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 reliable, it will 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 on the first day would be different than their answers on the second day. Reliability is a necessary, but not sufficient precondition for validity. Validity requires looking not just at the content of the survey but also includes, how the survey is conducted and how it relates to other surveys.

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. It is one of the most widely used measures of depression. 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.

Inaccurate reporting by teens on self-report surveys such as SWYS can arise through a number of mechanisms. The amount of social stigma that teens perceive with a given measure and social desirability effects are thought to play a particularly important role in the accuracy of substance use reported by adolescents. For example, a greater willingness of adolescents to report drug use in school-based settings than at home suggests that perceived confidentiality of responses and the acceptance of peers influence adolescents' willingness to report substance use truthfully. Adolescents also have been found to revise their pasts as their current behavior changes. Studies have also found that violence including victimizing other

students can be accurately measured with self-report questionnaires but has an increased variance possibly due to social stigma.\(^5\)

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.\(^6\)

In order to detect a more sophisticated source of error, we compared response patterns across related questions. Two such scales were developed to check the reliability of responses to questions regarding alcohol and sexual intercourse. In both 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, 5,528 participants consistently stated their level of sexual experience across all questions asked. This produced an inter-question reliability of 96 percent. In addition, 5,056 participants answered consistently about their level of alcohol use. This gives us an inter-related reliability score of 88 percent. 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.\(^7\) 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.

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. 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 will the measure and the data it produces be used?” If the objective is to predict or evaluate the behavior 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.

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 increase the reliability of that measure. For this study, we developed eight such scales designed to measure dimensions such as family rules, parental communication, and monitoring; the teen’s satisfaction with their school, how they use their time, and their attachment to the community; a general measure of the teen’s risk factors and protective factors; and a measure of the teen’s overall self-esteem. The range for each scale was then categorized into mutually exclusive and exhaustive quartiles for comparison.

Three separate scales were developed to measure the teen’s descriptions of their relationships with their parents. The first scale was Parental Monitoring. This scale summed the responses from six questions. Higher scores corresponded to teens responding their parents had greater awareness of the teen’s on-going behaviors. The second scale was parental communication. This scale was an overall measure of the teen’s perception of their communication with their parents. It included four questions.

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 two questions 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 affected by the answers to seven questions. Higher scores correspond to greater satisfaction with the school. The second scale measured the teen’s attachment with their larger community. This scale was derived from the sum of six questions. Higher scores reflect strong connection to the community.

A scale was created to measure how teens use their time. It aggregated the responses from six questions and higher scores indicated more constructive use of time. The next scale was a general measure of the teen’s self-esteem. It was computed by combining four questions. Higher scores correspond to more positive reports of self-worth and positive attitudes about themselves.

---

The final scale measured the teen’s experience with factors that have been shown to put them at risk and factors that have been shown to protect them from risky behaviors. The results related to this scale can be found in Chapter 13: Indicators of Positive Youth Development. This scale was created by aggregating the results of 31 questions. A high sum score corresponded with increased number of protective factors.

Generally the scales were used to determine if the issue they measured made a difference in a particular behavior or attitude. At times responses had to be reverse scored to be sure that higher numbers represented a response that would indicate a higher level of the characteristic being measured.

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.