ELICITING STUDENT EXPECTATIONS OF THE RETURNS TO SCHOOLING

Jeff Dominitz
Institute for Social Research
University of Michigan

and

Charles F. Manski
Department of Economics
University of Wisconsin

November 1994

This research is supported by grant SBR-9223220 from the National Science Foundation and by grant 91ASPE236A from the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. We are grateful to Charles Palit and Dawn Palit for making available the CAPI software. We are grateful to the Madison Metropolitan School District, Carolyn Taylor, Michael Harrington, Richard Steckleberg, W. Lee Hansen, and James Walker for enabling us to administer the survey to several classes of high school and college undergraduate students.

Abstract

We report here on the design and first application of an interactive computer-administered personal interview (CAPI) survey eliciting from high school students and college undergraduates their expectations of the income they would earn if they were to complete different levels of schooling. We also elicit respondents' beliefs about current earnings distributions. Whereas a scattering of earlier studies have elicited point expectations of earnings unconditional on future schooling, we elicit subjective earnings distributions under alternative scenarios for future earnings. We find that respondents, even ones as young as high school sophomores, are willing and able to respond meaningfully to questions eliciting their earnings expectations in probabilistic form. Respondents vary considerably in their earnings expectations but there is a common belief that the returns to a college education are positive and that earnings rise between ages 30 and 40. There is a common belief that one's own future earnings are rather uncertain. Moreover, respondents tend to overestimate the current degree of earnings inequality in American society.

1. Introduction

We report here on the design and first application of an interactive computer-administered personal interview (CAPI) survey eliciting from high school students and college undergraduates their expectations of the income they would earn if they were to complete different levels of schooling. The expected returns to schooling is central to economic thinking on schooling behavior. Hence it might be anticipated that our work would fit into a substantial literature collecting and analyzing data on student expectations. But there is almost no precedent literature and we have had to break considerable new ground.

The most relevant studies to date seem to be Smith and Powell (1990) and Blau and Ferber (1991). Smith and Powell asked a sample of college seniors to forecast their "anticipated annual income in 10 years" and their "expected earnings" in the first year of their first job. They also asked respondents to provide similar forecasts for the average member of their class. Blau and Ferber asked a sample of college seniors to forecast "how much they would expect to earn initially and after 10 and 20 years if they were to be *continuously* employed in their preferred occupation after leaving school." Our work differs in two main ways:

Elicitation of Subjective Earnings Distributions: The Smith-Powell and Blau-Ferber studies sought point forecasts from their respondents, using the phrases "anticipated annual income" and "expected earnings". We want to characterize students' uncertainty about the outcomes they will experience. The standard economic model of decision making under uncertainty supposes that agents place subjective probability distributions on their outcomes. We elicit subjective earnings distributions from our respondents.

Elicitation of the Expected Returns to Schooling: The Smith-Powell and Blau-Ferber studies asked their respondents to forecast their future incomes unconditional on their future schooling. We want to learn how students

perceive the returns to further schooling. This requires us to ask each respondent for two subjective income distributions, one predicated on a scenario in which the respondent leaves school at a specified time and the other supposing that he or she obtains further schooling.

There are only a few other studies eliciting student earnings expectations, and these are even more distant from our concerns. Freeman (1971) and Betts (1994) asked college undergraduates about the average earnings of persons in various occupations or major fields – they did not ask respondents about the earnings they themselves would expect to receive if they were in these fields. And there is a social psychological literature interpreting youths' responses to loosely worded questions as indicators of their expectations. Berndt and Miller (1990), for example, asked their sample of junior high school students to respond, on a five-point scale, to the question "How valuable do you think your education will be in getting the job you want?" Mickelson (1990) asked her sample of high school seniors to express their degree of agreement with the statement "Studying in school rarely pays off later with good jobs."

The dearth of studies of student earnings expectations reflects the longstanding antipathy of economists toward subjective data of all kinds. Rather than collect data on student expectations, economists have preferred to impose assumptions. In an analysis of the major field decisions of male college undergraduates, for example, Freeman (1971) assumed each student to believe that, should he select a given field, he would obtain the mean income realized by the members of a specified earlier cohort who did make that choice. In an analysis of the college enrollment decisions of male veterans of World War II, Willis and Rosen (1979) assumed that each veteran knows the actual stochastic process generating lifetime incomes and uses this knowledge to make rational-expectations predictions of his future income should he enroll or not enroll in college. In an analysis of the college enrollment

decisions of high school seniors, Manski and Wise (1983, Chapter 6) assumed each youth to believe that the return to enrollment at a particular college depends on his or her own ability and on the average ability of students enrolled at that college.

Our decision to collect data on student expectations reflects two considerations. First, we find unattractive the conventional practice of imposing assumptions on expectations. In principle, we do not see how a researcher would introspectively know what information students possess and how students use this information to form expectations. In practice, the absence of expectations data makes it impossible to assess the realism of the different, mutually incompatible assumptions that various researchers have made about student expectations. Moreover, incorrect expectations assumptions can yield incorrect inferences about the way that students make schooling decisions. See Manski (1993) and Manski (1995, Section 5.2).

Second, examination of the history of economic thinking about expectations data shows that the profession's antipathy, although longstanding, is based on only a narrow foundation. Over thirty years ago, economists used both empirical findings and logical arguments to conclude that the answers to certain vaguely worded qualitative questions reveal little about respondents' expectations (see Juster, 1966, and Manski, 1990). But the weakness of vaguely worded qualitative questions in measuring expectations implies nothing about the usefulness of more tightly worded probabilistic questions of the type answered by our student respondents. The available evidence does not support a condemnation of all expectations data.

Because of the lack of precedent, our work is necessarily exploratory. What we have learned about the design and administration of expectations surveys may well be of more lasting importance than the specific empirical findings that we report, so we explain our procedures at some length in Section 2.

We have administered the survey to 110 students, 71 of them high school students enrolled in social studies courses at a Madison high school and 39 of them college undergraduates enrolled in two economics courses at the University of Wisconsin-Madison. Given the absence of prior experiences with elicitation of probabilistic expectations from adolescents and given the entrenched skepticism with which subjective data are viewed, an important objective of our work is simply to learn whether the data we collect from these respondents reveal something meaningful about their thinking. There is no way to answer this question definitively but we try to shed some light on it in Section 3. There we consider the internal consistency of the answers to our questions, the prevalence of response patterns suggesting that respondents are answering our questions in a perfunctory manner, and the comments made by respondents in debriefing sessions following administration of the survey.

We present a variety of empirical findings in Sections 4 and 5. Section 4 examines respondents' earnings expectations as elicited in the survey. Section 5 presents respondents' beliefs about the current distributions of earnings among persons with different levels of schooling and analyzes the relationship between these beliefs, actual current earnings distributions, and respondents' earnings expectations.

We draw generally encouraging conclusions from this exploratory study. Section 6 offers some of our thinking about what steps should be taken next and where this work may eventually lead.

2. Design and Administration of the Survey

2.1. Survey Medium and Structure

The survey is implemented on personal computers using CAPI software (Palit and Palit, 1986). We have found the CAPI medium to be well-suited to

the task of eliciting probabilistic expectations. CAPI software has enabled us to design a survey that appears straightforward to respondents but that actually incorporates an extensive question-branching algorithm (see Section 2.2). CAPI has also enabled us to incorporate into the survey several tools intended to aid respondents in expressing meaningful expectations. In particular,

training screens explain basic probabilistic ideas through examples before the respondent begins the actual survey.

help screens may be accessed by the respondent at any time.

error checks inform the respondent if a response to a probabilistic question is not a proper probability or if the response is logically inconsistent with earlier responses.

review-and-revise screens show the respondent his or her responses to each completed sequence of questions and allow the respondent to revise these responses if desired.

In practice, our student respondents have appeared comfortable with the CAPI medium and they have seemed to enjoy the process of expressing their expectations. Completion of the survey generally takes 30 – 35 minutes.

Following the introductory and training screens, the CAPI software leads the respondent through the five parts of the survey. These are

1. Unconditional Earnings Expectations: Each respondent is queried about expected earnings (i.e., his or her subjective probability distribution of real earnings) at ages 30 and 40.[1]

2. Expected Earnings Under Alternative Schooling Scenarios: Each respondent is asked about his or her expected earnings at ages 30 and 40 under the hypothetical scenario that the respondent continues in school at least through receipt of a bachelor's degree. The respondent is also asked about expected earnings at ages 30 and 40 under an alternative scenario assuming that less schooling is completed. In the version of the survey administered to high school students, the alternative scenario supposes that the respondent attains a high school diploma but no further schooling. In the version administered to college undergraduates, the alternative supposes that the respondent completes the current semester and then leaves school permanently.[2]

3. Schooling Expectations: High school respondents are asked to state the probabilities that they will (a) attend college before age twenty-one, (b) receive a bachelor's degree before age 30, and (c) still be in school at age 30. College undergraduate respondents are asked to respond to questions (b) and (c).

---

[1] Respondents receive the following on-screen instruction before the first earnings question is posed: "Also, please ignore the effects of price inflation on earnings. That is, assume that one dollar today is worth the same as one dollar when you are 30 years old and when you are 40 years old." Our debriefings of respondents following administration of the survey make us confident that this instruction was understood and adhered to by most students.

[2] We gave some thought to asking college respondents about their earnings expectations under a scenario supposing that they had attained a high school diploma but no further schooling. We decided not to pose this or other counterfactual scenarios to respondents.

4. Beliefs About Current Earnings Distributions: Male (female) respondents are asked about the current distributions of earnings among American men (women) of age 30 who have attained at least a bachelor's degree and among those who have attained a high school diploma but no further schooling.

5. Background Questions: Respondents are asked to report their race/ethnicity, sex, grade point average, the levels of schooling completed by their mother and father, and their family income.

Two aspects of the survey require more extended discussion. Section 2.2 describes the way we elicit subjective earnings distributions. Section 2.3 discusses the specification of hypothetical schooling scenarios.

2.2. Eliciting Subjective Earnings Distributions

We considered eliciting earnings expectations by asking respondents to report quantiles of their subjective distribution of future earnings or, alternatively, probabilities that earnings would exceed various thresholds. We decided to use a hybrid approach. First we elicit the median of a subjective distribution and then the probabilities that earnings would exceed a sequence of thresholds, posed in increasing order. The thresholds about which a given respondent was queried were determined by the respondent's elicited median.

The discussion in this section uses our questions about unconditional earnings expectations to illustrate ideas. We use the same approach to elicit expectations under alternative schooling scenarios and beliefs about current earnings distributions.

ELICITING PROBABILITIES THAT EARNINGS EXCEED THRESHOLDS: It is easy to pose readily understood questions eliciting probabilities that earnings would exceed specified thresholds. Here is the format we use to elicit unconditional earnings expectations:

> When you are x YEARS OLD, what do you think is the percent chance that you will earn more than Y thousand dollars per year?
> That is, what are the chances out of 100 that you will earn more than Y thousand dollars per year?

Each respondent answers this question with x taking the values 30 and 40 and Y taking three values that depend on the respondent's elicited median. Elicitation of probabilities for three thresholds, rather than more or fewer, was our resolution of the tension between our desire to collect as refined data as possible and our concern that too much attempt at refinement would sap respondents' attention.

Effective selection of the threshold values Y is a non-trivial matter. Asking questions about a range of thresholds spanning the support of a respondent's subjective distribution should yield more information about the shape of the distribution than would the same number of questions asked about a narrower or wider range of thresholds. Morgan and Henrion (1990) warn that the specific thresholds chosen may influence respondents' beliefs, a phenomenon called *anchoring* by psychologists. Suppose, for example, that a respondent expects his earnings to be no greater than $30,000. Psychologists fear that, if the first question asked concerns the probability that earnings will be greater than $50,000, the respondent may be influenced to think that this amount is objectively reasonable and so may report a higher probability than believed a priori.

ELICITING MEDIAN EARNINGS: To select threshold values, we ask the respondent to state the median of the subjective distribution. Eliciting the median determines the probabilistic center of a respondent's expectations and provides a natural *self-anchor* for subsequent selection of threshold values in probability elicitation.

At the beginning of the survey, respondents receive this on-screen instruction defining the median of a distribution:

> The first question will ask you about the MEDIAN AMOUNT of money that you think you will earn at some time in the future.
>
> > The MEDIAN is the amount of money for which there is a 50 percent chance that you will earn more than it and a 50 percent chance that you will earn less than it.
>
> So, to answer this question and others like it, you should try to pick the amount of money that you think there is just as good a chance you will earn more than it as less than it.

The question we later use to elicit the median of the distribution of unconditional earnings has this format:

> Look ahead to when you will be x years old. Think about the kinds of jobs that will be available for you and that you will accept. What is the MEDIAN AMOUNT of money that you think you will earn per year by the time you are x YEARS OLD?

Our question-branching algorithm uses the response to this question to select threshold values for elicitation of probabilities. The algorithm typically selects at least one threshold value to the left of the elicited median and one to the right.

It may be asked why, after eliciting the median, we change the structure of our questions to ones eliciting probabilities. Why do we not instead elicit a sequence of quantiles of the subjective distribution? We do think the latter approach is appealing in principle, but we have found it difficult to implement in practice. Thus, we decided to implement the hybrid approach.[3]

CHARACTERIZING THE CENTRAL TENDENCY AND SPREAD OF EXPECTATIONS: Let m_i be the elicited median of respondent i's subjective earnings distribution. Throughout this paper, we use m_i to characterize the central tendency of respondent i's expectations.

Let (Y_{i1}, Y_{i2}, Y_{i3}) be the three threshold values at which probabilities are elicited from respondent i. Let the elicited probability values be denoted (G_{i1}, G_{i2}, G_{i3}). To characterize the spread of respondent i's expectations, we use the interquartile range (henceforth, IQR) of a respondent-specific parametric probability distribution fitted to the elicited probability values. The approach is described here.

Let $F(Y; m_i, r)$ denote the log-normal distribution function with median m_i and IQR r, evaluated at any point Y. Let $G(Y; m_i, r) \equiv 1 - F(Y; m_i, r)$.

---

[3] We attribute our success in eliciting medians to the on-screen definition of the median received by respondents at the beginning of the survey and to reminders placed at later points in the survey. These aids are easy to provide in a CAPI survey but may not be feasible in surveys administered through other media.

We found it difficult to elicit medians in pretests of our telephone survey eliciting the one-year-ahead income expectations of American households (Dominitz and Manski, 1994). In that case, we eventually decided to use an alternative approach suggested by Morgan and Henrion (1990). There we ask respondents for the "lowest possible" and "highest possible" income that they might experience. We do not interpret the answers to these preliminary questions literally as minimum and maximum incomes but rather use them more loosely to indicate the general region of the respondent's subjective support of future income.

For each respondent i, we find r_i that yields the log-normal distribution giving the best fit, in the sense of least squares, to the elicited probability values (G_{i1}, G_{i2}, G_{i3}). That is, we solve the minimum-distance problem[4]

$$\min_{r} \sum_{k=1}^{3} [G_{ik} - G(Y_{ik}; m_i, r)]^2.$$

We use r_i to measure the spread of respondent i's expectations.

2.3. Posing Schooling Scenarios

Economists analyzing schooling decisions assume that youth compare the expected outcomes of completing different levels of schooling and choose the best option. It is particularly common to assume that youth compare the distributions of lifecycle earnings that they associate with different schooling choices. This section describes our questions eliciting earnings expectations under alternative schooling scenarios. We focus here on the version of the survey administered to high school students. The same considerations arise in the version administered to college undergraduates.

Let s denote a level of completed schooling. Let y(s,x) denote the income that a respondent in our survey would earn at age x if he or she were to complete schooling level s. Let Ψ denote the information available to a respondent at the time of the survey. We would ideally like to learn P[y(s,x), all $x \geq \Psi$], the distribution of lifecycle earnings that the

[4] The only exception occurs in a small number of cases (13 of the 660 subjective earnings distributions elicited) in which the responses are $G_{i1} = G_{i2} = G_{i3} = .50$. In these cases, we set $r_i = \infty$.

respondent associates with various schooling options s. Elicitation of entire distributions of lifecycle earnings is impractical so we focus on earnings at ages 30 and 40 and limit attention to two schooling options: s = 0 indicating attainment of a high school diploma but no further schooling and s = 1 indicating attainment of at least a bachelor's degree. Thus, we seek to learn $P[y(s,x) \geq \Psi]$ for x = 30, 40 and s = 0, 1.

To elicit these distributions, we first give respondents this on-screen introduction:

> The next sets of questions ask you to put yourself in one of two hypothetical situations. In the first situation, you assume that you continue in school until you finish your senior year of HIGH SCHOOL and obtain your diploma, and you DO NOT continue in school after that. In the second hypothetical situation, you assume that you continue in school AT LEAST until you finish your senior year of college and obtain your COLLEGE DIPLOMA (a Bachelor's degree).
>
> When responding to these questions, please attempt to fully place yourself in the hypothetical situation as it is described.

After this, respondents receive these more detailed instructions concerning the first scenario (i.e., s = 0):

> In the first hypothetical situation, assume that you continue in school until you complete your senior year of high school and obtain your high school diploma. Please respond under the assumption that you DO NOT RETURN to school at any time after high school. Remember, this is a hypothetical situation. Just think about the kinds of jobs that would be available for you and that you would accept. Think about the amount of money you would make on these jobs. Again, you should ignore the effects of price inflation on earnings.

Following this, respondents are asked to state their median earnings at age 30 under scenario 0 and the probabilities that earnings would exceed three thresholds. These questions have the same format as the ones described in Section 2.2.

EXPECTATIONS UNDER SCHOOLING SCENARIOS AND CONDITIONAL ON SCHOOLING CHOICES: Observe that the wording we use to describe the two schooling scenarios offers respondents no reason why one or the other scenario might be realized. This is intentional – we did not want respondents to draw from the descriptions of the scenarios information that might influence their expectations. In particular, we did not want respondents to state expected earnings conditional on a specified choice of schooling level.

To understand this point, let z indicate the schooling level that a respondent will eventually choose. We say "eventually" because, in general, respondents do not know z at the time of the survey; that is, z is not part of the information Ψ. Suppose that, in describing a scenario, we were to instruct respondents to assume that they choose schooling level s. Then respondent i should report earnings expectations $P[y(s,x) \geq \Psi_i, z = s]$ rather than $P[y(s,x) \geq \Psi_i]$. But our concern was to elicit the latter distribution, not the former one.

The difference between $P[y(s,x) \geq \Psi_i, z = s]$ and $P[y(s,x) \geq \Psi_i]$ will be familiar to anyone aware of the *selection problem* in the empirical analysis of the returns to schooling. Let P^* denote a frequentist probability distribution, as distinct from a subjective one. A standard objective of studies of the returns to schooling is to compare $P^*[y(s,x)]$ across alternative schooling levels s. The problem is that one is only able to observe realizations of y(s,x) among those persons who choose schooling level

s. Thus the data do not reveal $P^*[y(s,x)]$, but do reveal $P^*[y(s,x) \geq z = s]$. See, for example, Manski (1995), Chapter 2.

3. Data Collection

3.1. Description of the Respondents

We administered the survey to 71 high school students and 39 college undergraduates. The high school respondents were drawn from two social studies courses at Memorial High School in Madison, Wisconsin during the spring of 1992. The teachers of the two classes offered students extra credit for completing the survey in the school's computer lab during a study period or after school, and nearly two-thirds of the enrolled students chose to do so. The first column of Table 1 describes the 71 respondents, all but four of whom were in their sophomore year at the time of the survey. The descriptive statistics in the table show that, although the respondents included persons from a variety of family backgrounds, they were predominately the children of college-educated parents with incomes above the U.S. median.

The college respondents were drawn from two economics courses at the University of Wisconsin-Madison during the 1992-93 academic year. The instructors in these courses assigned students the task of completing the survey at a university computer lab, and nearly all complied. The second column of Table 1 describes the 39 respondents. At the time of the survey, 18 percent were freshmen, 13 percent were sophomores, 46 percent were juniors, and 23 percent were first term seniors.[5] The family backgrounds of the

[5] The sample of respondents excludes second-term seniors or graduate students enrolled in the courses. For such students, we could not meaningfully pose a scenario in which they would leave school following the current semester but without completing the B.A.

college respondents were broadly similar to those of the high school respondents.

Table 1: Characteristics of the Respondents

	High School (71 respondents)	College (39 respondents)
Year in School		
Freshman	0.00	0.18
Sophomore	0.94	0.13
Junior	0.06	0.46
Senior	0.00	0.23
Female	0.59	0.28
Race/Ethnicity		
Asian/Pacific Islander	0.01	0.13
Black, Not Hispanic	0.03	0.03
Native American	0.01	0.00
Hispanic	0.01	0.05
White, Not Hispanic	0.93	0.79
Mother's Education		
<12 years	0.01	0.05
High School Diploma	0.20	0.28
Some Postsecondary	0.18	0.18
Bachelor's Degree	0.31	0.26
Graduate Degree	0.29	0.20
Don't Know	0.00	0.03
Father's Education		
<12 years	0.01	0.03
High School Diploma	0.11	0.21
Some Postsecondary	0.15	0.18
Bachelor's Degree	0.21	0.15
Graduate Degree	0.50	0.38
Don't Know	0.01	0.05
Family Income (10^3 dollars)		
[0, 20)	0.02	0.08
[20, 35)	0.06	0.11
[35, 50)	0.21	0.18
[50, 75)	0.27	0.26
[75, 100)	0.20	0.00
[100, 200)	0.15	0.33
Don't Know	0.08	0.05
mean GPA	2.95	3.01

Note: two college respondents did not report their GPA.

3.2. Are the Expectations Responses Meaningful?

Before commencing substantive analysis of the survey data, we attempt to judge whether respondents were able to report their expectations in a meaningful way. First, we describe our qualitative impressions formed from discussions with respondents following administration of the survey. Then we assess the prevalence of response patterns suggesting that respondents are answering in a careless or perfunctory way.

DISCUSSIONS WITH RESPONDENTS: The process of designing the survey instrument was iterative. We pretested various early versions on 23 high school students in the Madison area and debriefed these students for ten to fifteen minutes following their completion of the survey. These debriefing sessions influenced the design of the final form of the survey and helped to form our impressions of how students respond to the questions. We also received less structured feedback from some of the 110 students who responded to the final version of the survey.

Our overall impression of student comprehension of the questionnaire and ability to articulate a coherent set of expectations is very positive. We found that, although many of the students were not previously familiar with the formal concept of the median, they were able to grasp it without difficulty from the information provided in the training and help screens. They appreciated the interactive features of the computer software, and reported that they greatly preferred this interview mode to such alternatives as pencil-and-paper or personal interview. The series of consistency checks, reminders, and error screens generally helped them to focus on and comprehend the tasks required of them. These CAPI tools did not irk or confuse the respondents.

The debriefing sessions did reveal various concerns, some of which we were able to address prior to finalizing the survey. Perhaps the foremost problem uncovered by the pretests is that respondents, when reporting expectations of their own future earnings or beliefs about current earnings, tend to report earnings conditional on full-time employment rather than unconditional on employment.

In early versions of the survey, we elicited expected earnings at age 25 and the probability of attending school at that age. Even when respondents reported a high probability of attending school at age 25, they tended not to report low expected earnings. We learned from the debriefing sessions that respondents were not adjusting their earnings expectations in accordance with their expectations of being a full-time student and, hence, not working full time. In an effort to mitigate this problem, we made two changes to the survey. First, we added text suggesting that the respondent consider the chance of working part-time or not at all. Second, we decided to ask about earnings not at age 25 but instead at age 30, when the probability of being in school would presumably be much lower.

LOGICAL RESPONSE ERRORS: The consistency checks and error screens embedded in the CAPI-administered survey ensure that the medians and probabilities elicited from each respondent must ultimately be logically consistent. How often are respondents' initial responses coherent and, contrariwise, how often must respondents be prompted to change initially inconsistent response patterns? Because our software captures as output the entire sequence of a respondent's responses and revisions, we are able to answer this question.

A respondent may make three types of logical errors, each of which yields an error message requiring the respondent to revise a response. The respondent may:

(1) report a probability that does not fall within the unit interval

(2) report a sequence of probabilities that violates monotonicity of the cumulative distribution function

(3) report a probability that is logically inconsistent with the previously reported median.

No member of the sample ever reported a probability outside the unit interval, but respondents sometimes did make errors of types (2) and (3).

The prevalence of such errors is described in Table 2. Table 2A gives the frequency with which all respondents made errors of type (2) and (3) in each section of the survey. Table 2B gives the corresponding error frequencies for the subset of respondents who made the same error in the preceding section.

Table 2: Frequencies of Logical Response Errors

A. All Respondents

Section of Survey		Violation of Monotonicity	Probability Inconsistent with Median
Expected earnings	age 30	.07	.40
	40	.06	.21
	30, no B.A.	.03	.16
	40, no B.A.	.03	.11
	30, B.A.	.03	.25
	40, B.A.	.06	.25
Current earnings	30, no B.A.	.02	.15
	30, B.A	.03	.16

B. Respondents Making Error in Preceding Section

Section of Survey		Violation of Monotonicity	Probability Inconsistent with Median
Expected earnings	age 30	---	---
	40	.14	.32
	30, no B.A.	.17	.17
	40, no B.A.	.67	.17
	30, B.A.	.33	.50
	40, B.A.	.67	.48
Current earnings	30, no B.A.	.00	.22
	30, B.A	.00	.25

Two patterns are noteworthy in Table 2A. First, errors of type (3) are far more common than are errors of type (2). Second, the error rates in the first section of the survey (i.e., .07 and .40) are much higher than the error rates in later sections (e.g., .03 and .16 in the final section). It is apparent that respondents learn from their logical errors early in the survey, although they still make some such errors even at the end of the survey.

Are errors statistically independent or are some respondents more error-prone than other? This question may be answered by comparing the entries in Table 2B with the corresponding entries in Table 2A. We find that the error

rates in Table 2B are almost always higher, and often substantially so, than those in Table 2A. Thus, it appears that some respondents are considerably more error-prone than others.

BUNCHING OF RESPONSES: Sequences of responses may be logically consistent but perfunctory. There is no definitive way to assess the seriousness with which respondents answer our questions, but we can look for response patterns that may indicate a lack of care. In particular, we can examine the extent of bunching of responses at round numbers.

Table 3 gives the frequency with which different values of median earnings were reported in all eight sections of the survey. We interpret the table as showing remarkably little bunching. At income levels below fifty thousand dollars, we obtain some responses at almost every thousand dollar interval. The spikes that do occur indicate that some respondents round to the nearest multiple of two and others to the nearest multiple of five. At income levels between fifty and one hundred thousand dollars, rounding to the nearest multiple of five is prevalent. At income levels between one hundred and two hundred thousand dollars, respondents round to the nearest multiple of ten. Thus, through the entire range, it appears that respondents generally round their reported medians by less than ten percent.

Table 3: Frequencies of Values for Elicited Medians

Reported value (10 dollars)	Frequency	Reported value (10 dollars)	Frequency
0	38	55	38
1	40	60	93
2	32	65	66
3	76	70	78
4	1	75	132
5	125	80	108
6	104	85	78
7	135	90	114
8	154	95	54
9	120	100	12
10	82	110	20
11	83	120	75
12	115	125	195
13	78	130	
14	2	140	
15	38	150	
16	5	200	
17	8	225	
18	24	250	
19	2	other values	
20	66		
21	4		
22	7		
23	7		
24	8		
25	75		
26	7	Total	
27	7		
28	10		
29	3		
30	83		
31	1		
32	7		
34	3		
35	96		
36	3		
37	3		
38	10		
39	1		
40	64		
41	1		
42	3		
44	1		
45	48		
46	1		
48	2		
50	66		
52	2		

Note: 2640 probability values were elicited (110 respondents x 8 survey sections x 3 probabilities per section). The table excludes 217 cases in which the reported median was the threshold at which a probability was elicited. In these cases, the elicited value was necessarily 50 percent.

We have also analyzed the extent of bunching of the probability responses, which were elicited in "percent chance" form. We obtained some responses at almost every value between 0 and 100 percent, but most respondents round to the nearest multiple of five. Table 4 gives the frequencies with which the various multiples of five were reported throughout the survey, as well as the frequencies with which very low and high probabilities were reported. A fear commonly expressed by researchers skeptical of probability elicitation is that respondents will concentrate their responses on the values 0, 50, and 100 percent. We do not find excessive bunching at these values.

4. Earnings Expectations

Having some confidence that the expectations responses are meaningful, we report empirical findings in Tables 5 and 6. Table 5 presents several quantiles of the empirical distribution of the central tendency of respondents' earnings expectations, as expressed through the subjective median m. Table 6 reports corresponding findings concerning the spread of respondents' expectations, as expressed through the IQR r. Here and elsewhere, the units are thousands of dollars.

Table 5: The Central Tendency of Earnings Expectations

Table 6: Subjective Uncertainty of Future Earnings

A. Median Earnings at Age 30

A. Interquartile Range of Earnings at Age 30

Respondent Group	Unconditional on Schooling Empirical Quantile .10	.50	.90	High School Diploma or no B.A. Empirical Quantile .10	.50	.90	At Least B.A. Empirical Quantile .10	.50	.90
Female HS	25	40	60	8	15	25	30	40	70
Female Col	20	32	60	10	20	30	25	35	60
Male HS	25	40	70	15	20	30	25	45	56
Male Col	25	35	100	15	20	40	27	40	100
Male HS	22	39	69	13	24	43	28	38	58
Male Col	16	39	60	12	23	41	17	37	55

B. Median Earnings at Age 40

B. Interquartile Range of Earnings at Age 40

Respondent Group	Unconditional on Schooling Empirical Quantile .10	.50	.90	High School Diploma or no B.A. Empirical Quantile .10	.50	.90	At Least B.A. Empirical Quantile .10	.50	.90
Female HS	30	45	80	10	20	30	37	50	80
Female Col	35	40	80	20	25	35	35	40	85
Male HS	30	50	95	20	25	40	32	50	75
Male Col	35	48	200	18	26	50	35	50	200
Female HS	18	40	109	10	30	77	18	49	294
Female Col	19	42	71	11	29	78	34	48	66
Male HS	24	43	99	14	28	54	24	44	84
Male Col	21	42	70	14	28	53	21	44	83

C. Earnings at Different Ages (Age 40 vs. Age 30)

Respondent Group	Unconditional on Schooling Empirical Quantile .10	.50	.90	High School Diploma or no B.A. Empirical Quantile .10	.50	.90	At Least B.A. Empirical Quantile .10	.50	.90
Female HS	0	5	20	0	5	10	1	10	20
Female Col	3	10	25	0	7	10	4	10	25
Male HS	2	8	40	0	5	15	5	10	20
Male Col	5	11	100	0	5	15	5	15	100

D. Returns to Schooling (At Least B.A. vs. High School Diploma or No B.A.)

Respondent Group	Age 30 Empirical Quantile .10	.50	.90	Age 40 Empirical Quantile .10	.50	.90
Female HS	14	25	50	18	30	60
Female Col	8	15	35	12	20	55
Male HS	10	20	30	13	22	45
Male Col	8	15	80	10	25	170

Recall that we elicit six subjective earnings distributions from each high school respondent: earnings at two ages (30 and 40), under three schooling scenarios (unconditional on schooling, the scenario with high school diploma, the scenario with at least a B.A. degree). We similarly elicit six distributions from each college respondent, the only difference being that the scenario with high school diploma is replaced by one with termination of schooling after the current semester.

The reader should be aware that each entry in the tables refers to two distinct probability distributions. To illustrate, consider unconditional earnings expectations at age 30. Respondent i has a *subjective earnings distribution* described by $[m_i(30), r_i(30)]$, where $m_i(30)$ is the elicited median and $r_i(30)$ is the IQR fitted from the elicited probabilities. There is an *empirical distribution* of $[m_i(30), r_i(30)]$ across the sample of respondents. Thus, we learn from Table 5A (6A) that, among our female high school respondents, ten percent report their subjective median (IQR of) earnings at age 30 to be no greater than 25 (20) thousand dollars.

Although our sample of respondents is small and somewhat idiosyncratic, we find a number of interpretable empirical patterns. These include the following:

WITHIN AND BETWEEN GROUP VARIATION IN EXPECTATIONS: The tables present findings for each of four respondent groups, defined by gender and current schooling level (high school or college). We find that earnings expectations vary little between these four groups, but substantially within each group.

The close correspondence between the earnings expectations of males and females is intriguing. Consider, for example, the entries in Table 5B for subjective median earnings at age 40 with at least a B.A. degree. Our female and male high school respondents have very similar empirical distributions of m(1,40), but each empirical distribution shows wide within-group variation in

expectations: the .10, 50, and .90 empirical quantiles are (37, 50, 80) for females and (32, 50, 75) for males.

It is also interesting that the earnings expectations of our high-school and college respondents are so similar, despite the fact that the latter respondents have already self-selected into college. This similarity presumably reflects the fact that our high school respondents largely view themselves as college bound. Asked to state their subjective probabilities of college attendance before age twenty-one, ninety percent of the females (males) reported attendance probabilities .80 (.75) or higher. Fifty percent of the females (males) reported attendance probabilities .99 (.95) or higher.

UNCONDITIONAL EARNINGS EXPECTATIONS : Earnings expectations unconditional on schooling are quite similar to those under the scenario assuming that the respondent obtains a B.A. degree by age 30, and quite different from those under the scenario assuming that the respondent obtains less schooling. This presumably reflects the fact that most of our respondents report high subjective probabilities of obtaining a B.A. degree by age 30. Table 7 describes these schooling expectations.

Table 7: Subjective Probabilities of Obtaining a B.A. Degree Before Age 30

Respondent Group	Empirical Quantile .10	.50	.90
Female HS	.80	.95	1.00
Female Col	.85	1.00	1.00
Male HS	.55	.90	1.00
Male Col	.90	1.00	1.00

EARNINGS AT DIFFERENT AGES: Table 5C describes the empirical distribution of the difference between respondents' subjective median earnings at ages 30 and

40. All respondents expect their earnings to increase with age. Moreover, they all expect earnings to increase more should they obtain a B.A. degree than otherwise. Consider, for example, the male high school respondents (the response patterns for other groups are quite similar). The .10, .50, and. 90 empirical quantiles of m(1,40) – m(1,30) are (5, 10, 20), but the corresponding quantiles of m(0,40) – m(0,30) are only (0, 5, 15). Thus, our respondents share the common perception of labor economists that the life-cycle earnings paths of college graduates tend to be more steeply sloped than those of high school graduates.

RETURNS TO SCHOOLING: We are especially interested in student perceptions of the returns to schooling. Table 5D describes the empirical distribution of the difference between respondents' subjective median earnings assuming they obtain a B.A. degree and assuming that they do not do so. Recall that the lesser schooling scenario posed to high school respondents supposes they obtain a high school diploma and that posed to the college respondents supposes that schooling ends after the current semester.

All respondents perceive positive, non-negligible returns to a B.A. degree. Respondents vary considerably, however, in the magnitude of the returns that they perceive. Consider, for example, the female high school respondents. The .10, .50, and .90 empirical quantiles of m(1,30) – m(0,30) are (14, 25, 50), and the corresponding quantiles of m(1,40) – m(0,40) are (18, 30, 60). Thus, the ten percent of respondents who perceive the highest returns to a B.A. degree view those returns to be more than three times as large as do the ten percent who perceive the lowest returns to a B.A. (i.e., 50 vs. 14 and 60 vs. 18).

SUBJECTIVE UNCERTAINTY OF FUTURE EARNINGS: As explained in Section 2.2, we use the IQR r of a respondent's fitted subjective earnings distribution to measure his or her subjective uncertainty about future earnings. Because r is

obtained indirectly and not elicited directly from respondents, it is prudent to view r as only a rough measure of respondents' subjective uncertainty. Nevertheless, Table 6 shows some highly suggestive patterns.

Perhaps the most striking finding is that respondents appear rather uncertain of their future earnings: the interquartile ranges of the fitted subjective earnings distributions are the same order of magnitude as the subjective medians reported by respondents. Consider, for example, the unconditional earnings expectations at age 30 elicited from female college respondents. The .10, .50, and .90 empirical quantiles of m(30) are (29, 32, 60). The corresponding quantiles of r(30) are (20, 38, 83).

A second noteworthy pattern is that respondents tend to be more uncertain about their earnings with a B.A. degree than about their earnings with no B.A. Again consider the female college respondents. The .10, .50, and .90 empirical quantiles of r(1,30) are (20, 34, 75), while the corresponding quantiles of r(0,30) are (16, 24, 45).

A third finding of interest is that respondents tend to be similarly uncertain about earnings at ages 30 and 40. Before administering the survey, we had anticipated that our respondents, typically 15 to 21 years old at the time of the survey, would report more uncertainty about earnings at age 40 than at age 30. That turned out not to be the case.

5. Beliefs About Current Earnings Distributions

In addition to reporting their own earnings expectations, male (female) respondents report their perceptions of the current distribution of earnings among 30-year-old American men (women) who have attained at least a bachelor's degree and among those who have attained a high school diploma but no further schooling. Tables 8A and 8B describe respondents' beliefs about the medians and IQRs of the current earnings distributions. Table 8C presents Current

Population Survey (CPS) estimates of the actual earnings distributions of persons age 25–34 in the year 1992.

Table 8: Beliefs about Current Earnings Distributions

A. Beliefs About the Central Tendency of Current Earnings Distributions: Median Earnings of Persons of the Same Sex, Age 30

Respondent Group	High School Diploma: Empirical Quantile .10	.50	.90	At Least B.A.: Empirical Quantile .10	.50	.90
Female HS	10	15	25	25	35	50
Female Col	8	20	25	20	30	35
Male HS	15	20	30	28	35	45
Male Col	10	20	25	25	35	50

B. Beliefs About the Spread of Current Earnings Distributions: IQR of Earnings of Persons of the Same Sex, Age 30

Respondent Group	High School Diploma: Empirical Quantile .10	.50	.90	At Least B.A.: Empirical Quantile .10	.50	.90
Female HS	13	33	53	23	55	295
Female Col	5	25	37	21	40	134
Male HS	15	24	50	27	47	83
Male Col	5	24	46	22	44	71

C. The Actual Earnings Distribution: Persons Age 25–34

	High School Diploma: Median	IQR	At Least B.A.: Median	IQR
Female	12.3	17.1	24.7	22.1
Male	20.0	17.8	32.0	26.2

Source for Table 8C: U.S. Bureau of the Census (1993), Table 29.

WITHIN AND BETWEEN GROUP VARIATION IN BELIEFS: We found earlier that earnings expectations vary little between our four groups of respondents, but substantially within each group. Much the same pattern holds for respondents' beliefs about current earnings distributions. Consider, for example, the entries in Table 8A for beliefs about the median earnings of persons currently age 30 with at least a B.A. degree. The .10, .50, and .90 empirical quantiles are (25, 35, 50) for female high school respondents, (20, 30, 35) for female college respondents, (28, 35, 45) for male high school respondents, and (25, 35, 50) for male college respondents. Thus, within each group, the ten percent of respondents with the highest beliefs about the current median earnings of college graduates view those earnings to be perhaps twice as large as do the ten percent with the lowest beliefs about current median earnings.

ACCURACY OF BELIEFS ABOUT CURRENT EARNINGS DISTRIBUTIONS: How closely do respondents' beliefs correspond to actual current earnings distributions? Even without looking at the CPS data on actual earnings distributions in Table 8C, we know that the correspondence between beliefs and reality cannot be high. After all, we have seen that there is substantial heterogeneity in respondents' beliefs about current earnings.

If we focus on the empirical median beliefs of the respondents, some interesting findings emerge. The median male respondent has quite accurate perceptions of the current median earnings of males, but the median female respondent overestimates the current median earnings of females. The CPS estimates for the median earnings of male high school and college graduates are 20.0 and 32.0, and the median male respondent perceives these values to be 20 and 35. The CPS estimates for the median earnings of female high school and college graduates are 12.3 and 24.7, but the median female high school (college) respondent perceives these values to be 15 and 35 (20 and 30).

We speculate that the tendency of female, but not male, respondents to overestimate current median earnings reflects a tendency of all respondents to report their beliefs about the earnings of full-time, year-round workers rather than of all persons. More 30-year-old females than males are not full-time, year-round workers. If respondents ignore these persons in reporting their beliefs, female respondents will tend to overestimate the median earnings of females more that male respondents overestimate the median earnings of males.[6]

Comparison of Table 8B and 8C shows that respondents correctly perceive that the spread of earnings among college graduates, as measured by the IQR, is larger than the spread of earnings among high school graduates. But respondents substantially overestimate the degree of current earnings inequality. Consider, for example, the actual IQR of earnings among males with a B.A. degree. The CPS estimate for males aged 25-34 is 26.2. The median male high school (college) respondent believes that the current IQR of earnings among 30-year-old males with a B.A. degree is 47 (44).

EARNINGS EXPECTATIONS AND BELIEFS ABOUT CURRENT EARNINGS: There are at least two reasons why respondents' earnings expectations may legitimately deviate from the beliefs we elicited about current earnings distributions. First, respondents may think that future earnings distributions will differ from current ones. Second, respondents should form expectations conditional on all the information available to them, and that information presumably includes personal attributes in addition to their sex.

It is, nonetheless, of interest to compare respondents' earnings expectations and their beliefs about current earnings. We do so in Tables 9

[6] In fact, the CPS estimates for the median earnings of female high school and college graduates who work full-time, year-round are 17.6 and 28.3. The beliefs of our median female respondents correspond much more closely to these figures than to the CPS estimates unconditional on labor force status.

and 10. These tables present findings only for the high school respondents, not for those in college. Recall that the latter respondents were not asked to report earnings expectations under a scenario assuming no schooling beyond a high school diploma.

Table 9 tabulates the frequencies with which respondents report varying pairs of values for the median of the current earnings distribution and for the subjective median of their own future earnings. It is striking how the responses cluster near the diagonal of each table. Respondents who believe current median earnings to be high (low) tend also to expect their own earnings to be high (low).

Table 9: Subjective Medians of Expected Earnings and Beliefs About the Medians of Current Earnings Distributions

Earnings at Age 30 with a High School Diploma

Females	Subjective Median of Expected Earnings					
Beliefs About Current Median	[0,10)	[10,15)	[15,20)	[20,25)	[25,30)	[30,∞)
[0,10)	3	0	0	0	0	1
[10,15)	2	7	4	1	0	0
[15,20)	1	1	5	2	0	0
[20,25)	0	2	2	6	0	0
[25,30)	0	1	0	0	0	1
[30,∞)	0	0	0	0	1	2

Males	Subjective Median of Expected Earnings					
Beliefs About Current Median	[0,10)	[10,15)	[15,20)	[20,25)	[25,30)	[30,∞)
[0,10)	0	0	0	0	0	0
[10,15)	0	1	1	0	0	0
[15,20)	0	0	5	2	0	0
[20,25)	0	0	2	4	3	0
[25,30)	0	0	0	2	2	2
[30,∞)	0	0	1	0	0	4

Median Earnings at Age 30 with at least a B.A.

Females	Subjective Median of Expected Earnings						
Beliefs About Current Median	[0,25)	[25,30)	[30,35)	[35,40)	[40,50)	[50,60)	[60,∞)
[0,25)	0	0	1	0	0	0	0
[25,30)	1	2	1	2	1	1	0
[30,35)	0	0	0	5	2	1	0
[35,40)	0	0	1	5	1	1	3
[40,50)	0	1	0	0	3	1	1
[50,60)	0	0	0	0	1	2	1
[60,∞)	0	0	0	0	0	0	4

Males	Subjective Median of Expected Earnings						
Beliefs About Current Median	[0,25)	[25,30)	[30,35)	[35,40)	[40,50)	[50,60)	[60,∞)
[0,25)	0	1	0	0	0	0	0
[25,30)	0	1	0	1	0	0	0
[30,35)	0	2	0	0	3	0	0
[35,40)	0	0	0	2	3	2	1
[40,50)	0	0	0	1	5	5	0
[50,60)	0	0	0	0	0	1	1
[60,∞)	0	0	0	0	0	0	0

6. Conclusion

We have learned a great deal from this exploratory study. We are now confident that adolescents from backgrounds similar to those of our respondents, even ones as young as high school sophomores, are willing and able to respond meaningfully to questions eliciting their earnings expectations in probabilistic form. We also think that the particular approach we use to elicit expectations works well.

Although our sample of respondents was small and somewhat idiosyncratic, we have found a number of striking empirical patterns that seem to us likely to hold more generally. These include the following:

(1) There is substantial within-group variation in earnings expectations and in beliefs about current earnings distributions.

(2) Despite the within-group variation in expectations, there is a common belief that the returns to a college education are positive and that earnings rise between ages 30 and 40.

(3) Despite the within-group variation in expectations, there is a common belief that one's own future earnings are rather uncertain.

(4) Despite the within-group variation in beliefs about current earnings distributions, there is a common overestimation of the degree of earnings inequality in American society.

Other empirical findings reported in Sections 4 and 5 are suggestive but perhaps not quite as compelling.

If we are to improve our understanding of adolescent decision making regarding schooling, fertility, employment, and other behaviors, we need to learn what expectations adolescents have about the outcomes associated with alternative actions. Our hope is that, eventually, well-formulated

expectations modules will appear regularly in major national surveys of adolescents.

With this exploratory study completed, several next steps seem warranted. First, there is a need to determine whether adolescents from backgrounds less advantaged than those of our respondents are able to respond to probabilistic questions of the type we pose in as meaningful a manner as our respondents did. Second, larger samples of respondents should be drawn in ways that permit formal statistical analysis of expectations among adolescents in specified populations. Third, respondents should be queried about their expectations regarding important outcomes other than earnings, and under scenarios other than the ones we have posed.

References

Betts, J. (1994), "What do Students Know about Wages? Evidence from a Survey of Undergraduates," Department of Economics, University of California at San Diego.

Berndt, T. and K. Miller (1990), "Expectancies, Values, and Achievement in Junior High School," Journal of Educational Psychology, 82, 319-326.

Blau F. and M. Ferber (1991), "Career Plans and Expectations of Young Women and Men," Journal of Human Resources, 26, 581-607.

Dominitz, J. and C. Manski (1994), "Using Expectations Data to Study Subjective Income Expectations," Social Systems Research Institute Working Paper 9444, University of Wisconsin-Madison.

Freeman, R. (1971), The Market for College-Trained Manpower, Cambridge, Mass.: Harvard University Press.

Juster, T. (1966), "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," Journal of the American Statistical Association, 61, 658-696.

Manski, C. (1990), "The Use of Intentions Data to Predict Behavior: A Best Case Analysis," Journal of the American Statistical Association, 85, 934-940.

Manski, C. (1993), "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" in C. Clotfelter and M. Rothschild, eds., Studies of Supply and Demand in Higher Education, Chicago: University of Chicago Press.

Manski, C. (1995), Identification Problems in the Social Sciences, Cambridge, Mass.: Harvard University Press, forthcoming.

Manski, C. and Wise, D. (1983), College Choice in America, Cambridge, Mass.: Harvard University Press.

Mickelson, R. (1990), "The Attitude-Achievement Paradox Among Black Adolescents," Sociology of Education, 63, 44-61.

Morgan, G. and M. Henrion (1990), Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Risk and Policy Analysis, New York: Cambridge University Press.

Palit, C., and D. Palit (1986), "Computer Assisted Survey Systems: Questionnaire Development Module," University of Wisconsin-Madison.

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

U.S. Bureau of the Census (1993), Money Income of Households, Families, and Persons in the United States: 1992, Current Population Reports, series P60, no. 184. Washington, D.C.: U.S. Government Printing Office.

Willis, R. and Rosen, S. (1979), "Education and Self-Selection," Journal of Political Economy, 87, S7-S36.