejecting an offer from outside the training occupation. This represents the expected cost of remaining unemployed and waiting for an offer from inside the training occupation.

Then, displaced workers will only accept an offer from outside of the training occupation when $\xi_{Bi} - \xi_{Ai} > \alpha H(1 - \gamma) - U$, so that behaviour will only mirror that under A1 for those workers with large U (in other words, when there is no option but to move). Under this more general assumption, an obvious strategy is to find proxies for U that will enable us to instrument the decision to leave the training occupation. These instrumental variables estimates could then be interpreted as if they were derived under assumption A1. Another option is to use the answers to a survey question regarding the usefulness of skills learned during the apprenticeship that we interpret as abstracting from any matching considerations. We discuss both of these strategies in section 7.

4 German Apprenticeship Training

To understand the structure of GAT, we begin by outlining the 1969 Vocational Training Act, which remains the foundation stone of GAT. This Act explicitly defined a number of occupations in which school-leavers could apprentice. Whilst these currently number 375, fewer than the 600 that could be apprenticed in the 1970s, they are defined very narrowly. For example, within the class of electrical occupations (a two-digit category) school-leavers can apprentice in 15 different occupations.¹⁴ This provides a rationale for our assumption that the skills acquired by apprentices are occupation- rather than firm-specific.

Although the length of apprenticeship depends on the apprenticeship occupation, GAT typically lasts between two and three and a half years. A crucial part of the curriculum for every training occupation involves training firms releasing their apprentices for one day per week to attend a local vocational school. These are organized around one of five vocational fields (industry, commerce, home management, agriculture and other occupations) and are designed to fill any gaps in general education and to prepare apprentices for the final examination (see Steedman, Green, Bertrand, Richter, Rubin, and Weber (1997) for an example of such a curriculum). Training is completed when apprentices pass the final examinations. These typically consist of several written examinations in the subjects laid down by the training regulations, with many including an oral or practical component.

It has sometimes been suggested that the centralised German wage bargaining structure limits the degree of post- and pre-apprenticeship competition among firms. In fact though, the wages bargained centrally are more like minimum wages, with firms free to increase wages above these minimum levels. In any case, firms can increase wages by changing the job titles of workers and in the case of apprenticeships, by offering different fringe benefits. Casey (1991) reports evidence of this practise.

5 Data Issues

For our estimates of returns to and transferability of apprenticeship we use data from a 1% sample of German social security records (see the IAB Data Appendix for a fuller description of the data set). The data are available for the years 1975-1995, and are supplemented by data on the firms to which workers are attached. Due to a change in the notifying procedure, we use wage information from 1984 onwards. The data are well suited to the task

¹⁴These include occupations such as ‘electronic specialist, telecommunications’; ‘electronic specialist, communications (telecommunication systems)’; ‘electronic specialist, communications (information systems)’ and ‘electronic specialist, communications (radio engineering)’. See Federal Ministry of Education and Science (1992) for more details.

at hand, since the wage information and timing of employment spells is very accurate.

5.1 The Sample

Only German males are retained for analysis, and our sample consists of two groups: apprentices and non-apprentices. In order to exclude those engaged in short training spells, internships and the like, apprentices are defined as those having been observed training for greater than 450 days. We further restrict the sample of apprentices to those without the *Abitur* (usually completed by those that will eventually attend University) and those starting their apprenticeship aged 19 or under. The age restriction is designed to include those that take their military service after leaving school, but exclude those training after a spell in the labour market. We exclude those with an *Abitur* as the labour market for apprentices with this qualification will be significantly different to that for those without an *Abitur*. In any case, this group is relatively small.¹⁵ To make the sample of non-apprentices as comparable as possible, we include only those whose first spell is observed aged 19 or under and who do not have the *Abitur*.

Table 1 displays some descriptive statistics for our sample. The most noticeable feature of the Table is the gradual ageing of the sample. Since the maximum age for a person in the sample increases from 19 in 1975 to 39 in 1995, this is to be expected. Note also that the proportion of workers in the sample with apprenticeship training is increasing over the observation window. This reflects an increase in the proportions of young school-leavers undertaking apprenticeship training over this period. Due to the fact that the sample is relatively young, we do not observe many apprentices with the Meister certificate (an advanced vocational qualification typically undertaken by apprentices with several years of labour market experience).

5.2 Displacement

A key part of our empirical strategy involves the construction of a sample of ‘exogenously’ displaced workers. In this respect, the fact that plants are given a unique identifier in the IAB data helps, although we cannot assume that the disappearance from the data of a plant identifier implies that a plant has closed. This can happen for a variety of reasons, including closure, takeover or a merger. To deal with this problem, we construct three subsamples of ‘separations’:

‘Displaced’ First, we use a sample of workers who experience an unemployment spell after separation. We further restrict this unemployment spell to

¹⁵This group makes up less than 20% of all apprentices according to the authors’ calculations with the German Socio-Economic Panel (GSOEP).

be greater than one month to avoid including those workers that quit their previous firm and exclude workers with unemployment spells of greater than one year to avoid problems regarding the scarring effects of unemployment. Although this upper limit is somewhat arbitrary, experiments suggest that it does not impact much on our results. Whilst this sample does not enable us to disentangle those workers displaced exogenously and those displaced for ‘cause’, displacements for cause are only a problem in equation (8) when they are based on unobserved and transitory components of earnings which are correlated with the decision to move out of the training occupation. Hence results based on this subsample are robust to dismissals for cause based on permanent components of earnings (observed or unobserved).

‘Close’ We can compare our results using the ‘displaced’ subsample to those obtained by further restricting this sample to those workers who separated, experienced an unemployment spell and whose plant identifier disappeared from the data. If we assume that plants that merge or reorganise lay workers off on a ‘last-in-first-out’ basis, this group will contain a higher proportion of workers displaced for exogenous reasons. Since we do not know the exact date at which the plant closed in the IAB data, we generate two samples of workers displaced because of ‘closure’: those whose plant identifier disappeared within one and two years of the separation date (‘close1’ and ‘close2’ respectively).

‘Quits’ Finally, for the purpose of comparison, we present the results for a sample of workers separating firms but not experiencing an intervening spell of unemployment.

Table 2 presents some descriptive statistics for the four different groups. Focusing on the pre-displacement characteristics, we find significant differences between these groups. In the second and third rows, we see that displaced workers are slightly younger than the other groups, although there are no significant differences between the ‘quits’ and the ‘close’ samples. The most marked differences occur with respect to pre-displacement tenure. The finding that this is lowest amongst the ‘displaced’ workers is consistent with a layoff policy of ‘last-in-first-out’. This is very important in Germany (see Bender, Dustmann, Margolis, and Meghir (1999) for details) and may also explain the differences between the ‘close’ and ‘quit’ samples, since those ‘close’ workers that were displaced prior to the actual closure of the plant will also have been subject to the ‘last-in-first-out’ rule. Pre-separation wages reflect these differences, and it is interesting to note that in a regression of pre-displacement wages on pre-displacement characteristics and dummy variables representing the groups ‘displaced’, ‘close2’ and ‘close1’, the estimated co-efficients (standard errors) on these variables were -0.00374 (0.00616) for displaced workers, and -0.0234 (0.0267) and -0.0251 (0.0438)

for ‘close2’ and ‘close1’ respectively. Hence, controlling for pre-displacement characteristics, we can not reject the hypothesis that these workers represent a random sample of pre-displaced workers.

6 Occupational Mobility

Before analysing the returns to and transferability of GAT, we provide a brief overview of the occupational mobility of German apprentices. To assess the mobility of apprentices out of their training occupations, Figure 1 plots Kaplan-Meier estimates of the nonparametric survival functions of post-apprenticeship spells in the 3-digit training occupation. That the probability decreases sharply upon completion of apprenticeship training implies that a significant proportion of apprentices leave the apprenticeship occupation immediately. Although the hazard decreases at a much slower rate after this point, it remains the case that after 20 years in the labour market, 75% of apprentices have left the training occupation at the 3-digit level. The graph for the probability of leaving the 1-digit occupation shows a similar pattern but a slightly higher proportion of workers remaining within the 1-digit occupation (about 35%). This implies that two-thirds of German apprentices eventually leave the apprenticeship occupation at the 1-digit level.

To compare this with the mobility of non-apprentices, Table 3 presents occupation and firm mobility statistics for apprentices and comparable non-apprentices for the years 1985, 1991 and 1995. For non-apprentices, we use the first occupation/firm worked in upon labour market entry in place of the apprenticeship occupation/firm. For apprentices, Table 3 shows that in 1995, 60% have left their apprenticeship occupation with the highest mobility taking place at the 1-digit level. For non-apprentices however, the figure is 80%. Using mobility from the training industry rather than the training occupation produces a similar pattern: large number of apprentices leave the apprenticeship industry, but even larger numbers of non-apprentices leave the first industry worked in. This would suggest that there is some specificity in the apprenticeship training investment that deters apprentices from switching occupation/industry. In contrast, the differences in firm mobility between apprentices and non-apprentices are far less pronounced. This adds weight to our assumption that apprenticeship training is occupation- rather than firm-based.

7 Estimates of Returns and Transferability

Our objective in this section is to estimate an approximate lower bound to the return to apprenticeship training within the training occupation $\alpha(q^+)$ and an upper bound to the costs of transferring training.

7.1 Returns to Apprenticeship

In order to obtain our approximate lower bound to the return to apprenticeship (α), we begin by assuming that this value does not depend on school-leaver ability a , and simply split apprentices according to whether they are working inside or outside the training occupation. From the top panel of Table 4, we see that estimates of αH based on equation (7) are approximately 0.15. With the conservative assumption that apprenticeship training lasts for an average of 2.75 years, and that an average of two-thirds of the apprenticeship is spent training, this gives an annualised average return of approximately 8.2%. This is comparable to estimates of the rate of return to schooling found in the literature. For example, Angrist and Krueger (1991), using US Census data, find estimates of the rate of return between 5% and 7% when estimating by OLS, and between 6% and 10% when estimating by IV.¹⁶ Our estimates of the return to additional years of experience is large, although this is likely to reflect the fact that we observe these workers over their first few years in the labour market, when much of the information regarding occupational match productivity is revealed. The low returns to tenure are consistent with those reported in Dustmann and Meghir (1999).

As noted in section 3, the value of apprenticeship is likely to depend on worker ability q . To obtain our approximate lower bound to the return to apprenticeship $\alpha(q^+)$, we split the group of apprentices according to the size of their training firm. As a check that apprenticeship firm size proxies school-leaver quality, the first column of Table 5 presents self-reported school test scores (Harhoff and Kane (1997)). As commonly assumed in the GAT literature, we see a clear correlation between training firm size and school-leaver ability, although the group with the lowest scores are actually those trained in the second smallest firm size. The second column calculates the proportion of apprentices trained in firms of this size, whilst the third column reports the co-efficients estimated from a version of equation (7) in which the apprenticeship variable is disaggregated according to firm size, and we pool across years (and include year dummies). To obtain estimates of the rate of return to an additional year of training, we again adjust these co-efficients according to the length of the training program (again assumed to

¹⁶For the purposes of comparing our results with the previous literature, the second panel of Table 4 presents the results of estimating equation (7), in which we pool the two types of apprentice (those working inside and outside the training occupation) to estimate the ‘return to apprenticeship training’ as commonly reported. This co-efficient averages roughly 0.19 over the sample period. Based on a similar specification, Winkelmann (1994), using the German Socio-Economic Panel (GSOEP), obtains co-efficients that average approximately 0.2 over the period 1985-1990. Werwatz (1998), using the Qualifications and Careers (QaC) data, obtains a co-efficient for 1985 of 0.137. That this is slightly lower than our estimate may be attributable to the fact that Werwatz (1998) includes additional covariates (such as industry dummies) in his earnings equation.

be 2.75 years for each firm size group) and the average time spent training (column 5).

Rates of return across the different size groups are again in the broad range of estimates of the return to an additional year of schooling. Taking the return to one year of apprenticeship training for those workers trained in firms with between two and nine employees as the return to those on the margins of apprenticeship and work, this return is approximately 5.87%. As we would expect, this is lower than the return found for higher ability school-leavers training in larger firms, but it is still within the broad range of estimates of the rate of return to schooling. Moreover, each of these estimates represent lower bounds on the true returns to apprenticeship training. We now attempt to estimate the transferability of this training.

7.2 Transferability

We now turn our attention to the transferability of apprenticeship training between occupations, and in Table 6, we present estimates of the costs of transferring training obtained from estimating equation (8). Before turning to the results for the groups of ‘displaced’ workers that we are interested in, we begin with our comparison group of ‘quits’. Looking at the left-hand column of the Table, we see that amongst the group of quits, the wage penalty associated with moving out of the training occupation is very close to zero. Whilst this would suggest that training is entirely transferable, these estimates are obviously biased because of the match-driven nature of this mobility. Turning then to column 2, we find that for displaced workers, the point estimate of the wage penalty is now negative and significant at the 10% level. This suggests that there are some costs to transferring occupations, although since these are small in comparison to the total value of apprenticeship training inside the occupation ($\alpha H = 0.15$ from Table 4), this suggests that transferability is high. Moreover, in the final two columns, our estimates are not significantly different from zero.

The finding of only very small wage penalties associated with leaving the training occupation might suggest that training is transferable. However, it may be that training is more transferable between closely related occupations than between occupations which are a greater occupational ‘distance’ apart. To investigate this possibility, we disaggregate moves out of the training occupation and consider the wage penalties associated with different kinds of moves. We begin by measuring distance according to the occupational codes. That is, we say that a move at only the three-digit level is a move into an occupation more closely related to the training occupation than one involving a move at the two-digit level. The penalties to moving out of the training occupation according to the distance moved are presented in the second panel of Table 6. Looking first at the left-hand column, for the quits, only the wage penalty associated with a move out of the training

occupation at the 1-digit level is negative, although it is very small. Again, since the majority of these moves are selective, we would not expect to find large penalties to moving for this group.

For displaced workers, the wage penalty associated with moving out of the training occupation at the 1-digit level is slightly larger, and significant at the 1% level. However, it is still small when compared with estimates of the total value of apprenticeship training within the occupation (0.15 from Table 4), hence transferability is still high. Amongst the ‘close’ samples, these estimates are also negative. For those leaving a plant that closed down within one year, the point estimate is larger in absolute value and approximately two-thirds of the value of apprenticeship training. Taking this estimate as the worst-case scenario suggests that whilst training is completely transferable within a 1-digit occupation, only one-third of the training can be transferred outside of the 1-digit occupation.

Since these codes may not be an adequate measure of occupational distance, we look in a next step only at moves out of the training occupation that also involve changes in industry at the two-digit level (the classification is produced in Table A1). The idea here is that reported changes in occupation are less likely to be spurious if they are also accompanied by changes in reported industry. Looking at the third panel of Table 6, we find a similar set of results, with the exception that amongst the ‘close’ sample there is some evidence of wage penalties incurred for moves at the 2-digit level. Like most of the estimates relating to movements at the three- and two-digit level, these estimates are very unstable and often imprecisely estimated. This reflects the small cell sizes associated with these movements (as seen in Table 3).

7.3 Results from a Question regarding Skill Use

It was observed in section 3.3 that if apprentices select the new occupation by trading off improved match values with the costs of transferring skills (as opposed to accepting the first offer and searching on the job), we may not be estimating an upper bound to the cost of transferring training. One solution to this problem is to instrument the decision to move out of the training occupation. Two instruments that we considered were whether the worker was married (assumed to be correlated with the value of leisure) and employment levels in the training occupation (assumed correlated with the arrival rate of offers from within the training occupation when unemployed). Whilst these variables entered the second-stage regressions with the right sign, they were rarely significant and so the first-stage estimates were extremely unstable.

Hence as an additional check on the robustness of our results regarding transferability, we use the answers from a question contained in the Qualifications and Careers Survey (QaC) data. The survey is cross-sectional, but it

asks workers a number of retrospective questions that enable us to identify the training occupation of the worker (see appendix for details). In particular, the survey asks workers: “How much of the occupational knowledge and skills you acquired during apprenticeship can you still apply in your current work?” The answer can be “very little or nothing at all; a little; some; quite a lot, a lot”. Since it is hard to see how workers could interpret this as anything other than a question concerning the actual value of apprenticeship skills in the new job, it should act as a direct measure of transferability that is not affected by the value of the match in the new occupation.

Table 7 presents the answers to this question based on a sample of similar workers (German male apprentices without the *Abitur* under the age of 35). From the top panel, we see that overall, almost two-thirds respond that they are using ‘many’ or ‘very many’ of their apprenticeship skills, with the remaining third using ‘some’, ‘few’ or ‘very few or none at all’. When we split this group into those working inside and outside of the training occupation, the results are very interesting. Amongst those working inside their training occupation, almost 85% claim to be using ‘many’ or ‘very many’ of the skills acquired during apprenticeship. The figure for those outside of the apprenticeship occupation is just under 40%. Hence it is clear that apprenticeships are occupational. But it is interesting to note that even amongst those outside of the training occupation, only one-quarter claim to be using ‘very few or none’ of their skills.

In Panel B we break the movers down according to the distance moved. The results are very dramatic. Amongst those that move at a 3-digit level, only 1.4% claim to be using ‘very few or none’ of their skills, whilst 45.7% claim to be using ‘very many’. However, for those moving at the 1-digit level, the pattern is exactly reversed, with 30.31% claiming to use ‘very few or none’ and only 16.74% claiming to use ‘very many’. We find a similar pattern of results when interacting occupational moves with switches in two-digit industry (Panel C). Overall, these skill use results reinforce the results based on earnings: apprenticeship training is completely transferable within a broad occupational group (e.g. a 1-digit occupation), but perhaps only one-third of apprenticeship training is transferable outside of this group.

8 Conclusions

The paper began by stressing that workplace training has indirect as well as direct costs: namely, that it can prevent productive job-shopping. However, our model provided conditions under which firms will take account of expected future mobility and provide training that is to some extent transferable. In the empirical part of the paper we painted a positive picture of GAT. Trainees receive training in an occupation, with a lower bound to annual returns we estimated around 6%. The training is sufficiently trans-

ferable to enable them to transfer all of these skills across a wide range of occupations and in a worst-case scenario, one-third of this skills outside of this group. In line with these findings, patterns of occupational mobility suggest mobility from the training occupation is the norm rather than the exception, although mobility outside of the 1-digit occupational group is lower than that among non-apprentices.

Given the picture we have painted, it would appear that workplace training is a promising alternative to traditional classroom-based routes to skills. This begs the question of why countries such as the UK and US do not have similar programs. Two sources of market failure suggest themselves. First, it may be that young school-leavers in other countries do not have the means to finance this kind of training since, for example, living in the parents' home into young adulthood is not the norm. In other words, our assumption that young workers are not credit-constrained and are willing and able to pay for their own training (via the acceptance of lower wages during training) may need to be relaxed in other contexts. Another source of market failure is the possibility that training can not be written into a contract. Then, interesting work by Malcomson, Maw, and McCormick (2000) shows that there may be a role for the regulation of contract length to ensure that the period in which workers earn lower wages is sufficiently long to allow firms to recoup training expenses. Only then will they have incentives to train. Since German apprenticeship training contracts are regulated in this manner, this presents an interesting agenda for policy-makers interested in the design of workplace training programs for non-college-bound youth.

A Mathematical Appendix

We begin by deriving expressions for the three components of the social returns described on the right-hand side of equation (5). The stay/leave choice faced by the newly trained worker (as described in section 2.2) gives rise to the following *expected* probability of leaving:

$$P(\textit{leave}) = P(\xi_A - \xi_B > \alpha H(\gamma - 1))$$

where we suppress the dependence of α on q . Hence we need to derive the distribution of $D = \xi_A - \xi_B$. Recall that $\xi_j \sim U[-\frac{\bar{\varepsilon}}{2}, \frac{\bar{\varepsilon}}{2}]$ where $j \in \{A, B\}$. Then it is relatively easy to show (using a method similar to that in Mood, Graybill, and Boes (1974) chapter V) that:

$$f_D(d) = \begin{cases} \frac{\bar{\varepsilon}+d}{\bar{\varepsilon}^2} & \textit{if} \quad -\bar{\varepsilon} \leq d < 0 \\ \frac{\bar{\varepsilon}-d}{\bar{\varepsilon}^2} & \textit{if} \quad 0 \leq d \leq \bar{\varepsilon} \end{cases}$$

so that D has a triangular density function, and that:

$$F_D(d) = \begin{cases} \frac{(d+\bar{\varepsilon})^2}{2\bar{\varepsilon}^2} & \textit{if} \quad -\bar{\varepsilon} \leq d < 0 \\ \frac{\bar{\varepsilon}^2+2\bar{\varepsilon}d-d^2}{2\bar{\varepsilon}^2} & \textit{if} \quad 0 \leq d \leq \bar{\varepsilon} \end{cases}$$

Now we can use this distribution function to derive:

$$P(\textit{stay}) = P(d > k) = \frac{\bar{\varepsilon}^2 - 2\bar{\varepsilon}k - k^2}{2\bar{\varepsilon}^2}$$

where $k = \alpha H(\gamma - 1)$. Recall that we assumed that firms will never provide enough training to prevent some (expected) turnover. This implies that $-\bar{\varepsilon} < k \leq 0$. Intuitively, when $k \rightarrow -\bar{\varepsilon}$ (training is not transferable), we have that $P(\textit{stay}) \rightarrow 1$. When $k = 0$ (transferable training), we have that $P(\textit{stay}) = \frac{1}{2}$. Turning to expected match values, we have that:

$$\begin{aligned} E(\xi_A|\textit{stay}) &= E(\xi_A|d > k) \\ &= \frac{2\left(\frac{\bar{\varepsilon}^3}{12} - \frac{k^3}{6} - \frac{\bar{\varepsilon}k^2}{4}\right)}{\bar{\varepsilon}^2 - 2\bar{\varepsilon}k - k^2} \\ E(\xi_B|\textit{leave}) &= E(\xi_B|d < k) \\ &= -\frac{k}{3} + \frac{\bar{\varepsilon}}{6} \end{aligned}$$

Hence when $k \rightarrow -\bar{\varepsilon}$, $E(\xi_A|d > k) \rightarrow 0$ and $E(\xi_B|d < k) \rightarrow \frac{\bar{\varepsilon}}{2}$. In other words, when training is not transferable, the fact that workers stayed in the training occupation tells us little about their match to that occupation, whereas workers who leave the training occupation have an excellent match in their new occupation. When $k = 0$ (transferable training), we have that

$E(\xi_A|d > k) = E(\xi_B|d < k) = \frac{\bar{\varepsilon}}{6}$, hence expected matches are positive and identical for movers and stayers.

We can now substitute all of these expressions into the social returns function (the right-hand-side of equation (5) without costs) to derive:

$$\begin{aligned} SW + C(H, \gamma) &= \alpha H + kP(\text{leave}) \\ &\quad + [P(\text{stay})E(\xi_A|\text{stay}) + P(\text{leave})E(\xi_B|\text{leave})] \\ &= \alpha H + k \frac{(k + \bar{\varepsilon})^2}{2\bar{\varepsilon}^2} + \frac{\frac{\bar{\varepsilon}^3}{6} - \frac{k^2\bar{\varepsilon}}{2} - \frac{k^3}{3}}{\bar{\varepsilon}^2} \end{aligned} \quad (\text{A1})$$

Recall that the second term is the loss of skills due to expected turnover. It can be shown that this loss follows an inverted U shape over the range $-\bar{\varepsilon} < k \leq 0$. When $k \rightarrow -\bar{\varepsilon}$, there is no loss since there is no turnover. When $k=0$, there is no loss since training is general ($\gamma = 1$). In between these two values this term is negative. Turning to the third term - the expected gains from matching - when $k = 0$, and training is transferable, the expected gain is $\frac{\bar{\varepsilon}}{6}$ (as already noted). As k decreases (training becomes less transferable), this term decreases, until it tends to zero as $k \rightarrow -\bar{\varepsilon}$ and workers become ‘trapped’ in their training occupation. Finally, in proving the propositions, we will make use of the following results:

$$\begin{aligned} SW_k &= \frac{d}{dk} \left\{ k \frac{(k + \bar{\varepsilon})^2}{2\bar{\varepsilon}^2} + \frac{\frac{\bar{\varepsilon}^3}{6} - \frac{k^2\bar{\varepsilon}}{2} - \frac{k^3}{3}}{\bar{\varepsilon}^2} \right\} \\ &= \frac{(k + \bar{\varepsilon})^2}{2\bar{\varepsilon}^2} > 0 \\ SW_{kk} &= \frac{k + \bar{\varepsilon}}{\bar{\varepsilon}^2} > 0 \\ SW_{k\bar{\varepsilon}} &= -k \frac{(k + \bar{\varepsilon})}{\bar{\varepsilon}^3} > 0 \end{aligned}$$

We are now in a position to prove Proposition 1.

A.1 Proof of Proposition 1

Differentiating the social welfare function with respect to H and γ we have the first-order conditions for a maximum:

$$\begin{aligned} SW_H &= \alpha + SW_k \frac{dk}{dH} - C_H \\ &= \alpha + SW_k \alpha (\gamma - 1) - C_H = 0 \end{aligned} \quad (\text{A2})$$

$$SW_\gamma = SW_k \alpha H - C_\gamma = 0 \quad (\text{A3})$$

We first show that $H \neq 0$ using SW_H . Suppose $\gamma \geq 0$. Then $k = 0$ and $SW_k = \frac{1}{2}$ hence $\alpha[1 + SW_k(\gamma - 1)] > 0$. But given our assumptions regarding the cost function, $C_H(0, \gamma) = 0$. Hence a contradiction. Now suppose that $\gamma = 0, H > 0$. Then $SW_k > 0$ (since $-\bar{\varepsilon} < k = -\alpha H$) but again, $C(H, 0) = 0$ hence a contradiction. *QED.*

A.2 Matching and Optimal Training Levels

Taking total differentials of both of the first-order conditions, dividing through by $d\bar{\varepsilon}$ and rearranging gives:

$$\frac{dH}{d\bar{\varepsilon}} = \frac{-SW_{H\gamma}SW_{\gamma\bar{\varepsilon}} + SW_{H\bar{\varepsilon}}SW_{\gamma\gamma}}{SW_{\gamma H}SW_{H\gamma} - SW_{HH}SW_{\gamma\gamma}} \quad (\text{A4})$$

$$\frac{d\gamma}{d\bar{\varepsilon}} = \frac{SW_{\gamma\bar{\varepsilon}}SW_{HH} - SW_{\gamma H}SW_{H\bar{\varepsilon}}}{SW_{\gamma H}SW_{H\gamma} - SW_{HH}SW_{\gamma\gamma}} \quad (\text{A5})$$

From the second-order conditions for a maximum, the denominator must be negative as must the terms $SW_{\gamma\gamma}$ and SW_{HH} . We can also show that:

$$\begin{aligned} SW_{H\bar{\varepsilon}} &= SW_{k\bar{\varepsilon}}\alpha(\gamma - 1) \\ &= -k\frac{(k + \bar{\varepsilon})}{\bar{\varepsilon}^3}\alpha(\gamma - 1) \\ &< 0 \end{aligned}$$

This says that more (expected) turnover induced by an increase in the importance of matching reduces the effectiveness of training investments (at given levels of transferability). And:

$$\begin{aligned} SW_{\gamma\bar{\varepsilon}} &= SW_{k\bar{\varepsilon}}\alpha H \\ &= -k\frac{(k + \bar{\varepsilon})}{\bar{\varepsilon}^3}\alpha H \\ &> 0 \end{aligned}$$

Hence when $SW_{H\gamma} < 0$, we have that ($\frac{dH}{d\bar{\varepsilon}} < 0$, $\frac{d\gamma}{d\bar{\varepsilon}} > 0$), whereas when $SW_{H\gamma} > 0$, we can not rule out that ($\frac{dH}{d\bar{\varepsilon}} < 0$, $\frac{d\gamma}{d\bar{\varepsilon}} < 0$) or ($\frac{dH}{d\bar{\varepsilon}} > 0$, $\frac{d\gamma}{d\bar{\varepsilon}} > 0$). We can however rule out the case of ($\frac{dH}{d\bar{\varepsilon}} > 0$, $\frac{d\gamma}{d\bar{\varepsilon}} < 0$) since simple manipulation of the numerators in (A4) and (A5) show that this will result in a violation of the second-order condition: $SW_{\gamma H}SW_{H\gamma} - SW_{HH}SW_{\gamma\gamma} < 0$. Note that:

$$SW_{H\gamma} = \alpha\frac{(k + \bar{\varepsilon})(3k + \bar{\varepsilon})}{2\bar{\varepsilon}^2} - C_{H\gamma}$$

so that the sign of this term will depend *inter alia* on the properties of the cost function.

A.3 Proof of Proposition 2

Recall that we assume that $\alpha = \alpha(q)$, where $\alpha'(q) > 0$. Now, taking total differentials of both of the first-order conditions, dividing through by $d\alpha$ and rearranging gives:

$$\begin{aligned} \frac{dH}{d\alpha} &= \frac{-SW_{H\gamma}SW_{\gamma\alpha} + SW_{H\alpha}SW_{\gamma\gamma}}{SW_{\gamma H}SW_{H\gamma} - SW_{HH}SW_{\gamma\gamma}} \\ \frac{d\gamma}{d\alpha} &= \frac{-SW_{\gamma H}SW_{H\alpha} + SW_{\gamma\alpha}SW_{HH}}{SW_{\gamma H}SW_{H\gamma} - SW_{HH}SW_{\gamma\gamma}} \end{aligned}$$

We can show that:

$$\begin{aligned} SW_{H\alpha} &= 1 + (\gamma - 1)(SW_{\gamma H}) \\ &= 1 + (\gamma - 1) \left[\frac{(k + \bar{\varepsilon})(3k + \bar{\varepsilon})}{2\bar{\varepsilon}^2} \right] > 0 \end{aligned}$$

and it is easily shown that $SW_{\gamma H}$ and $SW_{\gamma\alpha}$ must take the same sign. Then, using the second-order conditions, it must be the case that $\frac{dH}{d\alpha} > 0$ whilst the sign of $\frac{d\gamma}{d\alpha}$ depends on $SW_{\gamma H}$, which we have shown can not be signed. *QED.*

A.4 Proof of Proposition 3

We can use the envelope theorem to evaluate a change in q on social welfare at optimum levels of H and γ :

$$\begin{aligned} \frac{dSW(H, \gamma)}{dq} &= H + SW_k(\gamma H - H) \\ &= H[1 + (\gamma - 1)SW_k] \\ &= H[1 + (\gamma - 1)\frac{(k + \bar{\varepsilon})^2}{2\bar{\varepsilon}^2}] \\ &> 0 \end{aligned}$$

Hence social welfare is a continuous (by inspection of equation (A1)) and increasing function of q . Hence there must exist a number q^+ such that $SW(q) > 0$ for $q > q^+$ and $SW(q) < 0$ otherwise. *QED.*

B Data Appendix

B.1 IAB Data

We use data from the German Institute for Employment Research (IAB) for the years 1975-1995. The basis of the IAB employment subsample is the integrated notifying procedure for health insurance, statutory pension scheme and unemployment insurance which is regulated through German legislation. The procedure requires that employers report all information of their employees registered by the social security system to the social security agencies. One problem with the data is that prior to 1984, firms were not obliged to report extra payments such as Christmas and holiday bonuses. Since these are an important part of compensation in Germany, all of our earnings equations are estimated using data from 1984 onwards. Employers have to notify the beginning and the end of an employment spell and have to give an annual notification for each employee. The employment statistics include all employees obliged to pay social insurance contributions. The employment statistics do not include, among others, civil servants, family

workers, those in marginal employment, and students enrolled in higher education (Cramer (1985)). For 1995, the employment statistics cover nearly 79.4% of all employed persons in Western Germany (Bender, Haas, and Klose (2000)).

The notification provides information on individual characteristics as gender, year of birth, number of children and qualifications. Furthermore it reports information on the employment including information on the occupational code, the occupational status, the establishment number of the employer with information on the size and the industry of the employer, and finally the gross earnings of the employee over the past employment spell which served as the basis for social security contributions. Importantly, this allows us to infer the occupation trained in and the size of the training firm. This information is passed on from the social insurance agencies to the Federal Employment Services and collected in the so called *historic file*. The IAB employment subsample is an anonymised 1% sample from the historic file. Details of the anonymisation procedure are described in Bender, Haas, and Klose (2000). One effect of the anonymisation procedure is the top coding of the data. But the top coding affects only a tiny proportion of the young apprentices in our sample. Due to the fact that the information for East Germany is only available for the time after unification we use only the information of notifications for people working in Western Germany. The employment subsample contains a total of 7,847,553 notifications with 6,711,153 notifications for Western Germany. On the basis of the final notifications in each case, the file provides information of 483,327 Western Germans (Bender, Haas, and Klose (2000), p.2).

Apart from information in the historic file the IAB employment subsample contains information from two other data sources. The benefits recipients file contains person-related information on periods in which the Federal Employment Service paid benefits like the status of the unemployed and the type of benefit payments (unemployment benefit, unemployment assistance or maintenance payments for participating in training or re-training programs). But not all spells of registered non-employment were covered (Bender, Haas, and Klose (2000)). The second file which adds information to IAB employment subsample is the establishment file. The file provides additional information on the notifying establishment as the date of birth and death of the establishment as well as generated information on the pattern of skill levels of employees within the establishment.

B.2 QaC Data

We use the 1991/1992 Qualifications and Careers Survey (QaC) data of the Bundesinstitut fuer Berufsbildung (BiBB) and the Institut fuer Arbeitsmarkt- und Berufsforschung (IAB). This data set asks a random sample of the working population (excluding persons currently enrolled in an apprenticeship,

people on military or civil service, and helping family members) about their qualification, job career, workplace conditions, job satisfaction as well as activities in formal and informal education. Similar surveys exist for the years 1979 and 1985/86, but all surveys are cross-sections. The data set collects in total information on 34277 individuals - 24090 for West Germany and 10187 for East Germany. We choose a sub-sample of male employees residing in West Germany, without an academic degree, who completed an apprenticeship, which lasts longer than 24 months. In addition we restrict our focus on employees younger than 35 years of age.

References

- ACEMOGLU, D., AND J.-S. PISCHKE (1998): "Why Do Firms Train? Theory and Evidence," *Quarterly Journal of Economics*, 113(1), 79–119.
- ANGRIST, J. D., AND A. B. KRUEGER (1991): "Does Compulsory Schooling Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106, 979–1014.
- BAILY, N. M., G. BURTLESS, AND R. LITAN (1992): *Growth with Equity: Economic Policymaking for the Next Century*. Washington DC: Brookings Institution.
- BASSI, L. J., AND J. LUDWIG (2000): "School-to-Work Programs in the United States: A Multi-Firm Case Study of Training, Benefits and Costs," *Industrial and Labor Relations Review*, 53(2), 219–239.
- BECKER, G. S. (1993): *Human Capital*. University of Chicago Press, Chicago, 3rd edn.
- BENDER, S., C. DUSTMANN, D. MARGOLIS, AND C. MEGHIR (1999): "Worker Displacement in France and Germany," The Institute for Fiscal Studies: Working Paper Series: No. W99/14.
- BENDER, S., A. HAAS, AND C. KLOSE (2000): "IAB Employment Subsample 1975-1995: Opportunities for Analysis Provided by the Anonymised Subsample," IZA Discussion Paper No. 117.
- BURTLESS, G. (1994): "Comment on Heckman et al, "US Education and Training Policy: A Re-Evaluation of the Underlying Assumptions Behind the New Consensus",", in *Labor Markets, Employment Policy, and Job Creation*, ed. by L. C. Solomon, and A. R. Levenson, pp. 123–139. Milken Institute Series in Economics and Education. Boulder and London: Westview Press in cooperation with the Milken Institute for Job and Capital Formation.
- CARD, D. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labour Economics Volume 3A*, ed. by O. Ashenfelter, and D. Card. Elsevier, Amsterdam.
- CASEY, B. (1991): "Recent Developments in the German Apprenticeship System," *British Journal of Industrial Relations*, 29(2), 205–222.
- CHANG, C., AND Y. WANG (1996): "Human Capital Investment under Asymmetric Information: The Pigovian Conjecture Revisited," *Journal of Labor Economics*, 14(3), 505–519.

- CRAMER, U. (1985): "Probleme der Genauigkeit der Beschäftigtenstatistik," *Allgemeines Statistisches Archiv*, 69, 56–68.
- DUSTMANN, C., AND C. MEGHIR (1999): "Wages, Experience and Seniority," Institute for Fiscal Studies (IFS) Working Paper 99/1.
- FEDERAL MINISTRY OF EDUCATION AND SCIENCE (1992): *Vocational Training in the Dual System in the Federal Republic of Germany* Federal Ministry of Education and Science, Bonn.
- HARHOFF, D., AND T. J. KANE (1993): "Financing Apprenticeship Training: Evidence from Germany," NBER Working Paper No. 4557.
- (1997): "Is the German Apprenticeship System a Panacea for the US Labour Market?," *Journal of Population Economics*, 10, 171–196.
- HECKMAN, J. J. (1999): "Doing It Right: Job Training and Education," *The Public Interest*, Spring, 86–107.
- HECKMAN, J. J., R. L. ROSELIUS, AND J. A. SMITH (1994): "'U.S. Education and Training Policy: A Re-Evaluation of the Underlying Assumptions Behind the 'New Consensus.'", in *Labor Markets, Employment Policy, and Job Creation.*, ed. by L. C. Solmon, and A. R. Levenson, pp. 83–121. Milken Institute Series in Economics and Education. Boulder and London: Westview Press in cooperation with the Milken Institute for Job and Capital Formation.
- HUGHES, K. L., T. R. BAILEY, AND M. J. MECHUR (2001): "School-to-Work: Making a Difference in Education: A Research Report to America," Institute on Education and the Economy, Teachers College, Columbia University.
- KRUEGER, A., AND C. ROUSE (1998): "The Effect of Workplace Education on Earnings, Turnover, and Job Performance," *Journal of Labor Economics*, 16(1), 61–94.
- LYNCH, L. M. (1992): "Private Sector Training and the Earnings of Young Workers," *American Economic Review*, 82(1), 299–312.
- MALCOMSON, J. M., J. W. MAW, AND B. MCCORMICK (2000): "General Training by Firms, Apprentice Contracts, and Public Policy," Working Paper, University of Southampton.
- MCCALL, B. P. (1990): "Occupational Matching: A Test of Sorts," *The Journal of Political Economy*, 98(1), 45–69.
- MILLER, R. (1984): "Job Matching and Occupational Choice," *The Journal of Political Economy*, 92(6), 1086–1120.

- MOOD, A. M., F. A. GRAYBILL, AND D. C. BOES (1974): *Introduction to the Theory of Statistics*. McGraw-Hill, Auckland.
- MORTENSEN, D. (1986): "Job Search and Labour Market Analysis," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and R. Layard, pp. 849–919. Elsevier Science Publishers BV.
- NEAL, D. (1995): "Industry-Specific Human Capital: Evidence from a Sample of Displaced Workers," *Journal of Labor Economics*, 17(4), 237–261.
- (1999): "The Complexity of Job Mobility Amongst Young Men," *Journal of Labor Economics*, 17(2), 237–161.
- RYAN, P. (2001): "The School-to-Work Transition: A Cross-National Perspective," *Journal of Economic Literature*, 39(1), 34–92.
- SCOONES, D. (2000): "Matching and Competition for Human Capital," *Labour Economics*, 7, 135–152.
- STEEDMAN, H. (1993): "The Economics of Youth Training in Germany," *The Economic Journal*, 103(420), 1279–1291.
- STEEDMAN, H., A. GREEN, O. BETRAND, A. RICHTER, M. RUBIN, AND K. WEBER (1997): "Assessment, Qualifications and Standards: The UK Compared to France, Germany, Singapore and the US," London.
- STEEDMAN, H., F. GREEN, AND P. RYAN (1998): "Apprenticeship - a Strategy for Growth," London.
- STEVENS, M. (1994): "A Theoretical Model of on-the-Job Training with Imperfect Competition," *Oxford Economic Papers*, 46, 537–562.
- TOPEL, R. H., AND M. P. WARD (1992): "Job Mobility and the Careers of Young Men," *Quarterly Journal of Economics*, 107(2), 439–479.
- WERWATZ, A. (1998): "Mobility After Apprenticeship - How Effective is German Apprenticeship Training?," SFB 373 Discussion Paper 97-55, Humboldt University, Berlin.
- WINKELMANN, R. (1994): "Training, Earnings Mobility and Skill Acquisition," CEPR Discussion Paper No. 982.

Table 1: Descriptive Statistics

	1985	1987	1989	1991	1993	1995
Daily Wage (1995 DM)	102.6	114.1	122.1	131.9	136.6	139.9
Age	23.1	24.2	25.4	26.6	27.9	29.2
Experience	4.82	5.70	6.74	7.76	8.96	10.17
Tenure	2.49	2.85	3.21	3.64	4.29	4.79
Apprenticeship	0.772	0.791	0.785	0.798	0.822	0.830
Meister Qualification	.00690	0.0103	0.0141	0.0178	0.0221	0.0242
N	23725	28919	34279	37670	37981	38765

Table 2: Descriptive Statistics for Sample of Workers Inside 3-digit Apprenticeship Occupation Prior to Separation

	'Quits'	'Displaced'	'Close 2'	'Close 1'
Pre-Separation Tenure	2.26	1.41	1.65	1.87
Age	24.61	24.34	24.54	24.68
Experience	5.47	5.23	5.46	5.74
Pre-Separation Wage	114.82	102.98	105.12	105.34
Average Length of Unemployment (Yrs)	---	0.296	0.309	0.314
N	14279	4893	408	161

Notes: See text for definitions of 'Quit', 'Displaced', 'Close 2' and 'Close 1'. Each column is based on sample of workers observed in the apprenticeship occupation pooled across years 1984-1995.

Table 3: Comparison of Occupational Mobility amongst German Apprentices and Non-Apprentices

	App.	Non-App.	App.	Non-App.	App.	Non-App.
	1985	1985	1991	1991	1995	1995
Proportion Out of Training/First Occupation:	0.481	0.6853	0.570	0.7825	0.603	0.8067
...at 3-digit Level	0.0516	0.0285	0.0540	0.0262	0.0543	0.0270
...at 2-digit Level	0.0974	0.0921	0.0966	0.0988	0.0950	0.1101
...at 1-digit Level	0.332	0.5647	0.419	0.6575	0.454	0.6696
Proportion Out of 2-digit Training/First Industry and Training/First Occupation:	0.344	0.5721	0.430	0.6742	0.461	0.6925
...at 3-digit Level	0.0255	0.0143	0.0271	0.0140	0.0281	0.0141
...at 2-digit Level	0.0601	0.0683	0.0651	0.0727	0.0666	0.0819
...at 1-digit Level	0.259	0.4895	0.338	0.5876	0.367	0.5965
Proportion leaving Training/First Firm	0.633	0.7066	0.714	0.8074	0.753	0.8336
Amongst those with Training/First Firm:						
<i>Proportion Out of Train./First Occupation:</i>	0.201	0.1894	0.240	0.2459	0.266	0.2758
Amongst those left Training/First Firm:						
<i>Proportion Out of Train./First Occupation:</i>	0.643	0.8912	0.702	0.9105	0.713	0.9127
N	18312	5399	30063	7582	32174	6582

Notes: We impose the same age and education (with the exception of apprenticeship of course) as for the apprenticeship sample on selection of the non-apprentices sample. We consider only non-apprentices who we do never observe completing an apprenticeship (even at a later stage of their work life).

Table 4: Returns to Apprenticeship Training

	1985	1987	1989	1991	1993	1995
Panel A: Return to `Intensive' Apprenticeship Training						
Apprenticeship Inside Occupation	0.122	0.127	0.166	0.144	0.155	0.163
Panel B: Estimates of `Standard' Return to Apprenticeship Training						
Experience	0.0958	0.113	0.104	0.119	0.106	0.0988
Experience ²	-0.00263	-0.00468	-0.00370	-0.00443	-0.0037	-0.00325
Tenure	0.00419	0.009	0.00879	0.0104	0.0124	0.0140
Apprenticeship	0.168	0.167	0.213	0.195	0.189	0.184
Meister	<i>-0.019</i>	0.159	0.152	0.176	0.144	0.155
N	22591	27578	32516	35739	35941	37140

Notes: Estimates in the first panel are based on the same equation as those presented in the first panel, with the apprenticeship variable referring only to those with apprenticeship inside the occupation worked in. All estimates with the exceptions of those in italics significant at the 1% level, where t-ratios based on robust standard errors.

Table 5: Returns Broken Down by Training Firm Size

Training Firm size	Propn with good Math. scores	Propn trained in this size of firm	Return Inside Training Occupation	Propn of Time spent Training	Estimated Annual Return to Training
1	0.110	0.023	0.0726	0.56	4.71%
2-9	0.107	0.225	0.0904	0.56	5.87%
10-19	0.136	0.141	0.0900	0.61	5.37%
20-49	0.136	0.148	0.109	0.61	6.50%
50-99	0.136	0.089	0.154	0.69	8.12%
100-499	0.154	0.168	0.202	0.69	10.65%
500-999	0.160	0.063	0.190	0.81	8.53%
>1000	0.172	0.144	0.270	0.81	12.12%

Notes: Data in the second column from Harhoff and Kane (1993), Table 7. Their table uses an identical size breakdown with the two exceptions: in their Table, rows 1 and 2 correspond to firms of size 1-4 and 4-9 respectively. Secondly, their data aggregates firm sizes 10-49. We assume the figures for sizes 10-19 and 20-49 are identical. Figures reported in the third and fourth columns are based on data pooled over the years 1984 to 1995. The return inside the apprenticeship occupation is derived from an equation identical to that used to estimate these returns in Table 4, except that the apprenticeship variable is interacted with firm size and year dummies are included (sample size 374,710). All estimates significant at the 1% level. Estimates of the proportion of apprenticeship time spent training are based on unpublished results from a study on the costs of apprenticeship for firms in West Germany in 1991, by the Federal Institute for Vocational Education (BiBB). We thank Ursula Beicht of the BiBB for making this information available. Estimates of the rate of return are calculated by dividing the estimated coefficient by the average numbers of years spent training and the estimated proportion of apprenticeship time spent training (inside or outside of the training firm). The average numbers of years spent training is assumed as 2.75 years.

Table 6: Wage Penalty to Moving Out of the Apprenticeship Occupation

	‘Quits’	‘Displaced’	‘Close 2’	‘Close 1’
Panel A: Wage Penalty to Leaving the Apprenticeship Occupation				
	---	-0.132 ^{***} (0.0318)	-0.107 (0.111)	-0.352 ^{***} (0.0134)
Change in Experience	0.116 ^{***} (0.00714)	0.137 ^{***} (0.00785)	0.0899 ^{***} (0.0288)	0.174 ^{***} (0.0552)
Change in Experience ²	-0.00488 ^{***} (0.000539)	-0.00691 ^{***} (0.000622)	-0.00373 ^{**} (0.00209)	-0.00579 [*] (0.00399)
Change in Tenure	0.00268 ^{***} (0.000904)	0.00787 ^{***} (0.00323)	-0.00149 (0.0114)	-0.0101 (0.0175)
Move Out	-0.000552 (0.00619)	-0.0165 [*] (0.0117)	0.0300 (0.0381)	-0.0349 (0.061)
N	14279	4893	408	161
Panel B: Wage Penalty to Leaving the Apprenticeship Occupation by Distance (measured by Occupational Codes)				
3-digit only	0.000641 (0.0128)	0.0208 (0.0242)	0.374 ^{***} (0.122)	0.360 ^{***} (0.136)
2-digit only	0.0484 ^{***} (0.0124)	0.0424 ^{***} (0.0216)	0.0690 [*] (0.0632)	0.033 (0.0965)
1-digit	-0.0128 ^{**} (0.00738)	-0.0352 ^{***} (0.0132)	-0.0114 (0.0420)	-0.103 [*] (0.0708)
Panel C: Wage Penalty to Leaving the Apprenticeship Occupation and Switching 2-digit Industry (occupational distance measured by Industrial and Occupational Codes)				
Move Out (All)	-0.0199 ^{***} (0.00730)	-0.0397 ^{***} (0.0124)	-0.0239 (0.0414)	-0.0619 (0.0705)
3-digit only	-0.00235 (0.0196)	0.00496 (0.0307)	0.283 ^{***} (0.108)	0.459 ^{***} (0.132)
2-digit only	0.0326 ^{***} (0.0164)	0.0172 (0.0263)	-0.0359 (0.0738)	-0.106 (0.205)
1-digit	-0.0320 ^{***} (0.00834)	-0.0541 ^{***} (0.0137)	-0.0400 (0.0449)	-0.0950 [*] (0.0764)

Notes: see text for definitions of ‘Quits’, ‘Displaced’, ‘Close 2’ and ‘Close 1’. Panel B and C are estimated using the same equation as Panel A, with moves out of the apprenticeship occupation disaggregated by distance moved. Statistical significance at the 1% (5%, 10%) level denoted *** (**, *).

Table 7: How many Apprenticeship Skills used in the Current Job?

	Very Few or None	Few	Some	Many	Very Many	N
	Panel A: Analysis of Skill Use according to whether Working in Apprenticeship Occupation					
All	11.06	9.50	15.01	21.74	42.69	3317
Working in Training Occupation	0.87	3.80	10.79	25.87	58.68	1844
Working Out of Training Occupation	23.83	16.63	20.30	16.56	22.67	1473
	Panel B: Analysis of Movers according to Distance Measured by Codes					
3-digit Level	1.44	10.58	14.42	27.88	45.67	208
2-digit Level	12.24	12.24	25.00	19.90	30.61	196
1-digit Level	30.31	18.62	20.58	13.75	16.74	1069
	Panel C: Analysis of Occupational Switchers and Industry Switchers (2-digit level) according to Distance Measured by Codes					
All	33.41	20.26	21.14	12.92	12.27	913
3-digit Level	1.43	24.29	17.14	30.00	27.14	70
2-digit Level	19.15	17.02	26.60	15.96	21.28	94
1-digit Level	38.18	20.29	20.83	10.95	9.75	749

Notes: Column headings are answers to question ‘How much of your occupational knowledge and skills, which you have obtained in your apprenticeship, can you actually use in your current job?’ from QaC data. See text for details.

Table A1: Industry Classification as Used in Empirical Analysis

	Industry	Two digit ES-classification*	PSID equivalent
1	Agriculture, Forestry	00-02	17-27
2	Fishing	03	28
3	Energy	04	377, 378, 467-479
4	Mining	05-08	47-57
5	Chemical	09	347, 357-369
6	Synthetics	10-13	348, 349, 379-387
7	Earth/Clay/Stone	14-16	119-138
8	Iron/Steel	17-21	139-169
9	Mechanical Engineering	22-32	177-198, 219-238
10	Electrical Engineering	33-39	199-209, 239-259
11	Wood/Paper/Printing	40-44	107-118, 328-339
12	Clothing/Textiles	45-53	307-327, 388-398
13	Food industry	54-58	268-299
14	Construction/ Construction related	59-61	
15	Trade	62	507-588, 607-698
16	Train system	63	407
17	Postal system	64	447-449, 907
18	Other transport	65-68	408-429
19	Financial institutions	69	707-709, 717
20	Restaurants, Service Industry	70-73	777-809
21	Education/Sport	74-77	857-869
22	Health Service	78	828-848
23	Legal Services	79	718, 849
24	Other Services	80-86	727-759, 888-897
25	Non Profit (Voluntary/Church/Private Households)	87-90	877-887, 769
26	Public Institutions (Regional Authority/ Social Security)	91-94	917-937

Notes: Classification in the list of industries used for the statistics of the Federal Employment Service in Germany (1973 edition).

Figure 1: Probability of Staying in Apprenticeship Occupation

