Do college students make better predictions of their future income than young adults in the labor force?

Abstract

Several studies have considered whether American college students' hold "realistic" wage expectations. The consensus is that they do not – overestimation of future earnings is in the region of 40 to 50 percent. But is it just college students who overestimate the success they will have in the labor market, or is this something common to all young adults? In this paper I analyze National Educational Longitudinal Study (1988) data to consider whether twenty year old college men are more realistic about their future income than their peers (of the same age) who are already in the labor force. My findings suggest that young people in employment actually make worse predictions of their future income (on average) than certain student groups, so long as the latter successfully obtain a university degree.

Key Words: Income expectations, occupational expectations, college students

JEL codes: I20, I21, I28

1. Introduction

The expectations that people hold for the future play a critical role in many micro-economic models of decision and choice. For instance, the wage premium that one expects to receive for completing a college degree is a key determinant of young people's willingness to invest in college (Attanasio and Kaufmann 2009). Yet despite the theoretical importance of individuals' subjective expectations, for most of the twentieth century they were the subject of relatively little empirical analysis. This has started to change, however, over the last fifteen to twenty years, with the study of subjective data becoming increasing common in economic research.

One of the most widely studied topics in this area surrounds the expectations that young people hold for the future. Economists have been particularly interested in whether college students are able to accurately predict their future income and how this varies with characteristics such as gender, race, institutional quality and college major (Smith and Powell 1990, Betts 1996, Rouse 2004, Jerrim 2011, Webbink and Hartog 1999). Using small scale convenience samples, often collected from students enrolled in just one or two institutions, researchers have typically found the predictions made to be rather poor. With regards to the American literature, average expected earnings typically exceed later average actual earnings by 40 to 50%. Arcidiacono et al (2012) have recently shown how such misalignment of young people's expectations can seriously influence the decisions that they make. Specifically, they suggest that one in twelve college students would pick a different college major if their expectations were "correct" (contained no forecast errors).

It is this literature (on the accuracy of young people's wage expectations) that I shall build upon in this paper. As noted in the paragraph above, expectations of future income streams have almost exclusively focused on those who have made the decision to enter college. Almost no work (to my knowledge) has considered the same issue for those young adults who have chosen to enter the labor market instead. Although this group has been largely ignored in the existing literature, it is both large in size and of substantive interest. For instance, by investigating this group one can establish whether unrealistic labor market expectations is something specific to young people still in education, or if it is a more general phenomena that occurs across a wide body of young American adults.

But why might young people who are already working make better predictions about their future income than their peers in college or school? One possibility is that they will have contact with older workers who may, either formally or informally, pass on sector-specific

details of future pay and progression. Alternatively, organizations themselves could make information on career progression and pay freely available to their staff. Another factor is that workers have been through the job search process at least once. They should have found out about wages and career opportunities during this time. Indeed, these individuals may have held unrealistic expectations before this experience, but actually going out and trying to find a job may have taught them about the realities of the labor market[1]. Many university students will not have yet been through a similar process. Self-selection into the labor market or college may also play a role. Educational attainment is linked to migration (see Borjas 1999). The less educated (who have self-selected into work) are more likely than students to stay in their age 20 location. Workers may therefore gather information about wages in the local labor market that they incorporate into their expectations. Students, on the other hand, may well expect to work in other areas of the country, making the local labor market less informative for many of them. Each of the above thus suggests that the expectations of college students are likely to be less realistic than those of young adults already in employment.

One may, of course, put forward arguments as to why the opposite might hold true. In particular, those who enrolled in college may be less myopic than their peers who chose to enter the labor force, or are more focused on one particular career. College students may therefore collect more specific information about certain types of jobs, while also putting more effort into learning about long-term labor market outcomes. Similarly, with career councilors available in almost all college institutions, one could argue that students should actually be quite well-informed about the labor market. Indeed, each of the above could hold for certain groups of college students but not for others; those enrolled in an Education, Nursing or Engineering course are training for a particular career and thus have incentive to collect very specific information, while the same may not be true for those majoring in the Social Sciences, Humanities or Liberal Arts.

It is therefore clear that the accuracy of college students' wage expectations may differ to those of young adults in the labor force, making comparison between the two groups of substantial interest. The aim of this paper is thus to assess whether young people who are working in the labor market make better predictions of their future income than those who

[1] Recent work in the sociological literature by Morgan (2005) depicts young adults as "Bayesian learners". In particular, he illustrates the accuracy of a "fast" and "slow" learner's expectations over time. Morgan suggests the difference between fast and slow learners could be to do with the different timing of key life events. This could include entry into the labor market, a period when young adults should receive a lot of information that will lead them to quickly (and more accurately) updating their expectations.

chose to enter college, and how this differs between various sub-groups within these two populations. This includes an investigation of students enrolled in different majors and young employees working in different occupations. Moreover, a distinctive feature of the National Educational Longitudinal Study (NELS) data that I analyze is that it follows the same individuals over time. Hence I am able to investigate whether the wage expectations of students who successfully graduate from university are more accurate than those who enrolled in college but never completed their course.

The paper now proceeds as follows. In section 2 I describe the NELS data and my empirical methodology. Results are then presented in section 3, with conclusions to follow in section 4.

2. Data

The data that I analyze is the National Educational Longitudinal Survey (NELS) of 1988. This study's aim was to provide information about American adolescents at critical points in their development and later into their careers. Children were initially interviewed in 1988 when most were 14 years of age. They were then followed up four further times, at ages 16, 18, 20 and 26 (the last wave having taken place in 2004). Parents and teachers of the pupils also completed the first three rounds of the survey.

In the first wave (age 14), a two-stage stratified sampling design was employed, with schools as the primary sampling unit, and probability of selection (of schools) proportional to size. 1,052 schools participated in the survey, including some oversampling of private institutions. A random sample of 26 students was then selected from each school. 26,432 students were eventually selected with 24,599 taking part (93%). In each of the next two waves (age 16 and 18) students who participated in the initial survey were followed up. The sampling process added some newly selected students (1,043 at age 16 and 244 at age 18)[2]. This was done to create a valid probability sample (a nationally representative cross-section) of students in each of the respective years. In total, 20,923 18 years olds took part in the third wave. The fourth wave took place when students were 20 years old. To reduce costs, a sub-sample was selected based on demographic characteristics and response history. It is important to note that this reduction is not the result of sample attrition, but from a conscious

[2] These students were not randomly selected, but drawn from schools where there were other second and third wave respondents. More details can be found on page 56 of Curtin et al (2002).

effort of the survey design to limit burden and cost[3]. This led to the age 20 sampling frame being reduced to 15,964 individuals. In total, completed responses were available from 14,915 (93%) 20 year olds. The final survey took place when most sample members were 26, with information for 12,144 individuals available (76% of the age 20 sub-sample). Throughout my analysis I apply the sampling weights provided by the survey organizers. These weights help to correct for the oversampling of certain groups (e.g. low income and minority students) and attrition from the survey between the first and the final survey wave[4].

The sample that I analyze in this paper is subject to three further restrictions. Firstly, my analysis will focus on men only due to the added complexity of labor market selection for women, reducing the number of observations available to 5,782. Secondly, 743 individuals had missing or problematic data on income expectations and were therefore dropped from the analysis. Finally, 605 individuals had no information on actual income by the time of the last survey (at age 26) and are also excluded[5]. This leaves me with a working sample of 4,434 observations. In Table 1 I investigate the extent of selection this induces into the sample in terms of observable characteristics. The left hand column illustrates the characteristics of the initial 5,782 male observations in the complete NELS sample, while the column on the right shows the characteristics of the 4,434 individuals who are not missing any key information. Reassuringly, the distribution of observable characteristics remains reasonably similar[6].

Table 1

As part of the age 20 survey, all respondents were asked the following question about their future income (with responses to be provided in an open text field):

"What do you expect your total annual income to be when you are 30 years old?"

Although respondents are clearly asked to predict their future income, and the wording used

[3] Around 5,000 individuals were dropped from the study. 2,000 of these were classed as "poor responders", who were basically excluded because of the low chance of future contact. Hence it may be more appropriate to consider these 2,000 observations as non-respondents. The other 3,000 individuals dropped were not classified as poor responders, but excluded purely to lower costs.

[4] The substantive conclusions of this paper remain unchanged whether or not one applies these survey weights in the analysis.

[5] Further analysis (which can be found in Jerrim 2010) shows that most of these observations were graduate students who had never entered full-time employment.

[6] Jerrim (2010) further investigates this issue, and demonstrates that results are robust to methods that attempt to correct for any selection induced in the sample in terms of observable characteristics.

is similar to that in many of the existing studies on students' wage expectations, this type of question is not without its problems (Manski 2004 provides an overview of the difficulties with using such open questions to elicit individuals' expectations). For instance, when this question is considered in isolation it is not immediately apparent as to whether respondents should take into account inflation, report gross or net earnings, and if they should provide figures conditional on having a full-time job.

To shed some light on these issues, it is important to consider the ordering of survey questions. *Directly before* they were asked for their income expectations, respondents were given the following question:

*"What was **YOUR** total income from all sources, before taxes, in 1993?* [i.e. the year prior to interview] *This figure should include salaries, wages, pensions, dividends, interest, unemployment compensation, grants, financial aid, scholarships, government assistance (AFDC) and all other income"* [Capitalization in original question]
Write in dollar amount, write in 0 if no income $.....

Throughout the course of this paper, I assume that respondents use this question as an anchor on which to base their income expectations, and provide reports following the same criteria. For example, I assume respondents use their answer to the question above as a reference point and thus report their income expectations in current prices (all figures are therefore reported in 1994 dollars)[7]. Similarly, based upon the question above, I also assume that respondents report income expectations in gross (pre-tax) figures. Moreover, note the emphasis here on *income.* This would suggest respondents should not only take into account future wages, but also additional receipts such as benefit or interest payments. Another point to note is that the income expectation question asks for the "your" expected income. This is made even clearer in the preceding question, with **"your"** in bold, capital letters. It seems that the respondent should only be considering their own, personal income, and not their partner's or other family members. Following Dominitz and Manski (1996), I also assume that the expectations reported are conditional upon holding a full-time job. Given that I have restricted the sample just to men, and this refers to their anticipated income as 30 year olds,

[7] Ideally, students would have been formally instructed not to consider inflation in the wording of the question, as per Dominitz and Manski (1996). They report that students generally adhere to this, and do not consider inflation in their wage expectations. Moreover, Brunello et al (2004) use similar wording to the question asked in NELS, in that students are not directly informed how to deal with inflation. They also assume students report their expectations in current prices, and find inconsistencies in their data with the idea that respondents make an adjustment to their responses to try and account for inflation.

this seems a reasonable assumption to make. The final issue is how respondents interpret the word "expect". I assume that this variable is truly reflecting what respondents *think* will happen (their expectations) rather than what they *hope* will occur (their aspirations). Yet there is also the additional issue of what point on respondents' belief distribution their reported "expectations" correspond to (i.e. do the reported figures correspond to the mean, median or mode of the earnings distribution that young people believe that they will face). Ideally, this would have been made clear in the wording of the question, as per the methods of Dominitz and Manski (1996) and Manski (2004). Throughout the course of this paper, I assume that respondents are reporting their median expectation, and hence use this as my preferred measure of central tendency.

NELS also contains data on respondents' wage history. As part of the survey at age 26, they were asked various questions about their income over the past twelve months (including separate questions about earned and non-earned income) along with their earnings from work for each of the three previous years. Hence data is available for each of the sample members' earnings from work between the ages of 23 and 26, along with their unearned income in the last of these years.

At this point, however, one should note that there are two significant problems with comparing expectations and realizations using the NELS data:

(a) At age 20, respondents were asked what they expect their income (and occupation) to be at age 30. However, data on labor market realizations is only collected between ages 23 and 26.
(b) Respondents are asked about their expected *income.* Data on realizations focuses on *wages*.

In other words, respondents' actual age 30 income is unobserved and must be estimated from the data available. In Jerrim (2010) two methods of predicting age 30 income are described and applied[8]. The results presented here are based upon prediction "method 2" from the aforementioned working paper. These predictions are based upon a fixed effects regression model, following the methodology of Carneiro and Heckman (2003). Firstly, the NELS survey (which only contains wages until 26) is pooled with a second data source (the

[8] Jerrim (2010) is a working paper version of this paper. It contains a set of additional results, along with further details of the NELS data that I analyze. This includes an explanation of how I predict individuals age 30 income from the information in the dataset that is available. Throughout this paper, I present results that refer to prediction "method 2".

National Longitudinal Study of Youth NLSY 1979) that follows individuals to the point of interest (up to age 30). The following wage equation is then estimated (using this pooled dataset) for five educational groups:

$$Y_{ia} = \alpha + \beta_0 A_{ia} + \beta_1 A_{ia}^2 + \eta_i + \varepsilon_{ia} \quad \nabla \in E$$

Where:

Y_{ia} = log earnings of individual i at time a

A_{ia} = Age of individual i at time a

η_i = Individual (or fixed) effect

ε_{ia} = Error term, assumed to be normally distributed

E = Highest qualification achieved by age 26 (Less than high school, high school, associate degree, bachelor's degree, MSc/PhD)

The estimated coefficients then enter a prediction equation for age 30 wages:

$$\hat{Y}_{i30} = \alpha + \hat{\beta}_0 A_{i30} + \hat{\beta}_1 A_{i30}^2 + \hat{\eta}_i + \hat{\varepsilon}_{i30} \quad \nabla \in E$$

Where:

$\hat{Y}$ = Predicted log wage at age 30

$\hat{\varepsilon}_{30}$= Random draw from the distribution of errors at age 26 (assumed to be normally distributed)

Jerrim (2010) provides further details, and illustrates that substantive conclusions remain intact when using (i) a different method to predict age 30 income and (ii) no further prediction at all (i.e. when expected age 30 income is compared to actual observed income at age 26).

At the end of the following section I shall also present some additional results referring to other aspects of young adults' anticipated labor market success as an additional robustness check. In particular, I put forward the argument that if young people are unrealistic about their future income, they are also likely to be unrealistic about other aspects of the labor market, like their future occupation. Analogous to finding excessive income expectations, individuals may expect to be in a professional occupation when they turn 30, but actually end up working in a relatively low paying job. As part of the NELS survey at age

20, individuals were asked what occupation they thought they would be working in at age 30[9]. Then, in the final survey wave (age 26), individuals were asked what occupation they currently hold. Therefore I also compare expected and realized occupation to support my main analysis surrounding young adults' income expectations.

I conclude this section by defining my three broad groups of interest: (a) workers (b) students who go on to graduate (c) students who do not complete college. The first group ("workers") refers to sample members who were economically active, employed and had no attachment to any college institution at age 20. Group two ("student – graduates") refer to young adults who were enrolled in college at age 20 and had obtained a degree by the time of the last survey wave at age 26[10]. Finally, "students who do not complete" refer to sample members who were enrolled in college at age 20, but who (at age 26) were working full-time without holding a degree. In the following section, I shall present results for these three groups before further sub-categorizing each depending on either the subject they study or the occupation held (when they were making their predictions of future income).

3. Results

I now turn to my comparison of young adults' expected and "actual" income at age 30[11]. To begin, I consider the results for those who were enrolled in a college course at age 20 (without, for the moment, distinguishing between those who went on to successfully graduate and those who did not). Results can be found in the left hand panel of Figure 1. The red dashed line illustrates the distribution of college students' expected income for when they turn 30, while the solid black line refers to my prediction of "actual" income at the same age. Both sets of estimates are presented on the log scale.

Figure 1

It is immediately clear that 20 year old students tend to overestimate their income ten years into the future. The expected income distribution is some distance to the right of the "actual" income distribution, with average expected earnings ($45,000) being around $15,000 dollars

[9] The exact wording of the question was: "What job do you expect or plan to have when you are 30 years old?" Respondents were asked to write in an occupational description into an open text field.

[10] It is, of course, possible for students to also hold a job while they study. In auxiliary analysis not presented, I investigated whether the employment status of students was associated with the accuracy of their wage expectations. I found little evidence that this was the case.

[11] As explained at the end of section 2, "actual" age 30 income is unobserved in NELS and so must be estimated from the data available. See Jerrim (2010) for full details.

more than that for actual wages ($30,200). This result is consistent with much of the existing US literature on college students' wage expectations; Smith and Powell (1990), Rouse (2004) and Betts (1996) have all similarly found that college students over-estimate their future income by large amounts (up to 50 percent).

In the right hand panel of Figure 1 I present analogous estimates for the 20 year olds who were in employment. Again, the red dashed line refers to the expected income distribution and the solid black line the actual income distribution. Interestingly, the general pattern seems to be very similar. Note, in particular, that the expected income distribution is again to the right of the predicted actual income distribution. Moreover, the median expected income ($40,000) is significantly below the actual median income ($24,400), with overestimation typically standing around $16,000. One can see how similar these results are to those for college students, with one unable to reject the null hypothesis that the two groups (students and workers) overestimate future income by the same amount. Auxiliary analysis (not presented for brevity) indicated that this result remains intact even when using different prediction methods for age 30 income and alternative sample selections. Interestingly, I also find that low wage earners at age 20 (bottom quartile of the income distribution) hold significantly less realistic expectations than other groups. Specifically, the difference between median expected and median actual age 30 income for low wage earners is around $20,000 (compared to roughly $12,000 for workers in the second, third and top age 20 income quartiles).

Further detail is added to this result by dividing the student sample into two separate groups – those who successfully obtained a bachelors degree by age 26 and those who did not. This can be found in Figure 2.

Figure 2

The contrast between the two sets of results is striking. There is quite a reasonable degree of overlap for the two distributions in the right hand panel (graduates), with the average expected wage ($50,000) being roughly $15,000 more than the average actual wage ($34,400). The right hand side of Figure 2 (for the subset of college non-completers) tells a rather different story. There is little overlap of the expected and actual age 30 income distributions, with the medians being $21,700 apart ($45,000 compared to $23,300). In other words, college students who do not complete their course end up earning (on average) roughly half the amount of income they once anticipated. This is notably more than both their

peers who were in the labor force at age 20 and those college students who successfully graduated.

One must, however, exercise caution when interpreting this result. In particular, I am unable to say whether the inaccuracy of non-completing college students' income expectations is a cause or a consequence of their decision to leave university. One possibility is that these students stated their income expectation on the assumption they would obtain a certain level of human capital and a valuable labor market signal. However, they did not go on to actually receive the outcomes they initially anticipated from their human capital investments, thus causing their apriori expectations to be incorrect. Alternatively, these individuals could have dropped out of university *because* of their overly ambitious expectations. For instance, they may have gone to university thinking they would earn a high wage (i.e. their high expectation observed at age 20). But through their later experiences, they may have revised down their expectations substantially (i.e. if one were to observe their expectations again at age 21, they would perhaps be much lower). On the basis of this revision, they may have decided that the benefits from obtaining a degree are not worth their continued investment, and hence leave university before the end of their course. It could thus be these students unrealistically high expectations that is driving their decision to leave university, rather than their expectations being unmet because they drop out (i.e. the decision to not complete college is potentially endogenous with respect to expected income).

Thus far, I have treated both college students and workers as largely homogeneous groups. There is, of course, also variation within both. In Table 2 I now consider how the accuracy of young adults' income expectations varies between students majoring in different subjects (Panel A) and workers who are employed in different occupations (Panel B). The figures presented in this table refer to the difference between average expected and "actual" age 30 income in thousands of 1994 US dollars. The stars in Panel A indicate where overestimation for students enrolled in that particular subject is significantly different to the overestimation made by their peers who were in the labor force (with the latter being broadly defined)[12].

Table 2

The first striking point is once again the stark difference between college graduates

[12] Standard errors are based on 200 bootstrap replications. The NELS complex sample design (children clustered within schools) has been taken into account by clustering the bootstrap by school.

and college non-completers; in most subjects the overestimation of future income is a lot lower for the former than it is the latter. One such example is the Business Management group; those who successfully complete such a course overestimate their future income by (on average) $12,800 compared to almost double this amount for those who dropped out ($23,100). Secondly, it is interesting to see that some groups of students actually make quite good predictions of age 30 income (conditional upon successful completion of their course). Note, in particular, that one can not reject the null hypothesis that there is no difference between average expected and average actual income at age 30 for Computer Science, Maths, Engineering, Physical Science and Agricultural students (at the 5% level). Hence, against the main thrust of the existing literature, these results suggest that certain groups of students can make reasonable predictions of their labor market future. It is also interesting to see that students who graduate with a major in these subjects make significantly better predictions of future income than young adults in the labor force (the same is also true for those who graduate with a degree in Education). Together, this adds support to the notion that workers do not make better estimates of their future income than those enrolled in a college course and, compared to certain groups of students who go on to graduate, their predictions are sometimes significantly worse.

I investigate this finding further by estimating an OLS regression model where the dependant variable is defined as the difference between expected and actual age 30 income[13]. The sample of students included is restricted to only those who go on to successfully graduate from college by age 26. Formally this model is specified:

$$Y_{ij} = \alpha + \beta_1.S_i + \beta_2.G_i + \beta_3.R_i + \beta_4.F_i + \beta_5.A_i + \beta_6.W_i + \varepsilon_{ij}$$

Where:

Y = Expected minus actual (predicted) income at age 30

S = Subject student studied (reference group: working in labor force)

G = Gender

R = Race

F = Family background (parental education and occupation)

A = Academic achievement in secondary school (math, English and science test scores)

W = Whether the student worked while studying (a proxy for work experience)

ε = Error term

[13] I have investigated the sensitivity of my findings to using quantile regression (focusing upon the median) rather than OLS. The substantive conclusions that I reach remain largely unchanged.

i = student i

j = school j[14]

Table 3 presents estimates from five model specifications, with variables entering in the sequential order outlined above. My goal is to investigate whether the more realistic expectations of Math, Engineering, Computer and Physical Science students can be "explained" by these potential confounding factors. Specification 1 presents estimates from the base model, where only the variable of interest (the subject the student is studying compared to workers as the reference group) is included. Notice that the parameter estimates for Computer Science, Math, Engineering and Physical Science are negative and statistically significant – as previously noted students who graduate in one of these subjects hold more realistic expectations than young people already in the labor force. Specification 2 and 3 illustrate that this finding holds when one controls for gender, race and socio-economic status (in-fact the absolute magnitude of the difference actually increases). The inclusion of prior academic achievement in specification 4 does seem to have some impact – for instance the difference between Computer Science / Math students and workers falls from $13,300 to $9,900. However, large and statistically significant differences remain. Adding an additional control for work experience (whether the student also held a job while at college) does little to change this result. Indeed, specification 5 continues to show how Computer Science, Math, Physical Science and Engineering students hold significantly more realistic expectations than workers, while the opposite is true for those enrolled in Social Science, Humanities, Law, Health and Biological Science courses. Clearly, differences in the demographic and academic composition of these subject groups are not the main driver of this result.

Table 3

In panel B of Table 2 I divide workers into different occupational groups. There is some evidence that the accuracy of young workers' income expectations is linked to the job that they hold. Yet it is also interesting to note that this association (between job type and the accuracy of workers' income expectations) is not particularly strong. For instance, the smallest difference between average expected and actual income is for craftsman ($9,800), which is not substantially different from clerical ($12,300), service ($12,500), laborer

[14] This refers to the school the respondent was in when selected in take part in the study at age 14. I have taken this clustering into account by making the appropriate adjustment to the estimated standard errors.

($14,700) and managerial ($15,500) groups. Indeed, the only occupations that are notably different from others are farm and military workers, for whom sample sizes are rather small (approximately 60 for the former and 30 for the latter). Thus, unlike college students (where there are large differences between young people enrolled in different subjects), there seems to be limited variation in the accuracy of income expectations for young people working in different jobs.

To conclude this section, I investigate whether similar patterns hold with respect to young adults' occupational plans. Specifically, the occupation that young adults' expect to work in at age 30 (as reported in the age 20 survey) is divided into 18 different groups. The actual occupation that the respondent is working in at age 26 is then assigned into one of the same 18 categories. If the "expected" and "actual" jobs are the same, then their occupational expectation is said to be "correct". In Table 4 below I present the proportion of NELS respondents where this is the case. The top panel refers to college students grouped by their major, while the bottom panel is for those who were in employment grouped by their occupation.

Table 4

Reassuringly, many aspects of Table 4 are consistent with the previously presented results. For instance, it is Education, Computer Science, Math, Engineering and Physical Science students who once more make the best predictions of the future (i.e. I find their occupational expectations are the most likely to be met, just as I found that they had ,on average, the most realistic income expectations). Similarly, it is college students' majoring in these subjects who make significantly better predictions of their future occupation than their peers who were in the labor force at age 20 (conditional upon successful completion of their course). Finally, also note that the occupational expectations of college students who go on to graduate are much more likely to be met than for those who drop out. This is once more consistent with my previous finding that the former tend to hold more accurate wage expectations than the latter. Together this provides further support for the findings and substantive conclusions drawn throughout this section. Specifically, this demonstrates that the results presented are unlikely to have been driven by the method used to predict age 30 income or the assumptions that I have made about the income expectation data (e.g. that it is reported under the assumption of working full-time and is given in current prices). The evidence from this section therefore suggests that college students do not make worse

predictions of their future wages than young men who are already active in the labor force. Indeed, the predictions made by certain sub-groups of the student population are actually significantly better.

4. Conclusions

The subjective expectations individuals hold for the future is a growing area of economic research. One topic that has received particular attention is whether college students are able to accurately predict their future income. The general consensus in the existing literature is that expected earnings typically exceed actual wage outcomes by 40 to 50 percent. There has, however, been little attempt to put this result into a comparative context. Is it just college students who overestimate the success they will have in the labor market? Or is this something that is common to all young adults?

My findings provide little evidence that 20 year olds in full-time employment make better predictions of their future income (or occupation) than those enrolled in college. In fact, for certain groups I find the very opposite to be the case. Conditional upon successful completion of their course, those with a major that contains a large mathematical component (e.g. Computer Science, Engineering, Physical Science and Math) over-estimate their age 30 income by less than those already in employment. Indeed, in contrast with the existing literature, I find that some of these groups actually make quite good predictions (on average) of their pay ten years into the future.

What can explain these findings (particularly that young workers do not make better predictions than college students)? As noted in the introduction, it maybe that young workers are not focused on a particular career and hence suffer from a lack of direction in the labor market. This may also help to explain the differences found between college subject groups – Engineering and Math students may have specific career paths in mind while those studying Social Science, Humanities or Biological Sciences do not. Alternatively, it might be that young workers are myopic and choose to collect information from those who are closer to them in terms of age and the next rungs on the career ladder, whereas those enrolled in college are more likely to consider longer term outcomes (including their earnings at distant points in the future). Another possibility is that workers have both "accurate" and "inaccurate" sources of labor market information that they struggle to distinguish between. For instance, a manager may be keen to retain a particular staff member who is considering employment elsewhere. Thus the manager may overstate the chances of pay and progression

within the firm. If the worker cannot tell that this is "bad" information, it may lead him to raise his future income expectations. Indeed, in situations where workers only receive relatively poor quality information, one would expect them to be no more (and possibly even less) realistic than their student peers.

Alternatively, some young adults may not realize the value of the information that they hold, or how it applies to them and their future; they may discard (or give less weight) to some important information as they see it as irrelevant. For example, a young worker may know what a 30 year old employee in his organization is paid. But he (perhaps unrealistically) views his current job as a stop-gap solution, and believes he will have entered an entirely different industry in a few years time. He therefore does not fully incorporate the information he holds on the wages of 30 year olds into his income expectations. Such an interpretation is consistent with Smith and Powell (1990) who found that although young adults understand what average wage levels are, they expect their own salaries to be a lot higher. Again this may also explain the differences I find between subject groups – although Social Science, Humanities and Biological Science students know what most graduates in these fields earn, they do not fully incorporate this information into their expectations. Finally, it is possible that most young people simply do not have any idea what they will be earning at age 30, and thus report their quite vague view of what average salaries are at this age. This may explain why when expectations are compared to realizations, students in certain subjects (e.g. Math, Engineering etc) make more accurate predictions than others – all young people anchor their (quite vague) expectations around the same (rather hazy) point and the greater accuracy in the expectations for such groups simply reflects the fact that they tend to earn more than others.

It is difficult with the data available to distinguish between each of these possible explanations. Indeed, much more detailed data on the information sources children and young adults hold is needed for economists and other social scientists to fully understand how such expectations are formed and their role in the decision making process. Although work in this area has already begun (see Zafar 2011 and Delavande 2008 for recent work exploring how expectations are developed and revised and Kaufmann 2009 for how this influences the educational decisions that young people make) there is also a clear need for future research.

Nevertheless, one should not lose sight of the contribution that this paper has made to the existing literature. Several studies have previously found college students to hold unrealistic expectations regarding their labor market future. This paper has extended this finding to show that this is a more general problem amongst American youth. A concern for

academics and policymakers is that young people may be basing important decisions on such miss-guided views. Intervention may hence be needed to ensure young people understand the realities of the labor market, including their likely future pay and possible careers.

References:

Attanasio, O. & Kaufmann. K. (2009) "Education Choices, Subjective Expectations and Credit Constraints", NBER Working Paper 15087

Arcidiacono, P., Hotz, J. & Kang, S. (2012) "Modelling College Major Choice using Elicited Measures of Expectations and Counterfactuals", Journal of Econometrics, volume 166, issue 1, pp 3-16

Betts, J. (1996) "What Do Students Know About Wages? Evidence from a Survey of Undergraduates", Journal of Human Resources, volume 31, pp27-56.

Borjas, G. (1999) "The Economic Analysis of Immigration", Handbook of Labor Economics, pp 1697-1760.

Brunello, G. Lucifora C. & Winter-Ebmer, R. (2004) "The Wage Expectations of European Business and Economics Students", Journal of Human Resources, volume 39, issue 4, pp 1116-1142

Curtin, T. Ingels, S. Wu,S. & Heuer, R. (2002) "National Education Longitudinal Study of 1988, Base-Year to Fourth Follow-up Data File User's Manual", National Center for Education Statistics

Delavande, A. (2008) "Measuring Revisions to Subjective Expectations", Journal of Risk and Uncertainty, volume 36, issue 1, pp 43-82

Dominitz, J. & Manski, C. (1996) "Eliciting Student Expectations of the Returns to Schooling", Journal of Human Resources, volume 31, pp 1-26

Jerrim, J. (2010) "Who Has Realistic Income Expectations: Students or Workers?", s3ri working paper A10/05

Jerrim, J. (2011) "The Wage Expectations of UK Students: Are They Realistic?", Fiscal Studies, volume 32, issue 4, pp 483 – 509

Kaufmann, K. (2009) "Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns to College", Mimeo. Bocconi University

Manski, C. (2004). "Measuring Expectations," Econometrica, volume 72, pp 1329-1376

Morgan, S. (2005) "On the Edge of Commitment: Educational Attainment and Race in the United States", Stanford University Press

Rouse, C. (2004) "Low Income Students and College Attendance: An Exploration of Income Expectations", Social Science Quarterly, volume 85, pp 1299-1317

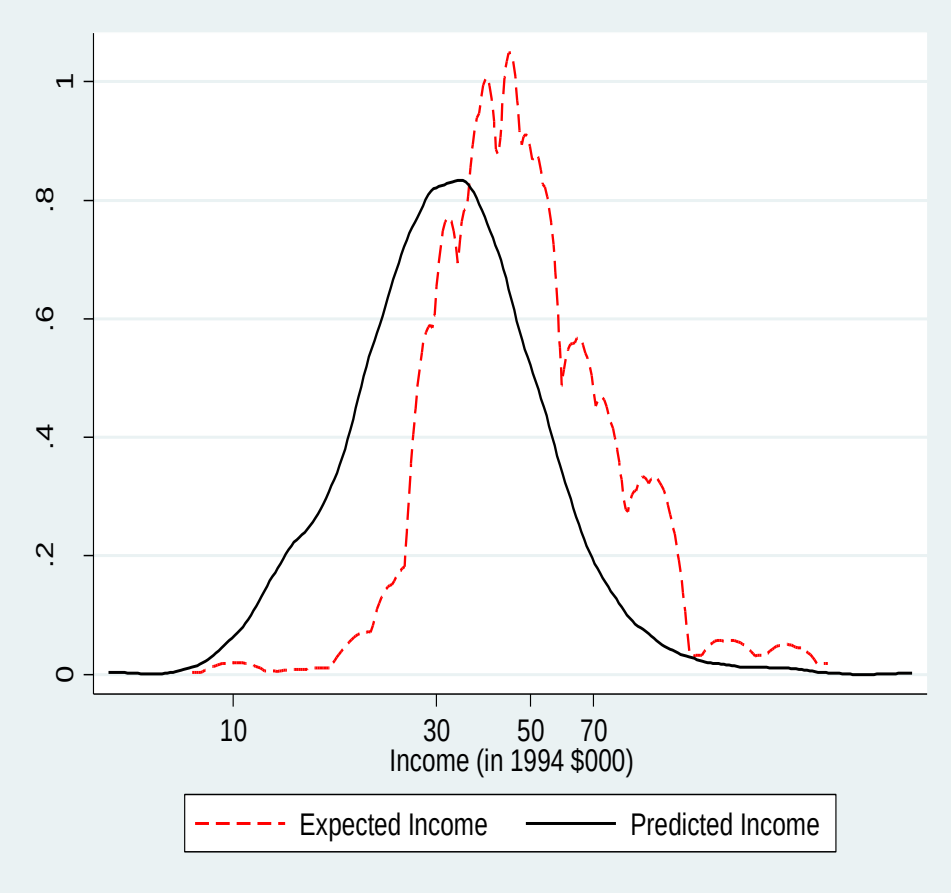
Smith, H. & Powell, B. (1990) "Great Expectations: Variations in Income Expectations Among College Seniors", Sociology of Education, volume 63, pp 194-207

Webbink, D. & Hartog, J. (2004) “Can Students Predict their Starting Salary? Yes!”, Economics of Education Review, volume 23, pp 103-113

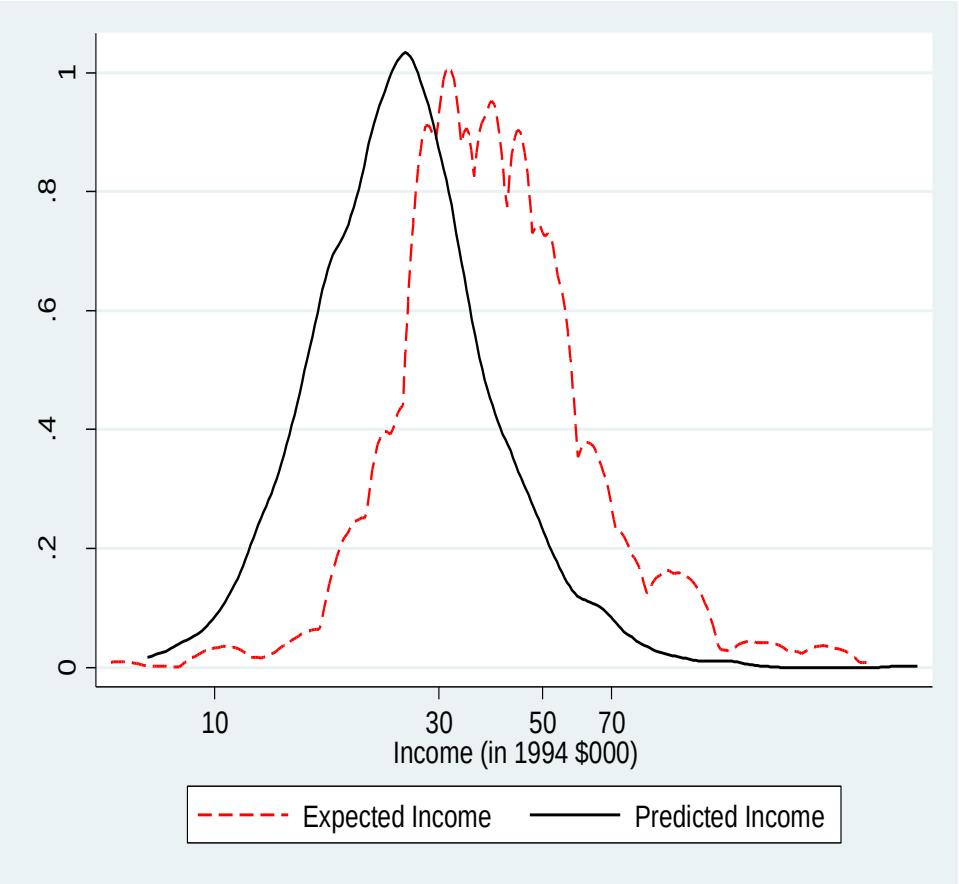
Zafar, B. (2011) “How Do College Students Form Expectations?”, *Journal of Labor Economics*, volume 29, issue 2, pp 301-348.

Figure 1. Distribution of expected and (predicted) actual income for students and workers

Students

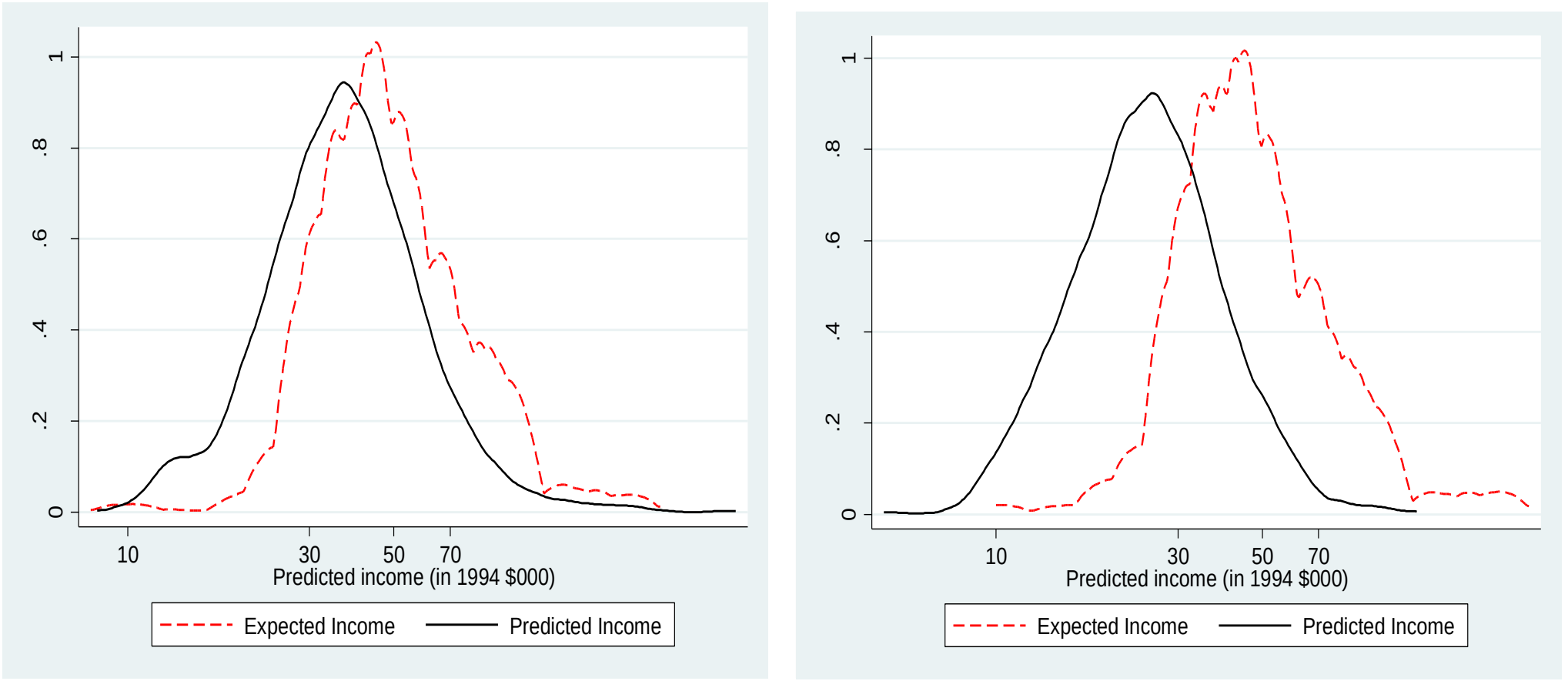


Workers



Notes:
The dashed red line is the distribution of expected age 30 income reported when NELS respondents were age 20. The solid black line refers to my prediction of actual age 30 income. All values presented are in real (1994) US dollars. Figures on the y-axis refer to probability density estimates. The optimal bandwidth has been used (0.08).

Figure 2. Distribution of income for university drop outs versus university graduates



Notes:
The dashed red line is the distribution of expected age 30 income reported when NELS respondents were age 20. The solid black line refers to my prediction of actual age 30 income. All values presented are in real (1994) US dollars. Figures on the y-axis refer to probability density estimates. The optimal bandwidth has been used (0.08).

Table 1. Summary statistics showing the NELS sample composition, before and after the exclusion of missing expectations and wage data

	Starting sample %	Final sample %
Labor force status at age 20		
Students who also have a job	26.6	26.0
Students who do not have a job	27.0	28.4
Working, not a student	35.0	35.8
Neither student or working	11.3	9.8
Highest qualification at age 26		
Less than high school	6.0	5.3
High school	55.6	56.2
Associates degree	7.1	6.9
Bachelors	28.2	28.6
MA/PhD	3.1	3.0
Race		
White	66.6	68.5
American Indian or Alaska Native	1.0	0.8
Asian or Pacific Islander	5.5	5.1
Black, not Hispanic	8.2	8.0
Hispanic or Latino	13.1	13.1
More than one race	3.0	2.6
Other	2.6	1.8
Family income student reported at age 18 ($1992)		
0-20000	18.5	18.1
20000-35000	19.8	19.8
35000-50000	17.7	18.4
50000-75000	16.0	16.5
75000+	12.3	12.4
Missing	15.7	14.9
University subject at 20 years old (If reported being a student)		
Agriculture	1.9	2.3
Accounting, Finance	6.1	6.2
Business Management	12.7	13.1
Journalism, Communication	3.3	3.6
Computer Science, Math	4.8	5.4
Education	5.1	5.4
Engineering, Physical Sciences	16.9	17.4
Languages	1.8	1.7
Health	6.8	6.2
Law	4.2	3.9
Biological Science	7.4	6.6
Social Sciences, Humanities	9.1	9.1
Arts	5.0	4.6
Other	14.9	14.6
Working full-time At age 26		
Yes	74.0	84.0
No (e.g. unemployed, student, working part-time etc)	26.0	16.0
Observations	5,782	4,434

Notes:
1 “Starting sample” refers to all men in the age 26 sweep of the NELS. “Final Sample” refers to the sample I use in my analysis, once I have excluded missing data

Table 2a. Difference between average expected and average "actual" age 30 incomes for those who were enrolled in college at age 20

	Graduates		Drop-outs	
	Difference	**SE**	**Difference**	**SE**
Computer science & Math	0.9*	4.8	14.8	11.4
Agriculture	3.7*	5.4	4.7	7.1
Engineering & Physical Sciences	5.8*	3.1	12.6	4.6
Education	9.8*	2.2	13.7	3.8
Journalism & Communication	11.2	5.5	41.2	5.7
Accounting & Finance	12.0	3.7	25.4	11.8
Business management	12.8	4.2	23.1	2.8
Arts	13.1	2.5	23.6	8.3
Law	14.9	4.5	33.6	16.7
Workers	**15.6**	**0.7**	**15.6**	**0.7**
Other	17.3	2.5	20.5	4.4
Social Sciences & Humanities	17.9	4.9	23.6	9.4
Biological Science	18.3	7.3	38.1	9.2
Languages	19.8	6.5	0.6	5.3
Health	24.5	5.3	20.9	5.1

Notes:
1 Figures refer to difference between median expected and median (predicted) actual age 30 income in thousands of 1994 US $.
2 The "graduates" column refers to NELS sample members who were enrolled in college at age 20 and who had successfully obtained a degree by age 26. The "drop-outs" column refers to those who were enrolled in college at age 20 but who at age 26 did not hold a degree and were working full-time.
3 * Indicate where the overestimation of age 30 income for that particular group is significantly different to the overestimation made by workers
4 The "workers" row refers to the NELS sample members who were in employment at age 20 (without distinguishing the type of employment in which they are in).

Table 2b. Difference between expected and actual age 30 income for young people in different occupations at age 20

	Difference	**SE**
Craftsman	9.8	1.8
Skilled operative	10.3	2.8
Clerical	12.3	1.6
Service	12.5	2.0
Other	13.6	7.0
Laborer	14.7	2.4
Manager	15.5	2.5
Sales	17.8	6.4
Farm	23.2	7.2
Military	29.0	6.2

1 See notes to Table 2a

2 Difference refers to difference between average expected and average "actual" age 30 income in thousands of 1994 US $.

Table 3. Parameter estimates from OLS regression model of the difference between expected and actual (predicted) age 30 income

	Specification 1		Specification 2		Specification 3		Specification 4		Specification 5	
	Difference	**SE**	**Difference**	**SE**	**Difference**	**SE**	**Difference**	**SE**	**Difference**	**SE**
Computer Science / Math	-11.0***	3.1	-11.8***	3.1	-13.3***	3.2	-9.9***	3.0	-9.1***	3.0
Engineering / Physical sciences	-8.7***	2.3	-8.8***	2.4	-9.9***	2.4	-7.1***	2.5	-6.4***	2.5
Accounting / Finance	-7.3**	3.1	-6.9**	3.1	-8.1***	3.2	-5.2*	3.1	-4.5	3.0
Education	-5.1*	2.7	-4.8*	2.8	-6.0**	2.9	-3.5	3.0	-2.7	3.0
Business /Finance	-3.4	2.3	-3.7	2.3	-5.2**	2.4	-1.9	2.4	-1.2	2.5
Arts	-3.2	3.7	-3.5	3.8	-4.8	3.7	-4.4	3.4	-3.5	3.6
Agriculture	-2.6	4.8	-2.5	4.5	-3.2	4.2	0.0	4.6	0.5	4.7
Other	-0.7	2.8	-0.5	2.8	-1.7	2.9	0.1	3.0	1.1	3.1
Languages	-0.1	5.4	0.3	5.4	-1.2	5.2	2.0	5.2	2.7	5.2
Workers	**REF**	**REF**	**REF**	**REF**	**REF**	**REF**	**REF**	**REF**	**REF**	**REF**
Journalism / Communication	0.4	3.2	0.3	3.2	-1.1	3.3	1.0	3.2	1.7	3.1
Social sciences / humanities	4.3	3.3	4.7	3.3	3.2	3.2	6.3**	3.1	7.0**	3.1
Law	5.7	3.9	6.2	3.9	5.5	3.7	6.9*	3.8	7.6**	3.9
Health	8.4**	3.4	8.2**	3.3	6.9**	3.4	9.8***	3.4	10.9***	3.5
Biological Sciences	8.9*	4.7	9.1**	4.7	7.6*	4.4	11.8***	4.7	12.4***	4.8
Race	-	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	-	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socio-economic status	-	-	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Prior academic achievement	-	-	-	-	-	-	Yes	Yes	Yes	Yes
Work Experience	-	-	-	-	-	-	-	-	Yes	Yes

Notes:

1 All figures refer to thousands of US dollars.

2 Negative figures suggest that students studying that particular subject hold more realistic income expectations than workers in the labor force.

3 *, ** and *** refer to a statistically significant difference (compared to workers) at the 10%, 5% and 1% level.

Table 4a. A comparison of whether college students were (by age 26) working in the occupation they had anticipated at age 20 (for different subject groups)

	All students		Graduates		Drop outs	
	Proportion	**SE**	**Proportion**	**SE**	**Proportion**	**SE**
Education	0.28	0.04	0.36	0.05	0.14	0.05
Computer science & Math	0.25	0.04	0.35	0.06	0.09	0.04
Accounting/Finance	0.28	0.04	0.35	0.05	0.06	0.04
Engineering and Physical Sciences	0.27	0.02	0.34	0.03	0.11	0.03
Journalism & Communication	0.24	0.05	0.31	0.07	0.09	0.06
Law & Liberty studies	0.24	0.05	0.31	0.08	0.05	0.03
Other	0.25	0.03	0.30	0.04	0.17	0.03
Arts	0.16	0.04	0.21	0.05	0.05	0.04
Business man/admin	0.20	0.02	0.21	0.03	0.19	0.04
Languages/ language based studies	0.16	0.06	0.19	0.08	0.09	0.09
Working	**0.18**	**0.01**	**0.18**	**0.01**	**0.18**	**0.01**
Social Sciences and Humanities	0.12	0.02	0.15	0.03	NA	NA
Agriculture	0.12	0.05	0.15	0.06	0.07	0.07
Biological Science	0.10	0.03	0.12	0.03	0.05	0.04
Health	0.07	0.02	0.09	0.03	0.05	0.03

Notes:

1 Figures refer to the proportion of respondents whose actual occupation at age 26 matched the occupational expectation they held at age 20 (e.g. 36% of Education graduates were in the career they expected at age 20 by the time they turned 26).

2 The "All students" column refers to all NELS sample members who were enrolled in a college course at age 20. The "graduates" column refers to NELS sample members who were enrolled in college at age 20 and who had successfully obtained a degree by age 26. The "drop-outs" column refers to those who were enrolled in college at age 20 but who at age 26 did not hold a degree and were now working full-time.

3 Data is sorted by the "graduates" column

4 The "working" row refers to the NELS sample members who were in employment at age 20 (without distinguishing the type of employment in which they are in).

Table 4b. A comparison of whether young adults in employment were (by age 26) working in the occupation they had anticipated at age 20 (for different occupational groups)

	Proportion correct	Standard error
Military	0.37	0.09
Craftsman	0.27	0.03
Sales	0.21	0.04
Laborer	0.18	0.02
Skilled operative	0.18	0.03
Manager	0.18	0.04
Other	0.16	0.05
Farm	0.15	0.05
Service	0.12	0.02
Clerical	0.12	0.02

Notes:

1 Figures refer to the proportion of respondents whose actual occupation held at age 26 matched the occupational expectation they held at age 20

2 Rows refer to the occupations the workers were in at age 20