

3 EVALUATION PROCESS

3.1 The evaluation process

The RMA has developed a process for evaluating¹ risk assessment tools and this has been followed for each tool presented in this document. This evaluation process will also be applied in future to assess newly developed tools. The RMA researcher working on this update to RATED used a standard pro-forma to describe each tool based on the summarised evidence and the adopted evaluation framework. This provides individual assessments of the tool against each of the framework criteria and comments to support them.

The RMA welcomes applications for inclusion in the directory, as well as any comments its users.

Adopting an evaluation framework

The adopted framework comprises a number of criteria considered essential for the evaluation of risk assessment tools, building upon earlier work by [Mclvor, Kemshall and Levy \(2002\)](#) and the findings of other researchers ([Heilbrun, Rogers and Otto, 2002](#); [Rogers 2000](#)). Criteria considered as essential are:

- Empirical Grounding
- Inter-Rater Reliability
- Validation History

Compilation of evidence and evaluation

Evidence collated for the current audit, comprised of publications (including 'grey literature,' such as doctoral theses and dissertations for Master's Degree and other unpublished material) and evidence of ongoing research.

The following criteria were used to select research studies for inclusion in the directory:

¹ In this document, the word 'evaluation' is used to refer to the assessment of the tools' performance with respect to the criteria framework outlined.

- Authorship, i.e. whether the author(s) were, or were not, involved in the validation of the tool under consideration, with greater weight given to studies conducted by independent researchers rather than the author(s) of the risk assessment tool;
- Place of publication, with greater weight given to publications in peer-reviewed journals;
- Date of publication, with greater weight given to more recent publications;
- Sponsorship, i.e. whether the study has been commissioned and/or funded by a government or statutory agency, with greater weight given to independent studies;
- Relevance to UK population; i.e. whether the study had employed a sample of individuals from the nations within the UK;
- Target population, i.e. male/female offenders, forensic mental clients, and children and young offenders;
- Size of study population, with greater emphasis on evidence drawn from large sample populations;
- Focus of study, taking into account whether specific issues have been considered, such as predictive validity, inter-rater reliability, significance/ usefulness,² sensitivity and so on.

The following information was provided for each tool listed within the directory:

- The title and any acronym by which it is known;
- The publisher(s)/developer(s) and the year of publication (where known);
- A brief description of the risk assessment tool;
- The type of risk addressed, i.e. the risk of general, sexual and/or violent offending, or any combination of these;
- An outline of its validation history, considering studies carried out in the United Kingdom and internationally; this includes details of the study focus and setting (community, secure unit, prison or mental health institution), the size of the sample and type of population used;
- Evidence of inter-rater reliability, significance/usefulness;
- Strengths, practice considerations, and future research directions.

² Significance/Usefulness of a tool refers to its evidential basis, predictive power and clinical utility in enabling the practitioner to undertake risk formulation and risk management planning.

Tools are divided up into 'Validated' and 'Awaiting Validation.' Validated tools are those which have sufficient research studies evidencing their suitability to be used within the field. Due to the scope of research available on the, validated tools have more extensive entries than those that are 'awaiting validation.' A further distinction between these categories is that to rate a tool in as validated, the RMA considers the existence and quality of the validation evaluation studies, assessed on the basis of the availability of two or more independent papers written by different authors and published in peer reviewed journals.

Tools that are awaiting validation currently do not have the sufficient research evidence to be considered validated; although these have the potential to become validated in future.

The RMA liaised directly with tool authors and publishers wherever possible to compile this material accurately.

3.2 Criteria definitions and performance measurement scales

The following criteria have been used to identify the strengths and weaknesses of each risk assessment tool listed within the current directory. It is important to note that these criteria have not been applied to risk assessment tools awaiting further validation. This is due to the limited research available relating to the psychometric properties of these risk assessment instruments.

3.2.1 Empirical Grounding

Empirical grounding examines the scientific and theoretical underpinnings of the risk assessment tools presented in this document. For example, a risk assessment tool based on sound theoretical evidence and / or other extensive scientific findings observed in prior research would be considered to have a high level of empirical grounding. Higher levels of empirical grounding may, therefore, increase the utility of the instrument in assessing the risk posed by an individual.

3.2.2 Inter-Rater Reliability

Inter-rater reliability refers to the degree to which two or more assessors are consistent in their ratings when using the same risk assessment tool. There are two types of statistics used to estimate this: the Intra-Class Correlation (ICC) and Kappa

(K) Coefficients (see section 5 for further details). Researchers must be able to demonstrate that the risk assessment instruments are reliable; without reliability results, the risk assessment's performance cannot be replicated and validation cannot be attained. Even if a tool had predictive validity, its usefulness will be undermined if it is applied inconsistently by different practitioners ([Taxman, 2018](#)). It is thus of utmost importance for a risk assessment tool to have high inter-rater reliability whereby assessors using the same tools score the items similarly.

3.2.3 Validation History

To rate a tool in this area, the RMA considers the existence and quality of the validation evaluation studies, assessed on the basis of the availability of two or more independent papers written by different authors and published in peer reviewed journals. The papers are required to have examined the predictive validity of the tool and/or its practical usefulness for the assessment and management of risk of harm to others. This approach accommodates concerns that have been raised in the literature that different research designs may be appropriate for identifying the properties, strengths and limitations of various types of instruments.

The validation history criterion is split into four subsections which are as follows:

- a. General Predictive Validity This item refers to the capability of the risk assessment tool to discern between the recidivist and non-recidivist populations. The recognised statistical measure for this is the receiver operating characteristic curve (probability that a recidivist would receive a higher rating than a non-recidivist) ([Taxman, 2018](#)).
- b. Applicability: Female Offenders This item refers to the validity of the risk assessment tool for women and girls.
- c. Applicability: Ethnic Minorities This item refers to the validity of the risk assessment tool for ethnic minority populations.
- d. Applicability: Mental Disabilities. This item refers to the validity of the risk assessment tool for those with mental disorders and/or learning disabilities.

Tools are rated on these criteria using the following performance scale:

0 Bars - No evidence of or evidence indicating insufficient empirical grounding, inter-rater reliability or validation.



3 Bars – Preliminary evidence of empirical grounding, inter-rater reliability or validation.



6 Bars – Intermediate evidence of empirical grounding, inter-rater reliability or validation.



9 Bars – Sufficient evidence of sufficient empirical grounding, inter-rater reliability or validation.



For tools that are ‘awaiting validation,’ this section is entitled ‘Tool Development’ and covers findings of studies in relation to predictive accuracy. ‘Tool Development’ will also cover the empirical grounding of a tool and studies relating to inter-rater reliability where these are available.

3.2.4. Contribution to Risk Practice

This section in tool entries supports the tiered approach to risk practice promoted within the RMA’s Framework for Risk Assessment, Management and Evaluation (FRAME), in particular the first three standards relating to: (1) Risk Assessment; (2) Planning and Responding to Change; (3) Risk Management Measures.

This section does not incorporate the ratings scales observed in the previous criterion; rather, it provides a qualitative evaluation of the properties of the instrument and its contribution to risk practice.

3.2.5 Other Considerations

This final section documents other elements to be considered before using the tool, e.g. whether it has a clinical over-ride feature to allow the risk rating to be readjusted at the assessor's discretion. Also noted in this section are any other research findings about the tool that do not relate to its inter-rater reliability or predictive accuracy. In tools that are 'awaiting validation,' this section is entitled 'General Notes' and covers similar areas.