STEP rapport / report

ISSN 0804-8185

R-07

•

2002

Tore Sandven

The Educational Attainment Of Employees As An Indicator Of The Innovation Capacities Of Enterprises

Tore Sandven
STEP
Hammersborg torg 3
N-0179 Oslo
Norway

Rapport fra prosjektet ”Kompetanse, mobilitet og verdiskapning” finansiert av Norges Forskningsråd, FAKTA-programmet

Oslo, Januar 2002

STEP group

Studies in technology, innovation and economic policy
Studier i teknologi, innovasjon og økonomisk politikk

Hammersborg torg 3, 0179 Oslo, Norway
Telephone +47 2286 8010
Fax +47 2286 8049
http://www.step.no/

STEP publiserer to ulike serier av skrifter: Rapporter og Arbeids-notater.

Publikasjonene i begge seriene kan lastes ned gratis fra våre internettsider.

STEP Rapportserien

I denne serien presenterer vi våre viktigste forskningsresultater. Vi offentliggjør her data og analyser som belyser viktige problemstillinger relatert til innovasjon, teknologisk, økonomisk og sosial utvikling, og offentlig politikk.

STEP maintains two diverse series of research publications: Reports and Working Papers.

Both reports and working papers can be downloaded at no cost from our internet web site.

The STEP Report Series

In this series we report our main research results. We here issue data and analyses that address research problems related to innovation, technological, economic and social development, and public policy.

Redaktører for seriene:
Editors for the series:
Finn Ørstavik (1998-2000)
Per M. Koch (2000)

Preface

Purpose of the study, the logic of the argument, and summary of results

The purpose of this study is to explore to what extent data on the educational attainment of employees can be used as indicators of the innovation capacities of enterprises. The study is purely statistical: we investigate the statistical relationship between variables expressing the educational attainment of employees and more established indicators of innovation capacity and activity, both on a simple bivariate basis and when we control for other variables. The study thus uses simple correlation analysis as well as multiple regression analysis.

Of the established innovation indicators perhaps the most well known and most often used is expenditures on R&D. R&D expenditures have been used both as indicators of innovation capacity, innovation activity, and innovation performance more generally. Moreover, they have been used to measure the innovation activity and performance of different kinds of societal units, at different levels: enterprises, industries, nations, etc. To make the expenditures comparable across units it is customary to express them as intensities by dividing them on some measure of the total size or total activity of each of the units in question, most commonly as R&D expenditures as a proportion of value added (of GDP in the case of nations) or as a proportion of sales, or also R&D expenditures per employee or (for instance in the case of regions or nations or even larger areas) per inhabitant. In many countries R&D data have been collected on a regular basis for many years.

Other indicators of innovation activity or innovation performance have also been used, for instance the number of patent applications, technological balance of payments, etc.

More recently, many countries have carried out innovation surveys, where a large number of enterprises have been asked a wide variety of questions concerning innovation activities, including R&D expenditures. Notably, this has been done inside the framework of the European Community Innovation Survey (CIS), which was carried out in first in 1992, and then a second time in 1992. In this study we use the data from the Norwegian innovation survey of 1997, which is part of the second Community Innovation Survey (CIS II). These data thus give us several different kinds of innovation indicators, including R&D expenditures.

What we bring into the picture in this study is data on the educational attainment of the employees of each of the enterprises in the innovation survey. We have access to public register data on the highest attained education level of all employees of all Norwegian enterprises, and thus of all enterprises participating in the innovation survey. In both data sources, each enterprise has a unique identity number, allowing us to add the variables from one source to the variables from the other. We have education data for several of the most recent years. Since most of the data from the innovation survey refer to the year 1997 (those which do not refer to the three year period 1995-1997), we have here chosen to use the year 1997 also for the education data.

The units of observation in the innovation survey are enterprises. If an enterprise in 1997 has R&D expenditures of 5 million NOK and a total turnover of 100 million NOK, it has an R&D intensity of 5 per cent. Thus, R&D intensity is here a characteristic of the enterprises. One enterprise may have an R&D intensity of 5 per cent, another 2 per cent, yet another 0 per cent, i.e. no R&D expenditures at all. Thus the enterprises are the units of observation, R&D intensity is a variable expressing a characteristic of the different enterprises, along with other variables expressing other characteristics of the enterprises.

In the same way we make the educational attainment of employees into characteristics of enterprises. If an enterprise has altogether 100 employees and 5 of these are engineers (according to some definition to be more precisely determined), we may say that the engineer intensity of the enterprise is 5 per cent. Thus, both R&D intensity and engineer intensity are here characteristics of enterprises: one enterprise may have an R&D intensity of 5 per cent and an engineer intensity of 5 per cent, another an R&D intensity of 1 per cent and an engineer intensity of 4 per cent, and so on. Both R&D intensity and engineer intensity are variables classifying the enterprises according to different characteristics. Our education data allow us to construct many different education variables, classifying the enterprises according to the proportion of their employees which have different kinds of highest attained formal education, and according to whether or not they have employees with different kinds of educational characteristics at all. Likewise, we have many variables from the innovation survey, classifying the enterprises according to different characteristics relating to innovation activities.

<u>Research questions</u>

Given this background, the present study addresses two main research questions.

The first main research question asks to what extent the variables measuring the educational attainment of employees may be used as indicators of the innovative capabilities of the enterprises who employ them. Basically we make this into a question of bivariate relationships between education variables and innovation variables. Here we look at several innovation indicators. In particular we focus on the relationship between educational attainment and R&D. The reason for the special focus on R&D is twofold. In the first place, R&D is more widely used, has been used for a longer time, is more well known and hopefully also more well understood than the other innovation indicators from the innovation survey. In the second place, and largely for the reason just stated, R&D has a special role to play in the context of the second main research question, to be explained below.

A comment is in place here. This procedure, measuring the education variables against the variables from the innovation survey, so to speak, might be interpreted as implying that we take the latter as given, established, ‘true’ indicators of innovative capabilities. This is not how we should see it. The variables in the innovation survey must be seen as part of a larger project of developing better innovation indicators. They must be seen as provisional, in the course of being tested out, improved upon, superseded, but perhaps also discarded. There might easily be other indicators of innovation, perhaps equally valid, perhaps better, which would have a higher correlation with the educational attainment variables than what we find for the variables from the innovation survey. Even if we here use the latter variables as a yardstick for measuring the suitability of the educational attainment variables as indicators of innovative capability, we should rather see this as part of a process of reciprocal validation. If we find no or very few meaningful correlations between the education attainment variables and the innovation survey variables, this should not simply make us reject the former variables as indicators of innovative capacities, but also throw some doubts on the latter. Conversely, to the extent that the education variables do correlate in meaningful ways with the innovation survey variables, this may be seen as partial corroboration of the latter as innovation indicators.

It should be equally clear, however, that it would be highly problematic to go to the opposite extreme and simply take for granted that the education variables will be good indicators of innovative capability. There are at least two reasons for this. Firstly, what matters for the innovative performance of business enterprises is the real competence of employees in the specific situation of the concrete business enterprise. What our education variables measure, on the other hand, is the formal competence of employees in the form of formal educational attainment. It is an open question to what extent formal educational attainment is a reliable measure of the real competence which is at issue here. Secondly, there may be many kinds of competence essential to the successful running of business enterprises which, nevertheless, have little or nothing to do with the capacity for innovation. Indeed, there might even in part be opposition between competence required for the smooth running of efficient routines and competence required for breaking with established routines to launch new ones. Thus, the question of the suitability of the educational attainment variables as indicators of innovative capability should in essence be regarded as an empirical question.

The second main research question is closely related to one of the issues lying behind the introduction of innovation surveys like the CIS. This is the growing gaining of acceptance of the view that there is far more to innovation than R&D (an issue related to the critique of the so called ‘linear model’ of innovation). There are other types of competence essential to innovative capability, other dimensions of innovative capability, than what is reflected in R&D activity and R&D expenditures. The innovation surveys of the CIS type are thus in part devised to find out more about the different dimensions of innovative capability. What do the other dimensions than what is captured by R&D consist in, how do they differ from the dimension or dimensions captured by R&D, what characterizes more precisely the latter dimension (or dimensions).

The second main research question addresses this multi-dimensionality of innovative capability. This second question presupposes that the education attainment variables to some extent reflect innovative capability. It then asks to what extent the education variables express the same dimension of innovative capability as what is captured by R&D and to what extent they express other dimensions of innovative capability. This question essentially involves multivariate analysis. The R&D variables here enter the analysis as control variables: we look at the relationships between education variables and innovation survey variables other than R&D, and investigate to what extent and in what ways these relationships are modified when we control for R&D. We will explain the procedure in more detail below.

Until recently, detailed data on R&D expenditures and other indicators of innovation have only been collected systematically for the manufacturing sector. Most studies using these kind of data have consequently also largely been confined to manufacturing industries. For instance, the OECD classification of industries into high-tech, medium-tech and low-tech, based on their total R&D intensity (for instance, total R&D expenditures as a proportion of total value added in each industry), has only ranked industries within this sector. Only recently has systematic collection of detailed and comprehensive data of this kind been extended to industries outside of manufacturing. Consequently, we know far less about how the R&D indicators and the additional indicators from the innovation surveys function in industries outside of manufacturing. For this reason we have chosen to limit also the present analysis to the manufacturing sector, although our data also cover much of the rest of the economy. An extension of this analysis to sectors outside of manufacturing may be the subject of a subsequent study, but then one should also investigate more thoroughly how the R&D indicators and other innovation indicators function in these industries in the first place: Whether they function largely in the same way as for manufacturing or if there are important differences here.

We should also note an additional limitation of the population: The Norwegian innovation survey only cover enterprises with at least 10 employees.

<u>The logic of the argument</u>

First a short note on the educational attainment variables. We here use variables which measure the proportion of the employees of each enterprise who have different kinds of educational characteristics. In addition, we use dichotomous variables saying whether or not the enterprise in question has employees with each of these educational characteristic at all. Among the higher education we distinguish between four broad types: 1) engineering subjects, 2) other natural science subjects, 3) business administration, accounting, economics, etc., and 4) other higher education (social science, law, medicine, etc.). Normally, only a small proportion of employees will belong to any of these groups. We also wished to have a measure of a more general average level of educational attainment, and here we chose the proportion of employees with at least secondary education, a level which includes all those with higher education. Lastly, we also found it of interest to include a variable on the proportion of employees with craft education, a category which also is included in the category of those with at least secondary education. A more detailed definition and overview of these variables will be found in the main text below.

We start the statistical analysis by looking at the bivariate relationships between the higher education intensity variables and R&D intensity, at the enterprise level. We here find that the two natural science education intensities correlate substantially with R&D intensity. The correlation of R&D intensity with the proportion of employees with at least secondary education is clearly smaller but still of some substance. For the two remaining higher education intensity variables as well as for craft intensity there is either very low correlation or none at all with R&D intensity. Here we also use regression analysis to predict R&D intensity with education variables as independent variables. When we study the effects of the different education variables controlling for the other variables, it turns out that it basically is only the natural science education variables which have an impact. Moreover, the effect of these variables largely remain when we control for the background variables enterprise size and industry.

Here we thus have evidence that the natural science education variables quite clearly may be used as indicators of innovative capability, since they partly express the same dimension of innovative capability as the variables measuring R&D intensity. It remains to be seen whether there in addition is evidence that they express other dimensions of innovative capability. The other education variables do not seem to reflect this R&D dimension of innovative capability to any significant degree.

We have then looked at the relationship at the industry level between R&D intensity and the intensity of the four different higher education types which we distinguish. The total or average R&D intensity of an industry is often used as an indicator of the innovative performance of the whole industry, and thus for ranking the industries in terms of innovative performance. We look at whether we get the same or different rankings of industries if we instead measure the intensity of the different types of higher education at the industry level. We thus use average R&D intensity of each industry and average intensity of the different kinds of higher education to characterize the industries, and then we correlate the industry R&D intensities with the industry intensities of each higher education type. We find that the average intensities of the natural science educations give almost exactly the same industry ranking as the one based on R&D intensity, the correlation being almost 1 (0.96 and 0.97). For the two other higher education intensities the corresponding correlations at the industry level are far lower. However, the impression here is that we not so much have to do with fundamentally different rankings of the industries. The impression in both cases is rather that a couple of atypical outlier industries deviate from a ranking which in large traits is the same as the one based on R&D intensity.

We then go on to the multivariate analyses where we in essence use different innovation variables other than R&D as dependent variable, the education intensity variables as independent variables, and R&D intensity as control variables. As the dependent variable in all these cases is dichotomous, we here use logistic regression analysis. As dependent variable we use, first, whether the enterprise is innovative or not (to be explained more precisely in the main text). This analysis involves all enterprises, and we predict the probability of being innovative. The rest of the analysis involves only the enterprises with innovation. Among these we predict the probability of having product innovations (as opposed to only process innovations), then the probability of having engaged in innovation cooperation, and lastly the probability of having applied for patents.

The procedure in each case is the following. We first use only the education variables to explain the dependent innovation variable. We then control the effect of the education variables for R&D intensity. At an intermediate stage we also control for the background variables enterprise size and industry, so that the R&D variables are entered on top of these background variables. At each stage we assess the addition to the predictive success of the model accounted for by the education variables. Thus, focusing on the relationships between the education variables and the innovation variables other than R&D, we first we assess the effect of the education variables when entered alone, then we control for enterprise size and industry, then we also control for R&D intensity.

We then interpret the results basically in the following way. To the extent that the education variables correlate with the innovation variables, i.e. when entered alone as independent variables, they do express innovative capability. To the extent that their effects are reduced when we control for R&D intensity, they express the same dimension of innovative capability as that expressed by R&D. To the extent that their effects remain after we control for R&D intensity, on the other hand, there is indication that the education variables express other dimensions of innovative capability, not captured by R&D intensity.

We should here also note a simplifying assumption of the present analysis, namely the assumption that R&D intensity expresses one single dimension of innovative capability. A more detailed analysis would probably find that there are different dimensions also to R&D, for instance between the R and D part. To simplify again, we can here think of the difference between the relatively more straightforward application of results of scientific research in pharmaceuticals and the more interactive testing and revising development work in the production of different kinds of machine tools. However, testing out this would require more detailed data, allowing us to distinguish between different components of R&D.

Summary of results

We find a clear contrast between the natural science education variables, on the one hand, and the other education variables, on the other. The natural science education variables in general correlate more strongly with the innovation variables than the other education variables. They may thus clearly function as indicators of innovative capability, as already by their substantial correlation with R&D intensity. However, their effects also almost invariably become non significant when we control for R&D. In a couple of the cases they even become non significant already when we control for the background variables enterprise size and industry. They thus only express the R&D dimension of innovative capability. They belong to the R&D dimension. For instance, they are instrumental in bringing about R&D: to perform R&D, you tend to need engineers, chemists, etc.

Several of the other education variables also correlate with the innovation variables, although less strongly so than the natural science education variables. Thus, they also function as indicators of the innovative capability of enterprises. Furthermore, being not or only weakly correlated with R&D intensity, the effects of these education variables tend to remain even when we control for R&D. To the extent that they do express innovative capabilities, we thus have indication that they express other dimensions of innovative capability than the R&D dimension.

In conclusion, the adding of the data on the educational attainment of employees to the R&D and other innovation data would seem worthwhile and promising. We get reasonably strong and meaningful correlations with the innovation variables, which also serves as a partial validation of the latter: it strengthens our confidence that they do reflect important aspects of innovative capability and activity. The education variables seem partly to reflect the R&D dimension of innovative capability, partly other dimensions, and in a pattern which makes sense. Moreover, we should not forget that there are aspects of competence important for economic performance which may have little or nothing to do with the capability for innovation. Thus, the education data may be of importance also in investigations where we not focus primarily on innovation, but for instance more broadly on the role of innovation in economic performance, where other kinds of competence may be equally important.

Index

1. INTRODUCTION

In this paper we will look at the relationship between education and innovation, using data at the enterprise level. The data come from two sources. One is the Norwegian innovation survey of 1997 (part of CIS II). The other is data on highest achieved formal education of individuals in Norway in 1997. These data have been aggregated by enterprise, so that we for each enterprise have the proportion of employees with different kinds of highest attained education level. These data have then been merged with the data from the innovation survey.

We want here to explore the question of whether, or rather to what extent, education level in the above sense may be used as an indicator of the innovative capacity of business enterprises. A further question is whether this indicator roughly says the same as R&D intensity and other more familiar indicators, or whether education indicators may help uncover other dimensions of innovative capacity than R&D.

We will in the following focus exclusively on the manufacturing sector.

THE DATA

The Norwegian innovation survey of 1997 is a component of the European Community Innovation Survey (CIS II). Here a representative sample of enterprises have been asked a number of questions relating to innovation. In the Norwegian survey only enterprises with at least 10 employees are included. (The European data cover only enterprises with at least 20 employees.) A couple of introductory questions allow us to distinguish between enterprises with and without innovations. The definition refers to the three year period 1995-1997, and the questions are whether the enterprise during this period has developed or introduced any technologically changed *products*, and whether the enterprise during this period has developed or introduced any technologically changed *processes*. The enterprises who answer in the affirmative to one or both of these questions may thus be defined as innovative, those who answer no to both questions as not innovative. Roughly, a little less than half of the enterprises in the sample are innovative according to this definition (in the sample we will use here, the proportion is 44.3 per cent, see below). The enterprises who are defined as innovative have then been asked a number of questions regarding their innovative activities and the results of this activity, for instance on R&D expenditures and other expenditures relating to innovation activities, on innovation cooperation with other enterprises or institutions, on patent applications, on how large proportion of sales product innovations accounted for, etc. In addition, for all enterprises, both innovative and non innovative, there is background information regarding such data as industry classification, number of employees, sales and exports. All these data refer either to the three year period 1995-1997, or to the year 1997.

As for the educational characteristics of the employees, we have data on all Norwegian enterprises. Since we also have the same organization numbers for identifying the individual enterprises in both data sets, we have thus been able to add the education variables to all the observations in the Innovation survey. For the education data we have here chosen the year 1997, as this is the year which most of the Innovation survey data, and notably the R&D data, refer to.

Our sample is thus the sample from the Innovation survey of 1997, to which we have added data on the education level of the employees in each of the enterprises. Confining ourselves to the manufacturing sector we thus have a total sample of 1944 enterprises. For some of the analyses, our sample is further restricted to enterprises with innovation, in which case the number of observations is 861.

The innovation survey sample is a stratified sample, where the probability of selection varies across strata. This deviation from the case where the observations have been selected through simple random sampling creates complications for the statistical analysis of the data, especially for the estimation of standard errors and confidence intervals, and thus for the evaluation of statistical significance. To take this deviation from a simple random sample into account would have complicated the analysis of the data substantially, and we have chosen not to do so here. Our main purpose in this paper has been to investigate to what extent the education variables contribute to explaining the variation in different dependent variables over and above other variables, notably R&D variables, i.e. when we control for these other variables. We thus use our sample more as one uses one's sample in an experiment. Besides, the results of this investigation are not likely to be substantially altered if we take account of the complex data structure. This is especially so since we in each analysis also control for the variables which define the strata from which the observations were drawn, namely enterprise size and industry. Only substantial interaction between the education variables and other independent variables in the effects on the dependent variables would substantially alter the conclusions if we took the complex data structure into account, and this is not likely. We may on a later occasion investigate this question explicitly through a more complicated analysis which aims to take the complex data structure into account.

THE EDUCATION VARIABLES

For all the enterprises we have data on the highest attained educational level of each employee. Thus, for each enterprise we have the number of employees with different kinds of highest attained educational levels. We will express these numbers as a proportion of total employees, to get an intensity measure of the different kinds of formal educational qualifications.

We will in this paper use the following education variables.

First there is the proportion of employees who have at least secondary education, whether this is their highest attained level or they have afterwards also attained higher levels. This variable will be referred to as *seceup*.

Of those who have gone further than the secondary level, one group is those who have some kind of craft education. The proportion of employees with such craft education is referred to as *craftp*. We here also use a dichotomous variable registering whether the enterprise in question has any employees with craft education or not. This variable is referred to as *craft01*.

We then have a number of variables classifying employees with different kinds of higher education. We have four different kinds of higher education, each at two different levels. Furthermore, for all this educational characteristics we have both an intensity variable registering the proportion of employees with the different kinds of characteristics, and a dichotomous variable saying whether the enterprise has any employees with the educational characteristic in question or not. This gives altogether 16 higher education variables.

The four different kinds of higher education are 1) engineering education, 2) other natural science education, 3) business administration, accountancy, economics, etc., and 4) other higher education. The two levels we use are the wide level and the highest level. The wide level includes all who have completed any kind of such education requiring at least one year of study beyond secondary education. The highest level includes only those who have completed an education requiring at least five years of study beyond secondary education. Note that we have chosen to let the wide level *include* the highest level.

The following table summarizes the 16 higher education variables and shows which variable names we use to refer to them:

type of education	wide level		highest level	
	dichotomous	intensity	dichotomous	intensity
engineering	*engin01*	*enginw*	*engin05*	*engin5p*
other natural science	*nres01*	*nresw*	*nres05*	*nres5p*
business adm., etc.	*econ01*	*econw*	*econ05*	*econ5p*
other higher education	*ores01*	*oresw*	*ores05*	*ores5p*

With the two craft education variables (*craft01* and *craftp*) and the proportion of all who have at least secondary education (*seceup*), we thus operate with altogether 19 education variables.

All the dichotomous variables are coded 1 if the characteristic is present, 0 if it is not. For all the intensity variables, the proportions are expressed as percentages.

2. R&D INTENSITY AND EDUCATIONAL ATTAINMENT AT THE ENTERPRISE LEVEL

Let us first look at the relationship between the education level of the work force and R&D intensity, which is perhaps the most used indicator of the innovative capacity of enterprises, industries and countries. The innovation survey has data on R&D expenditures for 1997, both internal R&D (R&D performed within the enterprise) and external R&D (acquisition of R&D services). We also have data on sales for 1997. Dividing R&D expenditures by sales we get a measure of R&D intensity. (We would have preferred R&D to value added, but the latter figure is not available.)

The R&D variables are very skewed. More than 75 per cent of the enterprises have no R&D expenditures at all. Among the enterprises which do have R&D expenditures most have low values, while a few have very high values. Ordinary averages therefore are problematic as measures of central tendency here. In general, the logged of the R&D intensity values function substantially better than the original R&D intensity values. For instance, using the logged values when R&D variables are independent variables tends to improve the prediction of the value of dependent variable. We use the logged versions of the R&D values. More precisely, R&D intensities have been expressed in per cent, and the figure 1 has been added. Then log values have been taken of these figures. In the table below, the averages reported are the averages of these logged values, transformed back into the original R&D intensity scale. The table also shows the proportion of enterprises with R&D expenditures.

Average R&D intensity, per cent (mean of logged values, then transformed back to original scale), and proportion of enterprises with R&D, per cent, all enterprises.

	Average intensity	Proportion of enterprises with R&D
Internal R&D	0.212	22.6
External R&D	0.081	15.8
Total R&D	0.257	24.4

We see that only 24.4 per cent of the enterprises have R&D expenditures. 22.6 per cent have internal R&D, only 15.8 per cent external R&D. For internal R&D intensity, the central tendency value given in the table is 0.2 per cent. The ordinary average is 0.5 per cent, while the 90th percentile has 1.1 per cent internal R&D intensity, the 95th percentile has 3.1 per cent and the 99th percentile an internal R&D intensity of 10.4 per cent.

The education variables are also skewed, but not nearly to the same degree as the R&D intensity variables. For the education variables the original variables seem to function better, i.e. give a better fit when used as independent variables, than when the logged values are used. Therefore we use the original variables and the averages reported in the table below are ordinary averages. For the higher education categories, the wide definitions have been used, i.e. all who have completed any kind of such education requiring at least one year of study beyond secondary education are included.

Average proportion employees with different kinds of education, per cent, and proportion of enterprises with employees with the different kinds of education, per cent, all enterprises. Wide definition for the higher education categories.

Type of education	Average intensity	Proportion of enterprises
Secondary or more	50.4	100.0
Craft	11.7	78.3
Engineer	3.9	51.4
Other natural science	1.1	34.0
Business adm.	2.7	55.0
Other higher education	3.7	55.2

There are no enterprises in the sample who do not have at least one employee with at least secondary education. A vast majority of the enterprises, 78 per cent, have employees with craft education. For each of the higher education categories, just over half of the enterprises have employees with such education, apart for the other natural science category, where the proportion is just over one third. For the higher education categories the average proportion of employees ranges from 3.9 per cent for engineers to 1.1 per cent with other natural science. Average craft intensity is 11.7 per cent, while average proportion of employees with at least secondary education is 50.4 per cent.

The characterisation of the education variables as skewed more specifically applies to the higher education variables. The craft intensity variable is only slightly skewed, while the proportion of employees with at least secondary education is not skewed at all.

We now go on to look at the correlations among these variables. We first look at the correlations among the R&D variables, which are shown in the following correlation matrix:

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	LOGRDIN	LOGRDEX	LOGRDT
LOGRDIN	1.00000	0.63879	0.97145
		<.0001	<.0001
	1924	1921	1921
LOGRDEX	0.63879	1.00000	0.77261
	<.0001		<.0001
	1921	1932	1921
LOGRDT	0.97145	0.77261	1.00000
	<.0001	<.0001	
	1921	1921	1921

Logrdin is the logged values of the internal R&D intensity variable, as explained above. *Logrdex* and *logrdt* are in the same way the logged values of external R&D intensity and total R&D intensity, respectively.

There is, naturally, a high correlation between internal and external R&D intensity. 24.4 per cent of the enterprises have R&D. More than half of these have both types of R&D.

We next look at the correlations among the education variables. We only look at the intensity variables, and for the higher education variables only in their wide definition. The coefficients are shown in the following correlation matrix:

Pearson Correlation Coefficients, N = 1944
Prob > |r| under H0: Rho=0

	SECEUP	CRAFTP	ORESW	ECONW	NRESW	ENGINW
SECEUP	1.00000	0.46708	0.35799	0.29304	0.33583	0.41680
		<.0001	<.0001	<.0001	<.0001	<.0001
CRAFTP	0.46708	1.00000	-0.15990	-0.16557	-0.06575	-0.02405
	<.0001		<.0001	<.0001	0.0037	0.2892
ORESW	0.35799	-0.15990	1.00000	0.31934	0.10783	-0.00554
	<.0001	<.0001		<.0001	<.0001	0.8072
ECONW	0.29304	-0.16557	0.31934	1.00000	0.18315	0.17409
	<.0001	<.0001	<.0001		<.0001	<.0001
NRESW	0.33583	-0.06575	0.10783	0.18315	1.00000	0.49470
	<.0001	0.0037	<.0001	<.0001		<.0001
ENGINW	0.41680	-0.02405	-0.00554	0.17409	0.49470	1.00000
	<.0001	0.2892	0.8072	<.0001	<.0001	

There is a quite high correlation between engineer intensity and the intensity of employees with other natural science education (0.49). There is also a reasonably high correlation between buiness administration intensity and other higher education intensity (0.32). Craft education intensity correlates negatively with higher education other than natural science, not significantly with engineer intensity and slightly negatively with other natural science education intensity. The proportion with at least secondary education correlates positively and quite substantially with all the other variables, both the craft variable and the four higher education variables.

We can then look at the correlation of R&D intensity with the different education intensities. For the higher education variables we use the wide definition. We start with internal R&D intensity, where we use the logged values as explain above. The coefficients and their p-values are shown in the table below.

Correlation of internal R&D intensity (*logrdin*) with education intensities (wide definition for higher education), all enterprises (N=1921).

logrdin	coefficient	p-value	
seceup	0.227	<0.0001	***
craftp	-0.045	0.0482	*
oresw	0.030	0.1905	
econw	0.131	<0.0001	***
nresw	0.377	<0.0001	***
enginw	0.446	<0.0001	***

We see that the engineer intensity in particular, but also other natural science intensity correlates highly with internal R&D intensity. The correlation is much higher than for business adm. intensity. For the rest of higher education it is virtually 0 and not significant. For the total of secondary education or higher, i.e. including all the other categories in addition to those who have only secondary education, the correlation is higher than for business administration intensity but lower than for the natural science intensities.

The corresponding correlations between the education variables and external R&D intensity are shown in the following table:

Correlation of external R&D intensity (*logrdex*) with education intensities (wide definition for higher education), all enterprises (N=1921).

logrdex	coefficient	p-value	
seceup	0.197	<0.0001	***
craftp	-0.006	0.8045	
oresw	0.047	0.0402	*
econw	0.097	<0.0001	***
nresw	0.300	<0.0001	***
enginw	0.310	<0.0001	***

Basically the same picture emerges here as for internal R&D intensity, but the coefficients are somewhat lower. The correlation is substantially higher with the intensity of the natural science educations (engineers and other natural science) than with the intensity of business administration subjects and other higher education.

This picture is confirmed by a simple factor analysis of these variables. We have used seven of the eight above variables (excluding the secondary education variable). After rotation of the initial factors (we have here used the Promax option in SAS) we get the following factor structure:

```
             The FACTOR Procedure

           Rotation Method: Promax

       Factor Structure (Correlations)

                   Factor1         Factor2

    CRAFTP        -0.02377        -0.56339
    ORESW          0.03824         0.75336
    ECONW          0.22615         0.73315
    NRESW          0.68985         0.23343
    ENGINW         0.73750         0.10592
    LOGRDIN        0.83309         0.05961
    LOGRDEX        0.75450         0.02186
```

Here there emerge two factors with an eigenvalue of more than 1, and they are quite clearly set out from the rest. The first factor is highly correlated with both the two R&D intensity variables and the two natural science education intensities. The second factor is highly correlated with the two other higher education intensities, i.e. business adm. intensity and the intensity of other higher education.

Thus, engineer intensity and other natural science intensity are most likely to correlate with innovation variables. At the same time, it may be that they give little in addition to the R&D variables. I.e., it may be that when we control for R&D, the effects of these variables will largely disappear.

The other education variables here are not highly correlated with the R&D variables. It is less clear that they will be correlated with other innovation variables. But to the extent that they are meaningfully correlated with other innovation variables, they are more likely to constitute a different dimension of innovative capacity. In that case they will be more likely to contribute something in addition to R&D variables.

Using education variables to predict R&D intensity

We have seen that many of the education variables correlate substantially with R&D intensity bivariately, i.e. when we correlate just two variables with it other at the time.

Let us now look at this multivariately, using R&D intensity as dependent variable and the education variable as independent variables, to see which education variables are still significantly correlated with R&D intensity when we control for the other education variables. We will then also introduce enterprise size and industry two see how much the education variables contribute with over and above these background variables. In the following we will only look at internal R&D intensity (*logrdin*) as dependent variable, not external R&D.

First we use only the education variables as independent variables, and concerning higher education we use only the wide definition variables. We use an ordinary least squares regression model with internal R&D intensity with logged values (*logrdin*) as dependent variable. After excluding variables which are not significant at the 5 per cent level, we get the following model:

The REG Procedure

Dependent Variable: LOGRDIN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	117.10979	39.03660	201.69	<.0001
Error	1920	371.60368	0.19354		
Corrected Total	1923	488.71347			

Root MSE	0.43994	R-Square	0.2396
Dependent Mean	0.19267	Adj R-Sq	0.2384
Coeff Var	228.33082		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.00542	0.01551	0.35	0.7269
ORES01	1	0.09592	0.02039	4.71	<.0001
NRESW	1	0.03834	0.00445	8.61	<.0001
ENGINW	1	0.02365	0.00162	14.61	<.0001

It turns out that when we enter the education variables together, i.e. control the effect of each for the other variables, three of the variables contribute significantly to explaining internal R&D intensity. These are, first, the proportion of employees with engineering education and the proportion with other natural science education. This is to be expected, as these variables were the ones most highly correlated with R&D intensity. But also the dichotomous *ores01*, which says whether or not there are any employees with other higher education, i.e. higher education but neither engineer, other natural science or business administration, is very clearly significant and positive, even when we control for the other variables.

Here R^2 is 0.24. We get some improvement by adding variables for the highest education level:

The REG Procedure

Dependent Variable: LOGRDIN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	125.68470	25.13694	132.81	<.0001
Error	1918	363.02877	0.18927		
Corrected Total	1923	488.71347			

Root MSE	0.43506	R-Square	0.2572
Dependent Mean	0.19267	Adj R-Sq	0.2552
Coeff Var	225.79867		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.01467	0.01560	0.94	0.3472
ORES01	1	0.06599	0.02090	3.16	0.0016
NRESW	1	0.01869	0.00588	3.18	0.0015
ENGINW	1	0.01850	0.00181	10.23	<.0001
NRES5P	1	0.04855	0.00978	4.96	<.0001
ENGIN05	1	0.12755	0.02734	4.67	<.0001

We here get two more highly significant and positive variables. These are *nres5p*, the proportion of employees with the highest level of other natural science education, and *engin05*, the dichotomous variable saying whether or not there are employees with the highest engineering education. R^2 here rises to 0.257.

R&D intensity varies with enterprise size and industry. For enterprise size we use the logged values of the number of employees (we call this variable *logemp*).

We have divided the manufacturing sector into 13 industries, defined at the NACE 2 digit level, as shown in the following table:

Industry	NACE 2	Dummies
Food, beverages, tobacco	15-16	reference
Textiles, leather, wood	17-20	di1720
Pulp and paper	21	di21
Printing, publishing	22	di22
Chemicals, chemical products	24	di24
Rubber, mineral products	25-26	di2526
Basic metals	27	di27
Metal products	28	di28
Machinery and equipment	29	di29
Electronics, instruments	30, 32-33	di3033
Electrical machinery	31	di31
Transport equipment	34-35	di3435
Other manufacturing, recycling	36-37	di36

The table also shows the names we have given to the respective dummy variables used in the regression models below.

Let us now first use only enterprise size and industry to predict R&D intensity. We then get the following results:

The REG Procedure

Dependent Variable: LOGRDIN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	93.29808	7.17678	34.67	<.0001
Error	1910	395.41539	0.20702		
Corrected Total	1923	488.71347			

Root MSE	0.45500	R-Square	0.1909
Dependent Mean	0.19267	Adj R-Sq	0.1854
Coeff Var	236.14853		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.19294	0.04349	-4.44	<.0001
LOGEMP	1	0.06962	0.00937	7.43	<.0001
DI1720	1	0.03423	0.03821	0.90	0.3704
DI21	1	0.00710	0.06691	0.11	0.9155
DI22	1	-0.01100	0.03940	-0.28	0.7801
DI24	1	0.56803	0.06359	8.93	<.0001
DI2526	1	0.13337	0.04645	2.87	0.0041
DI27	1	0.07923	0.06423	1.23	0.2175
DI28	1	0.02944	0.04173	0.71	0.4806
DI29	1	0.29278	0.04374	6.69	<.0001
DI3033	1	0.78760	0.05516	14.28	<.0001
DI31	1	0.48363	0.06583	7.35	<.0001
DI3435	1	0.05158	0.04238	1.22	0.2237
DI36	1	0.15053	0.04697	3.20	0.0014

Both size and industry are very significant. R^2 here is 0.191. The positive effect of enterprise size (*logemp*) here first and foremost reflects that the probability of having R&D expenditures at all increases sharply with enterprise size.

Controlling for enterprise size, we see that R&D intensity is high in electronics etc. (*di3033*), chemicals (*di24*) and electrical machinery (*di31*), it is low in printing and publishing (*di22*), Food and beverages (the reference group) and pulp and paper (*di21*).

We can then enter both size, industry and the education variables in the same model to investigate how much of the total variation in the dependent variable the education variables contribute with over and above the background variables size and industry. Again we start with only the wide definitions of our four types of higher education. After excluding non significant variables we get the following model:

The REG Procedure
Dependent Variable: LOGRDIN
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	139.13281	8.18428	44.62	<.0001
Error	1906	349.58066	0.18341		
Corrected Total	1923	488.71347			

Root MSE	0.42826	R-Square	0.2847
Dependent Mean	0.19267	Adj R-Sq	0.2783
Coeff Var	222.27335		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.20191	0.04451	-4.54	<.0001
LOGEMP	1	0.05883	0.01179	4.99	<.0001
DI1720	1	0.02668	0.03614	0.74	0.4606
DI21	1	-0.03267	0.06315	-0.52	0.6050
DI22	1	-0.03212	0.03744	-0.86	0.3910
DI24	1	0.33238	0.06178	5.38	<.0001
DI2526	1	0.09592	0.04425	2.17	0.0303
DI27	1	0.01798	0.06107	0.29	0.7685
DI28	1	-0.00960	0.03957	-0.24	0.8084
DI29	1	0.16840	0.04309	3.91	<.0001
DI3033	1	0.35017	0.05935	5.90	<.0001
DI31	1	0.29363	0.06412	4.58	<.0001
DI3435	1	0.00443	0.04023	0.11	0.9124
DI36	1	0.13097	0.04430	2.96	0.0031
ORES01	1	0.06410	0.02310	2.78	0.0056
NRES01	1	-0.11090	0.03309	-3.35	0.0008
NRESW	1	0.04524	0.00573	7.90	<.0001
ENGINW	1	0.01605	0.00179	8.94	<.0001

Both size and industry are highly significant when we control for the education variables. In addition, four of the education variables are significant. Again, this applies to the proportion of employees with engineering education (*enginw*) and the proportion with other natural science education (*nresw*), as well as the dichotomous *ores01*, which says whether there are any employees with other higher education (i.e. other than engineer, other natural science or business/economics. But here also the dichotomous *nres01*, which says whether there are any employees with other natural science education, is significant. Moreover, the coefficient is *negative*, giving a perhaps somewhat curious non-linear relationship between the intensity of other natural science and R&D intensity.

R^2 here rises to 0.285, up from 0.191 in the model with only size and industry as independent variables.

Again we get some improvement by introducing the highest level of the four higher education types. After excluding non significant variables we get the following model:

The REG Procedure

Dependent Variable: LOGRDIN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	144.46233	7.22312	39.93	<.0001
Error	1903	344.25114	0.18090		
Corrected Total	1923	488.71347			

Root MSE	0.42532	R-Square	0.2956
Dependent Mean	0.19267	Adj R-Sq	0.2882
Coeff Var	220.74631		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.11538	0.04774	-2.42	0.0157
LOGEMP	1	0.03337	0.01295	2.58	0.0100
DI1720	1	0.02484	0.03590	0.69	0.4891
DI21	1	-0.04297	0.06280	-0.68	0.4939
DI22	1	-0.03505	0.03733	-0.94	0.3479
DI24	1	0.28416	0.06212	4.57	<.0001
DI2526	1	0.09128	0.04397	2.08	0.0380
DI27	1	-0.01035	0.06120	-0.17	0.8657
DI28	1	-0.00643	0.03931	-0.16	0.8701
DI29	1	0.17897	0.04292	4.17	<.0001
DI3033	1	0.33553	0.05906	5.68	<.0001
DI31	1	0.29572	0.06375	4.64	<.0001
DI3435	1	0.00830	0.03998	0.21	0.8355
DI36	1	0.12993	0.04400	2.95	0.0032
ORES01	1	0.05936	0.02299	2.58	0.0099
NRES01	1	-0.10712	0.03654	-2.93	0.0034
NRESW	1	0.02946	0.00775	3.80	0.0002
ENGINW	1	0.01227	0.00197	6.24	<.0001
NRES05	1	0.09197	0.04341	2.12	0.0343
NRES5P	1	0.02841	0.01154	2.46	0.0139
ENGIN05	1	0.09133	0.03179	2.87	0.0041

R^2 now rises further to 0.296. Again this should be compared to 0.191 from the model with only size and industry as independent variables. Three of the highest level education variables are significant at the 5 per cent level, all of them are positive. Two are dichotomous: whether or not there are employees with the highest level of other natural science (*nres05*) and whether there are employees with the highest level of engineering education (*engin05*). Lastly, there is the proportion of employees with the highest level of other natural science (*nres5p*).

There is thus a clear relationship between the education variables and R&D intensity at the enterprise level, also when we control for the background variables size and industry. The overall impression is that only the natural science educations matter here, i.e. engineering educations and other natural science. Only one other education variable is significant when the other variables are controlled for, namely the dichotomous *ores01*, whether there are employees with other higher education.

3. EDUCATION AND R&D INTENSITY AT THE INDUSTRY LEVEL

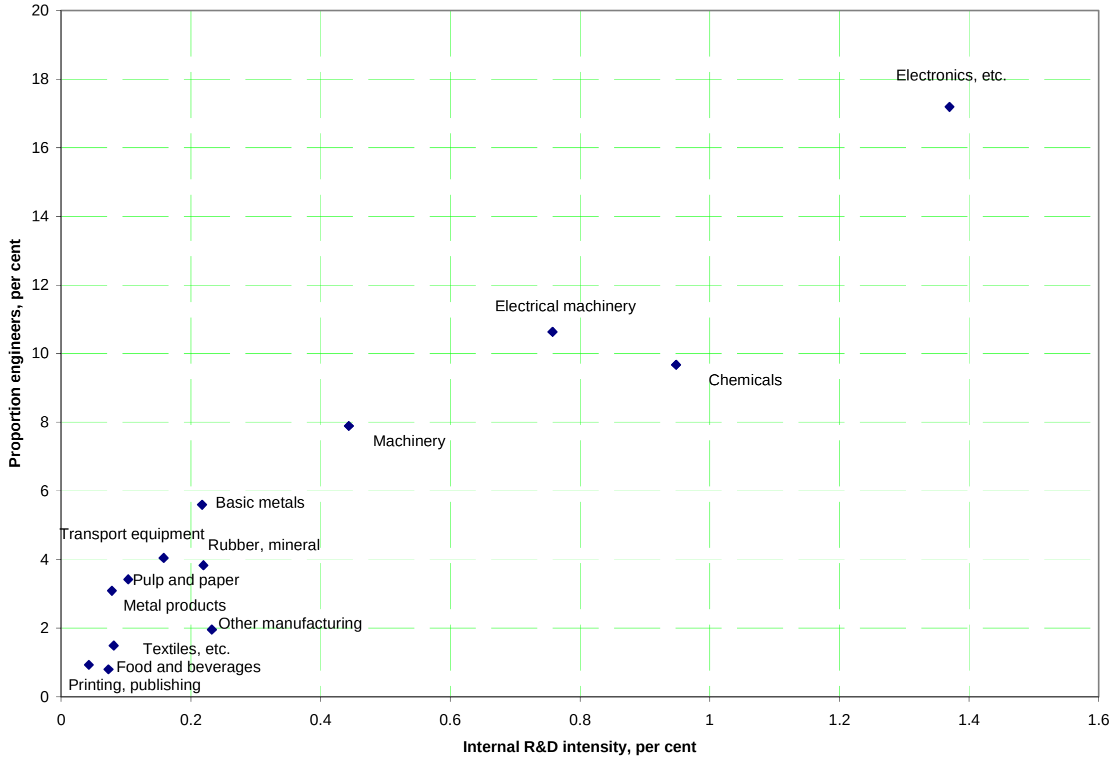
Industries are often characterized in terms of R&D intensity, where R&D intensity is used as a measure of the innovation activity and innovation capacity of the industries. The characterization of industries as high-tech, medium-tech and low-tech has often been based simply on the R&D intensity of the industries.

Let us now instead characterize the industries in terms of the intensity of the four types of higher education. We will here simply use industry averages. We will use the wide definition of higher education. For each of the four average higher education intensities we will compare to the industry averages for internal R&D intensity. Here we use the logged variables, as explained above.

To what extent will the different education intensities give the same ranking among industries as R&D intensity, and to what extent will they give an alternative ranking?

We first look at the relationship between internal R&D intensity and the proportion of engineers among the employees. This is shown in the following figure:

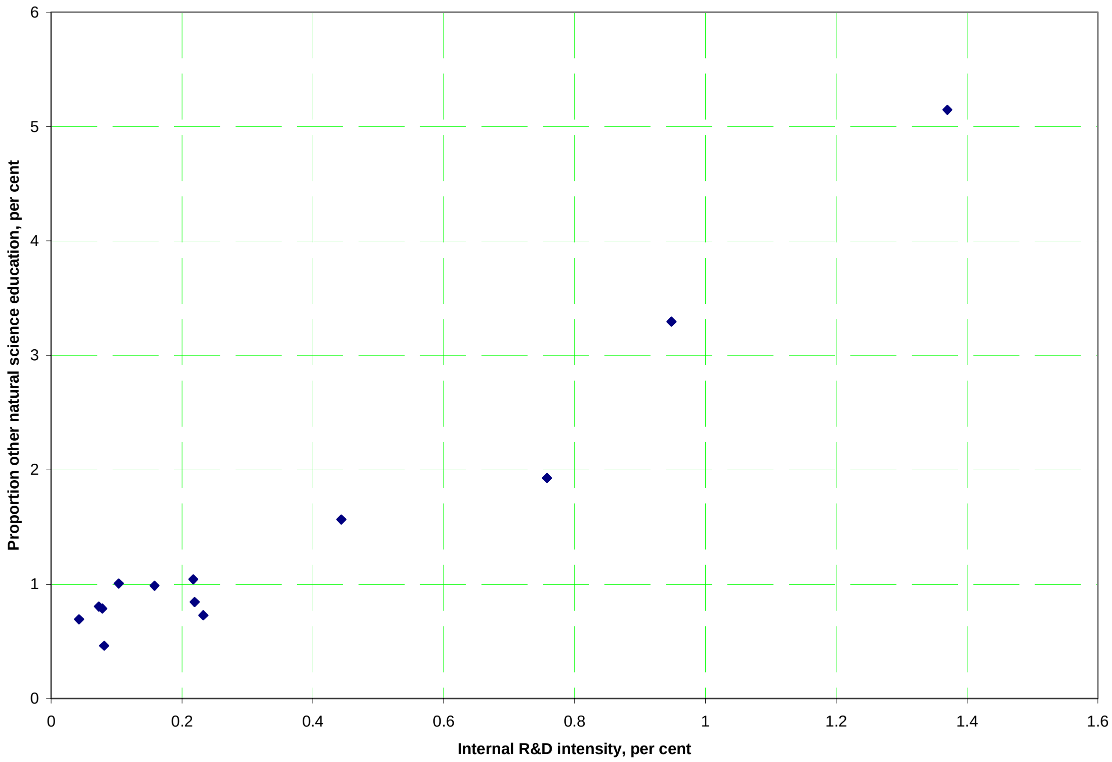
Average proportion engineers among employees (y-axis), average internal R&D intensity (x-axis), by industry.



We see that at the industry level the ranking by engineer intensity closely resembles the ranking by R&D intensity. As measured by the Pearson's *r* correlation coefficient, i.e. treating the intensities as quantitative variables, we get a correlation of 0.960, in other words almost perfect. Using Kendall's *tau*, i.e. treating the rankings of the industries only as rankings, ordinally, we also get a high correlation, 0.744.

We go on to look at the relationship between R&D intensity and other natural science education intensity at the industry level. This is shown in the following figure:

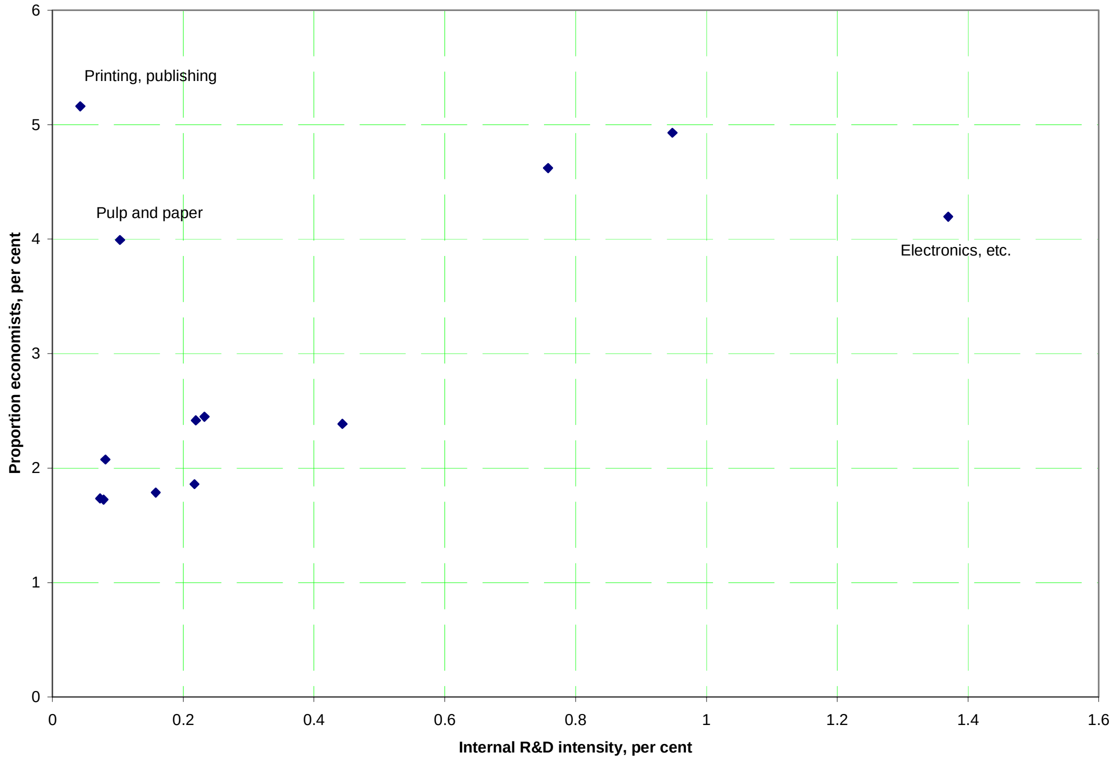
Average proportion other natural science education among employees (y-axis), average internal R&D intensity (x-axis), by industry.



The picture here very much resembles the picture from the relationship between R&D intensity and engineer intensity. Treating the intensities quantitatively, i.e. using Pearson's *r*, we get a correlation of 0.967, again an almost perfect relationship. Treating the variables ordinally, we get a high although not perfect correlation, with a *tau* correlation coefficient of 0.641.

Next we look at the relationship between business administrator, etc. intensity and R&D intensity at the industry level. This is shown in the following figure:

Average proportion employees with business administration, etc. among employees (y-axis), average internal R&D intensity (x-axis), by industry.

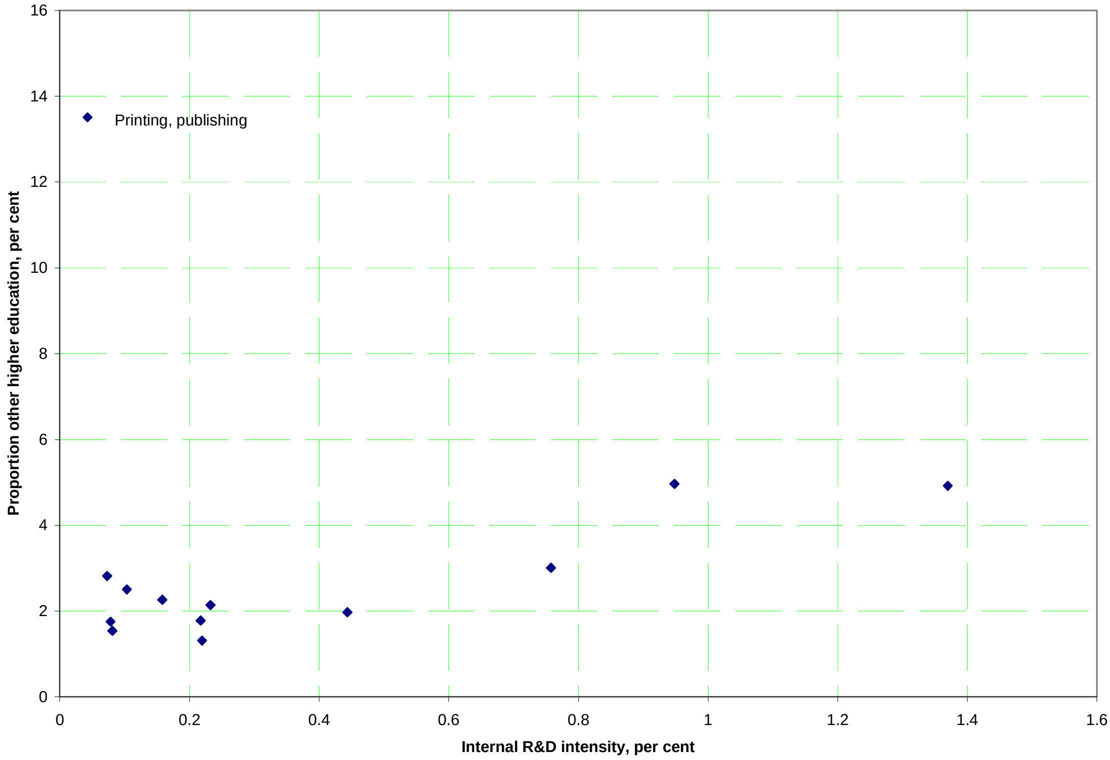


At the enterprise level the correlation between business administrator intensity and R&D intensity was much lower than between R&D intensity and the two natural science education intensities. As expected, this is also the case at the industry level. Pearson *r* is here 0.537, while *tau* is 0.385. Both coefficients are significant at the 10 per cent level, but not at the 5 per cent level.

It is possible to describe the relationship here by saying that we have three outliers from a linear relationship. Printing and publishing, in particular, and also Pulp and paper, in this perspective have a higher business administrator intensity than we would 'expect' from the R&D intensity, while Electronics, etc. has a lower business administrator intensity than expected in this sense. Without these three outliers Pearson *r* becomes 0.960, but this of course is a very *ad hoc* way of making a high correlation appear! (But removal of only two of the outliers also gives a very high correlation coefficient, 0.886.)

Lastly, let us look at the relationship between other higher education intensity and R&D intensity at the industry level. This is shown in the following figure:

Average proportion other higher education among employees (y-axis), average internal R&D intensity (x-axis), by industry.



At the enterprise level we found no correlation between other higher education intensity and internal R&D intensity, and this is also what we find at the industry level: the correlation is very low, with Pearson *r* of only 0.079 and *tau* of only 0.103, neither of which, of course, are significant.

However, in this case there really is one single outlier at the industry level, namely Printing and publishing. This industry has a very low R&D intensity but an extremely high intensity of other higher education, three times as high as the second highest ranked industry. Removing this single industry, we find a high positive correlation between other higher education intensity and internal R&D intensity, with a Pearson's *r* of 0.858.

The very high other higher education in Printing and publishing obviously reflects the fact that we here get all the consultants and editors etc. in publishing companies. Publishing in any case is quite special in relation to manufacturing in general, and one may question whether this activity should be classified as part of manufacturing at all.

Conclusion

We have looked at the relationship between R&D intensity and each of the four types of higher education intensity at the industry level. The intensity of employees with either of the two natural science educations, engineering education and other natural science, correlates strongly with R&D intensity. The industries which have high R&D intensity tend very strongly to have both high engineer intensity and high other natural science intensity, and correspondingly low intensity on these variables go together.

For the two other higher education intensities the correlation with R&D intensity is much weaker. However, the impression here is not so much of an alternative ranking of the industries, expressing a different dimension of capacities and competence, as of deviation from the main pattern of a few outlier industries. This is very clear in the case of the intensity other higher education, where the correlation with R&D intensity is virtually zero, but where this deviation from a high correlation may be seen as the result of the strongly atypical combination of values of one single outlier industry. Removal of this industry makes the correlation with R&D intensity a strong one also here.

This impression the relationship between business administrator intensity and R&D intensity also gives, though less clearly. The correlation at the industry level is moderate and significant only at the 10 per cent level. However, removal of two outlier industries makes the coefficient very high, and removal of three outliers makes it almost unity.

4. THE RELATIONSHIP BETWEEN EDUCATION AND INNOVATION

We will now look at the proportion of the enterprises who have introduced at least one product or process innovation in the course of the three year period 1995-1997. In the present context, this is what is meant by having innovation or being innovative (see the introduction above).

We will now look at the extent to which the probability of being innovative in this sense depends on the education variables. We will here use logistic regression analysis.

Let us first use only education variables as independent variables, and let us use only the wide definition of the higher education variables. The model we then end up with is the following:

```
                     The LOGISTIC Procedure
         Response Variable              INNO01
         Number of Response Levels      2
         Number of Observations         1944
                     Model Fit Statistics
                                           Intercept
                              Intercept       and
            Criterion           Only       Covariates
            AIC               2671.549      2408.173
            SC                2677.121      2447.181
            -2 Log L          2669.549      2394.173
            Testing Global Null Hypothesis: BETA=0
    Test                 Chi-Square       DF     Pr > ChiSq
    Likelihood Ratio       275.3760        6         <.0001
    Score                  260.3355        6         <.0001
    Wald                   227.5495        6         <.0001
           Analysis of Maximum Likelihood Estimates
                                Standard
Parameter    DF    Estimate        Error    Chi-Square    Pr > ChiSq
Intercept     1     -1.1691       0.0876      178.2871        <.0001
ORES01        1      0.5716       0.1238       21.3139        <.0001
ORESW         1     -0.0255      0.00900        8.0329        0.0046
ECON01        1      0.7408       0.1388       28.4877        <.0001
ECONW         1     -0.0435       0.0170        6.5766        0.0103
NRES01        1      0.5725       0.1144       25.0593        <.0001
ENGINW        1      0.0586      0.00937       39.1589        <.0001
```

We see that only the higher education variables are significant here, i.e. neither the proportion of employees with craft education nor the wider category of the proportion of employees with at least secondary education, when the other variables are controlled for.

Both for the proportion of employees with other higher education and the proportion with business administration, etc. education we get a special non-linear relationship:

to have employees with these kinds of education increases the probability of being innovative, but from there the probability decreases the higher the proportion is.

For an enterprise with no employees with higher education (all independent variables have the value 0), this model predicts a probability of being innovative of 23.7 per cent. For an enterprise lying at about the 9th decile on all higher education variables (engineers 10 per cent, other natural science 3.6 per cent, business administration, etc. 7.2 per cent, other higher education 10 per cent) the predicted probability is 67.6 per cent. For all enterprises the proportion is 44.3 per cent.

Here we get a clear improvement in the prediction if we include the highest level of the higher education variables:

```
                 The LOGISTIC Procedure
       Response Variable            INNO01
       Number of Response Levels    2
       Number of Observations       1944
                 Model Fit Statistics
                                     Intercept
                       Intercept        and
          Criterion      Only       Covariates
          AIC            2671.549     2374.119
          SC             2677.121     2424.271
          -2 Log L       2669.549     2356.119
          Testing Global Null Hypothesis: BETA=0
     Test               Chi-Square      DF     Pr > ChiSq
     Likelihood Ratio     313.4304       8         <.0001
     Score                297.4896       8         <.0001
     Wald                 251.7335       8         <.0001
           Analysis of Maximum Likelihood Estimates
                              Standard
Parameter     DF     Estimate      Error    Chi-Square    Pr > ChiSq
Intercept      1      -1.0936     0.0878      155.1177        <.0001
ORES01         1       0.4733     0.1261       14.0958        0.0002
ORESW          1      -0.0238    0.00920        6.7043        0.0096
ECON01         1       0.5903     0.1428       17.0807        <.0001
ECONW          1      -0.0382     0.0173        4.8781        0.0272
NRES01         1       0.2692     0.1295        4.3187        0.0377
ENGINW         1       0.0308     0.0101        9.3763        0.0022
NRES05         1       0.6109     0.1954        9.7775        0.0018
ENGIN05        1       0.7077     0.1463       23.4071        <.0001
```

We see that all the variables from the former model are still significant. In addition, two of the eight higher level variables are significant, both of them dichotomous. They are whether there are employees with the highest level engineer education, *engin05*, and whether there are employees with the highest level of other natural science education, *nres05*.

This model predicts a probability of being innovative for enterprises with no employees with higher education of 25.1 per cent. For enterprises lying at about the

9th decile on all higher education variables in their wide definition (engineers 10 per cent, other natural science 3.6 per cent, business administration, etc. 7.2 per cent, other higher education 10 per cent), but no employees with the highest level, the predicted probability is 50.9 per cent. If in addition the enterprise has employees with the highest level engineer education and employees with the highest level other natural science education, the predicted probability is 79.5 per cent.

Also the probability of being innovative varies with enterprise size and industry, and it is of interest to see how much the education variables contribute to the probability of being innovative over and above what is given by size and industry. Let us first look at size and industry alone:

```
                    The LOGISTIC Procedure
            Response Variable            INNO01
            Number of Response Levels    2
            Number of Observations       1944
                    Model Fit Statistics
                                          Intercept
                            Intercept        and
               Criterion       Only       Covariates
               AIC           2671.549      2406.535
               SC            2677.121      2484.550
               -2 Log L      2669.549      2378.535
               Testing Global Null Hypothesis: BETA=0
       Test                 Chi-Square      DF     Pr > ChiSq
       Likelihood Ratio       291.0136      13         <.0001
       Score                  273.5444      13         <.0001
       Wald                   236.8658      13         <.0001
              Analysis of Maximum Likelihood Estimates
                                  Standard
 Parameter     DF    Estimate        Error    Chi-Square    Pr > ChiSq
 Intercept      1     -2.6129       0.2172      144.7757        <.0001
 LOGEMP         1      0.6288       0.0482      169.8813        <.0001
 DI1720         1     -0.3415       0.1825        3.5031        0.0613
 DI21           1      0.2201       0.3121        0.4974        0.4807
 DI22           1     -0.5183       0.1918        7.3059        0.0069
 DI24           1      1.0700       0.3174       11.3635        0.0007
 DI2526         1      0.3619       0.2159        2.8095        0.0937
 DI27           1     -0.0563       0.3024        0.0347        0.8522
 DI28           1     -0.2435       0.1974        1.5219        0.2173
 DI29           1      0.6057       0.2051        8.7262        0.0031
 DI3033         1      0.6466       0.2579        6.2870        0.0122
 DI31           1      0.5151       0.3070        2.8154        0.0934
 DI3435         1     -0.2964       0.1992        2.2144        0.1367
 DI36           1      0.2098       0.2168        0.9364        0.3332
```

The probability of being innovative increases strongly with enterprise size. Controlling for enterprise size, we find that the probability of being innovative is highest in chemicals (*di24*), electronics, etc. (*di3033*) and machinery (*di29*). It is lowest in printing and publishing (*di22*), textiles etc. (*di1720*) and transport equipment (*di3435*).

We can then control the education variables for enterprise size and industry:

```
                    The LOGISTIC Procedure

          Response Variable            INNO01
          Number of Response Levels    2
          Number of Observations       1944

                    Model Fit Statistics

                                         Intercept
                         Intercept          and
             Criterion      Only        Covariates
             AIC          2671.549       2366.148
             SC           2677.121       2455.308
             -2 Log L     2669.549       2334.148

             Testing Global Null Hypothesis: BETA=0

     Test                Chi-Square       DF     Pr > ChiSq
     Likelihood Ratio      335.4009       15         <.0001
     Score                 309.9961       15         <.0001
     Wald                  258.6208       15         <.0001

            Analysis of Maximum Likelihood Estimates

                                   Standard
Parameter    DF    Estimate        Error    Chi-Square    Pr > ChiSq
Intercept     1     -2.4946       0.2207      127.8014        <.0001
LOGEMP        1      0.5482       0.0546      100.6233        <.0001
DI1720        1     -0.3719       0.1835        4.1105        0.0426
DI21          1      0.0985       0.3151        0.0978        0.7545
DI22          1     -0.5878       0.1933        9.2456        0.0024
DI24          1      0.5879       0.3336        3.1059        0.0780
DI2526        1      0.2191       0.2212        0.9811        0.3219
DI27          1     -0.3113       0.3079        1.0223        0.3120
DI28          1     -0.3802       0.2009        3.5820        0.0584
DI29          1      0.2595       0.2181        1.4148        0.2343
DI3033        1     -0.1691       0.2992        0.3192        0.5721
DI31          1     -0.0135       0.3241        0.0017        0.9667
DI3435        1     -0.4716       0.2027        5.4122        0.0200
DI36          1      0.1464       0.2178        0.4515        0.5016
ORES01        1      0.2495       0.1159        4.6294        0.0314
ENGINW        1      0.0572       0.0101       31.8436        <.0001
```

Using only the wide definition of the higher education variables, we find that two of the education variables are significant. They are first and foremost the proportion of engineers among the employees (*enginw*), and the dichotomous variable saying whether there are employees with other higher education (*ores01*).

However, also here adding the highest level education variables gives a significantly better prediction:

The LOGISTIC Procedure

Response Variable	INNO01
Number of Response Levels	2
Number of Observations	1944

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2671.549	2350.917
SC	2677.121	2451.223
-2 Log L	2669.549	2314.917

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	354.6315	17	<.0001
Score	329.1944	17	<.0001
Wald	268.7706	17	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-2.0899	0.2394	76.1759	<.0001
LOGEMP	1	0.4174	0.0627	44.3192	<.0001
DI1720	1	-0.3552	0.1836	3.7415	0.0531
DI21	1	0.0600	0.3192	0.0353	0.8510
DI22	1	-0.5822	0.1946	8.9564	0.0028
DI24	1	0.3956	0.3435	1.3261	0.2495
DI2526	1	0.2115	0.2225	0.9041	0.3417
DI27	1	-0.4801	0.3185	2.2730	0.1316
DI28	1	-0.3638	0.2013	3.2681	0.0706
DI29	1	0.2840	0.2187	1.6854	0.1942
DI3033	1	-0.2222	0.3031	0.5373	0.4635
DI31	1	-0.00370	0.3273	0.0001	0.9910
DI3435	1	-0.4600	0.2043	5.0717	0.0243
DI36	1	0.1485	0.2181	0.4633	0.4961
ORES01	1	0.2363	0.1163	4.1255	0.0422
ENGINW	1	0.0348	0.0111	9.7995	0.0017
NRES05	1	0.5029	0.1866	7.2643	0.0070
ENGIN05	1	0.5071	0.1577	10.3479	0.0013

We see that again the two significant highest level variables are whether there are employees with the highest level engineer education, *engin05*, and whether there are employees with the highest level of other natural science education, *nres05*. All four education variables are positive. Three of them are natural science education variables, and two of these engineer variables.

We may note here that when we control for the education variables, the probability of being innovative is quite low in electronics, etc. (*di3033*). Thus, the high innovation frequency in this industry clearly reflects the high formal educational characteristics of its employees. This, of course, is a general point, but stands out particularly clearly in this industry, with its high proportions of employees with higher education.

Let us look at some probabilities of being innovative predicted by this model. Let us choose the machinery industry (*di29*), and 50 employees. For an enterprise with no employees with higher education, the probability of being innovative is then 45.7 per cent. For an enterprise with an engineer intensity of 10 per cent and with employees with other higher education, but with no employees with the highest level education in engineering or other natural science, it is 60.2 per cent. If the enterprise in addition both has employees with the highest level engineer education and employees with the highest level other natural science education, the predicted probability is 80.6 per cent.

We will now bring R&D expenditures into the picture. It is obvious that R&D expenditures will be highly significant here. Of the 474 enterprises in the sample with either internal or external R&D expenditures (or both), 459 or 96.8 per cent have introduced at least one product or process innovation during the period. The question is now to what extent the other variables contribute to predicting innovation when we control for R&D expenditures. The following model sums up the results:

The LOGISTIC Procedure

Model Information

Response Variable	INNO01
Number of Response Levels	2
Number of Observations	1943

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2669.920	1834.628
SC	2675.492	1929.352
-2 Log L	2667.920	1800.628

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	867.2915	16	<.0001
Score	698.0433	16	<.0001
Wald	263.6029	16	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-1.9610	0.2587	57.4740	<.0001
LOGEMP	1	0.3153	0.0667	22.3627	<.0001
DI1720	1	-0.5991	0.2114	8.0307	0.0046
DI21	1	-0.0432	0.3582	0.0146	0.9039
DI22	1	-0.3683	0.2060	3.1965	0.0738
DI24	1	-0.2337	0.4576	0.2607	0.6096
DI2526	1	-0.0654	0.2534	0.0666	0.7963
DI27	1	-1.2954	0.4763	7.3977	0.0065
DI28	1	-0.3277	0.2229	2.1624	0.1414
DI29	1	0.0183	0.2475	0.0054	0.9412
DI3033	1	-0.4065	0.3528	1.3281	0.2491
DI31	1	-0.2037	0.3936	0.2679	0.6047

DI3435	1	-0.5924	0.2366	6.2679	0.0123
DI36	1	-0.00159	0.2441	0.0000	0.9948
RDIN01	1	3.2921	0.2842	134.1757	<.0001
RDEX01	1	2.6623	0.4468	35.5000	<.0001
ECON01	1	0.2688	0.1366	3.8703	0.0491

We see that when we control for R&D expenditures, the education variables are hardly significant at all. Only one of the education variables is significant at the 5 per cent level, and just barely. This is typically not one of the natural science education variables, which correlate substantially with R&D intensity, but the dichotomous variable reporting whether any employees have higher education within the business administration field (*econ01*).

Given that almost all enterprises with R&D expenditures have innovation, it is logical that it is the dichotomous R&D variables which are significant. The R&D intensity variables give nothing in addition to these.

We may also note that industry is just barely significant at the 5 per cent level when we control for R&D. The set of 12 dummies representing the 13 industries gives an addition to likelihood ratio chi-square of 22.54, which gives a p-value of 0.0319.

Adding the highest level for the education variables gives no significant contribution to the prediction of the dependent variable.

This means that among the enterprises without R&D expenditures we do not find any significant relationship between the education variables and innovation, when we control for enterprise size and industry. This is not trivial. Of the 1469 enterprises without R&D expenditures, 401 or 27.3 per cent have innovation. And, to take one of the education variables as an example, of these 1469 enterprises, 616 or 41.9 per cent have employees with higher education in engineering. Even if we exclude enterprises without R&D expenditures there is thus plenty of variation left in both the dependent and independent variables.

5. ANALYSIS OF ENTERPRISES WITH INNOVATION

We now go on to look at different kinds of characteristics of enterprises with innovations. Consequently, we will exclude enterprises without innovation from the analysis. Our sample will in the following thus consist of 861 enterprises.

TYPE OF INNOVATION: PRODUCT OR PROCESS

We will first look at type of innovation, where the data distinguish between product innovations and process innovations. We may thus classify the innovative enterprises into three groups: those with only product innovations, those with only process innovations, and those with both types of innovations. However, previous work has shown that most important distinction here is between enterprises with product innovations, whether only product innovations or both types, on the one hand, and enterprises with only process innovations, on the other. The differences between enterprises with only product innovations and enterprises with both types of innovation are generally less important.

Of the 861 enterprises with innovation, 652 or 75.7 per cent have product innovation (either only product innovations ort both types of innovation). Thus, 209 enterprises or 24.3 per cent of the enterprises with innovation have only process innovation.

Let us first look at how the probability of having product innovation varies with the education variables among enterprises with innovation. After excluding the variables which are not significant at the 5 per cent level, we end up with the following model:

```
                    The LOGISTIC Procedure
      Response Variable            INPDT
      Number of Response Levels    2
      Number of Observations       861
                    Model Fit Statistics
                                              Intercept
                              Intercept          and
               Criterion        Only          Covariates
               AIC             956.365          880.271
               SC              961.123          899.304
               -2 Log L        954.365          872.271
               Testing Global Null Hypothesis: BETA=0
       Test                  Chi-Square       DF     Pr > ChiSq
       Likelihood Ratio         82.0935        3         <.0001
       Score                    57.8019        3         <.0001
       Wald                     53.0338        3         <.0001
              Analysis of Maximum Likelihood Estimates
                                  Standard
  Parameter    DF     Estimate       Error    Chi-Square    Pr > ChiSq
  Intercept     1       0.6658      0.1358       24.0322        <.0001
  CRAFTP        1      -0.0169     0.00677        6.2394        0.0125
  NRES01        1       0.5230      0.1803        8.4122        0.0037
  ENGINW        1       0.1217      0.0228       28.5141        <.0001
```

Three of the education variables are significant. The higher the proportion of employees with higher education in engineering (*enginw*), the higher the probability that the enterprise has product innovation. Also, this probability is higher among enterprises with employees with other natural science higher education than among enterprises where no employees have such education (*nres01*). The third education variable is negative: the higher the proportion of employees with craft education (*craftp*), the *lower* the probability that the enterprise has product innovation, i.e. the higher the probability that the enterprise has only process innovation. Nothing is here gained by adding the highest level of the higher education variables.

If we divide likelihood ratio chi-square by total –2 log likelihood (i.e., for "intercept only") we get a measure which is analogous to R^2. This here becomes 0.086.

To what extent do the education variables contribute to explaining the probability of having product innovations among innovative enterprises over and above what is given by the other variables we are dealing with here? Let us next look at how this probability varies with enterprise size and industry.

```
                  The LOGISTIC Procedure

      Response Variable            INPDT
      Number of Response Levels    2
      Number of Observations       861

                  Model Fit Statistics
                                         Intercept
                        Intercept          and
           Criterion       Only         Covariates
           AIC            956.365        897.613
           SC             961.123        964.226
           -2 Log L       954.365        869.613

           Testing Global Null Hypothesis: BETA=0

       Test              Chi-Square      DF     Pr > ChiSq
       Likelihood Ratio     84.7518      13         <.0001
       Score                80.2955      13         <.0001
       Wald                 70.3317      13         <.0001

          Analysis of Maximum Likelihood Estimates
                                  Standard
  Parameter    DF    Estimate       Error    Chi-Square    Pr > ChiSq
  Intercept     1     -0.2638      0.3598        0.5375        0.4635
  LOGEMP        1      0.2803      0.0751       13.9263        0.0002
  DI1720        1      0.0234      0.3062        0.0059        0.9390
  DI21          1     -0.2624      0.4389        0.3574        0.5500
  DI22          1     -0.7918      0.3060        6.6973        0.0097
  DI24          1      0.9067      0.4788        3.5852        0.0583
  DI2526        1      0.8705      0.3798        5.2537        0.0219
  DI27          1     -0.4447      0.4234        1.1034        0.2935
  DI28          1      0.2565      0.3406        0.5674        0.4513
  DI29          1      1.3721      0.3976       11.9109        0.0006
  DI3033        1      1.9241      0.6245        9.4915        0.0021
  DI31          1      1.0296      0.5671        3.2960        0.0695
  DI3435        1      0.5808      0.3499        2.7550        0.0970
  DI36          1     -0.2359      0.3253        0.5256        0.4685
```

The probability of having product innovation increases strongly with enterprise size. The industries with the highest proportion of enterprises with product innovation among enterprises with innovation are electronics, etc. (*di3033*), machinery (*di29*) and electrical machinery (*di31*). The industries with the lowest proportions are printing and publishing (*di22*) basic metals (*di27*) and pulp and paper (*di21*).

Again we can calculate our R^2 analogue, i.e. the likelihood ratio chi-square divided by total –2 log likelihood. This here becomes 0.089, practically the same as with only the education variables.

Let us then see what is left of the education variables when we control for enterprise size and industry. After discarding non significant variables, we get the following model:

The LOGISTIC Procedure

Response Variable	INPDT
Number of Response Levels	2
Number of Observations	861

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	956.365	868.312
SC	961.123	949.200
-2 Log L	954.365	834.312

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	120.0529	16	<.0001
Score	104.5632	16	<.0001
Wald	83.5287	16	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.9432	0.4694	4.0373	0.0445
LOGEMP	1	0.2523	0.0794	10.0859	0.0015
DI1720	1	0.1277	0.3135	0.1659	0.6838
DI21	1	-0.3998	0.4520	0.7822	0.3765
DI22	1	-1.0687	0.3580	8.9100	0.0028
DI24	1	0.1305	0.5237	0.0621	0.8032
DI2526	1	0.6903	0.3893	3.1436	0.0762
DI27	1	-0.5748	0.4526	1.6126	0.2041
DI28	1	0.0876	0.3558	0.0606	0.8055
DI29	1	0.9603	0.4411	4.7408	0.0295
DI3033	1	0.7261	0.6683	1.1806	0.2772
DI31	1	0.1436	0.6049	0.0563	0.8124
DI3435	1	0.4059	0.3798	1.1420	0.2852
DI36	1	-0.3512	0.3352	1.0980	0.2947
SECEUP	1	0.0237	0.00776	9.3672	0.0022
CRAFTP	1	-0.0354	0.00942	14.0892	0.0002
ENGINW	1	0.0642	0.0263	5.9753	0.0145

Three education variables are still significant. The two natural science education variables from the former model, *enginw* and *nres01*, correlate substantially with size and industry. Thus, the latter is no longer significant, and the former has got its coefficient and contribution to likelihood ratio chi-square considerably reduced. Instead, the proportion of employees with at least secondary education (*seceup*) is now significant. This is positive, like the dichotomous other natural science higher education variable which it 'replaces,' and like the engineer intensity variable. The proportion with craft education (*craftp*) is still both significant and negative. Adding the highest level of the higher education variables still does not give any significant increase in likelihood ratio chi-square.

The education variables contribute substantially to predicting the probability of having product innovation among enterprises with innovation also when we control for enterprise size and industry. The R^2 analogue is here 0.126, up from 0.089 without the education variables.

We then introduce the R&D variables, first replacing them for the education variables. We get the following model after excluding those that are not significant (at the 5 per cent level):

```
                    The LOGISTIC Procedure
        Response Variable            INPDT
        Number of Response Levels    2
        Number of Observations       839
                    Model Fit Statistics
                                               Intercept
                               Intercept          and
                 Criterion       Only         Covariates
                 AIC            932.753         788.350
                 SC             937.485         868.797
                 -2 Log L       930.753         754.350
                 Testing Global Null Hypothesis: BETA=0
         Test                 Chi-Square       DF     Pr > ChiSq
         Likelihood Ratio       176.4033       16         <.0001
         Score                  144.0558       16         <.0001
         Wald                    98.7230       16         <.0001
                Analysis of Maximum Likelihood Estimates
                                  Standard
    Parameter    DF    Estimate      Error    Chi-Square    Pr > ChiSq
    Intercept     1     -0.1938     0.3898        0.2471        0.6191
    LOGEMP        1      0.1758     0.0842        4.3590        0.0368
    DI1720        1     -0.2429     0.3251        0.5583        0.4550
    DI21          1     -0.5120     0.4706        1.1837        0.2766
    DI22          1     -0.6385     0.3202        3.9754        0.0462
    DI24          1     -0.3259     0.5417        0.3618        0.5475
    DI2526        1      0.5210     0.3984        1.7105        0.1909
    DI27          1     -1.2489     0.4912        6.4650        0.0110
    DI28          1      0.1046     0.3589        0.0850        0.7707
    DI29          1      0.8707     0.4166        4.3687        0.0366
    DI3033        1      0.7699     0.6752        1.3002        0.2542
    DI31          1      0.0375     0.6380        0.0034        0.9532
    DI3435        1      0.4222     0.3845        1.2061        0.2721
```

DI36	1	-0.5904	0.3523	2.8084	0.0938
RDIN01	1	0.7980	0.3098	6.6376	0.0100
LOGRDIN	1	1.2513	0.4695	7.1029	0.0077
LOGRDEX	1	1.3969	0.6733	4.3046	0.0380

Some of the observations have missing values on the R&D variables. We here thus have 839 valid observation, instead of 861. The proportion who have product innovation among these is 75.7 per cent.

Three of the four R&D variables are significant: the dichotomous have or have not internal R&D (*rdin01*), as well as both R&D intensity variables, i.e. internal and external R&D intensity. All three are positive. Thus, the higher the R&D intensity among enterprises with innovation, the higher the probability of having product innovation, even when we control for enterprise size and industry.

The contribution of the R&D variables is here substantial. The R^2 analogue is 0.190, up from 0.089 without the R&D variables. Thus, the contribution from the R&D variables is clearly larger than from the education variables.

The question then becomes what effect there is left of the education variables when we also control for R&D. Eliminating the non significant variables, we end up with the following model:

The LOGISTIC Procedure

Response Variable	INPDT
Number of Response Levels	2
Number of Observations	842

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	934.421	780.875
SC	939.157	866.119
-2 Log L	932.421	744.875

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	187.5456	17	<.0001
Score	153.9922	17	<.0001
Wald	106.3856	17	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-1.0569	0.4781	4.8863	0.0271
LOGEMP	1	0.1958	0.0856	5.2340	0.0221
DI1720	1	-0.0990	0.3297	0.0901	0.7640
DI21	1	-0.4520	0.4726	0.9149	0.3388
DI22	1	-0.9719	0.3691	6.9334	0.0085
DI24	1	-0.3448	0.5489	0.3946	0.5299
DI2526	1	0.5537	0.4009	1.9078	0.1672
DI27	1	-0.9436	0.4913	3.6883	0.0548
DI28	1	0.1310	0.3675	0.1270	0.7216
DI29	1	0.8673	0.4535	3.6579	0.0558
DI3033	1	0.5051	0.6905	0.5350	0.4645

DI31	1	-0.2890	0.6433	0.2018	0.6533
DI3435	1	0.5340	0.4073	1.7195	0.1898
DI36	1	-0.5983	0.3567	2.8130	0.0935
RDIN01	1	0.8434	0.3104	7.3823	0.0066
LOGRDIN	1	1.4199	0.4622	9.4379	0.0021
SECEUP	1	0.0261	0.00779	11.1974	0.0008
CRAFTP	1	-0.0375	0.00990	14.3417	0.0002

We see that the engineer intensity variable is no longer significant. This is not surprising, as this variable correlates substantially with the R&D variables. However, both the two other education variables are still significant also when we control for R&D. We note also that the external R&D intensity variable (*logrdex*) is no longer significant when we control for education.

The R^2 analogue here becomes 0.201. Without the two education variables it is 0.182. However, the contribution of the two R&D variables is much larger: without these two variables the R^2 analogue is only 0.118.

Looking at the output from the analysis, with the chi-squares and p-values of the individual variables, it might seem that the education variables contribute more than the R&D variables, with chi-squares of 11.2 and 14.3 from the former, 7.4 and 9.4 from the latter. However, this reflects the fact that the R&D variables are more strongly correlated with each other than are the education variables, and thus the contribution of one of the variables in addition to the other is lower in the case of the R&D variables than in the case of the education variables. The contribution to likelihood ratio chi-square of the two education variables together is 18.043, which with 2 degrees of freedom gives a p-value of 0.0001. However, the contribution of the two R&D variables together is 79.596, which gives a p-value far below 0.0001,

Again we note that adding the highest level of the higher education variables does not make any significant contribution to explaining the variation in the dependent variable.

6. INNOVATION COOPERATION

We will now look at the probability of having innovation cooperation among enterprises with innovation. Here we have some missing values, so instead of all the 861 enterprises with innovation, we have 852 observations. Of these 445, or 52.2 per cent, report that they have had innovation cooperation with other enterprises or institutions.

We first look at how the probability of having innovation cooperation depends on the education variables. After excluding non-significant variables, we get the following model:

```
                          The LOGISTIC Procedure

            Response Variable            CO
            Number of Response Levels    2
            Number of Observations       852

                          Model Fit Statistics

                                                 Intercept
                                Intercept          and
                  Criterion       Only          Covariates
                  AIC            1181.427        1054.910
                  SC             1186.175        1092.891
                  -2 Log L       1179.427        1038.910

                  Testing Global Null Hypothesis: BETA=0

              Test                Chi-Square       DF     Pr > ChiSq
              Likelihood Ratio      140.5173        7         <.0001
              Score                 129.6388        7         <.0001
              Wald                  112.5735        7         <.0001

                 Analysis of Maximum Likelihood Estimates

                                         Standard
        Parameter    DF     Estimate        Error    Chi-Square    Pr > ChiSq
        Intercept     1      -0.8993       0.1522       34.9333        <.0001
        ORES01        1       0.5393       0.1886        8.1801        0.0042
        ORESW         1      -0.0522       0.0190        7.5283        0.0061
        ENGIN01       1       0.4086       0.1940        4.4362        0.0352
        ORES05        1       0.6573       0.2529        6.7546        0.0094
        ORES5P        1       0.2384       0.0915        6.7864        0.0092
        NRES5P        1       0.1324       0.0556        5.6707        0.0173
        ENGIN05       1       0.6603       0.1958       11.3757        0.0007
```

We here end up with 7 significant education variables. All of them are higher education variables, which means that neither the craft education variables nor the proportion of the workforce with at least secondary education are significant, when controlling for the other variables. Adding the highest level of the higher education variables contributes very significantly to predicting the occurrence of innovation cooperation: four of the seven variables are highest education level variables. The R^2 analogue is 0.119.

The probability of having innovation cooperation very clearly increases with education level. Only one of the seven variables has a negative coefficient, namely the proportion of employees with other higher education (*oresw*), where ‘other’ means higher education but not engineer, nor other natural science, nor business administration, etc. But the three remaining other higher education variables (*ores01*, *ores05* and *ores5p*) are also significant, and they are positive, unlike *oresw*.

Furthermore, both dichotomous engineer variables are significant and positive. Having employees with the highest level engineer education (*engin05*) increases the probability of having innovation, but even taking account of this effect, having employees with engineer education in the wide definition (*engin01*) increases this probability.

Lastly, the probability of having innovation cooperation increases with the proportion of employees with the highest education level in other natural science (*nres5p*).

To what extent do the education variables contribute to explaining the probability of having innovation cooperation also when we control for other variables. Let us first look at how the probability of having innovation cooperation varies with enterprise size, industry and type of innovation. This is summed up in the following model:

```
                   The LOGISTIC Procedure
Response Variable            CO                    Q4C INNL
Number of Response Levels    2
Number of Observations       852
                   Model Fit Statistics
                                         Intercept
                         Intercept          and
          Criterion        Only          Covariates
          AIC             1181.427        1026.263
          SC              1186.175        1102.225
          -2 Log L        1179.427         994.263
          Testing Global Null Hypothesis: BETA=0
  Test                 Chi-Square       DF     Pr > ChiSq
  Likelihood Ratio       185.1642       15         <.0001
  Score                  166.1019       15         <.0001
  Wald                   136.3616       15         <.0001
         Analysis of Maximum Likelihood Estimates
                                 Standard
Parameter    DF    Estimate        Error    Chi-Square    Pr > ChiSq
Intercept     1     -1.6083       0.3522       20.8573        <.0001
LOGEMP        1      0.4885       0.0702       48.3888        <.0001
DI1720        1     -0.4893       0.3078        2.5263        0.1120
DI21          1     -0.2830       0.4549        0.3870        0.5339
DI22          1     -0.7683       0.3395        5.1214        0.0236
DI24          1      1.8258       0.4933       13.6982        0.0002
DI2526        1      0.4641       0.3190        2.1163        0.1457
DI27          1      1.0839       0.4840        5.0150        0.0251
DI28          1      0.1046       0.3255        0.1033        0.7480
DI29          1      0.0420       0.2913        0.0208        0.8854
DI3033        1      0.6133       0.3706        2.7389        0.0979
DI31          1     -0.0882       0.4178        0.0446        0.8328
DI3435        1      0.1039       0.3083        0.1136        0.7361
```

DI36	1	-0.4080	0.3418	1.4250	0.2326
PDONLY	1	-0.5920	0.2112	7.8560	0.0051
PCONLY	1	-1.1809	0.1963	36.1803	<.0001

We here use all three type of innovation categories. Both types of innovation is the reference group. *Pdonly* and *pconly* are dummies for product only and process only, respectively. Enterprises with both types of innovation have a higher probability of having cooperation than the two other categories, but the difference from enterprises with only process innovation is clearly larger than from enterprises with only product innovation.

The probability of having cooperation very clearly increases with enterprise size. Controlling for enterprise size and type of innovation, we see that there is a relatively high probability of cooperation in chemicals (*di24*), basic metals (*di27*) and electronics etc. (*di3033*), while the probability is relatively low in printing and publishing (*di22*), textiles, etc. (*di1720*) and other manufacturing (*di36*).

The R^2 analogue is here 0.157, somewhat higher than when we only used the education variables as independent variables.

We now add the education variables to see how much they contribute to explaining the occurrence of innovation cooperation when we control for enterprise size, industry and type of innovation. After excluding non significant variables, we get the following model:

The LOGISTIC Procedure

Response Variable	CO	Q4C INNL
Number of Response Levels	2	
Number of Observations	852	

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1181.427	1006.206
SC	1186.175	1086.915
-2 Log L	1179.427	972.206

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	207.2215	16	<.0001
Score	180.7158	16	<.0001
Wald	143.7835	16	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-1.7531	0.3605	23.6443	<.0001
LOGEMP	1	0.4709	0.0713	43.6239	<.0001
DI1720	1	-0.3128	0.3127	1.0003	0.3172
DI21	1	-0.1025	0.4591	0.0499	0.8233
DI22	1	-1.0415	0.3734	7.7787	0.0053
DI24	1	1.8605	0.5026	13.7059	0.0002
DI2526	1	0.6237	0.3227	3.7358	0.0533
DI27	1	1.2499	0.4870	6.5879	0.0103
DI28	1	0.2860	0.3299	0.7514	0.3860

DI29	1	0.2415	0.2964	0.6637	0.4153
DI3033	1	0.7147	0.3765	3.6033	0.0577
DI31	1	0.1100	0.4217	0.0681	0.7941
DI3435	1	0.2778	0.3135	0.7851	0.3756
DI36	1	-0.2533	0.3462	0.5355	0.4643
PDONLY	1	-0.6040	0.2138	7.9791	0.0047
PCONLY	1	-1.1417	0.1989	32.9477	<.0001
ORES5P	1	0.3137	0.0866	13.1295	0.0003

We see that when we control for enterprise size, industry and type of innovation, only one of the education variables remain as significant. This is the proportion of employees with the highest other higher education level (*ores5p*). It is clearly significant, and positive.

The R^2 analogue here is 0.176, which should be compared to the corresponding figure for the model without this variable, 0.157.

We then introduce the R&D variables, first replacing them for the education variables. This gives the following model:

The LOGISTIC Procedure

Response Variable	CO	Q4C INNL
Number of Response Levels	2	
Number of Observations	851	

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1179.949	952.910
SC	1184.695	1038.345
-2 Log L	1177.949	916.910

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	261.0389	17	<.0001
Score	230.0035	17	<.0001
Wald	178.2265	17	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-1.5416	0.3694	17.4167	<.0001
LOGEMP	1	0.3493	0.0748	21.8305	<.0001
DI1720	1	-0.7560	0.3290	5.2801	0.0216
DI21	1	-0.6049	0.4850	1.5555	0.2123
DI22	1	-0.6540	0.3476	3.5402	0.0599
DI24	1	1.2772	0.5192	6.0512	0.0139
DI2526	1	0.2877	0.3415	0.7101	0.3994
DI27	1	0.4028	0.5175	0.6058	0.4364
DI28	1	-0.0164	0.3427	0.0023	0.9617
DI29	1	-0.2274	0.3099	0.5385	0.4630
DI3033	1	0.1087	0.3888	0.0782	0.7797
DI31	1	-0.2535	0.4429	0.3276	0.5671
DI3435	1	-0.0895	0.3229	0.0769	0.7816
DI36	1	-0.6683	0.3644	3.3633	0.0667

PDONLY	1	-0.5787	0.2246	6.6392	0.0100
PCONLY	1	-0.8554	0.2094	16.6873	<.0001
RDIN01	1	0.4960	0.1962	6.3875	0.0115
RDEX01	1	1.2714	0.2093	36.9182	<.0001

The R&D variables again contribute substantially to explaining the variation in the dependent variable. When all four R&D variables are considered, only the dichotomous variables are significant, not the quantitative intensity variables. Not surprisingly, having external R&D contributes more than having internal R&D to predicting the occurrence of innovation cooperation. The R^2 analogue is here 0.222, up from 0.157 for the model without the R&D variables.

We lastly look at how much the education variables contribute to predicting the occurrence of innovation cooperation when we also control for R&D. We then end up with the following model:

The LOGISTIC Procedure

Response Variable	CO	Q4C INNL
Number of Response Levels	2	
Number of Observations	851	

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1179.949	936.492
SC	1184.695	1026.674
-2 Log L	1177.949	898.492

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	279.4563	18	<.0001
Score	240.9667	18	<.0001
Wald	182.2938	18	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-1.6770	0.3772	19.7630	<.0001
LOGEMP	1	0.3355	0.0757	19.6315	<.0001
DI1720	1	-0.5855	0.3333	3.0864	0.0789
DI21	1	-0.4309	0.4884	0.7783	0.3776
DI22	1	-0.9285	0.3772	6.0587	0.0138
DI24	1	1.3665	0.5282	6.6928	0.0097
DI2526	1	0.4429	0.3462	1.6371	0.2007
DI27	1	0.5679	0.5195	1.1948	0.2744
DI28	1	0.1567	0.3466	0.2044	0.6512
DI29	1	-0.0338	0.3145	0.0116	0.9144
DI3033	1	0.2371	0.3945	0.3614	0.5478
DI31	1	-0.0556	0.4468	0.0155	0.9009
DI3435	1	0.0843	0.3276	0.0662	0.7969
DI36	1	-0.5167	0.3687	1.9640	0.1611
PDONLY	1	-0.5989	0.2265	6.9898	0.0082
PCONLY	1	-0.8207	0.2124	14.9253	0.0001
RDIN01	1	0.4794	0.1983	5.8428	0.0156

RDEX01	1	1.2592	0.2110	35.6283	<.0001
ORES5P	1	0.2873	0.0853	11.3454	0.0008

We see that the proportion of employees with the highest other higher education level (*ores5p*) is still clearly significant also when we control for R&D. This was to be expected, as the correlation between this variable and the R&D variables is at best very weak. The R^2 analogue here becomes 0.237, up from 0.222 without the education variable.

7. PATENT APPLICATIONS

The enterprises were also asked if they had applied for at least one patent, in any country, during the period 1995-1997. Of the 861 enterprises with innovation in our sample, 156 or 18.1 per cent had made at least one patent application.

We will first look at how the probability of having patent applications varies with the education level of the work force, ignoring other variables. After excluding non significant variables we get the following model:

```
                    The LOGISTIC Procedure
     Response Variable            PAT                 Q3C
     Number of Response Levels    2
     Number of Observations       861
                    Model Fit Statistics
                                         Intercept
                          Intercept         and
            Criterion        Only       Covariates
            AIC            816.825        736.671
            SC             821.583        769.978
            -2 Log L       814.825        722.671
            Testing Global Null Hypothesis: BETA=0
    Test                 Chi-Square       DF     Pr > ChiSq
    Likelihood Ratio        92.1534        6         <.0001
    Score                   98.5691        6         <.0001
    Wald                    81.2233        6         <.0001
           Analysis of Maximum Likelihood Estimates
                                  Standard
Parameter    DF    Estimate        Error    Chi-Square    Pr > ChiSq
Intercept     1     -3.2909       0.3206      105.4014        <.0001
SECEUP        1      0.0298      0.00692       18.5459        <.0001
CRAFTP        1     -0.0344       0.0107       10.3757        0.0013
ORESW         1     -0.0439       0.0191        5.2582        0.0218
ECONW         1      0.0630       0.0270        5.4322        0.0198
ECON05        1      0.8835       0.3154        7.8481        0.0051
ENGIN05       1      0.7682       0.2042       14.1569        0.0002
```

We are here left with six of the education variables. Four of them are higher education variables, and two of these higher education at the highest level. Besides, both the proportion of employees with at least secondary education (*seceup*) and the proportion of employees with craft education (*craftp*) are significant. The probability of having patent applications increases with the former, decreases with the latter. Having employees with the highest level engineer education (*engin05*) also clearly increases the probability of having patent applications. Besides, two of the business administration education variables are significant, and positive: the proportion of employees with business administration, etc. education at any level (*econw*) and the dichotomous variable registering whether there are employees with highest level business administration education (*econ05*). Lastly, the probability of having patent applications decreases with the proportion of employees with other higher education (*oresw*). The R^2 analogue is 0.113.

We then go on to look at how the probability of having patent applications varies with enterprise size, industry and type of innovation. This is summed up in the following model:

```
                The LOGISTIC Procedure
Response Variable            PAT                 Q3C
Number of Response Levels    2
Number of Observations       861
                Model Fit Statistics
                                       Intercept
                         Intercept        and
       Criterion           Only        Covariates
       AIC               816.825        718.149
       SC                821.583        794.279
       -2 Log L          814.825        686.149
       Testing Global Null Hypothesis: BETA=0
Test                 Chi-Square       DF     Pr > ChiSq
Likelihood Ratio       128.6753       15         <.0001
Score                  106.9774       15         <.0001
Wald                    80.9469       15         <.0001
          Analysis of Maximum Likelihood Estimates
                                Standard
Parameter    DF    Estimate        Error    Chi-Square    Pr > ChiSq
Intercept     1     -4.3138       0.5743       56.4161        <.0001
LOGEMP        1      0.3007       0.0782       14.7741        0.0001
DI1720        1      0.9970       0.5873        2.8822        0.0896
DI21          1      1.9457       0.6479        9.0184        0.0027
DI22          1      0.0675       0.8409        0.0064        0.9361
DI24          1      2.1521       0.5390       15.9422        <.0001
DI2526        1      1.7238       0.5296       10.5949        0.0011
DI27          1      2.3171       0.5952       15.1543        <.0001
DI28          1      2.2304       0.5245       18.0842        <.0001
DI29          1      2.1471       0.4857       19.5382        <.0001
DI3033        1      2.2818       0.5217       19.1325        <.0001
DI31          1      1.7992       0.5957        9.1226        0.0025
DI3435        1      1.8109       0.5054       12.8373        0.0003
DI36          1      1.6933       0.5699        8.8277        0.0030
PDONLY        1      0.3840       0.2248        2.9186        0.0876
PCONLY        1     -1.7663       0.4123       18.3512        <.0001
```

Concerning type of innovation, we see that the probability of having patent applications is considerably lower among enterprises with only process innovations (pconly) than among enterprises with only product innovations (*pdonly*) or with both types of innovation (reference group). The difference between the latter two categories is not significant. The probability increases with enterprise size.

Controlling for enterprise size and type of innovation, the industries with the highest probability of having patent applications are basic metals (*di27*), electronics, etc. (*di3033*) and metal products (*di28*), while the industries with the lowest probability are food and beverages (reference group), printing and publishing (*di22*) and textiles, etc. (*di1720*).

The R^2 analogue is here 0.160, substantially higher than in the model with only education variables.

We may then look at to what extent the education variables contribute to predicting the occurrence of patent applications when we control for enterprise size, industry and type of innovation. Excluding non significant variables, we get the following model:

The LOGISTIC Procedure

Response Variable	PAT
Number of Response Levels	2
Number of Observations	861

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	816.825	680.487
SC	821.583	770.891
-2 Log L	814.825	642.487

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	172.3376	18	<.0001
Score	145.6536	18	<.0001
Wald	106.8066	18	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-5.8618	0.6950	71.1407	<.0001
LOGEMP	1	0.3788	0.0833	20.6616	<.0001
DI1720	1	1.2711	0.5963	4.5440	0.0330
DI21	1	1.8847	0.6689	7.9390	0.0048
DI22	1	-0.8610	0.8946	0.9263	0.3358
DI24	1	1.7144	0.5667	9.1532	0.0025
DI2526	1	1.8579	0.5393	11.8690	0.0006
DI27	1	2.6373	0.6208	18.0500	<.0001
DI28	1	2.3632	0.5409	19.0867	<.0001
DI29	1	2.1743	0.5179	17.6285	<.0001
DI3033	1	1.2592	0.5793	4.7240	0.0297
DI31	1	1.3900	0.6128	5.1452	0.0233
DI3435	1	1.9449	0.5455	12.7121	0.0004
DI36	1	1.7769	0.5771	9.4788	0.0021
PDONLY	1	0.2614	0.2398	1.1889	0.2756
PCONLY	1	-1.5146	0.4180	13.1282	0.0003
SECEUP	1	0.0289	0.00782	13.6801	0.0002
CRAFTP	1	-0.0514	0.0128	16.0772	<.0001
ECONW	1	0.0744	0.0299	6.1808	0.0129

We see that three of the six variables which were significant when we only entered the education variables are still significant when we control for enterprise size, industry and type of innovation. We note that among the variables which are no longer significant is the engineering education variable (*engin05*).

Two of the variables which are still significant are not higher education variables. They are the proportion of employees with at least secondary education (*seceup*, which of course also *includes* all higher education) and the proportion of employees with craft education (*craftp*, also included among those with at least secondary education). The probability of having patent applications still increases with the proportion of employees with at least secondary education, decreases with the proportion of employees with craft education. The third education variable which is still significant, but clearly less so than the two others, is the proportion of employees with business administration education at any level (*econw*).

Together these three education variables contribute quite substantially to predicting the occurrence of patent applications. The R^2 analogue here becomes 0.212, up from 0.160 in the model without the education variables, i.e. with only enterprise size, industry and type of innovation.

We now introduce R&D, and first replace the education variables with the R&D variables. After excluding non significant variables we get the following model:

```
                The LOGISTIC Procedure
Response Variable            PAT
Number of Response Levels    2
Number of Observations       842
                Model Fit Statistics
                                    Intercept
                     Intercept         and
       Criterion       Only        Covariates
       AIC             791.076        637.976
       SC              795.812        723.220
       -2 Log L        789.076        601.976
       Testing Global Null Hypothesis: BETA=0
Test                 Chi-Square       DF     Pr > ChiSq
Likelihood Ratio       187.1002       17         <.0001
Score                  175.8097       17         <.0001
Wald                   118.7033       17         <.0001
       Analysis of Maximum Likelihood Estimates
                                 Standard
Parameter    DF    Estimate         Error    Chi-Square    Pr > ChiSq
Intercept     1     -5.0734        0.6461       61.6598        <.0001
LOGEMP        1      0.2744        0.0877        9.7922        0.0018
DI1720        1      1.0044        0.6299        2.5428        0.1108
DI21          1      1.8363        0.7213        6.4811        0.0109
DI22          1      0.2132        0.8918        0.0572        0.8110
DI24          1      1.4813        0.6026        6.0421        0.0140
DI2526        1      1.5988        0.5768        7.6840        0.0056
DI27          1      2.1201        0.6381       11.0402        0.0009
DI28          1      2.3516        0.5771       16.6072        <.0001
DI29          1      1.8324        0.5353       11.7159        0.0006
DI3033        1      1.0900        0.6113        3.1792        0.0746
DI31          1      1.0933        0.6517        2.8142        0.0934
```

DI3435	1	1.7306	0.5598	9.5575	0.0020
DI36	1	1.4425	0.6271	5.2914	0.0214
PDONLY	1	0.3847	0.2432	2.5021	0.1137
PCONLY	1	-1.1115	0.4322	6.6130	0.0101
RDIN01	1	0.7161	0.3131	5.2317	0.0222
LOGRDIN	1	0.8686	0.1843	22.2040	<.0001

We see that only the internal R&D variables are significant here, not the external R&D variables. The quantitative intensity variable is clearly the more important, but also the dichotomous variable contributes significantly. Again the contribution of the R&D variables as a whole is substantial, and larger than of the education variables. While, as we saw, the education variables increased the R^2 analogue from 0.160 to 0.212, the corresponding increase attributable to the R&D variables is from 0.160 to 0.237.

Lastly, we look at the contribution from the education variables to predicting the occurrence of patent applications when we also control for R&D, i.e. in addition to enterprise size, industry and type of innovation. After excluding non significant variables, we end up with the following model:

```
                    The LOGISTIC Procedure
         Response Variable              PAT
         Number of Response Levels      2
         Number of Observations         842
                    Model Fit Statistics
                                          Intercept
                             Intercept       and
                Criterion      Only       Covariates
                AIC          791.076       619.302
                SC           795.812       718.753
                -2 Log L     789.076       577.302
                Testing Global Null Hypothesis: BETA=0
        Test               Chi-Square      DF     Pr > ChiSq
        Likelihood Ratio     211.7743      20         <.0001
        Score                195.7550      20         <.0001
        Wald                 127.3162      20         <.0001
               Analysis of Maximum Likelihood Estimates
                                Standard
Parameter    DF    Estimate        Error    Chi-Square    Pr > ChiSq
Intercept     1     -6.2236       0.7657       66.0665        <.0001
LOGEMP        1      0.3272       0.0915       12.8008        0.0003
DI1720        1      1.2700       0.6393        3.9462        0.0470
DI21          1      1.7601       0.7434        5.6053        0.0179
DI22          1     -0.4296       0.9612        0.1997        0.6549
DI24          1      1.1938       0.6302        3.5883        0.0582
DI2526        1      1.7213       0.5895        8.5260        0.0035
DI27          1      2.4317       0.6622       13.4836        0.0002
DI28          1      2.4829       0.5929       17.5360        <.0001
DI29          1      1.9056       0.5648       11.3847        0.0007
DI3033        1      0.3534       0.6621        0.2849        0.5935
DI31          1      0.8520       0.6735        1.6000        0.2059
DI3435        1      1.8864       0.5998        9.8927        0.0017
```

DI36	1	1.6122	0.6280	6.5902	0.0103
PDONLY	1	0.3121	0.2556	1.4907	0.2221
PCONLY	1	-0.9255	0.4368	4.4901	0.0341
RDIN01	1	0.7174	0.3216	4.9748	0.0257
LOGRDIN	1	0.7742	0.1947	15.8133	<.0001
SECEUP	1	0.0213	0.00834	6.4964	0.0108
CRAFTP	1	-0.0387	0.0132	8.6151	0.0033
ECONW	1	0.0709	0.0321	4.8782	0.0272

The three education variables which were significant when we controlled enterprise size, industry and type of innovation are still significant when we also control for R&D. This reflects the fact that these three variables correlate either only moderately or weakly with the R&D variables. Adding the education variables increases the R^2 analogue to 0.268, from 0.237 without these variables.

8. CONCLUSIONS

We have here combined data on highest achieved education in Norwegian enterprises in 1997 with data on R&D and innovation from the Norwegian innovation survey of 1997. The purpose has been to investigate to what extent the educational characteristics of employees can be used as indicators of the innovation capabilities of business enterprises. We have done this by investigating to what extent the different education variables can be used to predict or explain different indicators of innovation and innovation activities from the Norwegian innovation survey. An important component of this investigation is the relationship between the education variables and R&D expenditures, which have been the most frequently used indicator of the innovation activity of enterprises, industries and countries.

A central concern has thus been to investigate to what extent the education variables express roughly the same dimension as the R&D variables, and to what extent they represent different kinds of innovation capabilities, different dimensions, than what is captured by R&D intensities. This investigation has consisted in, first, looking at the correlation of the education variables with R&D intensity. We have then investigated to what extent the education variables predict other innovation indicators, for instance whether enterprises are innovative at all, what kinds of innovations they tend to introduce (product or process or both), whether they have innovation cooperation, whether they have patent applications. Lastly, we have looked at to what extent the effect of the education variables on these other innovation indicators remain when we control for R&D expenditures.

We find here a clear difference between the natural science education variables and the other education variables. Natural science educations we have divided into engineering education and other natural science education. They are both higher educations. The other educations considered are higher education within the business administration field and other higher education, but we also use the proportion of employees with craft education, and, lastly, the proportion of employees with at least secondary education, i.e. including all higher education as well as craft education.

The natural science education variables generally correlate quite substantially with R&D. They generally also on their own contribute quite substantially to predicting other innovation indicators which we use as dependent variables. However, they in general do not contribute significantly over and above what is given by the R&D variables: in none of our analyses their effects are significant when we control for R&D. These variables thus seem to represent the same dimension as the R&D variables.

The other education variables correlate more weakly or not at all with R&D. In general, when entered alone, they contribute less than the natural science education variables to predicting the other indication indicators which we use as dependent variables. However, in contradistinction to the natural science education variables, to the extent that they do have an effect on these other innovation indicators, these effects tend to remain also when we control for R&D. In these other education variables we may thus see an indication of other dimensions of innovation capacity than what is reflected in R&D expenditures

STEP rapporter / reports

ISSN 0804-8185

STEP arbeidsnotater / working papers

ISSN 1501-0066

All reports and working papers are available for free download in full text from

http://www.step.no/

nmersborg torg 3, N-0179 Oslo, Norway
Telephone +47 2286 8010
Fax: +47 2286 8049
Web: http://www.step.no/

STEP-gruppen ble etablert i 1991 for å forsyne beslutningstakere med forskning knyttet til alle sider ved innovasjon og teknologisk endring, med særlig vekt på forholdet mellom innovasjon, økonomisk vekst og de samfunnsmessige omgivelser. Basis for gruppens arbeid er erkjennelsen av at utviklingen innen vitenskap og teknologi er fundamental for økonomisk vekst. Det gjenstår likevel mange uløste problemer omkring hvordan prosessen med vitenskapelig og teknologisk endring forløper, og hvordan denne prosessen får samfunnsmessige og økonomiske konsekvenser. Forståelse av denne prosessen er av stor betydning for utformingen og iverksettelsen av forsknings-, teknologi- og innovasjonspolitikken. Forskningen i STEP-gruppen er derfor sentrert omkring historiske, økonomiske, sosiologiske og organisatoriske spørsmål som er relevante for de brede feltene innovasjonspolitikk og økonomisk vekst.

The STEP-group was established in 1991 to support policy-makers with research on all aspects of innovation and technological change, with particular emphasis on the relationships between innovation, economic growth and the social context. The basis of the group's work is the recognition that science, technology and innovation are fundamental to economic growth; yet there remain many unresolved problems about how the processes of scientific and technological change actually occur, and about how they have social and economic impacts. Resolving such problems is central to the formation and implementation of science, technology and innovation policy. The research of the STEP group centres on historical, economic, social and organisational issues relevant for broad fields of innovation policy and economic growth.