And the Lord said unto them, "thou shalt validate thy survey instrument", and they validated and it was good.

Survey instrument[1] validation may be the most difficult task for a scientist starting to conduct educational research. A survey that is perceived to be inadequately validated will generally be rejected for publication, but there is little guidance on exactly how to validate a survey instrument. This paper is intended to provide an introduction to the process of survey instrument validation, for people whose expertise is in scientific research and teaching. The examples given are specific to my own interest in promoting ocean literacy (COSEE, 2005) and stewardship through courses in introductory oceanography, but the principles apply to any educational research in the sciences.

Formal definitions of validity include four parts: *face validity, content validity, construct validity* and *criterion-related validity* (CSU, 1993). A survey has face validity if it looks clear and well-organized; this is something a researcher determines before giving the survey to any students or colleagues. A survey has content validity if the questions fall into the area under study; in theory, experts in a given field will agree on what questions belong in that field. A survey with criterion-related validity is directly comparable with other measures of the same student attributes; for example, class grades should correlate with student responses to post-class surveys of content knowledge.

Construct validity, the requirement that the survey actually measures what it is intended to measure, is the most important requirement and the hardest to satisfy. The formal definition is that the theoretical concept matches the measuring device, and educational researchers have made the analogy that one cannot drive nails with a screwdriver. Neither statement really helps understand how one determines what students are actually thinking. A more helpful comment is that construct validation is an iterative process, analogous to building a log cabin and chinking the cracks as the wind and rain blow in (R. Beichner, personal communication).

---

[1] "survey instrument" is just educational jargon for a survey

The difficulty in establishing construct validity for even one survey question is illustrated by some potential responses to the question: "Did dinosaurs ever coexist with humans?" If a scientist were to include such a question in a survey, with potential answers of yes and no, s/he would probably be thinking, "No, Dinosaurs went extinct 65 million years ago, during a mass extinction caused by an asteroid impact. Humans have only been around for about 5 million years." But one student might be thinking "No, 'old' appearing dinosaur bones were buried by God, about 6000 years ago, to test our faith in His revealed world. Dinosaurs never existed." This student would get the correct answer for reasons quite unrelated to science. Another student might think "Yes, 'Jurassic Park' was a documentary, right?" while yet another might think "Yes, current paleontological research classifies birds as living, feathered, dinosaurs." One student got an incorrect answer due to spurious reasoning, and one student actually has a better answer than the scientist's original one! (Thanks to J. Libarkin for this example).

In my efforts to develop a valid survey instrument, by trial and error, I have found some key principles:

- start with qualitative data and work toward quantitative data
- listen to your students
- look at the data several ways and watch for surprises
- listen to experts, but trust yourself

It is best to start with qualitative data, such as open-ended written questions. There is no point in asking students whether they think 'a', 'b', or 'c' if they actually think 'd' and 'e'. Libarkin and Kurziel (2002a,b) advocate using qualitative data to establish the context for the study, analyzing the qualitative data using quantitative methods, and finally developing new surveys that produce quantitative data (Libarkin and Anderson, 2005). For example, the open-ended question "where have you learned about the ocean, before this class" produces qualitative data, but it is easy to count the number of students mentioning different sources of information. When I asked that question, most students mentioned formal education, about 1/4 mentioned conversations with friends and family, and almost none mentioned aquaria or museums (Cudaback, 2006). In later surveys, I listed all categories mentioned by students, and asked them to circle their main sources of information. This question is considered more quantitative, and also produced qualitatively different results -- the number of students choosing informal education increased dramatically. The number of students mentioning friends and family remained constant, but had I not started with an open ended question, I would not have known to include that category. The process of working from qualitative to quantitative surveys is crucial.

Listening to students is the best way to learn what they are thinking. The standard method for this is the "think aloud" interview (e.g. Adams et al, 2006). Students fill out the survey while thinking out loud. These interviews are taped for further study, and can reveal students' understanding of the questions. This process, however, is very time consuming. One shudders to think how many interviews must be conducted to find the very rare student who knows that birds are now considered to be dinosaurs. Fortunately, written surveys can allow students to comment on their thinking. J. Lambert and others actually provide two parts to each question -- the question itself, and a space for students' reasoning. In my own surveys, I have often learned about students' thinking from marginal notes. In all cases, it is important to start with a low-tech survey format that allows comments. Never start with a rigid format such as fill-in-the-bubble forms or online surveys.

It has been said that scientific discovery is announced, not with "eureka!", but with "that's strange ". Similarly, surprising data can greatly aid survey instrument validation. For example, on an attitude survey, one student with otherwise expert attitudes about the relevance of ocean science to the real world agreed with the statement "knowledge in oceanography consists of many disconnected topics". He responded to my query with a thoughtful discussion of the inter-disciplinary nature of the science, and I soon realized that many students were interpreting the question in that light. After discussing the question with my students, I re-worded it to say "topics in oceanography are not related to each other".

I have also been surprised by student responses to content surveys. My survey instrument alludes to the surprising fact that the feces of one dog contain enough bacteria to close down a beach in California. Students are very impressed with this, but most answered the question incorrectly on a post-class survey. It turned out that they did not know the meaning of the word "negligible". Similarly, when most students answered a version of this question correctly on later pre-class surveys, I learned that they were really thinking of multiple dogs instead of individual dogs. The multiple revisions of this question also remind us that it is insufficient to get student feedback on only the first version of a question or survey. The validation process is highly iterative.

Experts in your scientific field or in educational research can help validate your survey instrument, but you also need to trust yourself. I've asked dozens of scientists and educators simply to take my surveys, and most have been moved to comment on some aspect or another. Many significant improvements have come from this. However, a question I consider very important is whether household plumbing is affected by the Coriolis force. Many of my colleagues consider this pervasive myth beneath their notice, but if students think plumbing is affected by the earth's rotation, they really do not understand that the ocean is vastly bigger than their toilets.

For me, this is the most important point. The ocean is vastly larger than we, but we have tremendous power to harm it. My personal mission is to encourage my colleagues to teach introductory oceanography courses in a way that promotes both scientific understanding and informed stewardship of the ocean. The survey instrument I have developed is intended to test the hypothesis that teaching science in the context of stewardship improves students understanding of the nature and relevance of science. I hope that my experiences developing the survey are of use to other scientists embarking on their own educational missions.

Bibliography

Adams, W.K, K.K. Perkins, N. S. Podelefsky, M. Dubson, D. Finkelstein and C. E. Wieman, 2006 New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, *Physical Review Special Topics - Physics Education Research 2,* 010101, http://class.colorado.edu. (CLASS)

Centers for Ocean Sciences Education Excellence (COSEE), 2005: Ocean Literacy: A working definition. http://www.oceanliteracy.org/Ocean%20LiteracyDraft.pdf

CSU, 1993: Colorado State University Tutorial on Validation http://writing.colostate.edu/guides/research/relval/index.cfm

Cudaback, C.N., 2006: What do college students know about the ocean?, Eos, 87, 418.

Lambert, J., 2005: Students' conceptual understandings of science after participating in a high school marine science course. *Journal of Geoscience Education*.

Libarkin, J. C. and J. P. Kurdziel, 2001a: Strategies for Productive Assessment, *Journal of Geoscience Education,* 49(3), 300-304.

Libarkin, J. C. and J. P. Kurdziel, 2001b: Assessing Students' Alternative Conceptions, *Journal of Geoscience Education*, 49(4), 378-383.

Libarkin, J. C. and S. W. Anderson, 2005: Assessment of learning in entry-level geoscience courses: results from the Geoscience Concept Inventory, *Journal of Geoscience Education*, 53(4), 394-401. http://newton.bhsu.edu/eps/gci.html