Developing a Multi-mode, Longitudinal Study to Understand College Student Outcomes Using Becker's Human Capital Framework

Matthew Soldner
National Center for Education Statistics
matthew.soldner@ed.gov

Jennifer Wine and Natasha Janson
RTI International

Background

Following a cohort of first-time, beginning college students for a period of six years, the National Center for Education Statistics's (NCES) *Beginning Postsecondary Students Longitudinal Study (BPS)* is the source of the national graduation rate and is used extensively by researchers and policymakers to understand the relationship between student and institutional characteristics and their eventual persistence outcomes. Alternating administrations of NCES' *National Postsecondary Student Aid Study* (NPSAS), a cross-sectional study of how students finance education after high school conducted every four years, serves as BPS's base year data collection.

Over the past twenty-four months, NCES has worked with RTI, its primary data collection contractor, and a panel of technical experts to improve the analytic purchase of BPS through a series of revisions to NPSAS. The most notable change is a regrounding of the study's conceptual footing in Becker's [Bec93] Human Capital framework (HC), providing the opportunity to test questions and response formats that are, to our knowledge, uncommon in the federal statistical community. Secondary to that effort is the inclusion of items, informed by the work of Manski [Man04], that attempt to gather students' probabilistic estimates of key constructs, including completion likelihood, and the distribution of future wages. To provide context to our methodological concerns, we briefly introduce background on the relevant theories guiding our research endeavours.

Becker's Human Capital Framework In its most simple form, HC suggests that individuals make the decision to take additional education or training by maximizing a utility function with two components: a) the "costs" of that additional training, and b) the "benefits" that are expected to be obtained by doing so. Both monetary and non-monetary costs and benefits are considered, and might include things such as tuition and fee outlays (monetary costs), the psychic stress of studying (non-monetary costs),

higher future wages (monetary benefits), and the consumption value of college (non-monetary benefits). The rational actor, then, pursues additional education if and only if the anticipated benefits outweigh the anticipated costs [Bec93].

The complication, of course, is that not all student behavior *appears* rational. The field of behavioral economics has evolved at the intersection of psychology and choice theory, in part to help explain why this might be. Any number of explanations are plausible, and include poor information on costs and benefits, differing valuing functions, and unobserved constraints on choice. The challenge for NCES and RTI staff, then, was developing a student interview that not only captured those things that might be thought to influence educational decision-making, but also those that might influence apparently irrational action.

NPSAS Methodology

Sampling Sampling for NPSAS proceeds in two stages. First, a sample of postsecondary institutions participating in the Federal Student Aid program (Title IV of the Higher Education Act, as amended) is drawn from the census of all such institutions, collected via NCES's Integrated Postsecondary Education Data System. After institutions are selected, they provide RTI lists of all enrolled students between July 1 and June 30th of a given academic year. Based on demographic information contained on enrollment lists, students are sampled for participation in the study.

In NPSAS:2008, 1960 institutions were selected for participation in NPSAS, explicitly stratified to be nationally-representative by institutional sector (e.g., 4-year public doctoral-granting institution; 4-year private, nonprofit non-doctoral-granting institution; less-than 2-year private, for-profit institution) and other institutional characteristics of interest. From those institutions, 137800 students were sampled, explicitly stratified by undergraduate/graduate status and other student characteristics of interest. Unweighted institution and student interview response rates were 86% and 72%, respectively.

In NPSAS:2012, full-scale data collection (beginning in Spring, 2012) will sample 1670 Title IV-participating institutions and approximately 120000 students. Because NPSAS:12 will serve as the base year for the BPS study, first-time, beginning (FTB) students will be oversampled, ultimately reaching approximately one-quarter of the full sample. Results reported here are from the NPSAS:2012 *field test*, which fielded in Spring, 2011. The field test (FT) included 150 institutions with 4400 eligible students. Unweighted institution and student interview response rates were 97% and 63%, respectively.

Data Sources The focus of this paper will be on the NPSAS student interview, which can be completed via CATI or a self-administered web questionnaire. In the FT, 21% of interviews were completed via CATI, at an average of 42 minutes. The remaining 79% of respondents completed the NPSAS student interview via the web, at an average of 35 minutes. Both timing estimates are upwardly biased due to a longer interview for FTBs, approximately 48 minutes when completed via CATI and 40 minutes when completed via Web.

Although not discussed here, administrative data from several sources are joined with student interview responses to complete the NPSAS dataset. This includes information from institutional student information systems (e.g., registrar and financial aid databases), Department of Education data systems (e.g., the National Student Loan Data System and the Free Application for Federal Student Aid Central Processing System), and other extant sources. Because most key estimates from NPSAS can be generated using *only* administrative data, sampled students are considered *study members* when a subset of key data elements can be amalgamated from either the student interview or administrative data systems. As a result, the NPSAS study membership rate was 96% in 2008.

Methodological Challenges

Enhancing the NPSAS student interview to collect data on concepts central to the HC framework presented a set of methodological challenges to NCES and its contractor. First, because NCES wanted to collect constructs like "completion likelihood" using a response format that could elicit maximal variation, RTI sought to implement questions with visual analog scales (VAS). Because these scales are optimized for visual presentations, their performance characteristics in a mixed-mode context was unknown. Second, the addition of new items related to the non-monetary benefits and costs of attendance (e.g., satisfaction with peers and faculty or support from family) in a study that otherwise focused on seemingly objective topics (e.g., costs, hours worked, family size) raised the issue of social desirability in response, particularly among CATI respondents. Third, NCES's desire to elicit data on the monetary benefits of postsecondary education meant capturing students' anticipated future wage. Informed by Manski's [Man04] prior research on this point, this suggested not only a point estimate but also a probabilistic distribution of wage. Finally, because all costs and benefits were accrued at different points in time, NCES sought to capture a rough estimate of the function by which respondents conceptualized of the time value of money, known as a *discount rate*.

Challenge One: The Utility of Visual Analog Scales in Mixed-Mode Studies

Visual analog scales (VAS) replace radio buttons for ordinal (or better) responses with the option to select virtually any point along a number line. The granularity of response is controlled by adjusting the step-value of the slider as it moves along the number line. In the example in Figure 1, in which a number line ranges from one to five, a step of 1.0 results in a response set of $\{1, 2, 3, 4, 5\}$, a step of 0.5 results in a response set of $\{1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}$, and so forth.

The use of a VAS scale in a study that can be completed either via self-administered Web or CATI raises an empiric question related to mode effects. As its name implies, a *visual* analog scale is not optimized for aural-only modes of administration: phone respondents can be *told* they can respond via any point on a number line, but the extent to which that knowledge causes them to do so is unclear. Similarly, there are characteristics of the visual presentation of a VAS scale–such as where the slider is positioned–that may affect Web respondents but have no salience to those interviewed via phone. As

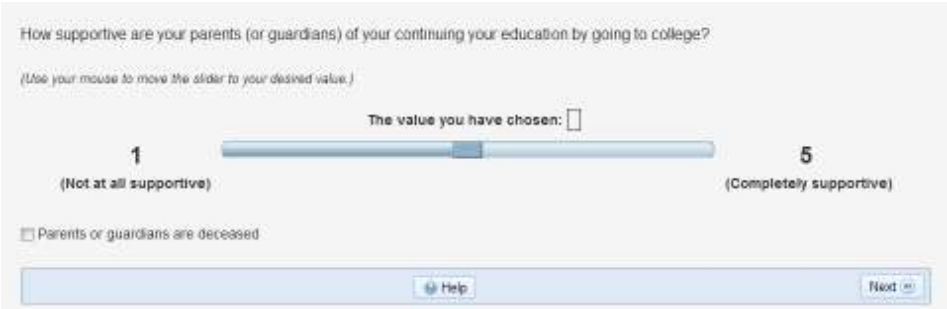

Figure 1: Sample Visual Analog Scale

Table 1: Completion Mode by Slider Starting Position

	Completion Mode		
Starting Position	Web	CATI	Total
Midpoint	510	140	650
Far left	500	140	630
Far right	520	140	660
Total	1540	410	1940

such, NCES explored three questions related to VAS scales. First, did respondents' use of "half-point" increments vary by mode? Second, for Web respondents, did response distribution vary by the starting position of the slider (i.e., far left, midpoint, far right)? Finally, for Web respondents, did item-nonresponse vary by starting slider position on VAS scales?

Although the NPSAS:12 FT used VAS-scaled items in three series, including completion likelihood and discount rates, a third series related to students' social experience on campus–which helps operationalize the consumption value of college, a potential non-monetary benefit–was used to explore these research question. In this set of questions, FTBs were asked to rate their agreement with three statements about their campus social experience on a scale from 1 (strongly disagree) to 5 (strongly agree). A step value of 0.5 resulted in 9 possible response options. All respondents were advised that half-point increments were available for their use when answering.

Respondents were randomly assigned to one of three slider starting position groups–the midpoint, the far left, and the far right–prior to contacting. Table 1 shows the distribution of responding students by mode and starting slider position. Note that details may not sum to marginals due to rounding. Pearson's chi-square suggested independence between starting position and completion mode ($\chi^2(2) = .050$, $p = .976$).

Table 2: Use of Half-steps by Mode

Variable	Web (SE)	CATI (SE)
SOCPEER	27.05	6.20
	(1.187)	(1.203)
SENSBEL	25.24	9.85
	(1.510)	(1.481)
SOCSAT	25.53	7.39
	(1.192)	(1.300)

Use of half-step increments by mode One benefit to VAS scaling over an ordinal scale is that respondents have access to the complete number line between the scale's anchors, conditioned on the step interval. This increased variability provides the opportunity for more nuanced response and turns categorical data in to data which are "quasi-continuous," increasing options for analysis. Despite interviewer instructions, however, NCES and RTI questioned whether CATI respondents would use the whole number line, in particular half-step values, at the same rate as their Web completing peers. For the sake of statistical testing, our *a priori* assumption was that the proportions of respondents who selected a half-step increment by mode would be identical.

Rates of half-point use by mode are detailed in Table 2. Test statistics comparing Web versus CATI respondents' use of half-point values were statistically significant across all questions in the series. For SOCPEER, F(1, 1803) = 152.12, P $>$ F = .000. For SENSBEL, F(1, 1831) = 67.37, P $>$ F = .000. For SOCSAT, F(1, 1842) = 143.75, P $>$ F = .000. From this, we conclude CATI respondents are less prone to use half-point values despite interviewer instructions, constraining potential variability. To the extent that the measurement priority is increasing response options, not specific scale maxima, minima and step values, further research should explore the variability evoked by using an VAS scale with nine response options with a step interval of one.

Response distribution by starting position Web respondents indicate their desired response option to a VAS-scaled item using a "slider" button that is moved along the number-line. Because that slider must have a starting position along the number line, either the far left, the center, or the far right, we explored whether the value of response appeared sensitive to that initial positioning. The mean values of SOCPEER, SENSBEL, and SOCSAT for web respondents are depicted in Table 3.

Statistically significant differences were noted between midpoint and right (F(1, 1400) = 11.47, P $>$ F = .001) and left and right (F(1, 1400) = 9.13, P $>$ F = .003) responses for SOCPEER and for midpoint and right (F(1, 1425) = 15.74, P $>$ F = .001) and left and right (F(1, 1425) = 8.80, P $>$ F = .003) responses for SENSBEL. A statistically significant difference was noted between the midpoint and right means for SOCSAT (F(1, 1436) = 6.89, P $>$ F = .01), but the difference between left and right

Table 3: Mean Response by Slider Position, for Web Respondents

	Slider Position		
Variable	Midpoint (SE)	Left (SE)	Right (SE)
SOCPEER	43.44	43.63	45.21
	(.385)	(.384)	(.353)
SENSBEL	40.11	40.76	42.82
	(.504)	(.521)	(.459)
SOCSAT	40.02	40.77	41.83
	(.503)	(.523)	(.472)

Table 4: Item Non-response Rates by Slider Position, for Web Respondents

	Slider Position		
Variable	Midpoint (SE)	Left (SE)	Right (SE)
SOCPEER	2.14	1.80	21.84
	(.639)	(.596)	(1.810)
SENSBEL	2.92	3.21	14.94
	(.743)	(.789)	(1.562)
SOCSAT	2.72	3.61	12.64
	(.719)	(.836)	(1.456)

means was not statistically significant at $\alpha = .05$. Not displayed here are results for CATI respondents, where no statistically significant differences in response by starter position were observed for any of these items. As a result, we conclude that slider positioning may be particularly influential in Web respondents' response behaviors, and may upwardly bias estimates for a subset of respondents.

Item non-response by slider position As noted above, when Web respondents encounter a VAS-scaled item the slider is in its pre-defined starting position: far left, midpoint, or far right. That position is not a "default," however: until the slider is moved, no response is recorded by the survey system. As a result, it was possible that respondents would encounter a VAS-scaled item with the slider positioned at their desired response and, thinking the question would default to that value, unintentionally fail to respond. In Table 4, we detail item non-response rates for Web respondents for each of our three items.

As can be seen in Table 4, item non-response appears larger when the starting

Table 5: Item Means, by Response Mode

	Response Mode	
Variable	Web (SE)	CATI (SE)
SOCPEER	44.02	45.67
	(.219)	(.398)
SENSBEL	41.17	42.68
	(.289)	(.528)
SOCSAT	40.74	42.98
	(.290)	(.505)

position is set to the far right side of a slider. Statistical testing confirms significant differences in itme-nonresponse between the midpoint and right (F(1, 1534) = 105.32, P $>$ F = .001) and left and right (F(1, 1534) = 110.52, P $>$ F = .001) settings for SOCPEER; between the midpoint and right (F(1, 1534) = 48.33, P $>$ F = .001) and left and right (F(1, 1534) = 44.97, P $>$ F = .001) settings for SENSBEL; and for the midpoint and right (F(1, 1534) = 37.32, P $>$ F = .001) and left and right (F(1, 1534) = 28.98, P $>$ F = .001) settings of SOCSAT. This may suggest that respondents who believe they are responding "5" on items set to a far right scaling are actually recording no response, potentially biasing later estimates.

Generally Taken together, the findings detailed above suggest that VAS scales are susceptible to mode effects. This finding is far from surprising. However, there is no evidence that those effects cannot be "managed" by common-sense form design (e.g., placing the slider in the center position) and judicious use of soft checks (e.g., warning respondents who skip items that they have not registered a mid-point response if they have not moved the slider). Efforts to enhance the variability of response should continue, and may be as simple as rescaling items from 1—5 to 1—10.

Challenge Two: Socially-desirable Response by Mode

The second methodological challenge identified earlier centered on whether new items measuring "non-objective" facets of students' educational experience, such as social support from family and peers, might be affected by social desirability bias. Specifically, we questioned whether CATI respondents would systematically answer more positively than their Web responding peers. To evaluate this question, we compared mean responses to SOCPEER, SENSBEL, and SOCSAT by response mode. The results of that analysis appear in Table 5.

One-tailed t-tests indicated that, for each item, CATI respondents had higher item means than respondents who completed the student interview via Web. For SOCPEER, t(1802) = -3.574, P(T $<$ t) = .000; for SENSBEL, t(1830) = -2.479, P(T $<$ t) = .007);

for SOCSAT, t(1841) = -3.678, P(T $<$ t) = .000).

Generally The data detailed above suggests that response to some NPSAS items may be influenced by students' need to answer in socially-desirable ways. The extent to which this biases resulting estimates depends on the nature of the function that determines choice of response mode. More investigation in to the form of this function, and additional attention to socially-desirable response, appears warranted.

Challenge Three: Probabilistic Estimates of Wage and of Completion

As noted earlier, Manski[Man04] and others have critiqued extant research on college completion for failing to capture respondent *uncertainty* around key estimates. Often, the estimate of interest is future wage. However, they also note their critique extends to researchers' assumption that students' prior probability for graduation is unity.

Future wage

Prior work by Manski[Man04] demonstrated that it was possible, with training, for respondents to define a probability density function associated with a range of expected future wages. Both mode and time constraints precluded our adoption of Manski's complete method, so RTI and NCES set about to identify one or more alternative formats for capturing uncertainty around expected future wage.

In cognitive testing that preceded the FT, RTI tested a three-question series to elicit probabilistic wage expectations.

1. Once you begin working as an OCCUPATION, how much do you think your yearly salary will be? Provide your best guess if you are unsure of the amount.
2. You just told me you expected to make ω as an OCCUPATION. On a scale from 0 to 10, how likely is you will make .75 $\times$ ω or less in that job?
3. You just told me you expected to make ω as an OCCUPATION. On a scale from 0 to 10, how likely is you will make 1.25 $\times$ ω or more in that job?

Interviews revealed that most respondents were unable to respond to the probability follow-ups in a logically consistent manner. Because the questions were unlikely to be successful (particularly in Web mode when an interviewer was not available and extensive cross-form validation would likely be off-putting), RTI fielded a less difficult, albeit less informative, series of questions in the FT.

1. Once you begin working as an OCCUPATION, how much do you think your yearly salary will be? Provide your best guess if you are unsure of the amount.
2. Once you begin working as an OCCUPATION, what is the least amount of money you would reasonably expect to make each year?
3. Once you begin working as an OCCUPATION, what is the most amount of money you would reasonably expect to make each year?

Wage estimates Neither NCES nor RTI have explored the "accuracy" of mean wage estimates, although because future occupations were coded via the SOC, this could be done by comparing estimates to data collected by the Bureau of Labor Statistics. However, we *can* explore mean estimates by a likely covariate: anticipated educational attainment. In Table 6, we summarize a regression of attainment on future wage (winsorized at the 5th and 95th percentiles). The reference category is baccalaureate attainment.

Table 6: Future Wage by Anticipated Attainment

Variable	Coefficient (Std. Err.)
Below Associates	-12981*** (3135.1)
Associates	-10798*** (2375.2)
Masters	9819*** (1704.5)
Doctorate	32927*** (2234.5)
Intercept	53071*** (1172.5)
Unweighted N	1820
R^2	0.1597
$F_{(4,1810)}$	86.16

Completion likelihood

In cognitive testing prior to the FT, NCES and RTI tested two formats for collecting data on completion likelihood. In the one series, respondents were asked about "chances in 10" that they would complete their course of study. In the second, respondents were asked "On a scale from 0 to 10, how likely is it you will finish your degree within five years from today," where 0 was anchored to "no chance at all" and 10 was anchored to "absolutely certain." Because respondents expressed a preference for the latter format, it was selected for field testing.

The only way to evaluate respondents' "accuracy" on their response to the completion likelihood question prior to an actual persistence event is to determine whether responses, in the aggregate, mirror trends seen in the actual completion behavior of students in prior BPS cohorts. Three hypotheses were tested.

Control of institution for bachelor's degree-seekers NCES hypothesized that students at nonprofit institutions would report completion likelihoods greater than their

Table 7: Completion Likelihood of BA Seekers by Institution Control

Control	Likelihood (SD)
Public	94.17
	(14.418)
Nonprofit	94.60
	(11.96)
For-profit	87.02
	(27.68)

Table 8: Completion Likelihood by Degree Program

Control	Likelihood (SE)
Associate's	92.42
	(.628)
Bachelor's	92.91
	(.685)

peers at public and for-profit institutions. Results of our analysis are summarized in Table 6.

Because Levene's test indicated heterogeneity of variance between groups, ($\chi^2(2)$ = 132.02, p = .000), the Kruskal-Wallis equality-of-populations rank test was used to test for differences. Based on that test, ($\chi^2(2)$ = 1.644, p = .440), we failed to reject the null hypothesis that the groups differ.

BA-seekers compared to AA-seekers NCES hypothesized that AA students would report lower completion likelihoods than their BA-seeking peers. Results of our analysis, summarized in Table 8, suggest no significant difference (t = -0.521, P(T $<$ t) = .301).

Educational aspirations NCES hypothesized that, among all degree-seeking undergradutes, greater levels of educational aspirations would be associated with greater completion likelihood. Using BA attainment as the reference category, a set of four dummy variables were regressed on completion likelihood: a) attainment below the AA, b) AA attainment, c) attaining a Master's degree, and d) attaining a doctorate or professional degree. Results are summarized in Table 9. While the model was statistically significant overall (F(4, 1280) = 11.16, P $>$ F = .000), it explained relatively little variance (R^2 = .03) in completion likelihood and our findings varied somewhat from

a priori assumptions. Statistically significant relationships were found between each attainment level and completion likelihood, save attainment at the Associates. However, significantly different aspirations for students who believed they would attain a doctorate degree than for those who anticipated attaining a masters degree were not noted ($F(1, 1284) = 0.67$, $P > F = .412$).

Generally As can be gleaned from Table 9, our respondents' estimates of their likelihood of completing their academic programs was high. Indeed, all groups, students believed the likelihood they would complete their program was 92 percent. Whether this is an accurate reflection of their estimates or an artifact of measurement error cannot be discerned here. Anecdotally, high completion likelihoods are consonant with findings from focus groups conducted prior to fielding the NPSAS field test. Longitudinal collection should shed some light on this phenomenon, when we are able to contrast estimates with observed behaviors.

Table 9: Completion Likelihood by Anticipated Attainment

Variable	Coefficient (Std. Err.)
Below Associates	-7.494** (2.4498)
Associates	-3.170 (1.7165)
Masters	4.050*** (1.1590)
Doctorate	5.394*** (1.6024)
Intercept	90.856*** (.7842)
Unweighted N	1290
R^2	0.0336
$F_{(4,1280)}$	11.16

Challenge Four: Elicitation of Discount Rates

Because the benefits associated with finishing a postsecondary degree are not immediately accrued, equating them to present costs requires some knowledge of how respondents value future rewards. In economics, this function is generally referred to as an individual's *discount rate*. To our knowledge, elicitation of discount rates in a nationally-representative sample survey has never been done; however, examples in experimental settings or relatively small-scale surveys are documented in the litera-

ture. Typically, the instrumentation mirrors that depicted in the figure below, used by Brunello and colleagues [BLWE04].

Figure 2: Field Test Discount Rate Item Example

In an effort to decrease respondent burden and to try new methods of data collection, RTI piloted a two-question method of eliciting discount rate. FTB respondents were randomly assigned to one of two versions of the questions that varied the pay-out period (i.e., one year versus six months, with an appropriately scaled pay-out amount).

Initially, respondents were first asked whether they would defer pay-out of a guaranteed gift (the "gate" question). Respondents who indicated they would defer were then asked to provide the *minimum* amount of money they would be willing to wait for, given the specified pay-out period, using a VAS scale running from $750 to either $1500 or $1125, as appropriate (the "rate" question).

1. Imagine you have a choice between receiving $750 in one month, or ($1500 one year) OR ($1125 six months) from today. This gift is guaranteed to be paid whether you would choose to take the $750 in one month or wait to receive ($1500 one year) OR ($1125 six months) from today. Would you prefer... a) $750 one month from today or b) ($1500 one year) OR ($1125 six months) from today

2. You just said you would wait a year to receive ($1500 OR $1125) rather than take $750 in one month. Starting with $750, whats the least amount of money youd be willing to wait a year for?

Gate question analysis *In theory*, the decision to defer pay-out and double one's money in only a year's time should be a simple one. However, 39 percent of respondents indicated that they would prefer to take $750 today over the future guaranteed payments. This finding was not wholly surprising: when asked about this choice in cognitive interviewing, several respondents reported financial exigency. Table 9 summarizes the percentage of students who chose not to defer pay-out by parental income band.

Table 10: Percentage of Dependent Students not Deferring Pay-Out

Parental Income	No deferral (SE)
Under $30,000	49.25
	(2.502)
Between $30,000 and $59,999	39.66
	(2.431)
Between $60,000 and $89,999	33.58
	(2.874)
Between $90,000 and $119,999	23.56
	(2.888)
$120,000 and above	22.11
	(3.018)

Statistically significant differences were noted between each pair of income bands except two: a) students whose parents made between $30,000 and $59,999 and students whose parents made between $60,000 and $89,999 (t = 1.61, P $>|$t$|$ = .107), and b) students whose parents made between $90,000 and $119,000 and students whose parents made $120,000 or more (t = .28, P $>|$t$|$ = .783). This is generally consistent with the hypothesis that lower-resourced students are more likely to heavily discount future cash flows.

There is evidence, however, that some respondents may have failed to fully understand the nature of the follow-up question. In it, respondents were asked to report the least amount of money they would be willing to wait year (or six months) to receive. RTI had set the scale minimum for both variants to $750, even though a response of $750 is generally illogical given the gate. Approximately 19 percent of respondents chose this seemingly illogical response.

Rate question analysis Concerns about the gate question notwithstanding, NCES and RTI had two questions related to VAS-scaled item designed to help elicit a student's discount rate. First, because discount rate shares conceptual ground with one's ability to delay gratification, was there a relationship between elicited discount rate and the pursuit of an advanced degree, a decision that presumably delays one's ability to maximize earning potential. Second, did respondents exploit the granularity of the VAS scale, set at a step of $25.

We began by exploring the relationship between discount rate and students' highest level of expected attainment. The regression of anticipated attainment on discount rate is displayed below in Table 11. As can be seen there, students in Master's degree programs report discount rates that are lower than those in baccalaureate programs (around 9 percentage points). Although all model coefficients were in the expected direction, the model's explanatory power was low, with an R^2 of only about 1 percent.

Table 11: Discount Rate by Anticipated Attainment

Variable	Coefficient (Std. Err.)
Below Associates	.0549
	(.04950)
Associates	.0332
	(.03750)
Masters	-.0913***
	(.02741)
Doctorate	-.0410
	(.03628)
Intercept	.7658***
	(.01878)
Unweighted N	1770
R^2	0.011
$F_{(4,1760)}$	4.82

Our second question regarded whether respondents' benefitted by having the ability to respond to the discount rate question using a VAS. In Figure 3, the response distributions of the two discount rate variants are plotted, by mode. As can be seen there, respondents seemed to cluster their responses around round numbers on the number line—particularly the end points of the scale—rather than drawing from the robust scale made available to them via the VAS. Given the cognitive complexity of the two-question format and the lack of variability it elicited in the FT, NCES returned its panel of technical experts for additional guidance on the discount rate question.

One panelist identified recent research by Hardisty and colleagues [HTKW11] that contrasted three methods: a) matching, a method similar to our two-question format in which the respondent is asked to equate two amounts across different time periods, b) titration, the method depicted in figure 3, and c) multiple staircase, a form of adaptive titration designed to quickly "narrow in" on the respondent's discount rate.

Resulting Actions

As a result of FT analyses, NCES and RTI took several actions in preparation for fielding NPSAS's full-scale administration.

1. Because a far-right starting position of the slider on VAS-scaled items appeared to upwardly bias estimates and result in greater item-level missingness, all VAS sliders will be centered.

2. To further reduce missingness on VAS-scaled items, presumably caused by respondents believing that a form would default to the slider's starting position if

Figure 3: Field Test Discount Rate Distribution, by Maxima and Mode

they took no action, RTI ensured that each form included a soft-prompt to confirm with the user that they meant to skip an item, rather than register a response.

3. Because of mode effects associated with half-point steps on short VAS scales (i.e., those scaled one to five), traditional radio buttons will be employed.

4. Because the revised wage expectation series appeared effective in the FT, both the point estimate and the wage bound questions were retained for full-scale.

5. Because the completion likelihood question appeared effective in the FT, it was retained for full-scale.

6. Based on concerns with the discount rate question from the FT, and given research by Hardisty and colleagues [HTKW11], NCES opted for the titration method. Because some panelists felt the choice of $750, equal to the cost of community college programs in some states, was sufficiently large as to artifically inflate the proportion of respondents who would not defer for a larger pay-out, NCES elected to use a six-question titration series with lower payout amounts, {250, 300, 350, 400, 450, 400}.

Figure 4: Titration Method Discount Rate Example

References

[Bec93] BECKER, Gary: *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago, IL : University of Chicago Press, 1993

[BLWE04] BRUNELLO, G. ; LUCIFORA, C. ; WINTER-EBMER, R.: The wage expectations of european business and economics students. In: *The Journal of Human Resources* 39 (2004), S. 1116–1142

[HTKW11] HARDISTY, David J. ; THOMPSON, Katherine J. ; KRANTZ, David H. ; WEBER, Elke U.: *How to measure discount rates? An experimental comparison of three methods*. `http://ssrn.com/abstract=1961367`. Version: 2011

[Man04] MANSKI, Charles F.: Measuring expectations. In: *Econometrica* 72 (2004), S. 1329–1376