
CHAPTER 14

Instrument Development, Reliability and Validity

A question sometimes asked about the TAP survey is *How valid and reliable is it?* In other words, how accurate is the information that was obtained? There is no simple answer to this question. In this chapter, we will try to clarify some of the relevant issues, describe the strategies employed to increase the survey's validity and reliability, and speculate about the data's accuracy and limitations.

Validity is usually defined by the question *Are we measuring what we intend to measure?* In other words, how accurate is the measure at assessing a given behavior or belief? Reliability refers to the consistency or reproducibility of a measure. If a measure is not reliable, it will not even agree with itself. For example, if students are administered a measure that has a low reliability on two consecutive days, it is likely that their responses would not be the same. Reliability is a necessary but not sufficient precondition for validity.

One way to increase the reliability and validity of a measure is to use a well-established measure that has demonstrated reliability and validity. Whenever possible, this was done in the TAP survey. Many of the items in the survey are from established measures that have demonstrated fairly high reliability and validity. Most of the drug and alcohol questions come from widely used national survey instruments, as do the questions dealing with suicide.

It should also be noted that most of the items developed specifically for this survey have been examined for their reliability and validity. Those items that did not measure up to this scrutiny in previous surveys were either dropped or redesigned for the present survey.

With any self-report survey, especially one aimed at teenagers, there is always the possibility that a small percentage of those surveyed will not take the survey seriously. Such surveys usually account for less than one to two percent of the total. Fortunately, most teenagers who do not take the survey seriously are not subtle about their responses. They typically exaggerate their responses to such an extent that their surveys are easy to spot and remove. In these data five consistency/exaggerator checks were run that eliminated 29 forms or 4% of the surveys. Examples of consistency/exaggerator checks include: students who reported that they used all drugs listed daily; students who reported time use activities none or all of the time; or students who reported on one question that they did not use alcohol and then reported on another that they did engage in binge drinking. This does not guarantee reliability, but it does eliminate detectably unreliable data.

All self-report surveys are susceptible to some bias in responding. For instance, there might be a slight under-reporting of socially undesirable behaviors (e.g., illicit drug use) and a slight over-reporting of behaviors that individuals perceive as socially desirable (e.g., sexual activity for early adolescent males) (Harrel, 1985). Studies of this phenomenon indicate that such under-reporting is usually small, ranging anywhere from two to ten percent, depending on the conditions under which the survey is administered and the questions asked.

In addition, due to problems remembering details, there is a tendency for individuals to be somewhat less accurate as the level of question specificity increases. For example, most students can accurately remember if they have ever smoked marijuana. However, their accuracy may decrease as the question becomes more specific, such as how frequently one has smoked it in the past year. This is primarily a result of problems in remembering details, rather than a lack of honesty.

Reliability and validity are usually discussed as if they are inherent in the survey instrument itself. However, much of what determines how accurate the data are depend on the conditions under which the survey is administered. Students are more likely to answer survey questions honestly if they feel that their responses are confidential and anonymous, if the atmosphere in which the survey is administered is serious and supportive, and if those who administer the survey are knowledgeable about the survey, believe the survey is important, and convey its importance to students (Nurco, 1985). Many steps were taken to ensure that local youth were comfortable that their responses were confidential and anonymous.

Another question often asked about surveys of this type is how representative are the findings for students in general. One factor to keep in mind is that the survey only represents the responses of students who were in attendance on the day the survey was administered. Studies have shown that students who are more frequently absent or truant are also more likely to use illicit drugs, drink alcohol, smoke and engage in potentially problematic and dangerous activities (Johnston & O'Malley, 1985). As a result, the current findings are likely to be a slight underestimate of the actual incidence of such problem behaviors in all youth who are currently enrolled in school. For drug use, Johnston and O'Malley (1985) found that these behaviors were underestimated from 1.4 to 2.7 percent.

It should also be noted that the numbers presented in this report reflect only adolescents enrolled in school, not those who have dropped out. There is some evidence to indicate that school dropouts are somewhat more likely than those enrolled in school to be users of illicit drugs and alcohol and to engage in other problematic behaviors (Beauvais & Chavez, 1996). Consequently, the numbers presented in this report probably underestimate the actual incidence of alcohol and other drug use for all local teens.

For a practical survey such as the present one, the issues of reliability and validity are only a means to an end. The real question is *How is the measure and the data it produces going to be used?* If the objective is the diagnosis of a particular individual, then the precision of the instrument is extremely important and imprecision can be a problem. In contrast, if the objective is to determine the prevalence of a particular behavior or behaviors for a given population (our current interest), then greater imprecision is usually tolerable. For instance, it will probably not matter much to the local community whether 25% or 30% of students are currently using cocaine. We can assume that a 5% under-or-over-estimate will make little difference and that such a high incidence of cocaine use will be viewed as a major problem. Finally, we are fairly certain that, in general, most youth responded in a way that closely paralleled their own situations, attitudes, behaviors and beliefs.