
How Do You Know If the Risk Assessment Instrument Works?

*Kelly Dedel Johnson
and
Patricia L. Hardyman*

By design, risk assessment instruments are easy to use: rate an offender on a small number of items, add up the score, check the scale, and identify whether the offender is low-, medium-, or high-risk. Easy! Fill out the form on a large number of offenders, and you can make impressive-looking charts as to the number of offenders in each risk group and do complex workload computations to estimate the number of probation/parole officers needed to manage the offender population.

But . . . how do you know if the instrument actually works? That is, how do you know if the offender who scored as high-risk really requires the intensive and costly interventions you've designed to mitigate the threat he or she poses to community safety? Conversely, how do you know if the offender who scored as low-risk—and who is therefore receiving only minimal levels of supervision—really poses very little threat to public safety? What if the instrument is wrong, and the offender who scored as low-risk will actually continue to victimize the community, given the low levels of service and supervision he is receiving? How can you be sure the instrument actually works?

What Does Validation Mean, and Why Is It Necessary?

The decision to use a risk assessment instrument isn't as easy as just pulling one off the shelf. It must be validated so that you are confident that the risk classification suggested by the instrument for a particular offender is an accurate representation of his or her risk to public safety. "Validation" is a research term that represents several common-sense features of an instrument. If these features are present, the instrument is said to be "valid."

- ◆ First, a valid instrument identifies discrete groups of offenders who pose different levels of risks to public safety. The level of risk is reflected in group recidivism rates, which refer to the portion of the group that experiences at least one failure (e.g., re-arrest, re-conviction) during a specified follow-up period. The group of low-risk offenders should have a statistically significant lower rate of recidivism than the group of high-risk offenders (i.e., approximately 30 percentage point difference between failure rates of the low- and high-risk groups).
- ◆ Second, an instrument cannot be considered valid if it is not reliable. There are two types of reliability: inter-rater and intra-rater reliability. Inter-rater reliability means that two different staff members applying the instrument to the same offender will reach the same score and risk classification. Intra-

rater reliability refers to whether the same rater will obtain the same score and risk classification with repeated assessments of the same offender, given no changes in the circumstances of the offender. If the instrument is not reliable, the risk classification assigned to the offender (and, consequently, the level of supervision he or she receives) varies depending on who completed the form or when it was completed.

Reliability needs to be established before an instrument’s validity is tested to ensure the integrity of the risk classifications. Reliability is often difficult to achieve, particularly when risk instruments include items that are somewhat subjective, that are poorly defined, or that require information that is difficult to access.

- ◆ Third, the instrument must be fair to all offender subpopulations. In particular, the instrument should assess women and racial minorities equitably, ensuring that they are subjected to supervision that is commensurate with their actual levels of risk to public safety.
- ◆ Finally, the instrument should be practical, efficient, and simple to implement. Instruments that are too complicated or too time-consuming to complete often suffer from inadequate reliability, and thus have little utility to staff.

Validating a risk instrument and ensuring its reliability, equity, and utility are serious undertakings, and they are essential for sound supervision practices. Differences in offender characteristics, laws, agency policies, and local supervision conditions mean that an instrument that creates accurate risk classifications in one jurisdiction won’t necessarily work well in another. A number of contextual factors can also suggest that a validation effort is needed, including changes in the characteristics of the offender population (such as average age or length of sentence), new sentencing legislation, or budget cuts and reductions in programs and services. Further, experience has shown that risk instruments do not always work equally well for different offender subpopulations (such as women, sex offenders, or mentally ill offenders).

Finally, validating an instrument is a key strategy to improve staff buy-in, instill public confidence in the effectiveness of community corrections, and defend the agency’s decision-making process in the event of an unfortunate, high-profile crime involving an offender on supervision.

How Do You Get Started?

Agencies need to make a number of decisions prior to undertaking a validation study. First, “recidivism” must be defined for use as an outcome measure. Risk assessment instruments group offenders according to their risk to public safety. This risk can take several forms: risk of a subsequent arrest, risk of a subsequent conviction, or risk of a return to jail or prison. Choosing among these various

measures of recidivism is an essential first step, as it will determine what data need to be collected.

There are compelling reasons for selecting either re-arrest, re-conviction, or re-incarceration as a measure of recidivism, but there are disadvantages to each choice as well. Variations in case processing time, completeness and availability of data, prosecutorial and plea bargaining practices, and the reliance on revocation in response to offender non-compliance should be considered for their impact on each potential measure of recidivism. Generally speaking, re-arrest and re-conviction both provide relatively accurate measures of harm to the community, while re-incarceration rates can be affected by non-compliance with technical conditions of supervision rather than new criminal behavior.

It is important to recognize that the selection of an outcome measure often limits the utility of the instrument for other purposes. For example, validating an instrument using re-arrest as the outcome measure means that the instrument may not be able to classify offenders according to their likelihood of re-conviction for a violent offense. It is therefore essential that the choice among recidivism measures be made thoughtfully.

Testing the instrument for its ability to classify according to risk of violence has appeal, but this is difficult to accomplish. This difficulty stems in large part from the fact that acts of violence are relatively rare. These low base rates mean that it is difficult to create an instrument that produces a group with a high rate of violent re-offending and to produce groups whose rates of violent re-offense are radically different from each other. As a result, most risk instruments are validated using broad categories, such as any re-arrest, or any felony re-arrest.

Another part of deciding on an outcome measure involves specifying the follow-up period that will be used, i.e., the risk that an offender will be re-arrested, re-convicted, or re-incarcerated during what length of time following the assessment. Most validation studies use a follow-up period of at least 12 months, and some extend as far as 36 months.

Tolerance for error also is an important consideration. Risk assessments are vulnerable to two main types of error: false positives (the instrument suggests that an offender will recidivate, but he or she does not) and false negatives (the instrument suggests the offender will not recidivate, but he or she does). Most jurisdictions try to minimize the false negatives produced by an instrument by adjusting the “cut” points of the scale to group more offenders into higher-risk categories. This subjects more offenders to higher levels of supervision and services. Ultimately, a jurisdiction’s tolerance for error, and the resulting adjustments to scale cut-points, are inextricably linked to the conservation or expenditure of supervision resources.

A final consideration is the agency’s readiness and commitment to modify its risk assessment process. The tasks of validation and implementing changes to the

current process are time-consuming and require substantial agency resources. If an agency has sufficient resources and staff to complete the validation, but lacks the commitment to change policies and procedures, the initiative should not proceed because it would waste limited resources and negatively impact staff morale and future willingness to participate in validation efforts.

What Are the Key Steps of a Validation Study?

The validation process includes four basic steps: reviewing the current risk assessment system and setting goals, conducting a detailed analysis of the risk assessment, developing an implementation plan for the new or revised system, and documenting the validation effort.

Step 1—Review the current system. A review of the existing system must examine the current risk assessment policies, practices, and issues; the positive and negative trends associated with these policies and practices; and the goals of the validation study. Specifying the goals of the validation is the most critical and the most difficult task. It requires the agency to define the specific problem(s) that will be addressed, to set realistic goals, and to define measurable objectives for the validation.

During the early stages of the study, the following information should be compiled and reviewed:

- ◆ Written risk assessment policies and procedures;
- ◆ Agency annual reports;
- ◆ Current risk assessment instruments;
- ◆ Current automated management reports, including relevant statistics regarding the risk assessment system;
- ◆ Agency staffing and budget; and
- ◆ Any recently enacted or pending legislation or administrative policies that may impact risk assessment or supervision practices.

In addition to reviewing the formal documentation of the system, the mechanics of the risk assessment process should also be considered by interviewing central office and line staff to understand their perceptions of the current risk assessment system and the issues that are of concern. It may not be possible to interview all or even a majority of the staff, so it will be necessary to carefully sample staff or conduct focus group meetings to ensure that a broad range of perspectives is captured. The interviews should be augmented by reviewing a small random sample of recently completed risk assessment instruments and the case files or information system screens used to score the instruments.

Step 2. Analyze the risk assessment system. Regardless of whether the agency has decided to develop a new system, to modify the current instruments, or simply to validate the current system, the following steps are required.

- ◆ **Draw representative samples of key offender populations.** The specific sampling procedures will need to be tailored to the agency's information system capabilities, supervision populations, and goals of the validation effort. Separate random samples of males and females will be needed, that consider their respective average daily population, number of admissions per year, and average length of supervision, to allow for separate analyses of these populations. In addition, stratification or over-sampling of special populations may be necessary. If there are concerns about the validity of the instruments for offenders with mental health problems, for example, this population may need to be over-sampled to ensure an adequate number of cases for statistical analysis. The size of the samples required for the statistical analyses should be adjusted according to the average daily population of the jurisdiction. At a minimum, 300 initial and 300 reassessment instruments should be completed for both male and female offenders, for a total of 1,200 cases.

- ◆ **Compile the data.** Depending on the sophistication, reliability, and accuracy of the data stored in the agency's automated information system, the information system or research staff will need to generate electronic data files regarding the sampled offenders' criminal history, demographics, and history of technical violations. A detailed request identifying the specific data to be included in the electronic files is critical to avoid misunderstandings and spurious conclusions. To ensure the accuracy of the data, an independent reviewer should audit any manually collected data to clarify inconsistencies and eliminate missing data.

- ◆ **Analyze the data.** Risk factors, scale cut points, and override factors need to be assessed to determine if they are valid and reliable for identifying offenders who pose a threat to public safety. It is important to conduct separate analyses by gender to ensure that the system is appropriate for both the male and female populations. Additional analyses can indicate whether separate instruments, scales, or risk factors are required for special populations.

At a minimum, the process should include the following statistical analyses:

- Examining demographic and offense characteristics of the samples and the agency's offender populations. These data describe the samples and the total offender population to ensure that the samples are representative.

- Reviewing frequency distributions, mean number of arrests/convictions/technical violations, and percent of offenders who successfully completed the follow-up period for the initial and reassessment risk

factors, separated by gender. These analyses provide insight as to the number of offenders falling within the respective categories of the risk factors. For example, the mean number of arrests per risk factor helps to identify risk categories that create distinct groups of offenders and to identify where further refinement of the risk factors and/or categories may be necessary.

- Reviewing the distribution of the scored risk levels and the mandatory and discretionary override factors. Analysts should examine the number of cases scoring within the risk levels based solely on their numerical points, along with the number of offenders whose risk level is derived from mandatory or discretionary overrides.
- Conducting a stepwise multiple regression analysis of the risk factors, the selected recidivism measure (re-arrests, re-convictions, technical violations, or incarceration), the total risk score, the scored risk level, and the final risk level. A stepwise regression analysis shows the contribution of the respective risk factors to the total score. Factors that are not statistically associated with the total score or supervision outcome should be refined or deleted from the instrument.
- Developing a correlation matrix. A correlation matrix is critical to assess the strength and direction of the relationships among the risk factors and the selected recidivism measure, total risk score, scored risk level, and final risk level. Special attention should be paid to the relationships among the risk factors to determine if any are duplicative or generate spurious relationships.
- Determining the risk scale cut points. Designating the cut points for the risk scale is a multi-step process. First, consider the rate of recidivism per total score to identify natural breaks in the distribution of cases. A significant increase in the number of arrests, for example, would suggest a shift from minimum to medium risk. These natural breaks should be further examined with an analysis of variance.
- Analyzing variance. An analysis of variance (ANOVA) is important to determine if the risk levels derived from the risk scale are statistically distinct. The ANOVA asks whether the offenders who scored at different risk levels actually represent a distinct group of offenders with respect to their risk of recidivism. If the analysis indicates substantial overlap between the respective risk levels (for example, if offenders who score as minimum do not differ statistically in their recidivism from those who score as medium), the risk scale is not valid. The total risk score may be statistically correlated with recidivism and thus be a valid indicator of the offender's risk, but if the risk scale does not create statistically distinct groups, the system is invalid because it does not

provide the decision-maker with accurate information about how to supervise the offender.

Step 3: Develop a comprehensive implementation plan. The agency's action plan for implementation must consider staffing; training; impact of any changes on key stakeholders; revisions to the information system screens, database, software, etc.; data required to monitor the system; and the estimated fiscal costs associated with implementation. An often overlooked but critical element of the action plan is system automation. Regardless of how easy the instruments are to score, automation ultimately is essential to enhance reliability, minimize staff workload, and facilitate ongoing monitoring of the system. The action plan should also include goals, objectives, and specific time lines for implementing any changes to the system.

Step 4: Document the validation effort. At the close of the validation study, the agency should prepare a written report documenting the development and evolution of the risk assessment system, the current validation process, and results. The report should be written in non-technical language and distributed to administrative, supervisory, and line staff. It should also provide baseline data for tracking any modifications to the system and assessing the impact of these changes.

Where Can We Get Help?

A validation effort is a technical endeavor requiring statistical expertise. This expertise is rarely available in-house, particularly among smaller jurisdictions. Fortunately, resources are available to assist agency administrators with this critical component of objective risk assessment. The National Institute of Corrections (NIC) offers short-term technical assistance that provides expertise to help plan a validation process.

Assistance can be maximized when agencies are able to commit staff resources to the task of manual data collection or can limit the scope of the specific research questions. Although time-consuming and resource-intensive, undertaking a validation study fortifies an overall supervision strategy by helping the agency make resource allocation decisions based on a valid risk assessment. ■

For more information:

Kelly Dedel Johnson, Ph.D.
President
One in 37 Research, Inc.
1531 S.E. Belmont Street
Portland, Oregon 97214
(503) 235-4053
kdedel@comcast.net

Patricia L. Hardyman, Ph.D.
Senior Associate
Criminal Justice Institute, Inc.
213 Court Street, Suite 610
Middletown, Connecticut 06457
(860) 704-6400
hardypurr@aol.com