

Economic returns to education and training for adults with low numeracy skills: Support document

LYNNE GLEESON

This document was produced by the author based on research for the report *Economic returns to education and training for adults with low numeracy skills*, and is an added resource for further information. The report is available on NCVER's website: <<http://www.ncver.edu.au>>

The views and opinions expressed in this document are those of the author(s) and do not necessarily reflect the views of the Australian Government, state and territory governments or NCVER. Any errors and omissions are the responsibility of the author.

© Australian Government, 2005

This work has been produced by the National Centre for Vocational Education Research (NCVER) on behalf of the Australian Government and state and territory governments with funding provided through the Australian Department of Education, Science and Training. Apart from any use permitted under the *Copyright Act 1968*, no part of this publication may be reproduced by any process without written permission. Requests should be made to NCVER.

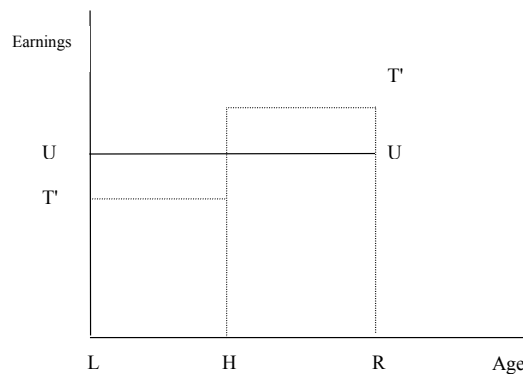
Appendix 1 - The human capital model

The human capital model (Mincer, 1962; Becker, 1962) extends the concept of investment beyond physical capital to examine investment in human capital, including education and training. Human capital is defined by the skills and knowledge embodied in people. The return on human capital investment lies in improving the skills and earning power of an individual, thereby increasing the efficiency and productivity of both the individual and the organisation. Therefore, these investment returns can accrue to both the individual and the organisation.

The decision to invest in human capital is assumed to be a function of the expected cost of education or training, the expected benefits of education or training, and the expected time horizon over which these benefits will be received (Royalty, 1996). Thus, a fully informed rational individual will make the decision to invest in further education or training when the rate of return to that training is greater than the market rate of interest. The increased earnings following investments in education or training are the fundamental components of analysis for human capital theory.

Some factors that influence these returns and thus the decision to invest include the age, wealth and ability of the individual. Different kinds of education or training will lead to different returns. Some training will benefit the individual directly, while other training will benefit both the individual and the organisation. Greater returns will occur if the education or training is completed when the individual is younger, since there is a longer period of higher earnings (Smith, 1994).

Figure 2: Relationship between Earnings and Age*



***Source: Adapted from Becker (1993)**

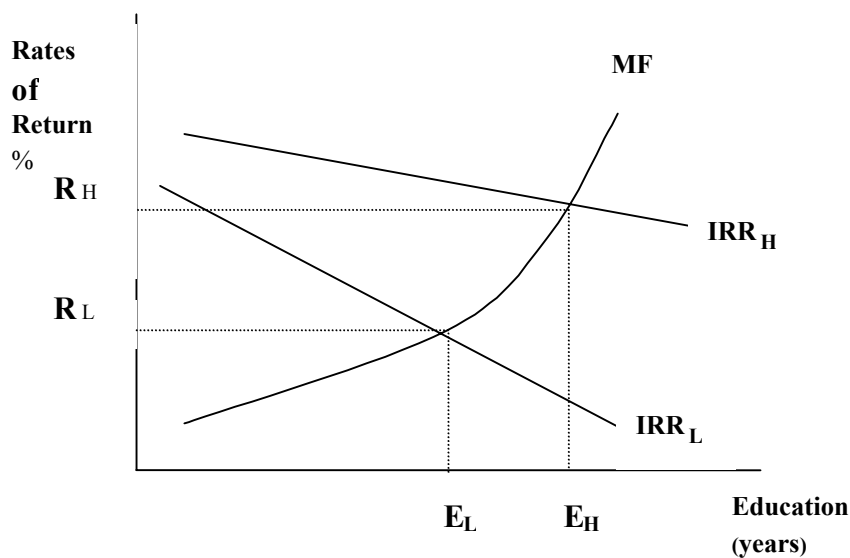
Figure 2 illustrates the relationship between earnings across time and age for individuals beginning employment at age L and working until retirement at age R . L represents individuals at a lower age (age L) and H represents individuals at a higher age (age H). An individual at age L receiving no education or training and no cost of living increases would have the age-earnings profile U . T' represents the age-earnings profile for a given education or training period ($H - L$), where earnings are lower during the initial training period and then higher following training. The opportunity cost of education or training is $(U - T') \times (H - L)$ and the benefits of the training would be measured as $(T' - U) \times (R - H)$ where R is the age at retirement. Thus the rational

individual would choose training if the benefits of training are greater than or equal to the costs of training, $(T' - U) \times (R - H) \geq (U - T') \times (H - L)$.

A second model of the returns to education and training allows us to understand the optimal human capital choice of individuals with different levels of ability. When considering the returns to education or training for individuals with differing ability, where ability includes both physical and intellectual skills, individuals with lower skill levels and thus lower productivity in the workplace tend to have lower returns to education or training.

As illustrated in Figure 3, the marginal financing cost curve (MF) represents the incremental costs of further education. MF is assumed to be increasing as longer periods of education result in greater forgone earnings. That is, the costs of education increase at an increasing rate, as there is less time to have higher earnings before retirement. The curve IRR_H represents the incremental returns to an additional year of education and training for individuals with high skill levels and curve IRR_L represents the incremental returns to an additional year of education and training for individuals with low skill levels. Thus, for any choice of years of education, adults with low skill levels have lower incremental returns than high-skill adults. Both IRR_H and IRR_L are decreasing with increasing levels of education, reflecting a basic assumption of the human capital model.

Figure 3: Rate of Return for Low and High Ability Individuals*



*Source: Adapted from Smith (1994)

Under the assumptions of the human capital model, an individual would make an optimal choice of education at the level where the marginal financing cost of the last unit of education or training consumed is equal to the incremental returns of that last unit. As indicated in the graph, adults with high skills (E_H) choose more education than adults with low skills (E_L), and they earn higher rates of return ($R_H > R_L$). The implication for this study is that it is expected that adults with low literacy or low numeracy skills will choose lower levels of education or training.

Appendix 2 – Selection and analysis of the data sets

Data requirements and selection of the data sets

Two data sets are examined in this study; the LSAY and the NLSY79. The LSAY is a longitudinal study of Australian youth and NLSY79 is a longitudinal study of U.S. youth.

The criteria for selection of the data sets included the availability of an objective measure of numeracy, and data regarding the type, amount and time periods spent in education and training. In addition, the data set needed to have data for wages and a vector of demographic and family background variables. To achieve this, the data set also needed to be longitudinal and include cumulative data.

The analysis section of this report presents the selection of the data sets, variables, data analysis and results using the LSAY data set first, followed by the NLSY79 data set.

Australian data set: LSAY

The cohort selected from the LSAY includes individuals born in 1975, and the LSAY 1975 cohort was originally part of the Youth in Transition 1975 survey. This cohort was first sampled in 1989, when the majority of respondents were in Year 9 at school. The sample included 5,657 participants, who were selected as being representative of the Australian youth population at this time. The participants have been surveyed annually until 2001, with the most recent data used in this study obtained from the 2000 survey. This data set includes data regarding educational levels, wages, gender, ethnicity, location, occupation codes, tenure on the current job, work experience, and full time or part time employment status. LSAY participants also completed an objective literacy and numeracy skills assessment in 1989, which is of particular interest for this study.

Defining and measuring numeracy: LSAY

The LSAY data set includes an objective measure of both literacy and numeracy, through the inclusion of literacy and numeracy subtests developed by the Australian Council of Educational Research (ACER) in the 1989 questionnaire. The numeracy skills subtest has been used as the basis of measuring numeracy skills in this study. Following McMillan and Marks (2003), the LSAY data set has been separated into four numeracy groups based on standard deviations away from the mean. These groups are therefore defined as:

- Very low: >1 SD below mean
- Low: Mean to 1 SD below mean
- High: Mean to 1 SD above mean
- Very high: > 1 SD above mean

Defining and measuring education variables: LSAY

The use of participation in educational programs in the LSAY data set provides a measure based on education, rather than the training variables that are used in the NLSY79 data set. In each

survey year, participants were asked if they were participating in any courses of study in October of the survey year. Following Marks, Hillman and Beavis (2003), this was used as a measure to indicate participation by individuals in educational programs in each survey year. Other Australian data sets include more data on training programs, but do not include the objective measures of literacy and numeracy that are available in the LSAY data set. The 1975 LSAY data set also includes some questions about training in both 1996 and again in 1998, but did not provide enough cumulative longitudinal data to effectively measure returns to training over time. As a result, educational variables are included in the regression equations as dummy variables, and are used to estimate returns to education. The educational variables that are used in the LSAY analysis are:

- Year 10 or lower at school
- Year 11 or 12 at school
- TAFE Certificate
- Traineeship or Apprenticeship
- Diploma or Associate Diploma
- Degree, Postgraduate or other qualification

Sample selection: LSAY

Interviews in the LSAY survey were completed in 2001, but the 2001 data was not included in this analysis as the occupation codes were different in this survey, and were not readily reconciled with occupation codes collected in previous years. Therefore, this analysis was limited to responses from interviews up to and including the 2000 survey. In addition, respondents needed to have completed the literacy and numeracy tests in 1989 to be eligible to be included in this study.

Data analysis: LSAY

A probit model and a regression model (with dependent variable log wage) can be estimated to examine the likelihood of receiving training and the returns to training for adults with low numeracy skills. A vector of demographic and work experience variables are included in order to control for other effects. The specific variables that are included in the LSAY analysis are summarised in Table 4 in the following section.

The returns to training, as measured by log wage, can be estimated in the model as functions of training type while controlling for numeracy level among other demographic and individual-specific variables. This regression equation also includes dummy variables for each year in order to control for other year-to-year growth, such as wage responses to changes in the minimum wage or inflation. With $\ln w$ measuring log-wage for the i^{th} individual in year t , the model is estimated as:

$$\ln w = \alpha_1 S + \beta B + \chi D + \delta E + \varepsilon$$

where S is a vector of education level dummies; B is a vector of background variables including gender, ethnicity and location; D represents demographic variables, including dummies for year-to-year growth; and E is a vector of employment variables including tenure, work experience, and occupation code dummies. ε is the residual, and is assumed normally and independently distributed across individuals.

Selection of variables: LSAY

This study includes the following specific variables in the probit analysis and regression equation based on the LSAY data set: gender, ethnicity, location, education, occupation, tenure, and work experience variables. The selection of specific variables follows Marks, Hillman and Beavis (2003), for selection of ethnicity, location, education and occupation variables. In addition, numeracy group variables are also included, as well as dummy variables for each year, as discussed in the previous section. For the log wage regression model, Table 6 presents the specific independent variable definitions using the LSAY data set.

Table 6: Independent Variable Definitions: LSAY

Variable Symbol	Variable Name	Variable Description*
S _{10orL}	Participation in Year 10 or lower	1 = Participation Yr 10 or lower
S _{11/12}	Participation in Year 11 or 12	1 = Participation Yr 11 or 12
S _{TAFE}	TAFE Certificate	1 = Participation TAFE Certificate
S _{App}	Apprenticeship or Traineeship	1 = Participation Apprenticeship or Traineeship
S _{Dip}	Diploma or Associate Diploma	1 = Participation Diploma or Associate Diploma
S _{Deg}	Degree	1 = Participation Degree, Postgrad study or Other Study
B _{gender}	Gender	1 = Female
B _{Ethnicity}	Ethnicity	1 = Fathers Country of Birth: Non-English speaking
B _{Location}	Location	1 = Non-Metropolitan (School location in 1989)
E _{Occup1}	Occupation Code 1	1 = Occ Code 1 (Managers and Administrators)
E _{Occup2}	Occupation Code 2	1 = Occ Code 2 (Professionals)
E _{Occup3}	Occupation Code 3	1 = Occ Code 3 (Para-Professionals)
E _{Occup4}	Occupation Code 4	1 = Occ Code 4 (Tradespersons)
E _{Occup5}	Occupation Code 5	1 = Occ Code 5 (Clerks)
E _{Occup6}	Occupation Code 6	1 = Occ Code 6 (Salespersons/Personal Serviceworkers)
E _{Occup7}	Occupation Code 7	1 = Occ Code 7 (Plant and Machine Operators)
E _{Occup8}	Occupation Code 8	1 = Occ Code 8 (Labourers and Related Workers)
E _{WExp}	Work Experience	Total hours worked in October of the survey year
E _{Tenure}	Tenure on the current job	Total number of months on the current job
N _{VLow}	Very Low Numeracy	1 = Very Low Numeracy
N _{Low}	Low Numeracy	1 = Low Numeracy
N _{High}	High Numeracy	1 = High Numeracy
N _{VHigh}	Very High Numeracy	1 = Very High Numeracy

*All dummy variables are coded as 0, 1. For ease of presentation, only the value of 1 is reported in Table 1.

The probit analysis: LSAY

In previous research, Lynch (1992) used a probit model and the NLSY79 data set to estimate the probabilities of groups receiving different types of training. Her study examined training for young workers who had not graduated from college. In Lynch's study, formal schooling (measured by years of schooling) was not significant. However, when the probit model was run with formal schooling broken into groups ("high school graduate" and "post high school but not college graduate"), she found that staying in school or attending some college significantly increased the probability of receiving off-the-job training and slightly increased the probability of receiving on-the-job training. Veum (1995) also used a probit model and the NLSY79 data set to estimate the probability of different groups receiving different types of training. Veum found that additional schooling was strongly significant with a positive effect on receiving training (excluding "some college" or "college graduates" who received apprenticeship training). He also found that workers with higher abilities, measured by the AFQT score, appeared to be significantly more likely to receive all forms of training.

In this study, the probit analysis using the LSAY data set examines the probability of individuals with different numeracy levels receiving any type of education. Both Lynch and Veum found that additional schooling increased the likelihood of receiving training, and Veum also found that workers with higher abilities appeared more likely to receive training.

The probit results: LSAY

Table 7: Results of the Probit Analysis using the LSAY data set

	Coefficient	z-stat	P-value
Female	0.2174	4.38	0.000**
Ethnicity	0.0349	0.57	0.570
Non-Metropolitan	-0.1733	-3.61	0.000**
Work Experience	0.0144	17.95	0.000**
Tenure on the current job	0.0396	19.80	0.000**
Occupation – Professionals	-5.0711	-16.46	0.000**
Occupation – Para – Professionals	-5.8674	-19.54	0.000**
Occupation – Tradespersons	-5.6787	-19.08	0.000**
Occupation – Clerks	-5.5050	-19.34	0.000**
Occupation – Sales/Personal Service Workers	-6.1609	-23.74	0.000**
Occupation – Plant and Machine Operators	-5.6033	(dropped)	(dropped)
Occupation –Labourers and Related Workers	-6.4463	-24.71	0.000**
Very Low Numeracy	-0.5539	-6.06	0.000**
Low Numeracy	-0.4160	-6.54	0.000**
High Numeracy	-0.1112	-1.99	0.047*
Constant	5.8212	21.94	0.000**
Number of Observations	6888		
LR chi2 (15)	1919.87		
Prob > chi2	0.0000		
Pseudo R2	0.3172		
Sensitivity Pr(+ D)	96.91%		
Specificity Pr(- ~D)	30.43%		
Correctly Classified	86.28%		

* Significant at the 0.05 level, and ** is significant at the 0.001 level

Fitting a probit function to these variables and controlling for gender, ethnicity, tenure, work experience and geographic location, the results show that, relative to adults with very high numeracy skills, adults in the very low or low numeracy groups are less likely to receive further education. Workers with greater job tenure or higher work experience are more likely to receive further education. The dependent variable takes the value one if the individual reported participating in an educational program in October activities of the interview year and zero otherwise.

The probit analysis in this study examines the probability of particular numeracy groups receiving any type of training, relative to those in the very high numeracy group. The probit model classifies a prediction of receiving training as positive if it has a probability greater than 0.5 and negative if it is less than 0.5. Sensitivity is the correct classification of a “success,” that is, the correct classification of when training has occurred. Specificity is the correct classification of a “failure,” that is, the correct classification of when has not occurred. This model indicates that 86.28% of training is correctly classified, and that the sensitivity is 96.91% and the specificity is 30.43%. Thus, the model is correctly classifying 96.91% of successes and 30.43% of failures.

Log wage regression analysis: LSAY

The log wage regression analysis uses a pooled data set with the natural log of wages as the dependent variable, and controls for other year-to-year growth, such as wage responses to changes in the minimum wage or inflation. ε is the random error. Following previous methodology from Lynch (1992) and Veum (1995), the regression equations are run in multiple models to estimate wages for the different numeracy groups. Model 1 includes all the independent variables, and models 2, 3, 4 and 5 in each of the regression analyses are based on separating the data set according to very low, low, high and very high numeracy levels respectively. Occupation codes include some non-responses, which result in implicit additional groups for these dummy variables. The profiles of these non-response individuals are captured by the constant.

Log wage regression results: LSAY

Selected results of the log wage regression analysis are summarized in Table 8.

All the models and results for this regression equation are presented in Appendix 3.

Table 8: Selected Results from the Log Wage Regression Analysis: LSAY

	Model 1		Model 2		Model 5	
	All independent Variables		Very Low Numeracy		Very High Numeracy	
	<i>Coeff.</i>	<i>t-stat</i>	<i>Coeff.</i>	<i>t-stat</i>	<i>Coeff.</i>	<i>t-stat</i>
Female	-0.0303	-2.10*	-0.0266	-0.43	-0.0337	-1.37
Non-Metropolitan	-0.0219	-1.55	-0.1345	-2.05*	-0.0340	-1.29
Participation in Year 11 or 12	0.1066	3.24**	-0.0030	-0.02	0.2180	3.23**
TAFE Certificate	0.1779	3.50**	-0.0517	-0.30	0.2901	2.26*
Diploma or Associate Diploma	0.2705	5.09**	0.1283	0.67	0.5473	4.09**
Degree/Post Grad	0.1783	3.69**	-0.0446	-0.25	0.4253	3.63**
Occupation – Professionals	0.1476	2.68*	0.3076	1.03	0.2310	2.72*
Occupation – Para Professionals	0.0708	1.20	0.5353	1.67	0.3410	3.37**
Occupation – Tradesperson	0.0897	1.57	0.3480	1.30	0.1897	1.99*
Occupation – Clerks	0.1482	2.73*	0.3237	1.18	0.2590	2.96*
Occupation – Sales/Personal Service Workers	0.1195	2.30*	0.2658	1.00	0.2797	3.39**
Occupation – Labourers and Related Workers	0.0400	0.74	0.1012	0.38	0.2182	2.53*
Tenure on the current job	-0.0008	-2.00*	-0.0007	-0.33	-0.0006	-0.75
Work Experience	0.0091	59.01**	0.0088	11.76**	0.0108	38.58**
Very High Numeracy	-0.0302	-1.09				
High Numeracy	-0.0062	-0.23				
Low Numeracy	-0.0180	-0.65				
Constant	3.9013	56.71	3.9161	13.01**	3.4744	27.82**
Number of observations	6546		467		1876	
F-statistic	F(20, 6515) = 264.38		F(17, 439) = 14.25		F(17, 1848) = 114.18	
Prob > F	0.0000		0.0000		0.0000	
R-squared	0.7337		0.6788		0.7614	
Adjusted R-squared	0.7325		0.6591		0.758	

* Significant at the 0.05 level, and ** is significant at the 0.001 level

In the overall log wage model (Model 1), the signs and significance of each of the coefficients are generally as expected. When dummy variables for are included in the overall log wage model for the very high, high and low numeracy levels, the coefficients are not significant. When examining the results for the very low numeracy group (Model 2) increased work experience was significant and positive, indicating individuals in the very low numeracy group with more work experience had higher wages. The results for the very high numeracy group are presented in Model 5 (models 3 and 4 for the low and high numeracy groups are included in the full results presented in Appendix 3). For adults in the very high numeracy group, the education coefficients were positive and significant, indicating that individuals with higher education levels have higher wages. In addition, several occupation groups and greater work experience had a positive and significant impact on wages for individuals in the very high numeracy group.

Fitting a logistic function to these variables and controlling for gender, race, formal schooling, union status, marital status, welfare received, health, SMSA, tenure, work experience and number of jobs, the results show that, relative to adults with very high numeracy skills, adults in the very low and low numeracy groups are again less likely to receive training. Estimates of odds ratios are presented in Table 9, and again, the dependent variable takes the value one if the individual participated in at least one episode of training between survey interviews and zero otherwise. Higher odds ratios indicate a greater likelihood of receiving training.

Table 9: Results of the Logistic Regression Analysis using NLSY79 data

	Odds Ratio	z-stat	P-value
Formal Schooling	1.0526	6.53	0.000**
Married	1.1459	3.49	0.000**
Female	1.3112	6.76	0.000**
Non-white	1.0505	1.19	0.235
Health	1.1622	1.44	0.150
Welfare Received	1.0000	-1.84	0.066
Union	1.3968	6.67	0.000**
Tenure on the current job	1.0001	1.66	0.097
Work Experience	1.0043	9.89	0.000**
Number of Jobs	1.0101	2.56	0.011*
SMSA - Not in Central City	0.9807	-0.47	0.642
SMSA - Central City	1.0265	0.48	0.634
Occupation - Sales Workers	0.5980	-5.09	0.000**
Occupation - Clerical	1.0009	0.02	0.987
Occupation - Craftsmen	0.8844	-2.65	0.008*
Occupation - Operatives	0.9735	-0.29	0.774
Occupation - Laborers, not farm	0.5953	-5.78	0.000**
Very Low Numeracy	0.3734	-13.65	0.000**
Low Numeracy	0.6491	-8.25	0.000**
High Numeracy	0.9347	-1.34	0.181

N= 31330; * Significant at the 0.05 level, and ** is significant at the 0.001 level

U.S. data set: NLSY79

The NLSY79 is a panel data set that includes a cumulative longitudinal survey of 12,686 participants aged between 14 and 22 years when they were first surveyed in 1979. The participants were selected as being representative of the U.S. population at the time, with some oversampling of military and minority groups being included. The participants were surveyed each year until 1994, and have been surveyed biennially since then. The most recent survey used in this study was completed in 2000. Participants were surveyed in 2002, but, similarly to the LSAY survey, occupation codes changed after 2000 and were not consistent with previous years. Therefore, this analysis includes data up to the 2000 survey.

The NLSY79 data set includes objective measures of literacy and numeracy skills, as well as education and training information. More specifically, the NLSY79 data set includes results from the Armed Services Vocational Army Battery (ASVAB), which was completed by 11,914 participants in 1980. Raw scores for each individual and each ASVAB subtest are available in the data set.

Defining and measuring numeracy: NLSY79

Within the NLSY79 data set, the ASVAB test consists of 10 sub-tests. Following Weltin and Popelka (1983), numeracy is measured in this study as the sum of three relevant ASVAB sub-test scores; arithmetic reasoning, paragraph comprehension and mathematical knowledge (AR + PC + MK). Arithmetic reasoning represents mathematical reasoning, paragraph comprehension represents understanding the meaning of paragraphs, and mathematics knowledge represents knowledge of algebra, geometry and fractions (Weltin and Popelka, 1983). Again, following the methodology used to separate the groups in the LSAY data set, the four numeracy groups are based on standard deviations away from the mean. Consistent with the LSAY data analysis, these groups are defined as:

- Very low: >1 SD below mean
- Low: Mean to 1 SD below mean
- High: Mean to 1 SD above mean
- Very high: > 1 SD above mean

Defining and measuring training: NLSY79

Using a wage growth regression model, Veum (1995) found that training variables defined as aggregated training hours were not significant. However, when training was represented by dummy variables, the results indicated a significant impact on wage growth. Based upon these results, Veum suggested that the duration of training (measured by hours) may have no effect on wage growth, but the incidence of training has a positive effect. He also suggested that there might be measurement error in the training variables since there may be high variability in the amount of training reported, and there may be increased potential for error when measuring and reporting aggregated training hours. Following Veum (1995), variables used in this data set represent the incidence of training, as each annual interview records up to four distinct training episodes and respondents may have multiple episodes for the year. The number of episodes over time are accumulated, thus creating a count variable for each individual. Therefore, in each year, current wage is permitted to co-vary with the number of current and previous training episodes in a way that allows current wages to respond to previous training.

This study follows Lynch (1991, 1992), Gardecki and Neumark (1998), Royalty (1996) and Veum (1995, 1998) and uses on-the-job, off-the-job and apprenticeship training classifications. On-the-job includes all company training and seminars at work, and off-the-job includes training completed in other locations (business college/school, vocational or technical institutes, correspondence courses, seminars outside work, vocational rehabilitation center and government training programs). Apprenticeships are classified as a separate training category. Consistent with Lynch, this model excludes any training received through formal schooling.

Sample selection: NLSY79

Interviews in the NLSY79 survey were completed in 2000 with 8,033 respondents. To be included in this study, respondents must also have completed the ASVAB test in 1980. Training data before 1987 was collected in different training categories, and is not consistent with data collected from surveys in later years. The 1987 data set does not include union status, and as this is a key variable for this analysis, the final data set included data from surveys from 1988 onwards. As discussed earlier, participants were also interviewed in 2002, but the occupation codes from the 2002 interview were not consistent with occupation codes used in previous surveys. In summary, this study includes survey responses from 1988 to 2000.

Data analysis: NLSY79

In contrast to the data analysis using the LSAY data set, the data analysis using the NLSY79 data set uses a logistic regression model and a fixed effects regression model (with dependent variable log wage). Again, a vector of demographic and work experience variables are included in order to control for other effects. The specific variables that are included from each of the data sets are discussed in the following section.

Following the methodology used for the analysis of the LSAY data set, the returns to training, as measured by log wage, can be estimated in the model as functions of training type while controlling for numeracy level among other demographic and individual-specific variables. Again, this regression equation also includes dummy variables for each year in order to control for other year-to-year growth, such as wage responses to changes in the minimum wage or inflation. With $\ln w$ measuring log-wage for the i^{th} individual in year t , the model is estimated as:

$$\ln w = \alpha_1 S + \beta T + \chi F + \delta D + \phi E + \varepsilon$$

where S is a measure of years of formal schooling, T represents the incidence of training by training type; B is a vector of background variables including race, gender, disability, marital status and welfare receipt; D is a vector of demographic variables, including unemployment rates and dummies for year-to-year growth; and E is a vector of employment variables including tenure, work experience, number of jobs and occupation code dummies. Again, ε is the residual, and is assumed normally and independently distributed across individuals.

Selection of variables: NLSY79

Following the precedent set by Lynch and Veum, this study included the following specific variables in the analysis: training type, gender, race, union membership, disability, marital status, county unemployment rates, Standard Metropolitan Statistical Area (SMSA), number of jobs held, tenure on the current job, work experience, years of formal schooling and occupation codes. In addition, the sum of welfare received and numeracy group variables were also included. For the log wage regression models, Table 10 presents the specific independent variable definitions using the NLSY79 data set.

Table 10: Independent Variable Definitions NLSY79

Variable Symbol	Variable Name	Variable Description*
S	Years of Formal Schooling	Highest grade completed at school
T _{on}	On-the-job training	Number of training episodes On-the-job training
T _{off}	Off-the-job training	Number of training episodes Off-the-job training
T _{app}	Apprenticeship training	Number of training episodes Apprenticeship training
F _{gender}	Gender	1 = Female
F _{race}	Race	1 = Non-white
F _{hlth}	Health Status	1 = Not Healthy
F _{nmr}	Marital Status	1 = Married
F _{union}	Union membership	1 = No union membership
F _{Welfare}	Welfare Received	Total sum of welfare funds received (AFDC, Food Stamps, Other Welfare, SSI) in the previous calendar year
D _{Unemp(3-5.9%)}	County Unemployment Rate	1 = Unemployment rate (3.0 – 5.9%)
D _{Unemp(6-8.9%)}	County Unemployment Rate	1 = Unemployment rate (6.0 – 8.9%)
D _{Unemp(9-11.9%)}	County Unemployment Rate	1 = Unemployment rate (9.0 – 11.9%)
D _{Unemp(12-14.9%)}	County Unemployment Rate	1 = Unemployment rate (12.0 – 14.9%)
D _{Unemp(>15%)}	County Unemployment Rate	1 = Unemployment rate (15.0% +)
D _{SMSA (Not SMSA)}	SMSA – Not in SMSA	1 = Not in SMSA
D _{SMSA (Not Central)}	SMSA – Not Central City	1 = SMSA - Not in Central City
D _{SMSA(Central City)}	SMSA – Central City	1 = SMSA - Central City
E _{Jobs}	Number of jobs held	Number of jobs held as of interview date
E _{Tenure}	Tenure of Current Job	Total Tenure (weeks) with current employer
E _{WExp}	Work Experience	Total number of weeks worked since previous interview date
E _{Occup1}	Occupation Code 1	1 = Occupation Code 1 (Professional)
E _{Occup2}	Occupation Code 2	1 = Occupation Code 2 (Managers)
E _{Occup3}	Occupation Code 3	1 = Occupation Code 3 (Sales Workers)
E _{Occup4}	Occupation Code 4	1 = Occupation Code 4 (Clerical)
E _{Occup5}	Occupation Code 5	1 = Occupation Code 5 (Craftsmen)
E _{Occup7}	Occupation Code 7	1 = Occupation Code 7 (Operatives)
E _{Occup8}	Occupation Code 8	1 = Occupation Code 8 (Laborers, not farm)
E _{Occup9}	Occupation Code 9	1 = Occupation Code 9 (Farmers)
E _{Occup10}	Occupation Code 10	1 = Occupation Code 10 (Farm Laborers)
E _{Occup11}	Occupation Code 11	1 = Occupation Code 11 (Service Workers)
E _{Occup12}	Occupation Code 12	1 = Occupation Code 12 (Private Household)
N _{VLow}	Very Low Numeracy	1 = Very Low Numeracy
N _{Low}	Low Numeracy	1 = Low Numeracy
N _{High}	High Numeracy	1 = High Numeracy
N _{VHigh}	Very High Numeracy	1 = Very High Numeracy

*All dummy variables are coded as 0, 1. For ease of presentation, only the value of 1 is reported.

As discussed earlier in this paper, the log wage fixed effects regression model in this study was run using variables that measured the incidence of training, following the methodology from Veum (1995). Again, SMSA, occupation codes and unemployment rates include some non-responses, which result in implicit additional groups for these dummy variables. The profiles of these non-response individuals are captured by the constant. For ease of presentation, only models 1, 2 and 5 which represent the model with all independent variables, and the models for the very low and very high numeracy groups are presented in this table.

All the results for this regression equation are presented in Appendix 4.

Table 11: Selected Results from the Log Wage Fixed Effects Regression Analysis - NLSY79

	Model 1		Model 2		Model 5	
	All independent Variables		Very Low Numeracy		Very High Numeracy	
	<i>Coeff.</i>	<i>t-stat</i>	<i>Coeff.</i>	<i>t-stat</i>	<i>Coeff.</i>	<i>t-stat</i>
On-the-job training	0.0186	4.47**	0.0354	2.12*	0.0157	1.72
Off-the-job training	0.0174	1.90	0.0092	0.26	0.0053	0.28
Apprenticeship training	0.0764	2.83*	0.0878	1.13	0.0365	0.67
Union	0.1109	9.50**	0.1611	6.25**	0.1099	3.16*
Welfare Received	0.0000	-5.54**	0.0000	-3.63**	0.0000	-0.26
Occupation - Sales Workers	-0.0861	-5.68**	-0.0561	-1.52	-0.0807	-2.00*
Occupation - Craftsmen	-0.0795	-9.34**	-0.0529	-3.02*	-0.1415	-5.64**
Occupation - Armed Forces	0.0934	2.84*	0.1413	2.28*	0.0864	0.92
Job Tenure	0.0003	13.01**	0.0003	4.97**	0.0002	4.08**
Experience (weeks)	0.0008	7.91**	0.0007	3.09*	0.0007	2.96*
Number of Jobs	0.0092	4.09**	0.0129	2.57*	0.0024	0.34
Constant	6.4279	270.39**	6.1921	118.34**	6.6498	97.77**
Number of observations	31130		6741		5298	
Number of groups	3911		902		682	
R-sq: Within	0.1642		0.1447		0.1617	
R-sq: Between	0.2538		0.2773		0.1604	
R-sq: Overall	0.1819		0.1848		0.1516	

* is significant at the 0.05 level, and ** is significant at the 0.001 level

In the overall log wage model (Model 1), the signs and significance of each of the coefficients are generally as expected. Consistent with previous research (Lynch, 1992; Veum, 1995), there are higher wages for individuals who are union members, have greater work experience, job tenure or a higher number of jobs. The amount of welfare received is significant and negative, indicating that individuals who receive higher amounts of welfare have lower wages. The coefficients for on-the-job training and apprenticeship training are significant and positive, indicating a positive impact on earnings.

Appendix 3
Log Wage Regression Results:
LSAY

Appendix 3: Log Wage Regression Results: LSAY

	All variables			Very low numeracy			Low numeracy			High numeracy			Very high numeracy		
	Coef.	t	P>t	Coef.	t	P>t	Coef.	t	P>t	Coef.	t	P>t	Coef.	t	P>t
Female	-0.0303	-2.10	0.036*	-0.0266	-0.43	0.669	-0.0719	-2.21	0.027*	-0.0158	-0.69	0.492	-0.0337	-1.37	0.172
Non-Eng Speaking	-0.0103	-0.59	0.555	-0.0966	-1.35	0.177	-0.0204	-0.56	0.576	-0.0190	-0.71	0.480	0.0210	0.64	0.524
Non-Metropolitan	-0.0219	-1.55	0.122	-0.1345	-2.05	0.041*	-0.0137	-0.49	0.628	-0.0146	-0.66	0.508	-0.0340	-1.29	0.197
Yr 11/12	0.1066	3.24	0.001**	-0.0030	-0.02	0.983	0.0957	1.58	0.113	0.0993	1.92	0.054	0.2180	3.23	0.001**
App/Traineeship	0.0176	0.32	0.748	-0.2785	-1.48	0.139	-0.1211	-1.21	0.226	0.0962	1.08	0.281	0.2377	1.86	0.063
TAFE Certificate	0.1779	3.50	0.000**	-0.0517	-0.30	0.765	0.1529	1.74	0.082	0.2042	2.48	0.013*	0.2901	2.26	0.024*
Diploma/Assoc Dip	0.2705	5.09	0.000**	0.1283	0.67	0.502	0.2394	2.63	0.009*	0.2180	2.55	0.011*	0.5473	4.09	0.000**
Degree/Post Grad	0.1783	3.69	0.000**	-0.0446	-0.25	0.802	0.2213	2.58	0.010*	0.1028	1.30	0.192	0.4253	3.63	0.000**
Occup: Professionals	0.1476	2.68	0.007*	0.3076	1.03	0.304	0.1858	1.26	0.206	0.0752	0.87	0.386	0.2310	2.72	0.007*
Occup: Para Prof	0.0708	1.20	0.232	0.5353	1.67	0.096	0.0708	0.50	0.616	-0.1021	-1.13	0.259	0.3410	3.37	0.001**
Occup: Tradesperson	0.0897	1.57	0.117	0.3480	1.30	0.194	0.1882	1.36	0.174	-0.0578	-0.64	0.524	0.1897	1.99	0.046*
Occup: Clerks	0.1482	2.73	0.006*	0.3237	1.18	0.239	0.1843	1.38	0.169	0.0687	0.81	0.418	0.2590	2.96	0.003*
Occup: Sales/Personal Service Workers	0.1195	2.30	0.022*	0.2658	1.00	0.320	0.1536	1.17	0.243	-0.0107	-0.13	0.895	0.2797	3.39	0.001**
Occup: Plant/Machine Operators	0.0706	0.90	0.368	0.5272	1.31	0.190	0.0472	0.25	0.802	-0.0090	-0.07	0.941	0.1705	1.35	0.176
Occup: Labourers and Related Workers	0.0400	0.74	0.457	0.1012	0.38	0.708	0.0790	0.59	0.556	-0.0956	-1.15	0.252	0.2182	2.53	0.011*
Tenure	-0.0008	-2.00	0.046*	-0.0007	-0.33	0.739	-0.0006	-0.69	0.490	-0.0009	-1.36	0.173	-0.0006	-0.75	0.456
Work Experience	0.0091	59.01	0.000**	0.0088	11.76	0.000**	0.0092	25.43	0.000**	0.0082	35.89	0.000**	0.0108	38.58	0.000**
Very High Numeracy	-0.0302	-1.09	0.275												
High Numeracy	-0.0062	-0.23	0.815												
Low Numeracy	-0.0180	-0.65	0.515												
Constant	3.9013	56.71	0.000**	3.9161	13.01	0.000**	3.8254	25.96	0.000**	4.1072	40.52	0.000**	3.4744	27.82	0.000**
Number of observations	6546			467			1439			2764			1876		
F-statistic	F(20, 6515) = 264.38			F(17, 439) = 14.25			F(17, 1411) = 61.58			F(17, 2736) = 120.35			F(17, 1848) = 114.18		
Prob > F	0.0000			0.0000			0.0000			0.0000			0.0000		
R-squared	0.7337			0.6788			0.7492			0.729			0.7614		
Adj. R-squared	0.7325			0.6591			0.7444			0.7264			0.758		

Appendix 4
Log Wage Regression Results: NLSY79

Appendix 4: Log Wage Regression Results using Fixed Effects Estimate: NLSY79

	All variables			Very low numeracy			Low numeracy			High numeracy			Very high numeracy		
	Coef.	t	P>t	Coef.	t	P>t	Coef.	t	P>t	Coef.	t	P>t	Coef.	t	P>t
Years of formal schooling	(dropped)			(dropped)			(dropped)			(dropped)			(dropped)		
On-the-job training	0.0186	4.47	0.000	0.0354	2.12	0.034	0.0158	2.04	0.042	0.0153	2.40	0.016	0.0157	1.72	0.086
Off-the-job training	0.0174	1.90	0.058	0.0092	0.26	0.795	0.0229	1.45	0.148	0.0195	1.26	0.209	0.0053	0.28	0.776
Apprenticeships	0.0764	2.83	0.005	0.0878	1.13	0.259	0.1011	2.04	0.041	0.0795	1.69	0.091	0.0365	0.67	0.505
Union	0.1109	9.50	0.000	0.1611	6.25	0.000	0.0885	5.06	0.000	0.1041	4.38	0.000	0.1099	3.16	0.002
Married	0.0097	1.09	0.274	0.0148	0.76	0.446	0.0077	0.55	0.581	0.0024	0.14	0.889	0.0114	0.48	0.632
Female	(dropped)			(dropped)			(dropped)			(dropped)			(dropped)		
Non-White	(dropped)			(dropped)			(dropped)			(dropped)			(dropped)		
Health	-0.0219	-1.34	0.180	-0.0265	-0.83	0.407	0.0120	0.48	0.629	-0.0825	-2.29	0.022	-0.0178	-0.35	0.727
Welfare Received	0.0000	-5.54	0.000	0.0000	-3.63	0.000	0.0000	-2.23	0.026	0.0000	-4.18	0.000	0.0000	-0.26	0.798
SMSA - Not in Central City	0.0000	0.01	0.996	0.0121	0.59	0.555	0.0194	1.33	0.183	-0.0396	-2.29	0.022	0.0054	0.23	0.821
SMSA - Central City	0.0100	0.80	0.425	0.0024	0.09	0.930	0.0314	1.55	0.120	-0.0400	-1.55	0.122	0.0566	1.77	0.077
Unemployment Rate 3.0-5.9%	-0.0183	-1.45	0.148	-0.0086	-0.29	0.771	-0.0445	-2.15	0.032	-0.0422	-1.79	0.073	0.0559	1.78	0.075
Unemployment Rate 6.0-8.9%	-0.0279	-1.93	0.053	-0.0089	-0.26	0.791	-0.0568	-2.39	0.017	-0.0653	-2.40	0.016	0.0642	1.78	0.074
Unemployment Rate 9.0-11.9%	-0.0225	-1.27	0.205	-0.0279	-0.69	0.490	-0.0470	-1.61	0.108	-0.0471	-1.40	0.161	0.0703	1.57	0.117
Unemployment Rate 12.0-14.9%	-0.0521	-2.41	0.016	-0.0830	-1.78	0.075	-0.0582	-1.61	0.108	-0.0758	-1.84	0.066	0.0487	0.88	0.380
Unemployment Rate >15.0%	-0.0687	-2.02	0.043	-0.0795	-1.18	0.236	-0.0893	-1.62	0.106	-0.0498	-0.74	0.460	-0.0432	-0.40	0.686
Occupation - Sales Workers	-0.0861	-5.68	0.000	-0.0561	-1.52	0.128	-0.1132	-4.72	0.000	-0.0632	-2.27	0.023	-0.0807	-2.00	0.045
Occupation - Clerical	-0.0131	-1.23	0.218	-0.0155	-0.51	0.611	-0.0320	-1.88	0.060	0.0021	0.11	0.909	0.0031	0.12	0.906
Occupation - Craftsmen	-0.0795	-9.34	0.000	-0.0529	-3.02	0.002	-0.0701	-5.18	0.000	-0.0880	-5.14	0.000	-0.1415	-5.64	0.000
Occupation - Armed Forces	0.0934	2.84	0.005	0.1413	2.28	0.023	0.0304	0.55	0.581	0.1338	1.96	0.050	0.0864	0.92	0.357
Occupation - Operatives	0.0294	1.87	0.062	0.0431	1.26	0.207	0.0159	0.61	0.539	0.0934	3.00	0.003	-0.0465	-1.20	0.231
Occupation - Laborers, not farm	0.0130	1.00	0.319	0.0262	1.10	0.272	0.0111	0.54	0.588	0.0293	1.00	0.318	-0.0324	-0.76	0.446
Occupation - Farmers	(dropped)			(dropped)			(dropped)			(dropped)			(dropped)		
Occupation - Farm Laborers	0.2712	1.58	0.114	(dropped)			-0.0004	0.00	0.999	0.3254	1.83	0.068	(dropped)		
Occupation - Service Workers	0.0246	0.08	0.940	0.052	0.15	0.877	(dropped)			(dropped)			(dropped)		
Occupation - Private Household	(dropped)			(dropped)			(dropped)			(dropped)			(dropped)		
Job Tenure	0.0003	13.01	0.000	0.0003	4.97	0.000	0.0003	9.18	0.000	0.0002	6.66	0.000	0.0002	4.08	0.000
Experience (weeks)	0.0008	7.91	0.000	0.0007	3.09	0.002	0.0011	5.92	0.000	0.0007	3.20	0.001	0.0007	2.96	0.003
Number of Jobs	0.0092	4.09	0.000	0.0129	2.57	0.010	0.0135	3.73	0.000	0.0062	1.52	0.128	0.0024	0.34	0.734
1989	0.0510	4.60	0.000	0.0242	0.98	0.326	0.0536	2.96	0.003	0.0562	2.64	0.008	0.0727	2.57	0.010
1990	0.0803	7.12	0.000	0.0716	2.82	0.005	0.1067	5.82	0.000	0.0570	2.66	0.008	0.0663	2.29	0.022
1991	0.1367	11.57	0.000	0.1167	4.43	0.000	0.1434	7.43	0.000	0.1240	5.52	0.000	0.1633	5.33	0.000

1992	0.1504	12.20	0.000	0.1037	3.77	0.000	0.1548	7.70	0.000	0.1444	6.15	0.000	0.2062	6.46	0.000
1993	0.1916	15.20	0.000	0.1436	5.03	0.000	0.2066	10.09	0.000	0.1785	7.48	0.000	0.2392	7.32	0.000
1994	0.2292	17.62	0.000	0.2028	6.80	0.000	0.2231	10.59	0.000	0.2261	9.25	0.000	0.2809	8.22	0.000
1996	0.2627	18.81	0.000	0.2058	6.48	0.000	0.2461	10.86	0.000	0.2850	10.78	0.000	0.3386	9.22	0.000
1998	0.3135	20.63	0.000	0.2988	8.62	0.000	0.2782	11.29	0.000	0.3413	11.85	0.000	0.3678	9.18	0.000
2000	0.3863	22.56	0.000	0.3380	8.72	0.000	0.3633	13.00	0.000	0.4064	12.46	0.000	0.4711	10.45	0.000
Low numeracy	(dropped)														
High numeracy	(dropped)														
Very high numeracy	(dropped)														
Constant	6.4279	270.39	0.000	6.1921	118.34	0.000	6.3243	165.01	0.000	6.6230	145.50	0.000	6.6498	97.77	0.000
Number of observations	31130			6741			11377			7714			5298		
Number of groups	3911			902			1423			904			682		
R-sq: Within	0.1642			0.1447			0.1743			0.1853			0.1617		
R-sq: Between	0.2538			0.2773			0.2232			0.1871			0.1604		
R-sq: Overall	0.1819			0.1848			0.1738			0.1745			0.1516		