Manchester Metropolitan University

Early testing and formative evaluation of the Enablers of Change assessment and sentence planning tool for adults with convictions

Kevin Wong and Rachel Horan

ABSTRACT

The Enablers of Change assessment and sentence planning tool has been designed to assess the risks, needs, strengths and protective factors of adults with convictions. Developed by Interserve, a Community Rehabilitation Company (CRC) provider in England, the tool is an innovation. The first of its kind in the United Kingdom (UK) to operationalise the risk needs and responsivity model with the 'good lives' model and desistance principles for the *general adult* population of low to medium risk of harm individuals managed by CRCs. This paper reports the development, early testing and formative evaluation of the tool and recommendations for it's onward development. Given that such integration is regarded by many as the 'holy grail' of probation practice, this article is of international significance and will make an original contribution to the limited evidence base on operationalising desistance in the management of adults with convictions in the UK and other jurisdictions.

INTRODUCTION

Integrating desistance research with the risk needs responsivity corrections model has been advocated as the 'holy grail' for advancing offender assessment and sentence planning in the United Kingdom (UK) and other jurisidictions (Canton 2014, McNeill and Weaver 2010). The Enablers of Change (EOC) assessment and sentence planning tool aims to do just this for the general adult offender population and therefore is of significance for policy and practice both in the UK and internationally. The tool has been developed by Interserve, a private company who run five Community Rehabilitation Companies (CRCs) established by the Ministry of Justice (MoJ) as part of the Transforming Rehabilitation (TR) changes to probation provision in England and Wales (Ministry of Justice, 2013). One of the innovations offered by Interserve as part of their bids to run the CRCs[i], the EOC tool is one of the more notable innovations arising from the TR changes and its utility for criminal justice case management practice goes beyond the shifting sands of politicised policy making (Raynor 2018, Senior, 2013,); which has taken a new turn with the review of TR announced by the

Justice Minister David Gauke, responding to the government acknowledged shortcomings of TR (Ministry of Justice, 2018).

This article is the first in a *series of related papers* which will collectively provide: an overall account of the development of the EOC tool; and examine the innovations in offender management which *the tool* and *how it is intended to be used* has been designed to operationalise: conceptually integrating RNR and desistance; more effective engagement with adults with convictions; and co-production.. *This paper will* provide an account of the development and testing of the tool, commencing with the rationale for the tool before considering the development of its first iteration and accompanying first phase evaluation. The modifications leading to the second iteration of the tool and accompanying second phase evaluation are then explored before a discussion of the fitness for purpose of the tool and a concluding consideration of its future development and testing.

The EOC tool is intended to support a whole system approach to Interserve's delivery of its CRC contracts for managing adults with convictions assessed as being low to medium risk of harm[ii] (Ministry of Justice, 2013). The delivery model "Interchange" is underpinned by six core modules: induction; assessment; plan; networks; review; and exit. These are intended to provide a structured approach to the service user's journey and a platform that enables Case Managers (CMs) and Service Users (SUs) to adopt a strengths approach to address the issues in the SU's lives that has caused them to offend (Barry, 2013); Interserve, 2016a). Collectively these modules are not dissimilar to the overarching structure of the National Offender Management Service[iii] (NOMS) ASPIRE[iv] offender management model (NOMS, 2006). The tool is intended to facilitate discussion between CMs and SUs to generate information which will inform:

- A sentence plan which is co-produced by the CM and SU which is intended to be dynamic and personalised to each SU;
- A risk assessment undertaken by the CM;
- Implementation of the sentence plan through the networks module; and
- The monitoring and review of SU progress against the sentence plan

The EOC tool has been designed to fulfil two aims. Firstly, to replace sections 2-13 of OASys[v], the needs and risk assessment tool used by probation and prison services in England and Wales since 2002, developed and tested by NOMS which, as summarised by Moore (2015), is based on the risk needs responsivity (RNR) model (Bonta and Andrews, 2007, Andrews and Bonta, 2010).[vi] Secondly, (the intended significant innovation) the EOC tool has also been designed to collect information on strengths, social capital and community networks and to assess SU's motivations to change. In so doing, the tool is intended to operationalise primary, secondary (McNeill and Weaver, 2010) and tertiary (McNeill, 2016) desistance processes and their 'zigzag' path (Barry, 2013). As observed by Maruna (2016) in an editorial about restorative justice and desistance, after thirty years, desistance (and restorative justice) is no longer new. That the EOC tool is the first attempt in the UK (and arguably other jurisdictions) to explicitly operationalise desistance as a fundamental part of mainstream probation practice is novel and welcome, but also arguably a long time coming. In aiming to operationalise desistance principles, conceptually, in England and Wales, it is aligned with the AssetPlus needs and risk assessment tool used with adolescents with convictions (Baker, 2012; Youth Justice Board, 2014) and the Active Risk Management System (ARMS) risk management tool used by probation and police for people who have committed sexual offences (Nicholls and Webster, 2014, Kewley and Blandford, 2017).

The development of the EOC tool has been undertaken under the close scrutiny of two bodies. Firstly, an ongoing consultation with the Ministry of Justice (MoJ) responsible for approving the use of the tool against a comprehensive framework devised by NOMS (2015) for all new risk and needs assessment tools intended for use in England and Wales. This concluded in June 2017 with a formal submission,by Interserve to the MoJ for approval to deploy the tool across one or more of their CRCs. Secondly through on-going independent evaluation undertaken by Manchester Metropolitan University (commissioned by Interserve) to validate the tool and inform the tool's development. It is the results from this ongoing evaluation which this paper draws on. Ethical approval was obtained through the university's ethics committee; and research approval for the methodology was obtained through Her Majesty's Prison and Probation Service (HMPPS) National Research Committee (NRC).

RATIONALE FOR THE ENABLERS OF CHANGE TOOL

It is widely recognised that the assessment of individuals with convictions is a key first step in their rehabilitation with the process providing an important and necessary multi-dimensional perspective of the individual to inform a suitable plan of rehabilitation (Canton, 2014; Moore, 2015; Council of Europe, 2010). A suitable assessment will identify needs and risks to determine what should be targeted and what and how any interventions should be delivered. Appropriate assessment is also an important element of progress review, enabling monitoring, if completed at regular intervals. The EOC tool has been developed by Interserve to fulfil these assessment and related functions and also to underpin a more personalised and co-productive approach to developing the sentence plan, providing a holistic approach to assessment and case management (Fox and Marsh, 2016; Interserve, unpublished). Importantly, the EOC assessment is not intended to be stand-alone but combined with information received including[vii]:

- offence and sentence information;
- Risk of Serious Harm Screening (from Case Allocation Screening);
- Risk of Serious Recidivism (RSR);
- the revised Offender Group Reconviction Score (OGRS 3); and,
- Risk of Serious Harm Analysis (if indicated by risk of serious harm screening).

It is planned that all SUs complete a self-assessment during the induction module, which will be the basis for the co-development of the assessment and plan. These areas are assessed through a review of all available information, a self-assessment questionnaire and assessment interviews (Interserve, 2016a). The language used in the new assessment tool is designed as being different; positive and enabling (Interserve, 2016a). The EOC questions are designed to be strengths based, and a move away from a focus on needs and deficits. The tool has been designed to be dynamic such that individual sections of the assessment can be updated at any time without the requirement for the whole document to be updated. It is planned to link with the Interchange Plan (see earlier) and if objectives are met or there is a change in circumstances recorded in the plan then this will trigger a notification or reminder to update the assessment and vice versa.

The EOC and process for using it represents a departure from previous general assessment practice of adults with convictions in England and Wales, represented by OASys (Moore, 2015). Throughout the EOC development process, Interserve have been in regular communication with the MoJ who have sought confirmation that the EOC tool would not compromise the 'gold standard' assessment of risk and risk management planning which OASys, based on risk needs responsivity principles (Bonta and Andrews, 2007) was regarded as providing (Moore, 2015). This was of particular concern given the stated innovation (designed into the EOC tool) of an integrative approach bringing together empirically evidenced RNR approaches with new thinking in desistance and strengths led approaches but which have a considerably less extensive empirical evidence base. That said, adopting desistance theory, emphasising the need for a holistic, flexible and person-centred approach to supporting people who have offended was clearly signalled in the NOMS Commissioning Intentions document (NOMS, 2014). Further support for the adoption of a more personalised approach to rehabilitation and to operationalising the desistance literature is provided by The European Probation Rule guidance on assessment and risk (Canton, 2014; Council of Europe 2010). Setting this within the context of promoting justice and human rights as the first duty of a criminal justice system, an individualised approach is encouraged, including additional factors in the development of appropriate assessment tools such as: strengths; aspirations; obstacles to desistance (including non-crime related needs); motivation; and, compatible interventions / resources that might be made available.

The formative and early evaluations of the EOC tool which we outline in the current paper have been guided by the quality and functionality thresholds for criminal justice assessment tools used in England and Wales, primarily practitioner perpectives of the desired characteristics of a risk and needs assessment tool again in part drawn from the design and use of OASys (Moore, 2015). These being:

- *Face validity:* It must be clear why each item is included.
- *Clear definitions:* Clear and unambiguous definitions of the items are required for consistency.
- *Simple scoring system:* Question scales that stretch beyond five points have been found to be difficult.
- *Evidence boxes:* Practitioners should have the opportunity to express their concerns and elaborate on their assessment.

- *Offender input:* Offenders should be provided with the opportunity to express their views.

- *Useable within limited time constraints:* Resource implications need to be considered.

- *Complements current practice:* The instrument needs to be continually developed.

The evaluation has sought to explore the extent to which the EOC tool fulfils these desired criteria which form seven research questions. Given the early phases of tool design and roll out, it is not within the scope of the presented evaluations to explore the tool's validity and reliability. This will be explored in subsequent evaluation phases in parallel to wider EOC roll out.

PHASE ONE DEVELOPMENT AND TESTING

Phase One iteration

The first iteration of the EOC tool was developed and tested in 2016. It comprised the following six *enabling areas*, i.e. the thematic constructs assessed by the tool (Interserve, 2016a)

1. Personal relationships and support networks.

2. Health and wellbeing (includes mental and physical health, drugs and alcohol).

3. Thinking, skills, attitudes and behaviour:
   - motivation;
   - interpersonal skills;
   - impulsivity;
   - ability to recognise problems;
   - problem solving skills;
   - awareness of consequences;
   - achieves goals;
   - understands other people's views; and,
   - concrete/abstract thinking.

4. Education and work (including employment, training and education (ETE)).

5. Home and money (including accommodation and finance, benefit and debt).

6. Friends, support and activities (including lifestyle and associates, networks and behaviours).

Each of these areas were intended to be scored by both the SU and the CM using a 0 to 5 Likert type scale. Initially, the SU was to complete a self-assessment form that should then be reviewed with the CM during the assessment. The CM was intended to score the SU using 'professional judgement' with any discrepancies between the SU and the CM being 'moderated' (Interserve, 2016b). Through this process CMs were intended to be able to form conclusions about:

•   future offending, risk of harm;

•   nature of previous offending;

•   SU's strengths;

•   sufficient information to develop a series of interventions; and,

•   understand any requirements around protected characteristics.

Phase One evaluation methodology

The Phase One evaluation tested the first iteration of the EOC tool. It was qualitative, with a scoping phase comprising: two observations of group inductions; two observations of individual inductions; and four observations of assessments between SUs and CMs. The main fieldwork phase comprised: 13 observations of assessments between SUs and CMs using the tool, followed by interviews with 10 SUs and 13 CMs. The scoping was conducted in one CRC area and the main fieldwork in two CRC areas. Sampling was opportune. The tool was alo assessed for internal coherency with each *enabling area* broken down into specific constructs to test for the 'singularity' of the questions. The ratings scales were also assessed for face validity comparing terminology in the scales across each of the constructs to test for uniformity both within each construct and between the constructs. There were considerable limitations of the methodology for this evaluation, principally in its scope and comprehensiveness. The findings were based on small numbers of interviews and observations in a restricted number of CRCs limiting SU and CM representativeness. However, 'saturation' appeared to have been reached within the limited fieldwork as the same issues with this iteration of the tool were consistently identified in the observations and interviews.

Phase One findings

Some positives were found regarding the operability of the EOC tool. The underlying principles of a strength-based approach seemed to be well received by CMs. Whilst some

CMs were confident in their expected use of the tool there was experience of a mixed 'roll-out' with many CMs under-trained and not sufficiently familiar with the tool, its theoretical underpinnings and how it should be used. SUs generally experienced the EOC positively, understanding what was being asked by their CM and feeling guided through their assessment.

This was the first iteration of the EOC tool and the first time that it had been tested. Therefore limitations of this iteration were anticipated and identified. The analysis of the internal coherency of the tool identified that the *enabling areas* may have shared the same scale but that the ratings were incompatible; aggregate ratings, if required, could not be obtained. All of the *enabling areas* measured multiple constructs which meant that rating scales differed between but also within *enabling areas;* sections rated SUs across different domains along the same scale. The combination of constructs rated on a rigid 0 to 5 scale prevented several possible combinations from being recorded, potentially biasing the assessment tool towards certain rating outcomes. There was insufficient evidence of internal consistency in this iteration.

The generalisability of this iteration of the tool was also not supported by the evaluation. Many CMs identified groups of SUs with whom the tool was deemed inappropriate, including those for whom English was not their first language and SUs with learning needs. This issue of generalisability and suitability across individuals exacerbated a limitation regarding the aspiration that the tool should support more individualised assessments.

This iteration did not provide a clear and specific account of risk. Some CMs expressed concern with the apparent lack of focus on risk. CMs were unable to form conclusions about future offending and risk of harm. This may, in part, be due to its reliance on the risk assessment elements of OASys to be used in conjunction with the EOC tool.

The importance of an integrative approach to the formulation of a suitable assessment tool for people with convictions, bringing together empirically evidenced RNR approaches with new thinking in desistance and strengths led approaches was not evidenced in this iteration of the tool or the way in which it was used.

Recommendations for tool improvement

The Phase One evaluation concluded that the first iteration of the tool did not have face validity nor the necessary attributes of best practice in the assessment of people with convictions. The following steps were suggested as necessary to improve the tool before further empirical research to assess its operability:

1. Review the operationalisation of the theoretical underpinnings and rationale of the tool.

2. Break down the constructs so that each area measures only one thing.

3. Ensure uniformity in the scoring system across and within the constructs.

4. Consistently reflecting the strengths-based focus of the tool for example by removing the word 'problems' from the assessment scales.

5. Ensure that the process of risk assessment and how the EOC tool enables this to be umabiguously articulated and understood by CMs.

6. Develop further training and guidance around how the tool, its components and the self-assessment are meant to be used within an assessment.

7. Review the language used for the tool to enable operability by participants. For example, allowing the constructs of concrete and abstract thinking to be operationalised in a more appropriate way than, for example, asking whether a SU is a concrete thinker.

PHASE TWO DEVELOPMENT AND TESTING

Phase Two iteration

The tool underwent significant revision by a small team of experienced CMs aided by academic advisers from the Phase One evaluation team. This included a large-scale review of the available literature featuring studies by (McNeill, 2006; Morton, 2009; McNeill, 2009; Andrews and Bonta, (2010); McNeill and Weaver, 2010), a practice review of learning conducted by Interserve and reformulations of the tool. The second iteration of the tool was finalised in August 2017. There was a clear separation of function and personnel between the academic development advisers and the independent team of researchers who undertook the Phase Two evaluation.

The second iteration of the EOC tool included the following *enabling areas* (i.e. constructs) assessed through a suite of questions (Interserve, 2017):

1.   Accommodation

2.   Employment

3.   Education

4.   Finance

5.   Confidence and life management

6.   Relationships and personal support networks

7.   Emotional and physical well-being

8.   Drugs and alcohol

9.   Thinking skills

10.   Attitudes

11.   My community networks

12.   Feeling included

Each of the enabling areas areas and the extent to which they were linked to risk, previous offending and strengths were scored on a five point Likert type scale by both the SU and the CM. Again, it was intended that the SU will have previously completed a self-assessment form (during induction) that is reviewed with the CM during the assessment. The CM scored the SU using their 'professional judgement' with any discrepancies between the SU and the CM being 'moderated' by the CM. Importantly, this iteration of the tool was designed to facilitate the collection of information which the CM could use to risk assess using all risk sections of OASys. Therefore at the time of testing, the way in which risk was intended to be assessed remained unchanged from existing practice.

Figure 1 illustrates how this version of the EOC tool was intended to fit within the six core modules of the Interchange offender management model.

Figure 1 The EOC tool and the Interchange model

<<Figure 1 goes here>>

Phase Two evaluation methodology

This Phase two evaluation, based on a theory-orientated methodology, (Fox et al., 2017; Stame, 2004) was undertaken between August and November 2017 and tested the operability

of the second iteration of the EOC tool and process. A Theory of Change (TOC) (Weiss, 1995; Vogel, 2012; Stein and Valters, 2012) was developed based on a document and literature review and interviews with key Interserve personnel. This provided a framework for constructing an evidence-based "story" about the tool, process and outcomes, against which the evaluation approach was determined (Stame, 2004). Initial cognitive testing (Willis, 2005) was undertaken with five SUs (opportunely sampled from a SU user group) to provide an in-depth exploration of concepts, processes and patterns of interpretation of the EOC tool. Fieldwork to undertake further cognitive testing, face validity and inter-rater reliability assessment took place in one CRC office comprising: 20 observations of EOC assessments; and post assessment, 20 semi-structured SU interviews and 15 CM interviews.[viii]; These aimed to capture:

- reflections on the EOC process and experience;
- levels of co-production and collaboration;
- congruence of assessment points and decisions;
- descriptive observation of the experience of each section and the relevance of EOC content.

Interview data were analysed thematically (Ritchie and Spencer, 1994) based on the above themes and observation data were analysed against the Working Alliance Inventory (Darchuck et al., 2000). Inter-rater reliability (IRR) analysis was conducted on three formulated low, medium and high need case studies completed by 16 staff (CMs and tool developers) and descriptively compared against a model EOC assessment for each. The sentence plans of the 20 observed EOC assessments were descriptively analysed to assess congruence.

The methodology was limited by sampling and the short time frame available for the evaluation. The cognitive testing was undertaken by SUs from a CRC user group, they were more engaged with the CRC and therefore may have been more positive about the tool. The TOC interview data were limited by the staff available to be interviewed at short notice. The fieldwork (including IRR assessment) was confined to a single CRC office, the only site where the second iteration of the tool was piloted and inevitably, limited to observations of assessments where the SU attended (the attendance rate was 50 per cent) and CM availability. The sampling of SUs was limited to these attenders (who may have been more engaged with

the CRC). While 15 CMs at the site were trained in using the tool and initially agreed to be observed, this reduced to a handful of CMs, latterly supplemented by a tool developer who was assigned to assist with assessments to attain the target of 20 observations.

Phase Two findings

Theory of change

A theory of change (TOC) logic model was built to identify how the EOC tool and its administration was intended to work and the inputs and activities required to generate intermediate outcomes for SUs and CMs. For the former, a more inclusive and improved experience of assessment, a better balance between support encouragement and enforcement, leading to greater ownership of the assessment (and sentence plan). For the latter, improved risk and sentence planning, improved satisfaction with the process and experienced as more dynamic and iterative. Ultimately these aim to facilitate primary, secondary and tertiary desistance outcomes (McNeil and Weaver, 2010; McNeil, 2016).

Cognitive testing

This explored whether the EOC assessment questions successfully captured the intent of the question and made sense to SU respondents, examining the question-response process, conceptualised by four stages: 1. Comprehension, 2. Retrieval, 3. Judgment and 4. Response (Centre for disease control and prevention 2017). Analyis of the cognitive interview data generally found minimal recall difficulties satisfying the requirement for *Retrieval*) and mostly complete responses (*Response*). However, for *Comprehension*,: language problems were experienced, i.e.– not knowing the meaning of words/phrase such as finances and networks; and logic problems - appropriately interpreting the scaled responses and placing responses on the Likert scales. In relation to *Judgment*, SUs tended to Yes/No answers rather than placing themselves on each Likert scale. SUs also found it challenging responding in intended ways to the summary ratings of the enabling areas and their links to risk, previous and current offending and strengths.

Usability

The observations and interviews confirmed that the EOC tool was well received and experienced by most CMs and SUs. For CMs, it was universally a positive experience yielding what was perceived as a true picture of the SU. SUs reported that the tool demanded

greater self-reflection which contributed to their generally positive experience. Both SUs and CMs valued the time spent drilling into issues and searching for solutions together with the tool's focus on strengths.

Most CMs and SUs experienced the assessment as a co-produced process but inevitable disagreements did occur. Disagreement, in some cases, yielded more comprehensive assessments with CMs challenging, probing and negotiating with the SU to reach agreement. However, some discussions did yield confusion as to whether an area was a strength, or whether positive factors had been identified.

Observations from SU and CM interviews indicated that the tool had elicited greater honesty and enhanced disclosures of information. For example, one CM reported that this was the first time that an SU had disclosed childhood sexual abuse. One CM commented that the specific inclusion of risk within the tool and the co-production ethos of its application allowed the CM to be more honest in relation to discussion of risk with the SU, where in other processes risk assessment was undertaken without the SU present.

Technology was experienced as a challenge. The piloted EOC version was provided on a spreadsheet. CMs reported that they wanted the flexibility to switch between sections, for information to auto-populate and sections or questions to be automatically identified as not applicable. Conditionality – 'if yes, move onto the next part'- was sought to guide the CM on where and when to answer specific questions. CMs noted that slower keyboard skills and unfamiliarity with IT could lengthen assessments but were assured by a longer term plan of a bespoke technology solution.

There was no evidence from the observations or interviews of self assessment information being used during the EOC assessment an intended feature of the process.

There was a lack of clarity amongst SUs and CMs of the purpose of the EOC.

There was also some confusion about the onwards application of a co-produced assessment to the criminal justice process, particularly evident where SU views had led the process, with CMs reflecting how that would be best communicated at National Probation Service case

transfer points[ix] and in any onwards risk management or allied processes. Concern was raised that the SU view could be assumed to be that of the CM opinion.

Challenges

Observed EOC assessments ranged from 80 minutes to 3 hours. Although experienced as lengthy by SUs and CM, this did not detract from their overall positive experience. Some felt it could be shortened and more targeted. Assessment length and requirement to conclude the interview forced a focus on content and section completion with consequent limited SUs needed time to respond to complex and personal questions and summation. SU were, on occasion, cut short, or answered-for by CMs rushing to complete the assessment.

The responsivity of the tool was well experienced by SUs and CMs, although on occasion it was less responsive to SUs with particular needs such as dyslexia, where the SU found the Likert scales difficult to read and understand. Due to sampling limitations it was not possible to determine tool responsivity to other SU needs.

The order of assessment garnered differing CM views. Some experienced it as starting negatively because of its initial exploration of offending and therefore perceived focus towards risks that set a negative basis for the remainder of the assessent. Other CMs felt that commencing with the offence gave the EOC assessment legitimacy.

The question that asks whether each enabling area is linked to previous or current offending, are in practice two questions that can yield two different responses, although, the question is intended to obtain a single 'yes' or 'no' answer. This was not well experienced by SUs and CMs. In many observations, two sub answers were provided. This led to confusion in agreeing ratings.

Confirming a finding from the cognitive testing (see above) SUs experienced frequent difficulties in understanding the concept of risk, and rating each enabling area to their risk of harm. SU interviewees acknowledged that they found this area the most difficult to understand. CMs reported that the co-production of a rating of risk for each enabling area was difficult, and in some instances created blockages in the assessment process.

SUs appeared to be rating whether an enabling area was a risk of harm to themselves, others whether an area led them to pose a risk to others. Others linked risk of harm to their offending behaviour and the risk that they may reoffend. In assessments where the CM concluded early that the SU was a low risk of harm, no enabling areas were rated as being associated with the risk of harm. The CM omitted these questions, moving onto other enabling areas. This pattern of assumptive ratings, led by some CMs, did not enhance SUs understanding and perhaps gave some inaccurate SU and CM overall ratings.

Face validity

The observations and interview data found that face validity - defined as clarity among the users about why each item is included in the tool (Moore, 2015), was experienced well by CMs and SUs for some enabling areas but less so for other areas.

Offences and *Accomodation* were well experienced and considered relevant.

Education - was experienced as a relevant. Struggles with learning, or learning support needs such as dyslexia were not always identified by this section.

Finance - was experienced as relevant. Interviewees commented that the questions were phrased to invite agreement if it was not a problem, contrasting with other sections which were orientated differently.

Confidence – was frequently misunderstood and harder for SUs to understand, requiring CMs to provide explanations which all differed. The section questions were well experienced but their relevance to confidence was unclear to the SUs.

Relationships – was well experienced and where CMs uncovered information about the SU that they did not previously know. A question about friends being a positive support often yielded good SU insight and reflection. CMs particularly valued questions about family life as they commonly felt this area was not normally discussed in such depth. A question regarding SU behaviour in relationships was commonly more difficult for SUs to understand.

Emotional and physical well-being - was experienced as a relevant section, although CM and SU feedback queried why the two areas were together in one section.

Drugs and alcohol - was generally experienced as a relevant section. However, most questions covered drugs and alcohol together, leading to complex responses.

Thinking - this was mostly a well experienced and relevant section with good questions, although some CMs felt there was some repetition.

Attitudes - feedback indicated that this section seemed very similar to the offence analysis section, nevertheless the section did generate frequent debate.

Network - generated interesting discussions and wide variations in SU and CM understanding of community both of which were helpful to the assessment. However, the relevance of this section needed to be clearer, with some SUs experiencing it as 'tokenistic'. Feedback also indicated that this section overlapped with the relationships section.

Feeling included - was often misinterpreted, commonly introduced as a section on 'feelings'. The content was generally experienced as relevant, but in parts, questions were long winded.

Inter-rater reliability (IRR)
Interesting trends were observed but caution should be taken in interpreting the findings as the number of assessments were low (especially for medium and high need case studies). Congruence appeared to generally increase through the level of need, with the low need case study having the least over-all congruence and high need case study having the most congruence. The *enabling areas* of accommodation and education saw the greatest Inter-rater congruence (IRC) which was less for finance, and health and well-being.

Low and high need cases generally saw less IRC in relation to strengths and links to previous / current offending. This supported the observations in SUs and CMs feedback which found these areas challenging. Across all three case studies, risk saw the largest rates of IRC. This suggested that CMs had a better understanding of risk, compared to the other two areas and that perhaps CMs had insufficient time to apply and embed their new learning regarding desistance and the integrated approach (RNR and desistance) to assessment.

Sentence planning quality

There were good elements of sentence plans that corresponded well to an individual EOC *enabling area*, but these were in the minority. Generally the content of sentence plans were sparse, actions were limited in detail and specificity (i.e not SMART[x]). There was an absence of clear milestones and targets, with unclear sequencing within the plan. Elements were included which seemed unrelated to the EOC identified needs and problem areas. There was a difficult balance between optimism for reform and realistic sentence plan (compounded by the EOC tool assessing previous and current offending behaviour together) making it unclear whether historic or current links to offending were being addressed. There was limited inclusion of strengths as a positive element of sentence plans to build upon.

DISCUSSION

As may be expected when testing a new assessment tool and process, some aspects of the EOC tool and processes worked well, others needed revision. The phases of evaluation have enabled consideration of the 7 research questions guided by practitioner perpectives of the desired characteristics of a risk and needs assessment tool (Moore, 2015).

1. *Face validity:* It must be clear why each item is included.

The evaluation has identified emerging evidence of the tool's face validity in its second iteration. This iteration was generally viewed as covering the concept it purports to measure. A number of enabling areas were experienced well, however, a number were less positively experienced. There was some identified lack of clarity of the purpose of the EOC which impacts its face validity. There was complexity experienced in enabling areas and queries as to whether the content of enabling areas is relevant in all cases.

It is important to consider that face validity relates more to what a test appears to measure to the person being tested than to what the test actually measures" (Cohen and Swerdlik, 2009) and evidence of its face validity, whilst important at a formative stage of evaluation, should be developed towards assessment of the tools content, criterion, and construct validity together with reliability and norming. Importantly, the face validity of the tool in the collection of risk information has not been explored.

2. *Clear definitions:* Clear and unambiguous definitions of the items are required for consistency.

Some SUs found it difficult to place themselves on the scales and did not understand the fine detail of the difference between items of the scale. There also needed to be improved consistency of response to the Likert scales – responses to each scale included both 'yes' and 'no' answers which reduced the clarity, and in some cases, potential meaning of responses.

As a consequence, verbal qualifiers have been developed with narrative ratings and definitions for each point value on the scale. Verbal qualifiers provide advantages that include ease of explanation and familiarity and to facilitate capturing normative judgments (Moxey and Sanford, 2000). This is offset by inferior measurement quality and that cultural factors might confound the data (Van de Vijver, 2001) and whether the interpretation of qualifiers is stable over time (Rohrmann, 1978).

3. *Simple scoring system:* Question scales that stretch beyond five points have been found to be difficult.

Whilst Likert scales were experienced as a necessary guide, the tool's Likert scales in its second iteration were considered to have limited utility. The majority of participants viewed scales as too long with too many available options.

As a consequence, Interserve have shortened the scales to a consistent 1-5 unipolar Likert scale question format, indicating respondents to think of the presence or absence of a quality – in this case the 'accuracy'of the statements applied to themselves

The simplicity of the tool's scoring system will be further explored in subsequent evaluation phases.

4. *Evidence boxes:* Practitioners should have the opportunity to express their concerns and elaborate on their assessment.

There was a mixed collection of EOC assessments, reflecting mainly the CM's views, others the SU's views, some were the SU's views ratified by the CM and others were generally a mutually agreed assessment. The recording of this would benefit from further guidance and training, given CM unfamiliarity with the co-produced nature of the tool and process.

5. *Offender input:* Offenders should be provided with the opportunity to express their views.

SUs were encouraged to express their views by CMs, in part an artefact of the design of the tool which required SUs to rate their responses as well as the way in which the tool was used by CMs. The EOC form was viewed by both parties at the same time, offering a novel experience and transparency to the process which was well experienced by SUs.

6. *Useable within limited time constraints:* Resource implications need to be considered. While the second iteration of the tool was experienced as lengthy by SUs and CM this did not detract from their overall positive experience, although they felt it could be shortened and more targeted. The relative unfamiliarity of the tool for CMs may have contributed to extending the time required to complete the assessment, as well as the way in which the tool was structured and its form – an electronic spreadsheet, which did not readily allow CMs to navigate to sections relevant to the SU and avoid sections that had no relevance.

Information about how Interserve have responded to this is below. The usability of the new format will be tested in future evaluation.

7. *Complements current practice:* The instrument needs to be continually developed. Learning from the evaluation of the first iteration of the tool was incorporated by Interserve into the design of the second iteration of the tool and more recently findings from the second phase evaluation have also been responded to. A third phase evaluation is planned for a wider roll-out which would allow for more rigorous testing.

It has not been possible within the space available to provide a comprehensive exposition of the development of the tool and evaluations, instead, the aim has been to relate the development journey up to the point where the Phase Two evaluation was concluded. The authors are aware that further development work has been undertaken by Interserve to

address the shortcomings of the tool and processes identified in the later evaluation. These include: integrating the EOC tool as part of an IT based case management system designed to meet MoJ IT security accreditation standards; designing the tool's touch screen interface to facilitate interaction by the CM and SU; providing pop up information to explain the purpose of each section in a standard way to SUs; having the functionality to allow non-linear completion of the tool; and information from the SU self assessment (where completed) appears in the tool.

The next phase of the EOC research would require a deeper and broader study to address the limitations of past evaluation methodologies; and to enable quasi-experimental design methods and quantitative analysis to be conducted. This would mean:

1. More extensively testing the face validity of the tool – with CMs and SUs drawn (ideally) from all five CRCs managed by Interserve.
2. More extensively testing the inter-rater reliability of the tool with a larger number of CMs drawn from all five CRCs.
3. Test the internal reliability of the tool, i.e. how well items within each 'enabling area' of the assessment tool measured various aspects of the same characteristics.
4. Test the tool for construct validity, i.e. how well the assessment distinguishes between discrete individual-level or social characteristics.
5. Test the use of the tool at each stage of the offender management process from assessment, through review to case closure examining relevance and usability.

In addition to these tests, there is a requirement to test the predictive validity of the EOC tool. A feasibility study is indicated in the first instance to guide the determination of the outcomes,the data required and the design of the study.

CONCLUSION

The EOC is a mostly well-experienced approach among the small number of case managers and service user who have tested it thus far. It has faced the inevitable challenges of a new, innovative and responsive assessment. The two evaluations of the EOC reported here tested the operability of the EOC tool and process: 'the what'; 'the way'; and 'the environment'. The research was guided by practitioner perpectives of the desired

characteristics of a risk and needs assessment tool. The evaluation has yielded rich data and insight into the EOC which offers learning for others attempting a similar enterprise. There is emerging evidence of the tool's operability and some evidence of face validity. The evaluation suggests that the EOC offers an assessment framework that is capable of delivering the short, medium and long-term outcomes that it seeks. The EOC does have remaining learning and refinements that are necessary to further develop the tool.

That some aspects worked well, others not so well, reflects the challenge of not just producing a new assessment tool but one which manages the integration of RNR and desistance approaches, ensuring that the more empirically based RNR practice is not diluted while drawing on the less extensive (but growing) empirical evidence for desistance (for example, see Savolainen, 2009; Skardhamar and Savolainen, 2014). There also appear to be challenges in managing the cultural transition required: CMs combining their more familiar RNR focus on managing risk and crime related needs, alongside promoting SUs strengths, protective factors and considering non-crime related needs - which is new. Evidence from the testing of the EOC tool suggests that this integrated approach is not necessarily easily adopted by CMs While CMs may ideologically welcome the incorporation of strengths based approaches as part of their case management processes, confidently operationalising this more positivist humanistic aspiration in a policy and practice context where the risk management is paramount is likely to be challenging. This is something that Interserve and indeed any body wishing to integrate desistance principles with RNR based approaches will need to address. The ultimate outcomes of the EOC or any tool which aims for this type of integration is worth striving for - supporting the progression of individuals towards desistance - alongside promoting justice and the human rights of the individuals and more broadly for those of society.

.

References

Andrews D and Bonta J (2010) Rehabilitating Criminal Justice Policy and Practice. *Psychology, Public Policy, and Law* 16(1): 39–55.

Baker K (2012) *AssetPlus rationale*. London. Youth Justice Board.

Barry M (2013) Desistance by design: offenders' reflections on criminal justice theory, policy and practice. European Journal of Probation 5(2): 47 – 65.

Bonta J and Andrews DA (2007). Risk-need-responsivity model for offender assessment and treatment. *User Report No*. 2007-06. Ottawa, Ontario: Public Safety Canada.

Canton R (2014) The european probation rules, assessment and risk. *EuroVista*, 3(2): 68–80.

Centre for Disease Control and Prevention (2017) Cognitive testing interview guide. Available at:
https://www.cdc.gov/nchs/data/washington_group/meeting5/WG5_Appendix4.pdf (accessed 30 October 2018).

Cohen RJ and Swerdlik ME (2009). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). New York, NY: McGraw-Hill.

Council of Europe (2010). Recommendation CM / Rec (2010)1 of the committee of ministers to member states on the Council of Europe probation rules, (adopted by the Committee of Ministers 20 January 2010). Available at: http://snipurl.com/261zfak (accessed 20 September 2018)

Darchuk A, Wang V, Weibel D, Fende J, Anderson T and Horvath, AO (2000). *Manual for the Working Alliance Inventory—Observer Form, 4th revision.* Unpublished Manuscript.

Fox C and Marsh C (2016) Operationalising desistance through personalisation. European Journal of Probation 8(3): 185-206.

Fox C, Grimm R and Caldeira R (2017) *An introduction to Evaluation.* London. Sage

Interserve (2016a) Development of research instruments. Interserve 13/05/16.

Interserve, (2016b). Guidance on risk and assessment approval criteria. Interserve.

Interserve (2017) Enablers of Change tester tool version 2. Interserve.

Kewley S and Blandford M (2017) The development of the active risk management system. *Journal of Criminal Psychology* 7(3): 155-167.

McNeill F (2006) A desistance paradigm for offender management. *Criminology and Criminal Justice* 6(1): 39-62.

McNeil F (2009) *Towards Effective Practice in Offender Supervision*. Glasgow: The Scottish Centre for Crime and Justice Research.

McNeill F and Weaver B (2010) Changing lives? Desistance Research and Offender Management. Available at: http://www.sccjr.ac.uk/publications/changing-lives-desistance-research-and-offender-management/ (accessed 29 September 2018)

McNeill F (2016) Desistance and criminal justice in Scotland. In: Croall H, Mooney G and Munro R (eds) *Crime, Justice and Society in Scotland*. London: Routledge, pp.200–216.

Ministry of Justice (2013) *Transforming Rehabilitation: A strategy for reform*. London. Ministry of Justice.

Ministry of Justice (2018) Justice Secretary outlines future vision for probation. London. Ministry of Justice. Available at: https://www.gov.uk/government/news/justice-secretary-outlines-future-vision-for-probation (accessed 2 October 2018)

Moore R (2015) *A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013*. London: Ministry of Justice Analytical Series, 2015.

Morton S (2009) *Can OASys deliver consistent assessments of offenders? Results from the inter-rater reliability study*, Research Summary 1/09. London: Ministry of Justice.

Moxey LM and Sanford AJ (1993) *Communicating Quantities: A Psychological Perspective*. Hillsdale: Erlbaum.

Nicholls CM and Webster S (2014) *Sex offender management and dynamic risk: Pilot evaluation of the Active Risk Management System (ARMS)*. London: Ministry of Justice Analytical Series.

NOMS (2014). *NOMS Commissioning Intentions from 2014*. London: National Offender Management Service.

Raynor P (2018) From 'nothing works' to 'post-truth': The rise and fall of evidence in British probation European Journal of Probation 10: 59 – 75.

Ritchie J and Spencer L (1994) Qualitative data analysis for applied policy research. In: Bryman A and Burgess RG [eds] *Analysing Qualitative Data.* London: Routledge, pp.173-194.

Rohrmann B (1978) Empirische Studien zur Entwicklung von Antwortskalen fuer die sozialwissenschaftliche Forschung. *Zeitschrift fuer Sozialpsychologie*, 222-245.

Savolainen J (2009) Work, family and criminal desistance. Adult social bonds in a Nordic welfare state. *British Journal of Criminology* 49: 285-304.

Senior P (2013) Probation: Peering through the uncertainty. British Journal of Community Justice 11 (2-3): 1-8

Skardhamar T and Savolinen J (2014) Changes in criminal offending around the time of job entry: a study of employment and desistance. *Criminology*, 52(2): 263-291.

Stame N (2004) Theory based evaluation and types of complexity. *Evaluation* 10(1): 58-76

Stein D and Valters C (2012) Understanding theory of change in international development. *The Justice and Research Security programme*, Paper 1.

Van de Vijver F (2001) The evolution of cross-cultural research methods. In Matsumoto D(eds) *The Handbook of Culture and Psychology,* New York: Oxford University Press, pp. 77-97.

Vogel I (2012) *Review of the use of "Theory of Change" in international development*. DFID, April 2012.

Ward T and Maruna S (2007) *Rehabilitation*. Oxon, UK. Routledge.

Weiss C (1995) Nothing as practical as good theory, exploring theory-based evaluation for comprehensive community initiatives for children and families. In Connell JP Kubisch AC Schorr LB and Weiss CH (eds) *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. The Aspen Institute, Washington DC.

Willis GB (2005) *Cognitive Interviewing: A Tool for Improving Questionnaire Design.* Thousand Oaks, CA: Sage Publications.

Youth Justice Board (2014) *AssetPlus Model document*. London. Youth Justice Board

---

[i] CRC bidders were required to offer up innovations in offender rehabilitation as part of their proposals to the MoJ

[ii] With high risk of harm offenders managed by the National Probation Service (MoJ 2013)

[iii] NOMS was renamed Her Majesty's Prison and Probation Service (HMPPS) in February 2017,

[iv] Assessment, Sentence Planning, Implementation, Review and Evaluation

[v] Sections dentifying the dynamic risk factors to be addressed in order to reduce the risk of re-offending

[vi] Following the TR changes, OASys has continued to be used by the NPS and by CRCs unless the CRCs have received approval by the MoJ to use an alternative assessment tool.

[vii] This was the case at the time of the first and second phase evaluations. It may change during the next phase of testing.

[viii] Some CMs were observed on multiple occasions. If the CM had been interviewed previously, interviews were only conducted if there was new information.

[ix] Under the TR changes, CRCs manager low to medium risk of harm offenders and the NPS high risk of harm offenders. Where risk of harm changes, cases are transferred between CRCs and NPS.

[x] Specific, Measurable, Achievable, Realistic and Timed