Appendix A. - Instrument Development, Reliability and Validity

A question sometimes asked about the SWYS 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, and speculate about the data's accuracy and limitations.

Validity is usually defined by the question, "Are we measuring what we intended to measure" or " 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 survey. Many of the measures in the survey are established measures that have demonstrated fairly high reliability and validity. For instance, the depression measure that was used is the short form of the Beck Depression Inventory one of the most widely used measures of depression.[1] 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 measures developed specifically for this survey have been examined for their reliability and validity. Those survey items that did not measure up to this scrutiny were either dropped or redesigned for this present survey.

Through reliability and validity tests of self-reported surveys, it has been found that teenagers are more likely to lie than adults. Of particular concern is self-reporting of age of first drug initiation. It had been found that on average teenagers will vary within two years of reporting when they first tried a particular drug. Males are also more likely to vary their responses. The low reliability of self-reported drug use has many implications on future research. The most important being that many age "cut-offs" for analysis purposes may not be as accurate as once thought.[2] Thus, we have limited our questions on initiation of substance use to alcohol and tobacco.

The reasons behind why teenagers lie are becoming clearer. Many researchers now report that teenagers lie to be more socially accepted. To them, what they perceive as the norm for level of involvement in a particular activity should be, is what they report on the surveys.[3,4] Typically, teenagers will inflate

[1] Beck, A.T., & Beck, R.W. (1972) Screening depressed patients in family practice-A rapid technique. *Postgraduate Medicine*, 52, 81-85

[2] Johnson, T.P., & Mott, J.A. 2001. The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction, 96*, 1187-1198.

[3] Champion, D.J. (2001). Measuring delinquency. In *The Juvenile Justice System: Delinquency, processing and the law. (3rd ed).* (pp. 40-88) Upper Saddle River, New Jersey: Prentice Hall.

[4] Rosenblatt, J.A., & Furlong, M.J. 1997. Assessing the reliability and validity of student self-reports of campus violence. *Journal of Youth and Adolescence, 26,* 187-202.

answers by 10% if given numbers of times engaged in. For this reason, we have eliminated questions that dealt with how many times a week a student has sexual intercourse.

While it is still being debated as to how to get students to report honestly, some researchers suggest impressing how important it is to tell the truth, which was emphasized through the training for survey administrators. Other suggestions include: having reliability checks within the surveys, controlling for social desirability as much as possible, and stressing that results will be anonymous.[3]

In order to detect a more sophisticated source of error, we compared response patterns across related questions. Three such scales were developed to check the reliability of responses to questions regarding alcohol, tobacco and sexual intercourse. In all three cases, responses to the target behaviors were compared across questions. Where possible, we compared questions that were spaced across several pages of the survey in order to best identify consistency in reporting. Overall, our analysis indicated a high level of reliability across the questions for these behaviors**.**

On the topic of sexual intercourse, 2,667 participants consistently stated they had never had sexual intercourse across all questions asked. Another 900 participants consistently answered that they had some level of sexual experience. This produced an inter-question reliability of .96. For tobacco related questions, 2,750 participants stated they had never smoked tobacco and another 431 participants consistently admitted to some tobacco use across these questions. This gives us a rating of above 0.85. Finally, 1,794 participants answered consistently they had never used alcohol while 1,520 participants were consistent in acknowledging their use. This gives us an inter-related reliability score of .89. It should be noted that the alcohol comparisons contained the largest range between questions, which should indicate the most accurate score among the three measures. Taken collectively, these measures suggest the teens were motivated to remain consistent in their responses throughout the survey.

Despite all prudent efforts, as with any self-report survey aimed at teenagers, there is always the possibility that a small percentage of those surveyed will not take the survey seriously. Fortunately, most teenagers who do not take the survey seriously are not subtle with their responses. They typically exaggerate their responses to such an extent that their surveys are easy to spot and remove.

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.[5] As a result, the current findings are likely to be a slight underestimate of the actual incidence of such problem behavior in all youth who are currently enrolled in school. For drug use, Johnson and O'Malley found that these behaviors were underestimated from 1.4% to 2.7%.[5]

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

---

[5] Johnson, L. & O'Malley, P. (1985). In B. Rouse, N. Kozel & L. Richards (Eds.). *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity.* Rockville, MD: National Institute on Drug Abuse.

in other problematic behaviors.[6] Consequently, the numbers presented in this report probably underestimate the actual incidences of alcohol and other drug use for all teens in Southwest Wisconsin.

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 to school officials whether 25% or 30% of students are currently engaged in 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 would be viewed as a major problem.

Use of Scales in Reporting Data

A series of scales were developed in order to ascertain general patterns for questions that are believed to measure a similar trait. Aggregating scores in this manner helps "smooth" out some of the noise in the system that can occur when individuals respond to rather specific questions. For this study, we developed nine such scales designed to measure dimensions such as parental rules, communication with their teen, and monitoring; the teen's satisfaction with their school and their involvement in the community; a general measure of the teen's values and beliefs; and a measure of the teen's overall self-esteem. Finally, a scale was developed to mimic the 40 Assets related to teen resiliency. A detailed explanation of this scale can be found in the section titled **Indicators of Positive Youth Development**. The range for each scale was then categorized into mutually exclusive and exhaustive quartiles for comparison. In each case, the higher categories correspond to an increase in the scale's dimension.

Four separate scales were developed to measure the teen's descriptions of their relationships with their parents. For all scales, teenagers who responded that a parent(s) was (were) absent from their home were excluded from that data set. The first scale was Parental Monitoring. This scale summed the responses from Questions 131 through 137. Higher scores corresponded to teen's responding their parents had greater awareness of the teen's on-going behaviors. There were two communication scales developed. An overall measure of the teen's perception of their communication with their mothers was determined by summing Questions 146 through 149. A scale was also created in a similar fashion to measure the teen's perception of their communication level with the father (Questions 150 through 153).

The final parental based scale measured teen's perception of their parent's communication and consequences for violating family rules. This scale summarized the responses to Questions 138 and 139 and higher scores corresponded to greater consistency in communicating and enforcing consequences for violating family rules.

Two scales were developed to measure the teen's perception of community involvement and satisfaction with the community and school. The first scale measured the teen's overall satisfaction with their school and was by subtracting Question 97 from the sum of Questions 92 through 96 and Question 98. The totals were then *reverse-scored*, in order to keep the scale consistent with the other scales. Higher

[6] Clayton, R., & Voss, H. (1982). National Institute on Drug Abuse. *Technical Review on Drug Abuse and Dropouts.* Washington, DC: US Government Printing Office.

scores correspond to greater satisfaction with the school. The second scale measured the teen's satisfaction with their larger community. Due to the nature of the questions and in order to keep the scale consistent with the previous scales, this measure actually measures the level of *dissatisfaction* with the community. This scale was derived from the sum of Questions 101, 102, 104, and 108 through 110. Higher scores reflect negative attitudes.

The next scale was a general measure of the teen's self-esteem. It was computed by summing Questions 7 through 9 and subtracting Question 10. Again, this scale was reverse-scored to maintain the consistency with the previous scales. Higher scores correspond to more positive reports of self-worth and positive attitudes about themselves.

The final scale measured the teen's belief system and values. It incorporated teen's attitudes towards alcohol and sexual behaviors as well as the importance of honesty, responsibility and treating others with respect and empathy. This scale was created by summing teen's responses (agree or disagree) to Questions 12 through 15, 17, 18, 111 and 112. The higher the sum score corresponded with greater agreement with these societal values.

Generally the scales were used to determine if the issue they measured made a difference in a particular behavior or attitude. For example, it was determined that students whose parents score high on the parental monitoring scale, were less likely to engage in smoking. This type of information can be of value to parents open to learning more parenting skills and tips.

Not every scale is used in this report. Occasionally quartiles have been collapsed for easier presentation of the data or because the number of students falling in a quartile was deemed too small to give meaningful data.