**Open-ended Questions in the NLS**
**Randy Olsen**

The National Longitudinal Survey dates back to about 1966. It tracks individuals rather than households such as the Survey of Income and Program Participation or family units such as the PSID, although PSID tracks individuals as well. The NLS focuses solely on the individual. This has traditionally been a socioeconomic panel. Recently we collaborated with ANES to ask questions on political ideology and activity. As for open-ended questions, we don't ask a lot of open-ended questions as Jon has said. The most consistent series of questions with verbatims generate industry and occupation codes for all the major jobs held by the respondent - which makes a whole lot of sense when you realize the study is primarily funded by the Bureau of Labor Statistics. That word "labor" kind of gives it all away. We use standard industry and occupation codes.

When I started doing this in 1987 I did a kind of first order check. This was back in the days of paper and pencil. We always asked industry and occupation in each interview. What I did was asked one of our people to take a look at everyone who had the same job in year K as they did in year K-1 and see how we coded up the industry. I got the bad news that the industries for the same company were just jumping all over the place. This was really not good news for me at all. So just the problem of inter-reader non-reliability which others have mentioned, creates a lot of problems. With longitudinal surveys you are focusing on the mobility that transitions of people over the entire life course. These false positives on mobility are bad news. One of the first things I did when I took over the NLS was to switch the entire project to computer assisted interviewing. This is back in the late '80's.  When we were able to make the change to computer assisted interviewing, we completely revamped the way we asked industry and occupation so once we got the verbatims fed into the preloaded files then we would talk to that particular respondent – "you are working for Nationwide Insurance…" – we would reread the old verbatim for both industry and occupation and ask them if it was still correct. If so, we wouldn't ask them again, we wouldn't code it, we would just carry the codes forward. We got rid of a lot of false positives on mobility by making this transition.

Let me say a few things about inter-coder reliability.  I came to be uncertain whether we were having the right people do industry and occupation coding. NORC, which does the survey I am going to be talking about the most - they have done the field work on that. They do an excellent, excellent job. But for some of the I&O coding when we would look at the codes when they came in they carried a "don't know" code. A few times we reworked it. Finally I said let's do a head to head comparison. We took 500 industry and occupation verbatims and sent them to NORC and the US Census Bureau and then I also took those same codes and gave them to a person at the Center for Human Resource Research at Ohio State (CHRR) who had a masters in sociology. She was way overqualified for a coder. I had them all code the very same verbatims. When they were done, I gave them to a person with a Ph.D. If you have more degrees then you are probably better at coding, at least we keep telling ourselves that. I gave it to her to determine who got it right. So, on the CHRR end we put far more effort into coding than what we would reasonably expect to do or what we should do if you have a finite budget constraint. This is what happens – for occupation coding – 401 cases – everyone agreed on 249 of them, 62%. NORC didn't code 58 cases. Of those 36 CHRR and the Census Bureau agreed. So if you take out the cases that NORC didn't feel they could code you get a crude 71% agreement rate. Some of the disagreements are what one might call close calls where people with goodwill could differ on how they are coded, especially for occupation. It is very important to be able to correctly classify the industry sometimes in order to classify the occupation. Some close calls would increase this agreement rate. Among the cases with disagreements – in 53 cases we got it right, in 35 cases the Census Bureau got it right and in 24 cases NORC got it right. Since we are adjudicating I'm not sure you can take the fact that we came out in first place seriously or not. It is the home field advantage. It helps if you are controlling the scoreboard.

Industry codes - here things are a little bit better. Everybody agreed in 73% of the cases. I'm not sure if you would consider that good news or bad news. I think probably most of us would say that it isn't such good news. NORC didn't code 56 cases and of those 40 CHRR & Census agreed so that is

83% crude agreement rate. Again, close calls would increase this a little bit more. Among the cases with disagreements – 33 cases CHRR got it right, 33 cases Census Bureau got it right, 20 cases NORC got it right. One thing to remember is back in 2000 none of this is being computer coded. This was all being done by professional coders. The Census Bureau does a ton of coding. They have little old ladies who sit down in the bowels of the building, sitting there doing nothing all day but industry and occupation coding. God Bless them. Then NORC also has a fairly substantial industry and occupation coding. These are two quality organizations that do a lot of this. But still on the reliability side, this is a little bit problematic. This is for things that where you think that you have a chance of getting it right – industry and occupation – at least is it is a bounded problem. It isn't – when the ANES asks "what did you think about this candidate?" I would hate to be in your shoes.

The verdict from all this is that it helps to have highly educated coders even if they aren't experienced. Now, experience is valuable. The bottom line is that you have to have smart people. I would give them a Wonderlic test to try to get a fix on their IQ. If their IQ isn't above 110 then move on to the next candidate. Even really well trained coders make mistakes. Error rates are high enough that panel studies will generate a lot of errors in industry and occupation coding due to coding error. In the case of ANES where you aren't really looking at transition so much that might be less of a problem but heaven help you if you look at why you like Barack Obama in an early wave and why you like Barack Obama in a late wave – what hits us will hit you.

The result of this is that we moved the coding to the Census Bureau. Their quoted prices were lower; their reliability rate was higher so sometimes you get more than what you pay for.

Other verbatim projects – we are recording interviews to code whether a respondent has a noticeably African American speech pattern to investigate whether such patterns are associated with different outcomes. To do this we asked them about the most positive and negative event in the past year which is an interesting question in and of itself but the idea was just to get them to talk and capture the recording so that one could turn around and code speech patterns.

Another area where verbatims can pop up in an unexpected fashion is in assets. We collect a fair amount of information on asset and liability holdings. Do you remember 10-15 years ago, Jon, when you guys did that project on coding the reliability of asset data - that was a real heart breaker. Assets data are just horrible in terms of noise. One of the things we often see – we ask a respondent what their house is worth and they say they don't know. Some people in this building founded what we call the unfolded brackets method – is it more than this – and you have different entry points and things of that nature. We tried something else that was more like a verbatim where someone said they didn't know what their house was worth – instead of going through a binary tree approach– we asked them to give us a self-reported range of values. It is kind of a verbatim in its mildest form. But what we often saw was something like this – "I have no idea what my house is worth". "Can you give us a range"? "It would be between $245,000 and $255,000". Good to know that you didn't know! We see that time after time after time – that if you give the person a chance to give a verbatim response to questions where you come across a really anal retentive respondent and you ask them any question and they will say "I don't know". If you ask them to give you a guess and just let them speak their mind they are quite precise.

Another area that is worthy of a bit of mention – most of our questions are closed ended and involve in factual information. We have found that even – can we talk? – we have asked women how many children they have ever had and they get it wrong. Now, how many women out there have forgotten instances of childbirth? It happens. I'm sure there is a very interesting story about how this goes on. The point is that even in the most salient events of a person's life, respondents screw it up. Then, if you up the ante – are you living in a marriage-like relationship with someone? What the hell is a marriage-like relationship? I have been married for 40 years. Sometimes I wonder if I am in a marriage-like relationship. Sometimes in order track the event history on people's co-habitation status and marital status and so forth – no, they are separated and then they say have never been married. It really pays to get their confidence. A professional set of interviewers is just going to suck all the information out of the respondent. We have found that real inadequacies in writing a question often

get bailed out with the verbatims that are provided by comments that the respondents volunteer. I am also coming around to the point of view that we probably don't ask enough open-ends.

Here's a question – "Thinking about your life since last year – what would you say was the worst thing that happened to you"? We ask this question primarily to listen to them talk and code the speech pattern. This was done in a really bad way so someone came up with a coding frame ahead of time and the interviewer records the verbatim and then codes the best answer. This is an interviewer coded item. What we usually do with questions of that sort – we always go back and review the answers. We have to change the answers because the interviewers get that coding scheme wrong. Notice that for six people the worst thing that happened to them was the military. Six people said the military was the best thing that happened to them. So my take on this is that if you are going to do open-ended coding it really makes a lot of sense to record the interview at that passage so one can return to the respondent's own words. We are trying to record respondents on laptops as well. The flash result that we are getting back is that the recordings really suck. The problem is that the microphone on the computer is on the front edge of the machine because these are all designed for telephony type applications.  If Jon is the interviewer and Skip is the respondent you can see by the way they are sitting this is not going to work out well.. It works a lot better with phone interviews. When we do phone interviews we use voice over IP and record all of the interviews. Then you get a really clear recording.

Jon, let me say something about this idea of recording and releasing the verbatims. The dirty little secret of longitudinal surveys and probably every other survey is that if you know someone who is in that survey, you can find them. The stuff about confidentiality, let's just give it up. If you know someone who is in a survey it is all over. The ability to find someone, if you don't know what state they are in, my guess is, if you have done things like taking out proper nouns things like Maserati, Gertrude and General Motors – what is left in the recording is not going to identify much of anyone unless you have supplied other information that is sensitive and would interact with the verbatims to re-identify the respondent. At ANES if you want to get the geographic data then this problem becomes really, really severe. But for my money - I am just making up a prescription – I would say don't put the money into coding – put the money into transcribing or redacting a few key words and just put it out. Every investigator is going to look at those comments with their research project in mind.  It is better not to impose ANES staff biases as to what is important rather when they try to guess what is right or not when it comes to coding. ANES staff don't know who is going to use what. Why should everyone pay for something which is essentially a good that only one investigator or two investigators are going to use? I'm not sure that is what ANES ought to be doing.