**BLOG PUBLISHED ON WEBSITE OF THE ASSOCIATION**

**FOR THE TREATMENT OF SEXUAL ABUSERS (ATSA)**

Leslie Maaike Helmus, Ph.D., Friday, November 13, 2015

**What Makes a “Good” Risk Assessment?**

**A Note on the Importance of Quality Control**

The Issue

Most risk assessment scales are developed and validated by researchers based on file information scored by research assistants, who may have extensive training in the behavioural sciences, but no field experience. An essential question is whether the accuracy of these research studies will be similar to the results of real cases assessed by front-line staff. In other words, how can we implement a scale so that it works with similar (or even better) accuracy as research studies? Before delving into concrete recommendations for conducting high quality assessments, I will first use an illustrative example and summary of research to demonstrate that the commitment of individual staff members and organizations can make a huge difference in how well a risk assessment scale will work.

Illustrative Example: A Tale of Two States

Static-99 (and its revised version, Static-99R) is the [most commonly used](http://cjb.sagepub.com/content/41/12/1406) sex offender risk scale and there are over 60 studies of the scale, which have found that on average, it has [moderate predictive accuracy](http://psycnet.apa.org/journals/pas/21/1/1/). Unfortunately though, there are very few studies of how it works in field settings.

Among many states to mandate the use of Static-99/R for imprisoned sex offenders are Texas and California. Texas found low predictive accuracy – [their results](http://psycnet.apa.org/journals/law/15/4/278/) were lower than most studies of the scale. In contrast, California [found remarkably high levels of predictive accuracy](http://psycnet.apa.org/journals/tam/1/2/102/) – their results were among the best of all studies conducted on the scale, or any other risk scale.

How could two American jurisdictions implementing the same risk scale achieve such remarkably different results?

There are many methodological or policy differences that could affect these findings, but at least part of the difference likely has to do with the quality of implementation. The study from Texas provides no information on how the correctional system maintains the quality of their risk assessments. In contrast, California has a remarkably rigorous implementation and quality control system. All staff who use the scale must be trained by someone certified by the Static-99R developers or a ‘super trainer’ who is certified by a certified trainer and has at least two years of scoring experience. All staff receives training from a detailed, standardized curriculum and by law, they must be retrained in the scale every two years. Additionally, they must pass scoring tests after training, and ideally, their first 10-20 cases are reviewed by a super trainer. Novice users are also encouraged to work with a mentor to maintain the quality of their assessments. With such diligent attention paid to the quality of their risk scale implementation, it is not surprising that California has found some of the highest predictive accuracy ever obtained in a field setting for a risk assessment scale.

What Does Research Tell us About Risk Assessment Quality?

In previous research, quality of risk scale implementation, defined either as [involvement of the scale’s developer](http://cjb.sagepub.com/content/38/5/413.short) to (hopefully) help ensure fidelity to the scale or [whether community supervision officers completed all the steps](http://cjb.sagepub.com/content/early/2015/09/04/0093854815602094.abstract) that were requested of them, was associated with substantial increases in predictive accuracy. Additionally, [quality of training has also been linked to quality of risk scores](http://psycnet.apa.org/journals/lhb/36/3/225/).

What Can Staff and Organizations Do to Promote High Quality Assessments?

The jurisdiction examples and research discussed above show that it is not enough merely to implement a risk assessment scale. Care should be taken to ensure the assessments are done well. This is particularly important given that any risk assessment conducted can be challenged in court.

Below is a list of policy and procedures that are recommended by the developers of the STABLE-2007 (Fernandez, Harris, Hanson, & Sparks, 2014) to keep quality high. This list was modified to be applicable to all risk scales, and is included with permission of the scale authors.

All risk assessment practices should have most, if not all, of the following components in place:

1. A Bring Forward system to cue when it is time to re-assess scorings, ensuring regular scorings (applicable for dynamic risk assessment scales).

2. A system of peer reviews so that everyone is working towards scoring calibration (i.e., all scoring the same case alike). Colleagues should meet on a regular basis and present their scorings to each other and discuss the scorings, working towards consensus.

3. Clinical supervision by a very experienced assessor so that those scoring have access to a resource person for tricky questions (this person may well organize the peer review sessions).

4. Mentorships with those who are more experienced in using the measure so that novice scorers have an identified person with whom they can discuss their cases and their risk scoring.

5. Participation in inter-rater reliability trials where about 10% of the cases are scored by more than one rater and the scores are compared. This technique leads to better calibration of scoring.

6. Your agency may wish to consider participating in webinars about scoring and other risk assessment issues.

7. When scoring risk assessments on any offender a jurisdiction should have a quality control process in place, either through regular professional development days, internal supervision by senior employees who are committed to the risk assessment process or possible “scoring clinics” run cooperatively within organizations.

L. Maaike Helmus, Ph.D., Forensic Assessment Group, Ottawa, ON, Canada