HM Prison & Probation Service

Developing a toolkit to measure intermediate outcomes to reduce reoffending from arts and mentoring interventions

Mike Maguire, Emma Disley, Mark Liddle, and Rosie Meek with Nina Burrowes

Ministry of Justice Analytical Series
2019

Preventing victims by changing lives

Her Majesty's Prison and Probation Service is committed to evidence-based practice informed by high-quality social research and statistical analysis. We aim to contribute to the informed debate on effective practice with the people in our care in prisons, probation and youth custody.

Disclaimer

The views expressed are those of the authors and are not necessarily shared by the Ministry of Justice (nor do they represent government policy).

First published 2019

© Crown copyright 2019

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to us at National.Research@noms.gsi.gov.uk

This publication is available for download at https://www.gov.uk/government/publications

ISBN 978-1-84099-820-7

Acknowledgements

The authors would like to thank all those who completed questionnaires at HMP and YOI Portland and HMP Moorland, or as part of their involvement with participating arts and mentoring projects across the country.

We are also grateful to a wide range of representatives from service provider organisations across the country, national networks, and other research organisations (and to key professionals working in the field) who participated in the consultation exercise and shared ideas with us at key stages of the research.

We are grateful to members of our Expert Panel – Shadd Maruna, Fergus McNeill, Chris Lewis, Peter Raynor, and Graham Beech – for providing critical and focused feedback throughout the work.

We would like to thank Judy Renshaw for her contributions to the development of the costing material, and Gwen Lewis from Royal Holloway, University of London who assisted in the statistical analysis of results from piloting the instrument. We owe thanks to members of the project Steering Committee, for providing useful feedback on draft reports, and during discussions at key meetings as the work progressed

We would also like to thank Anastasia Jablonska from Royal Holloway, University of London for assistance with data entry, and colleagues at New Philanthropy Capital for their involvement in joint events and discussions during the work.

Thanks to Kristy Kruithof, Joanna Miler and Jirka Taylor from RAND Europe and Adrian Button, Viv Francis and Lyndsay McAteer from ARCS, who provided vital support to the project and in preparing this report.

The authors

Mike Maguire is a Professor of Criminology at the University of South Wales. Emma Disley is an Associate Director at RAND Europe. Mark Liddle is Managing Director, ARCS LTD. Rosie Meek is a Professor of Forensic Psychology at Royal Holloway University of London. Nina Burrowes is a psychologists and independent researcher.

Contents

List of tables

List of figures

1. Summary **1**
- 1.1 Aims and background 1
- 1.2 Overview of the research process 2
- 1.3 Key findings 3

2. Introduction **5**
- 2.1 Aims of this study 5
- 2.2 Purpose and structure of this report 6
- 2.3 Potential uses of intermediate outcome measures 6
- 2.4 The predictive value of psychometrics with offenders 9
- 2.5 Arts and mentoring programmes with offenders 9
- 2.6 Limitations and caveats 11

3. Developing the Intermediate Outcomes Measurement Instrument **13**
- 3.1 Stage 1: Assessment of existing literature 13
- 3.2 Stage 2: Consultation with providers 15
- 3.3 Stage 3: Selecting intermediate outcomes for inclusion in an instrument 17
- 3.4 Stage 4: Designing questions for each intermediate outcome 20
- 3.5 Stage 5: Piloting the instrument with prisoners 22
- 3.6 Stage 6: Piloting in the community with providers 26
- 3.7 Stage 7: Test-retest with prisoners 30

4. Developing the guidance, costing tool and data entry tool **33**
- 4.1 Guidance document 33
- 4.2 Data entry tool 33
- 4.3 Costing tool 34

5. Summary of findings and implications **37**
- 5.1 Providers' interest in and capacity for measurement 37
- 5.2 'User-friendliness' of the instrument 38
- 5.3 Reliability and validity 38
- 5.4 Future testing and piloting 39
- 5.5 Policy and practice implications 40

References **43**

List of tables

Table 3.1: Intermediate outcomes identified during literature review – stage 1 14

Table 3.2: Dimensions included in the IOMI – stage 3 19

Table 3.3: Reliability coefficients – stage 5 piloting with prisoners 25

Table 3.4: Pearson correlations between scales – stage 5 piloting with prisoners 25

Table 3.5: Respondents completing the IOMI at T1 and T2 – stage 6 piloting with providers 29

Table 3.6: T1 and T2 reliabilities – stage 6 piloting with providers 29

Table 3.7: Summary of univariate MANOVA results – stage 6 piloting with providers 30

Table.3.8: Test re-test analysis results - stage 7 32

List of figures

Figure 3.1: Overview of research process 14

Figure 3.2: Overview of the IOMI 26

1. Summary

1.1 Aims and background

The aim of the research was to develop a framework for measuring the outcomes of mentoring or arts interventions for offenders. The focus was on 'intermediate outcomes', defined here as measurable changes in individuals that are directly or indirectly associated with reductions in reoffending. Such outcomes may also produce other social and individual benefits, independent of any demonstrable effect on offending.

The main objective was to design and develop a robust but user-friendly instrument to measure changes of the above kind. A theoretically informed 29-item questionnaire, named the Intermediate Outcomes Measurement Instrument (IOMI), was produced through an iterative process of literature searches, consultation, piloting and analysis.[1] IOMI has not undergone full validity and reliability testing, but the results of initial testing have been encouraging.

The IOMI forms part of a broader toolkit for service providers to assist them in evaluating their work. The toolkit includes: guidance on the administration of IOMI and how it could be used as part of wider evaluations; a data entry tool to facilitate the collection and analysis of IOMI data; and a tool for the collection and presentation of cost data.[2]

While the IOMI was developed for use in arts or mentoring interventions with adult male offenders, it is likely that it will have wider applicability to other intervention or groups. Its main purpose is to provide evidence of offenders' progress towards desistance, which is not necessarily reflected in standard reoffending rates. This could be particularly valuable for capturing any contribution to desistance made by individual interventions which are not expected to reduce reoffending on their own.

---

1 See Appendix A of the Technical Appendices, published alongside this report, for a copy of the questionnaire. The questionnaire is also available as a standalone document for download on gov.uk.

2 The guidance and costing tool are published alongside this report on gov.uk. The data entry tool is available on request from research@arcs-ltd.com.

1.2 Overview of the research process

Developing the Instrument

The development of the IOMI was undertaken in seven main stages, as follows:

1. Rapid evidence assessments (REAs) based on searches of existing literature on the effectiveness of arts and mentoring interventions.
2. Consultation with a range of providers and with academic experts.
3. Based on the above, selection of a candidate list of psychological dimensions in which arts and mentoring interventions aim to bring about change in their clients.
4. Creation, adoption or adaptation of possible questions for inclusion in a questionnaire to assess, and capture changes in, individuals in relation to each dimension.
5. Piloting of a 51-question instrument with 264 prisoners, analysis of the results, and subsequent reduction and revision of the instrument to a 29-item questionnaire (the IOMI) covering seven psychological dimensions (personal agency, wellbeing, impulsivity, interpersonal trust, motivation to change, hope, resilience) and one reflecting practical problems.
6. Piloting of the IOMI with community-based providers to begin to test its capacity to measure change, which involved completions of IOMI by 62 offenders both before and after participation in specific arts and mentoring interventions.
7. A test-retest exercise to confirm the stability of responses to the IOMI, in which 217 prisoners completed the instrument twice within a short period (8-21 days).

Developing the costing tool

The research team also developed a simple costing tool which could:

- Allow providers to estimate the costs of what they deliver
- Allow providers to calculate unit costs in a number of different ways
- Be used by providers to gather and present cost information in a transparent and consistent manner.

While not sufficient on its own to enable providers to undertake cost-effectiveness analysis, this tool may facilitate comparisons across areas and provide information useful for commissioners.

The tool is made up of four spreadsheets, three of which can be used by providers to input information, and one which is automatically populated using this information. The summary

sheet uses totals from the other sheets to calculate weighted annual costs by year, unit costs (by type of participant and by year), and an overall unit cost (a weighted average).

Guidance document and data entry tool

Another important part of the toolkit is the guidance document, the key aims of which are to provide organisations with:

- An understanding of the background to the development of the instrument and of the theory and evidence underpinning it.
- Information about the reliability and validity of the instrument.
- Information about the key dimensions being measured, and how they can be understood in relation to their own theories of change.
- Instructions for administration, data entry, scoring and reporting results.
- Broad advice about how data generated by the instrument might be understood and could be used as part of wider evaluation work.
- Awareness of issues concerning use of the instrument with particular sub-groups (for example, young people, participants at women-only projects, etc.).
- Guidance on ways of linking findings from the tool to a range of other kinds of data in order to validate or challenge the picture emerging from it.

In addition, the research team produced a data entry tool for the IOMI in database format. This allows the user to enter responses onto an on-screen version of the instrument. The package converts each response (such as ‘Strongly agree’ or ‘Disagree’) into a raw score, and groups and sums these by dimension (such as hope, or resilience). It then calculates a score for each dimension, which allows comparisons to be made across dimensions and reports to be generated on individual clients or on particular dimensions in aggregate.

1.3 Key findings

The preliminary research for the project found strong interest among providers in a tool of this kind, with providers admitting that they relied too much on anecdotal evidence when attempting to convince commissioners of the effectiveness of their work. At the same time, some resistance was encountered, particularly among front-line staff and volunteers, to the idea of introducing ‘measurement tools’. This was largely because of concerns that voluntary agencies often lack the capacity to collect and analyse significant amounts of data. So, while there was clearly an appetite for measurement, there was also a demand for simplicity and succinctness in any evaluation tool.

The toolkit produced (made up of the IOMI, the guidance document, the data entry tool and the costing tool) provides a user-friendly package which can be used by providers of arts and mentoring projects (and potentially many other kinds of intervention) to assess the impact of their own work in a relatively quick and straightforward way.

The IOMI itself is theoretically informed and is anchored in careful reviews of evidence from research literature, as well as extensive consultation with providers of mentoring and arts programmes. The results of all three stages of testing were highly encouraging. There were strong indications that the instrument has face validity, internal consistency, stability and the potential to reflect change across most dimensions. At the end of the research period, it was concluded that the IOMI is in a form that is likely to require little further amendment, apart from possible changes or additions to the questions in the 'resilience' dimension.[3]

It is however emphasised that an instrument of this kind inevitably requires extensive testing over longer periods of time (including the collection and analysis of reconviction data on those who have completed the questionnaires) before confident claims can be made for its validity and reliability as a measure of psychological change, or for its robustness as a proxy for reoffending.

Policy and practice implications

Until further testing is done, any results based on the instrument should be treated with caution, and should not be used to compare the effectiveness or performance of different providers. That said, the instrument has considerable potential as a tool for a number of policy and practice related purposes. These include:

- Alongside other evidence, helping bidders to make a case to commissioners for the effectiveness of particular interventions or practices.
- Helping organisations improve their services by finding out more about the outcomes that they are producing for clients.
- Offering an extra resource for research into what works in offender rehabilitation, especially into the effectiveness of small-scale and less mainstream interventions such as arts programmes.

[3] This dimension produced lower Alpha coefficients and appeared to be less sensitive to change than the other dimensions – see Chapter 3 for further details.

2. Introduction

2.1 Aims of this study

This report presents the findings of a research project, commissioned by Her Majesty's Prison and Probation Service (HMPPS, formerly the National Offender Management Service), which set out to develop a framework for outcome measurement which can be adopted by organisations delivering mentoring or arts interventions to adult offenders.

The focus of the study was on 'intermediate outcomes', defined here as measurable changes in individuals that are directly or indirectly associated with reductions in reoffending. These are 'intermediate' in the sense that they indicate that an offender is making positive changes towards an offence-free future, but has not necessarily at this stage stopped offending. Of course, such outcomes may produce other social and individual benefits, independent of any effect on offending.

The main objective of the research was to design and develop a robust but user-friendly instrument which could be used to measure changes of the sort referred to above. The Intermediate Outcomes Measurement Instrument (IOMI), a 29-item questionnaire, was produced through an iterative process of literature searches, consultation, piloting and analysis. A copy of the questionnaire is included in Appendix A of the Technical Appendices, published alongside this report; the questionnaire is also available as a standalone document for download on gov.uk. The instrument has not undergone full validity and reliability testing, and may need to be refined further. However, the initial stages of testing have been encouraging.

A second aim was to incorporate the instrument within a broader toolkit for service providers, to assist them in evaluating their work. The toolkit includes:

- A document which describes the IOMI instrument and gives guidance on how it should be administered and how it could be used in a wider evaluation. This guidance document is published alongside this report on gov.uk.
- A data entry tool to facilitate the collection and analysis of data from the IOMI instrument. This tool is available on request.[4]
- A tool for the collection and presentation of cost data, to assist organisations in measuring the cost of delivering their programmes, which can feed into wider cost-

[4] To access the data entry tool contact research@arcs-ltd.com for a link and login details.

benefit or cost-effectiveness assessments The costing tool is published alongside this report on gov.uk.

The IOMI instrument has been developed for use in arts or mentoring interventions with male offenders. However, it is likely to have wider applicability to other types of intervention and/or other groups. Issues related to future development of the instrument are raised toward the end of this report.

2.2 Purpose and structure of this report

This report has three main purposes:

1. It explains what the research team did at each stage of the process of developing and testing the IOMI instrument.
2. It presents the instrument and the other components of the toolkit and explains how they can be used by providers.
3. It sets out steps which could be taken to further test and validate the instrument.

The rest of this chapter outlines the main reasons for developing a toolkit and the uses to which it might be put. Some comments are also made about the use of arts projects and mentoring in rehabilitative work with offenders. The chapter ends with a brief overview of the methodology.

Chapter 3 describes how the instrument was designed, refined and piloted. Chapter 4 outlines the development of the guidance document, costing tool and data entry tool. Chapter 5 summarises the study, discusses the strengths and limitations of the instrument, describes progress to date in testing its validity and reliability, and sets out further steps needed to demonstrate these robustly.

2.3 Potential uses of intermediate outcome measures

Evidencing journeys towards desistance

The effectiveness of services to offenders is usually judged on the basis of rates of reoffending. How these are defined and labelled (for example, reconviction rate, proven reoffending rate) and what length of follow-up period is used have been subject to considerable variation. However, the basic concept – a measure of how many offenders in a

particular cohort have and have not reoffended within a set period[5] – has remained at the heart of official performance measurement for many years.

While such measures may be sufficient for many administrative and research purposes, they are only a starting point in understanding whether and how a particular form of intervention is truly working. Research into desistance from offending shows that, especially for offenders with deeply ingrained social and personal problems, desistance is often not a one-off event. It can be a long-term and difficult process of fundamental change in self-identity, values, and lifestyle, characterised by a zigzag pattern in which periods of abstention from offending are interrupted by relapses (Burnett and Maruna, 2004; Laub and Sampson, 2001; Maguire and Raynor, 2006). It is therefore important to seek evidence not only of 'primary' desistance (i.e. what may be only a temporary period of non-offending), but also evidence of whether offenders are making progress on a longer journey towards 'secondary' desistance (i.e. permanent change in both behaviour and self-identity, see Farrall and Calverley, 2006). Such progress might be detectable in: evidence of cognitive or attitudinal change; greater engagement with non-criminal groups or activities; amelioration of problems such as substance misuse; lack of suitable housing; or lack of skills and qualifications. Some of those who appear as failures in reoffending measures may actually have made considerable progress in other respects, while some of the successes may have changed little and may revert to offending in the longer term. Hence, instruments such as IOMI offer an alternative or an addition to short-term reoffending rates as indicators of offenders' progress towards desistance.

Possible proxy for reoffending rates

As well as providing an alternative measure of progress towards desistance, such instruments can potentially be used as proxies for reoffending rates. In other words, if a reduction in one-year reoffending rates is regarded as the target outcome of a particular intervention, it is very useful to have an instrument that can act as an early predictor of such a reduction rather than having to wait for the actual rates to emerge. IOMI's capacity to predict reoffending has not been tested. However, as there are theoretical and evidence-based links between its various dimensions and reoffending, there is a good possibility that positive change measured by IOMI would be significantly correlated with reduced reoffending, hence making the instrument suitable as a proxy measure.

[5] Of course, 'real' reoffending rates are unknowable, because many offences remain unreported or unrecorded and many perpetrators undetected. So, all reoffending measures are inevitably proxies.

Capturing individual interventions' contribution to desistance

Another important consideration is that few interventions are likely to bring about significant reductions in reoffending rates on their own. Rather, effective rehabilitation is most likely to be achieved by holistic approaches in which a number of interventions are delivered in a coordinated fashion, often by more than one provider.[6] In this situation, it is difficult to single out any individual component of the work undertaken as having been responsible (or not responsible) for reduced reoffending. Indeed, some services are designed as platforms on which others can build, rather than claiming to address offending behaviour themselves. For example, assisting offenders to find suitable accommodation has been described as a 'necessary but not sufficient' condition for rehabilitation (Humphreys and Stirling, 2008; Maguire and Nolan, 2007), and some mentoring services are aimed primarily at boosting prison leavers' confidence and motivation and helping them to access other services (Clancy et al., 2006; Maguire et al., 2010).

In this context, a before and after measure of a particular form of change which the provider seeks to bring about – preferably articulated in an explicit 'model of change' – provides a useful indication of the extent to which they have achieved their goal. So, if a mentoring service aims to enhance participants' motivation, or an arts project aims to increase their resilience, the extent to which these goals have been met might be measured through the motivation and resilience dimensions of IOMI.[7]

Evidence of this kind is important for a variety of reasons. It can help agencies understand the impact of their services and seek improvements where appropriate. If positive, it can enhance the morale of staff, volunteers and clients. It can also be valuable in bidding for funds, alongside other evidence, to help convince commissioners of the agency's effectiveness.

Contributing to the evidence base

A good intermediate outcome measure can contribute to the long term agenda of accumulating more reliable evidence about what works in offender rehabilitation. Despite the large amounts of international evidence about the effectiveness of some kinds of intervention

---

[6] There is also evidence that the coordinating functions of probation supervisors, combined with skills in helping offenders sustain motivation, can be key factors in effecting reduced reoffending or other positive change (Raynor et al., 2014; Sapouna et al., 2011; Shapland et al., 2012).

[7] This leaves open the question of whether short-term changes in resilience or motivation (or other selected 'intermediate outcomes') are associated with longer term changes in offending behaviour – a question that can only be answered with full confidence through long-term empirical research.

(notably structured programmes based on cognitive behavioural principles – see Lösel, 2012), there are others, including arts projects and mentoring, about which there is relatively little reliable empirical evidence (Burrowes et al., 2013; Taylor et al., 2013). The same is true of some important and popular theoretical approaches, such as the Good Lives Model[8] which has not been subject to large amounts of rigorous empirical testing. An instrument like the IOMI, used in the evaluation of several different interventions and service providers, could contribute to better evidence of the effects of interventions, especially small-scale and less mainstream interventions such as arts programmes, and could start to provide results which allowed comparisons to be made across programmes.

2.4 The predictive value of psychometrics with offenders

There is a large body of evidence showing that psychometric instruments can have some level of predictive validity in offender populations with regards to recidivism. However, the majority of these studies are concerned with sexual offenders and sexual recidivism (for example, Allan et al., 2007; Beech and Ford, 2006; Beech et al., 2002; Beech, 1998; Hanson, 1998) and often static factors have been found to be better predictors than psychometric indicators (for example, Wakeling & Barnett, 2014).

Generally, pre-treatment psychometric test scores seem more predictive than post-treatment scores[9] (Barnett et al., 2012). Positive change between scores on psychometric tests conducted before and after treatment has also been shown by some studies to be significantly correlated with reduced recidivism (for example, Hudson et al., 2002; Raynor, 1998), although the evidence here is more limited.

2.5 Arts and mentoring programmes with offenders

Arts projects are delivered to a wide range of offenders in the community throughout the prison estate and in secure mental health facilities. The types of activities delivered include theatre projects, dance, painting, drawing, creative writing, crafts, film and music. They may be delivered explicitly as rehabilitative, therapeutic or educational interventions, or purely as a creative activity. Typically, arts projects are delivered by VCSE (Voluntary, Community and

---

8 A useful and comprehensive list of publications focusing on the Good Lives Model can be found at: http://www.goodlivesmodel.com/publications.

9 This is potentially due to them being a purer measure of dysfunction than post-treatment scores. When psychometric tests are directly fed into an offender's risk reports and affect progression through the system, it may be that the results are negatively impacted by demand characteristics (desirable responding) resulting in a less direct relationship between scores and the constructs the measures are trying to capture and so diminished validity of post-treatment scores.

Social Enterprise) organisations, but also can be delivered either solely by or in collaboration with the public and private sectors.

Mentoring interventions usually involve a one-to-one relationship where the mentor provides support and assistance and acts as a role model for the mentee. While there may be an imbalance of experience or expertise between them, the aim is to develop a relationship that is anchored in mutual respect and trust. Mentoring also generally implies the provision of support that is ongoing rather than one-off. However, mentoring schemes vary considerably in their design and delivery, particularly in terms of who performs the mentoring function (a professional or a volunteer),[10] where it takes place (in prison, in the community or 'through-the-gate') and how it relates to other interventions (a standalone service or as part of a package of interventions). Mentoring interventions are often client-led, meaning that the objectives, ways of working and length of engagement are tailored to clients' expressed needs, so what is delivered differs significantly on an individual basis. As with arts projects, the sheer variety in the nature and intensity of mentoring poses major challenges for attempts to measure outcomes which may be attributed to it.

Can arts and mentoring programmes play a role in desistance from crime?

In the initial stages of this project, two rapid evidence assessments (REAs) were conducted to examine the evidence on arts and mentoring interventions (Burrowes et al., 2013, Taylor et al., 2013).[11] It was found that there is insufficient evidence to determine whether arts projects have a measurable impact on reoffending. Arts work with offenders has received only limited research attention,[12] and evaluations which have been conducted have had small sample sizes and looked mainly at short-term impacts. However, the REA illustrated ways in which arts projects may play an important part in the process of desistance. For example, they may provide opportunities for positive interaction with others or an outlet for difficult experiences and emotions. They may also help to foster a sense of achievement or new ways of seeing oneself and others, in some cases motivating offenders to engage with other services.

---

[10] One distinctive approach growing in prominence is peer mentoring, where the mentors have shared histories with the mentees (for example, having previously served a prison sentence or been addicted to drugs). It should be noted that some agencies do not use the term 'peer mentor', preferring 'peer advisor'.

[11] A rapid evidence assessment is a method of reviewing evidence which provides a balanced assessment of what is already known about a policy or practice issue, by using systematic review methods to search and critically appraise existing research. They aim to be rigorous and explicit in method and so systematic, but make concessions to the breadth or depth of the process by limiting particular aspects of the systematic review process.

[12] Although this has been increasing. For example, see Bilby et al., 2013; Cox and Gelsthorpe, 2008.

Although mentoring has received more research attention, there is similarly a shortage of good-quality research evidence on its impact. Available studies indicate that some kinds of mentoring may influence reoffending through: helping offenders maintain motivation to change, providing friendship and support, and supporting access to other services, especially 'through the gate' and after release from prison (see, for example, Clancy et al., 2006; Maguire et al., 2010).

It is not realistic to expect short-term single interventions, such as those commonly provided by arts or mentoring projects, to produce significant reductions in reoffending on their own. The process of desistance from crime is a long-term and complex process, requiring sustained motivation on the part of offenders and continuing support on a variety of fronts (Maruna, 2001; McNeill, 2006). On the other hand, such interventions may promote changes in individuals that make significant contributions to their progress towards desistance.

2.6 Limitations and caveats

Further validity and reliability testing is needed: As will be shown in this report, the IOMI instrument seems promising from the analysis undertaken, but it would need to be tested much more thoroughly before its validity and reliability in measuring psychological states and their change over time can be confidently asserted.

Link with reduced reoffending should be tested: The results of IOMI should be mapped against reoffending outcomes to assess its value as a predictor of, or proxy for, reoffending.

Relative stability of constructs and scores should be confirmed: It important to establish that the psychological constructs being measured are relatively stable in the short term and that the scores the instrument produces for each dimension measured (such as hope, or resilience) are not highly volatile over short periods. For example, this could be in relation to the circumstances in which it is administered or to rapid fluctuations in offenders' moods. Without such stability, increases or decreases in scores obtained before and after an intervention might be capturing only temporary states of mind (or peaks or troughs) which do not reflect the overall trajectory of change.[13] This can be investigated through a series of 'test-retest' exercises (building on the initial test-retest carried out during this study – see Section 3.7), conducted over different lengths of time, and in some cases retesting

[13] It is also important to investigate whether the instrument measures some outcomes more robustly than others. One might expect some psychological constructs or personality traits – for example, impulsivity - to be deeper rooted and slower to change than some states of mind such as optimism or levels of motivation, which may be more vulnerable to temporary mood swings. However, this needs to be tested properly.

individuals several times. The problem of distinguishing temporary from lasting change is not unique to this particular instrument. If one accepts that desistance from crime is typically a difficult and lengthy process peppered by setbacks and relapses (Farrall and Calverley, 2006; Maguire and Raynor, 2006; Weaver and McNeill, 2010), measurement of any kind of change in offenders faces similar challenges. This suggests that instruments such as IOMI may be most useful if administered periodically to the same individual over a substantial period of time, thereby producing longer-term trend lines (which, ideally, might be related to significant interventions or events in their lives). Again, this is an issue for further exploration.

3. Developing the Intermediate Outcomes Measurement Instrument

This chapter explains the process of developing and testing the Intermediate Outcomes Measurement Instrument (IOMI). The approach taken was motivated by the need to ensure that:

- The instrument drew on and was in accordance with theory and/or evidence linking the intermediate outcomes measured in the tool with reduced reoffending.
- The intermediate outcomes selected had conceptual clarity, and preferably were well-established and defined psychological constructs.
- Questions in the instrument which were intended to measure a particular outcome would be likely to do so.
- The instrument and the toolkit as a whole were user-friendly.

As set out in Figure 3.1, the development of the toolkit was undertaken in seven main stages. It began with reviews of available evidence and progressed through consultations with providers, the creation of a long list of possible questions for inclusion in the instrument, and the piloting and testing of these questions in prison and community settings. It is emphasised again that there remains much more research and testing that should be done before the instrument can be claimed to have been fully validated. The necessary future steps to do so are outlined in the final chapter.

3.1 Stage 1: Assessment of existing literature

Two rapid evidence assessments (REAs) were conducted to identify possible intermediate outcomes from arts and mentoring programmes which could be considered for inclusion in an instrument. The results of the REAs (including how the quality of available studies was assessed and a discussion of the limitations of the evidence base) can be found in Burrowes et al. (2013) and Taylor et al. (2013).[14] As noted earlier, little high-quality or conclusive evidence was found concerning the effectiveness of either arts or mentoring projects with offenders, although research into mentoring is more extensive. The sheer variety of both

[14] Sixteen studies were included in the review related to arts interventions. Two of these used randomised experimental design, six used comparison group experimental design, seven explored pre and post project scores for a treatment group only, and one was a post project qualitative study. Twenty-two studies were included in the review related to mentoring interventions, of which nine scored three or above on the Maryland Scale of Scientific Methods (Sherman et al., 1997). The Maryland Scale sets out five levels of methodological rigour, beginning with level 1 (the lowest standard, applicable to studies that focus only on correlations between programmes and particular measures at one point in time), and ending with level 5 (the highest standard, reserved for studies that involve random assignment of programme and control conditions to units).

mentoring and arts projects, combined with the generally low numbers of participants, makes outcome research in this area particularly difficult. Nevertheless, for mentoring, six possible intermediate outcomes were identified in the review, and three were identified for the arts (see Table 3.1).

Figure 3.1: Overview of research process

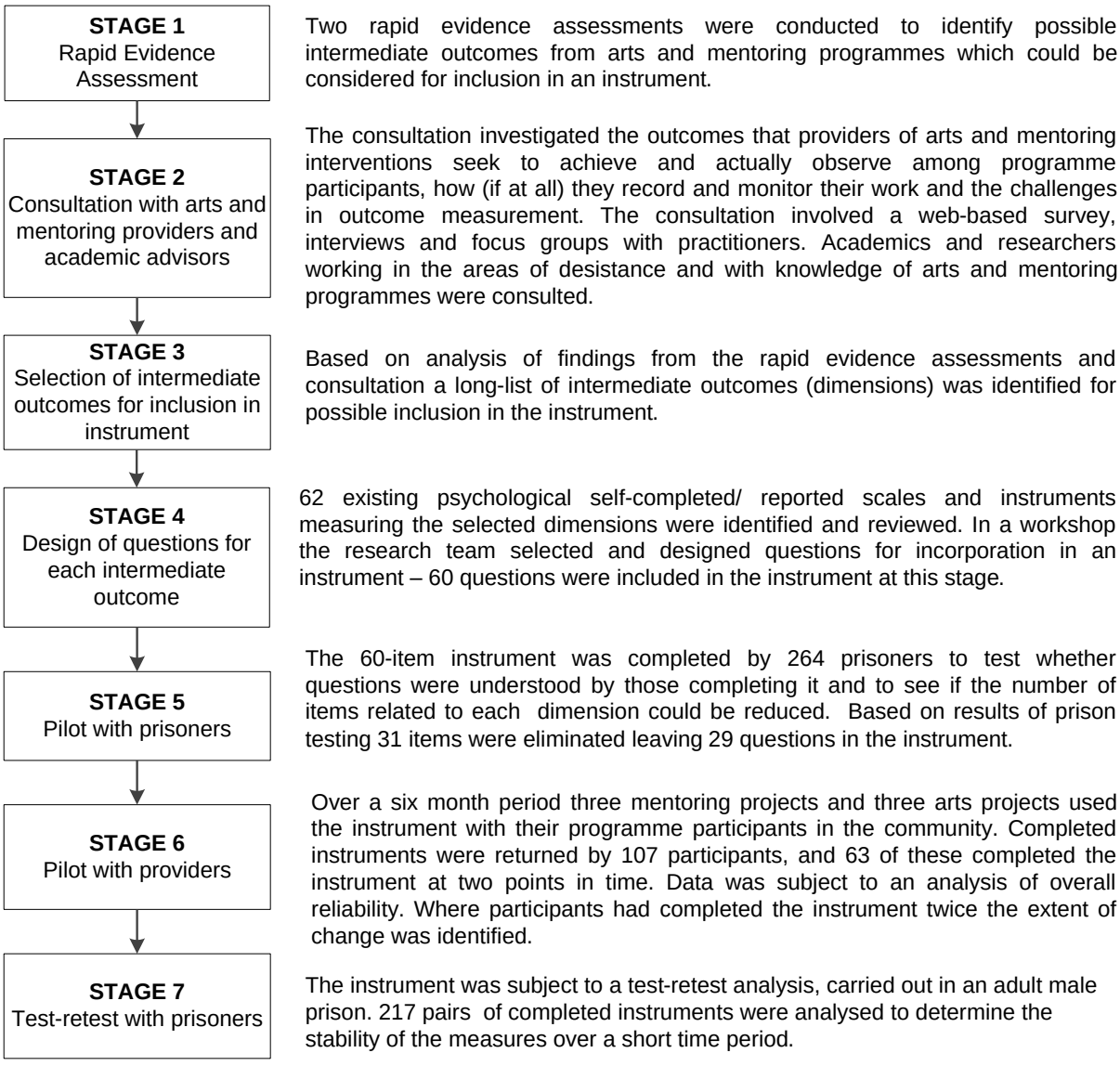


Table 3.1: Intermediate outcomes identified during literature review – stage 1

Mentoring	Arts
Employment outcomes	In-prison behaviour
Engagement in programmes and interventions	Individual psychological outcomes (such as optimism or sense of purpose)
Housing outcomes	Educational outcomes
Health outcomes	
Attitudinal, cognitive or motivational change	
Family and community relationships	

3.2 Stage 2: Consultation with providers

The second strand of work to identify candidates for intermediate outcomes involved consultation with provider organisations, conducted using three methods:

- Interviews with providers: 27 semi-structured interviews[15] – 10 with representatives from organisations delivering arts interventions, and 17 from organisations delivering mentoring work.
- Online survey: two surveys (one for mentoring and one for arts projects) were administered to generate feedback from a wider range of providers. The mentoring survey generated 54 responses, but the arts survey resulted in only five completed questionnaires.[16] This low response rate was disappointing, however the qualitative feedback from providers was useful, so the data was included in the analysis but analysed separately from data from the mentoring survey.
- Focus groups: seven focus groups were held with representatives from arts and mentoring projects. Some focus groups involved representatives from a single provider, and some involved representatives from a range of providers. The groups ranged from two participants to over ten.

The consultation exercises addressed three questions, and findings in relation to each are set out below.

What kinds of change do arts and mentoring programmes aim to bring about?

There was considerable overlap between the themes and theories identified in the REAs and those brought up by the interviewees, including the models of change discussed and ideas about the relationship between intermediate outcomes and desistance. Importantly, there were major overlaps between providers from both arts and mentoring projects in terms of specific outcomes they identified as likely to be produced by the two interventions. This strongly influenced the decision to produce just one instrument to measure the outcomes of both kinds of project (see Section 3.3). The main outcomes identified by interviewees in the consultation were:

- Practical problems: accommodation; substance misuse; health; employment; relationships; literacy and numeracy.

[15] The interviewees were selected on an opportunistic basis (see Appendix E), so clearly cannot be considered representative of all providers in the field. A much larger number of informal discussions were held with representatives from arts and mentoring organisations during this initial phase of work, which were not formal interviews but helped the research team familiarise themselves with the field.

[16] This was despite follow-up by email and telephone in some cases.

- Skills: communication; time management; problem solving; team work; creativity; money management.
- Psychological changes: self-efficacy; self-concept/ self-identity; self-esteem; wellbeing; hope; confidence; insight/ awareness; locus of control.
- Engagement with the intervention or programme and compliance with the programme.

Can (and should) these results and changes be captured or measured?

There was a strong interest in outcome measurement. A minority of participants (mainly among arts providers) were of the view that some of the changes which their intervention sought to achieve, such as increases in levels of hope, inspiration and confidence were intangible and could not be measured. There was some frustration with available approaches to outcome measurement, and a perceived lack of fit between these and what providers regarded as the key impacts of their work.

Providers reported already using a wide range of tools (both off-the-shelf and bespoke) and several had commissioned evaluations. In some cases providers were using tools for measurement that were not able to generate good quality data (some tools that had been designed in-house used overlapping categories for example, or sets of responses that were not jointly exhaustive).

What are the main barriers and facilitators to measurement?

Some interviewees had concerns that administering measurement instruments could have a negative impact on relationship building and engagement.

A number of practical barriers were identified, including:

- Time and resource constraints.
- The need to satisfy the reporting requirements of different funders.
- The small numbers of participants in some programmes.
- The length of participants' involvement in projects (both planned – some interventions only lasted a day – and unplanned – participants dropped out, were released from prison or moved to another prison).
- The difficulty of contacting participants for follow-up measurement once the programme had ended.
- Difficulties in accessing other key information (for example, about risk of reoffending) with which to correlate results.

Some concerns were expressed as to whether providers had the time, skills and resources to make sense of data generated from measurement. Some providers also stated that they had already accumulated significant datasets which had never been analysed.

3.3 Stage 3: Selecting intermediate outcomes for inclusion in an instrument

Having completed the evidence review and consultation, a long list of possible intermediate outcomes was identified, along with a sense of what would make an instrument attractive to providers (short, relatively quick to complete, easy to analyse and accessible to the client group). In workshops attended by all members of the research team (which included criminologists, psychologists and social scientists), each intermediate outcome that had been identified was discussed and assessed according to: whether it had been identified in both the evidence review and the consultation; whether there was an evidenced or theoretical link between the outcome in question and desistance from crime. The discursive and iterative nature of this selection process was important as it required research team members to clearly articulate a definition for each outcome, think about the mechanism through which arts and mentoring projects might bring about the outcome, and review the evidence linking each to desistance from crime. It was also an opportunity for the research team to identify possible gaps: outcomes which had not been identified in the REA or in the consultation, but which might be included on theoretical grounds.

As noted earlier, there was clearly a great deal of overlap in the kinds of outcomes which mentoring and arts programmes sought to achieve. This is perhaps not surprising since both aim to effect personal growth and development (albeit achieved through different mechanisms). For this reason it was decided to develop a single instrument to measure intermediate outcomes for both kinds of interventions, leaving open the possibility of adding extra items specifically for completion by arts or mentoring project clients (or indeed by those of other kinds of intervention) if appropriate.

As a result of the lengthy iterative process described above, eight dimensions (or 'domains' – we use the terms interchangeably) were eventually selected for inclusion in the instrument:[17]

- Resilience
- Agency and self-efficacy
- Hope

[17] Of the outcomes identified in the REAs, all were included in the instrument except for in-prison behaviour. This was principally because the instrument was designed for use in both community and prison settings.

- Wellbeing
- Motivation to change
- Impulsivity / problem-solving
- Interpersonal trust
- Practical problems.

Table 3.2 provides a description of each dimension included in the final instrument, including what poor scores and positive change might look like in relation to each. More details are provided in Appendix D.

Table 3.2: Dimensions included in the IOMI – stage 3

Dimension	Description	What poor scores mean	What positive changes might look like
Resilience	Capacity to recover from adversity, to move on in a positive manner or begin again. Related to individual coping skills and wider relationships and support networks.	More likely to give up in the face of setbacks ('what was I thinking – nothing will change for me'). More prone to depression.	Increased capacity to move on and continue to try, even in the face of setbacks and adversity.
Agency / self-efficacy	Whether one is able to make autonomous decisions about one's own life and to make things happen in the outside world as a result.	Passivity in relation to decision-making about one's own life ('things happen to me', rather than 'I make things happen). Prioritisation of luck or fate.	Increases in confidence in one's own ability to make decisions about the future, and to implement plans to bring about change.
Hope	A calculation about perceived scope for positive future change. Linked to motivation and self-assessments of efficacy.	A sense that the future is hopeless ('it is inevitable that things will not work out well for me; I should cut my losses and reduce my effort and commitment').	A new sense of hope – could be a catalyst for a number of other changes such as more flexible and positive perception of the future, internal motivation, and agency.
Wellbeing	General or overall mental/ emotional/ psychological health or balance. Linked to positive self-regard and confidence.	Low levels of positive self-regard, or self-esteem. Low levels of confidence.	Improvements in self-perception, estimations of self-worth. Increased levels of confidence.
Motivation to change	Linked to positive engagement. A key focus is on internal rather than external motivation.	Low levels of engagement with activities that may help with desistance (e.g. education, employment). Engagement only through external motivation. High levels of internal motivation to continue offending.	Shift from no motivation or external motivation to high levels of internal motivation. Increase in levels of engagement with interventions. Reduced motivation to continue with offending.
Impulsivity / problem-solving	Lack of reflection and planning and a disregard of the consequences of behaviour. People who are highly impulsive also generally lack problem-solving skills.	High levels of impulsive behaviour. Poor problem solving skills based on inaccurate perceptions, perceived limited range of options, no contingency planning.	Reduced impulsivity, increased ability to make conscious choices from a range of options. Increased planning and ability to think through options and consequences. Increase in focus and discipline – the ability to concentrate on one thing for a period of time.
Interpersonal trust	Positive attitudes toward and connectedness with others. Links to notions of social capital.	Other people are out to get me, dog eat dog attitude, people can't be trusted, no-one cares about me. Lack of interest in others and sense of being isolated.	Increase in positive attitude towards other people. Increase in connectedness (to a wider range of people, pro-social connections).
Practical problems	Extent to which respondents regard the key areas referred to as being problematic for them. The 8 areas listed are similar to the HMPPS pathways to rehabilitation.[18]	A multiplicity of areas in the individual's life appear to be problematic, to the extent that any efforts to engage the respondent in programmes or other interventions are seriously limited.	Reduction in or abstinence from drug use, controlled drinking or abstinence from drinking. Stable and secure accommodation. Improved health and access to health care. Employment or training which may lead to employment.

[18] Accommodation; Education, Training and Employment; Drugs and Alcohol; Finance, Benefits and Debt; Families and Children; and Attitudes, Thinking and Behaviour.

3.4 Stage 4: Designing questions for each intermediate outcome

Having identified eight dimensions for inclusion in the instrument, the next task was to identify questions which could be used to assess how respondents were positioned with respect to each dimension.

Question generation process

The research team reviewed 62 existing psychometric scales (these are listed in Appendix C) to identify a long list of possible questions for inclusion in the instrument. An expert evaluation was undertaken to select items most suitable for inclusion in the IOMI. Reasons for not including items from existing instruments included:

- That the wording was potentially confusing, for example 'there are people in my life who respect me for the steps I've taken to keep myself away from crime'.
- That they required a higher degree of literacy, for example 'thanks to my resourcefulness I know how to handle unforeseen situations'.
- That they could be interpreted in both a pro or anti-crime way, for example 'I can take charge of problems that require immediate attention'.
- That they did not relate sufficiently to the dimension of interest.

In some instances the research team modified questions into simpler language or made modifications to make the questions less categorical, for example, adding 'often' in 'I often do things without thinking of the consequences'. Some were re-worded to achieve a balance in the instrument between positively and negatively worded questions (for example, 'I feel as if nobody really understands me' became 'there are people who really understand me'). Where no questions were found in existing instruments that adequately captured the outcome of interest, the research team designed new questions.

Fifty-one questions were selected covering all the dimensions apart from 'practical problems'. For this dimension, eight questions were selected from the 15-item problem inventory of CRIME-PICS II.[19] CRIME-PICS is a widely used, validated questionnaire for examining, and detecting changes in, offenders' attitudes to offending, as well as in their perceived problems.

[19] The eight questions from CRIME-PICS II are third party copyright material and are reproduced with permission from M & A Research (Copyright M&A CRIME-PICS, 2013. All right reserved. No reproduction or use of permitted without the prior written consent of M&A Research).

Decision not to include practitioner judgements in the instrument

The instrument does not include any questions to be answered by practitioners (which might provide a different view on offenders' progress). This decision was taken primarily on the basis of findings from the consultation that staff have limited time, so requiring a practitioner to fill out an instrument for each participant would be too burdensome. Also, there is no guarantee that staff are able to make a more accurate or valid assessment of an offender's progress. There is a risk of positivity bias and the nature of the dimensions means that staff would have to know participants very well to make an assessment. Guidance on completing the IOMI advises projects to link the self-report data captured in the instrument to other data about participants – for example, information about changes in housing situation or drug use – to broaden the picture created through self-report alone.

Selecting Likert scales for responses to questions

As commonly used in psychometric instruments the IOMI uses a five-point Likert Scale for responses: strongly agree; agree; neither agree nor disagree; disagree; strongly disagree. This was selected following a review of literature about the strengths and weaknesses of possible response category options in general, and Likert scales in particular. These debates cover a range of issues including:

- The advantages and disadvantages of 5-point scales, which include a 'neutral' option such as 'neither agree nor disagree', and 4-point scales, which essentially force the respondent to choose between a negative or positive response) (Østerås et al., 2008).[20]
- Approaches to dealing with missing data when Likert scales are used, including the impact on data analysis and strategies for dealing with missing data.
- The advantages and disadvantages of different sets of scale categories (including the use of pictorial categories).

Questions about relationships with staff

The importance of the relationship between offenders and practitioners in the desistance process has been highlighted frequently in previous research (Burnett and McNeil, 2005) and was raised in the consultation. In order to capture the contribution of relationship building,

[20] A 5-point scale was selected because it can lead to greater variation in scores, which facilitates statistical analysis. A targeted review of the literature suggested that the middle term of the standard scale (neither agree nor disagree) is not as solid conceptually as the other terms, and that some respondents who do not want to take a clear stand on a particular question might opt for this answer (however, few studies were identified which quantified or explored the impact). It might not be possible to distinguish respondents who do not want to choose from those who have thought about their response and genuinely do not agree or disagree with the statement.

five items were developed by the research team after consulting experts in offender management and reviewing relevant studies (for example, Raynor et al., 2014; Shapland et al., 2012). The following were included as optional question, not considered as integral or essential to the instrument, which could be administered during or at the end of a programme to generate respondents' feedback on their perceptions of project staff:

- The staff here have treated me fairly
- The staff here have listened to me
- The staff here do what they say they'll do
- I feel able to trust the staff here
- The staff here have helped me to think differently about myself.

3.5 Stage 5: Piloting the instrument with prisoners

In the development of psychometric instruments there is always a balance to be struck between brevity (and so user-friendliness) and robustness. Many such instruments ask several questions about the same dimension, to demonstrate that any trait ascribed to an individual by the results is a consistent part of his or her personality. However, the messages coming from the field clearly put a very high premium on brevity and usability, and throughout the next stage of testing the research team bore firmly in mind the need to significantly reduce the number of items in the IOMI, without seriously compromising the validity of the instrument.

The IOMI was initially piloted in its 51-question form with sentenced young adult and adult males in an English prison, HMP Portland.[21] The instrument was circulated prison-wide to 521 prisoners, delivered to each cell along with weekly menu choices. Prisoners were asked to tick a box to indicate that they agreed to take part and understood that participation was voluntary. They were assured that responses would only be seen by the research team and that no identifying material would be used in any research report. Accompanying the instrument was an explanation that the researchers were in the process of developing new measures. Participants were asked to respond to each question using the five-point Likert scale. Respondents were also given an opportunity to provide qualitative feedback on specific questions. The pilot generated a total of 264 completed responses, a response rate of 50%.

[21] The practical problems dimension was omitted from the testing. This was in part because the CRIME-PICS questions were already validated, and partly in order to keep the length of the instrument short at this stage. The practical problems questions were tested at later stages, as described in Section 3.7. Questions about relationships with staff were also omitted, since these questions are designed to measure relationships with project (rather than prison) staff.

Method of analysis

Completed instruments from the prison testing were subject to four forms of analysis.

Identifying questions missed or not understood: Data analysis involved an initial stage of identifying those questions repeatedly missed or receiving unfavourable feedback.

Principal component analysis: Given that the instrument combined many newly constructed items along with established constructs that have not been administered together previously, a principal component analysis in the form of an exploratory factor analysis was conducted on the 237 complete responses. This is a form of exploratory factor analysis used to identify factor structures or clusters of variables.[22] This statistical process is widely used as a way of reducing a lengthy instrument to a more manageable size whilst retaining the necessary information. In this process of data reduction, factor analysis achieves what Field (2005) refers to as "parsimony by explaining the maximum amount of common variance … using the smallest number of explanatory concepts" (Field, 2005, p. 620). Although the technique originated in measuring personality traits, it is widely used in a range of contexts by looking at which variables cluster together in a dataset.

Calculating Cronbach's alpha: In identifying the most suitable items to contribute to each dimension, and in turn the consistency of the measures, a reliability analysis was conducted for each dimension by calculating Cronbach's alpha coefficients. Cronbach's alpha assesses the internal consistency of a scale and the closer Cronbach's alpha is to 1, the more reliable the associated items are in contributing to the domain to which they relate. Typically, the alpha for a cognitive test, for example, would be expected to be of at least .8 to be accepted, but as Kline (1999) argues, values of below .7 can be expected when dealing with personality constructs.

Calculating Pearson correlations: To examine the discriminate validity of the instrument, that is the degree to which items actually measure different domains, Pearson's correlations were calculated.

[22] An orthogonal (Varimax) rotation was used given the theoretical independence of the hypothesised constructs. A Kaiser-Mayer-Olkin (KMO) statistic of .88 was achieved indicating excellent sampling adequacy (Hutcheson and Sofroniou, 1999).

Results of the analysis

Questions missed or not understood: Only one question was identified as difficult for some respondents to understand ('I feel optimistic about the future'), and was excluded on that basis.

Principal component analysis: From the factor analysis, of the 51 items, 38 loaded onto five different factors as follows (see Appendix B for details of the loadings):

- Ten items loaded onto factor one. All of these derived from questions which aimed to measure resilience, wellbeing or agency / self-efficacy. Given that these are three theoretically important and distinct dimensions, it was decided to split factor one into three separate subscales.
- Six items loaded onto factor two. These all related to impulsivity.
- Seven items loaded onto the third factor. These all related to motivation to change.
- Eight items loaded onto the fourth factor. These all related to hope.
- Six items loaded onto factor five. These all related to interpersonal trust.

Only the 38 items loading onto the above factors were retained (i.e. 13 items were removed).

Calculating Cronbach's alpha: Reliability analysis identified the items with lower internal consistency. On the basis of this, a further 17 items were removed, thereby reducing the instrument to 21 items, with optimum factor solutions (i.e. sets of items) for each of the seven dimensions being measured. These factors contained between two and four questions each. Table 3.3 shows the coefficients of the dimensions. Grouped in this way, five of the seven clusters achieved an alpha reliability coefficient of .77 or above, including three – impulsivity, motivation to change and interpersonal trust – which scored above .8. The only two dimensions with relatively low scores were resilience (.56) and agency (.63). An option considered at this stage was whether to add more items for these dimensions with low alpha scores or to remove them. Adding more questions (at least from the long list of questions piloted) did not increase the alpha score – in fact, this decreased the score. Removing resilience and agency / self-efficacy was not considered desirable at this stage as they are important outcomes inherent to the theory of change of arts and mentoring programmes. For these reasons, and because it is not unusual to include items with such scores, resilience and agency / self-efficacy were included in the IOMI with two and three items respectively at this stage of development. However, as discussed in Chapter 5, the research team suggest that in any further development of the IOMI, some new questions relating to resilience and

agency / self-efficacy should be tested as possible additions to the instrument, in the hope that higher alpha scores can be obtained for these two dimensions.

Table 3.3: Reliability coefficients – stage 5 piloting with prisoners

Dimension	Number of items	Reliability coefficients (α) (n = 264)
Resilience	2	0.56
Wellbeing	3	0.79
Agency / self-efficacy	3	0.63
Impulsivity	3	0.82
Motivation to change	3	0.81
Hope	3	0.77
Interpersonal trust	4	0.82

Pearson correlation: Correlations between subscales were low to moderate, indicating that the vast majority of subscales were measuring independent domains, except for Resilience, Wellbeing and Agency / self-efficacy, which were each moderately to highly correlated with each other. This is not surprising given that these three subscales loaded onto the same factor in the factor analysis, but as noted, were separated on theoretical grounds.

Table 3.4: Pearson correlations between scales – stage 5 piloting with prisoners

	Wellbeing	Agency / self-efficacy	Impulsivity	Motivation to change	Hope	Interpersonal Trust
Resilience	.660**	.567**	-.310**	.089	.501**	.292**
Wellbeing		.649**	-.273**	.193**	.490**	.401**
Agency / self- efficacy			-.360	.095	.425**	.227**
Impulsivity				-.133*	-.513**.	-.011
Motivation to change					.347**	.226**
Hope						.226**

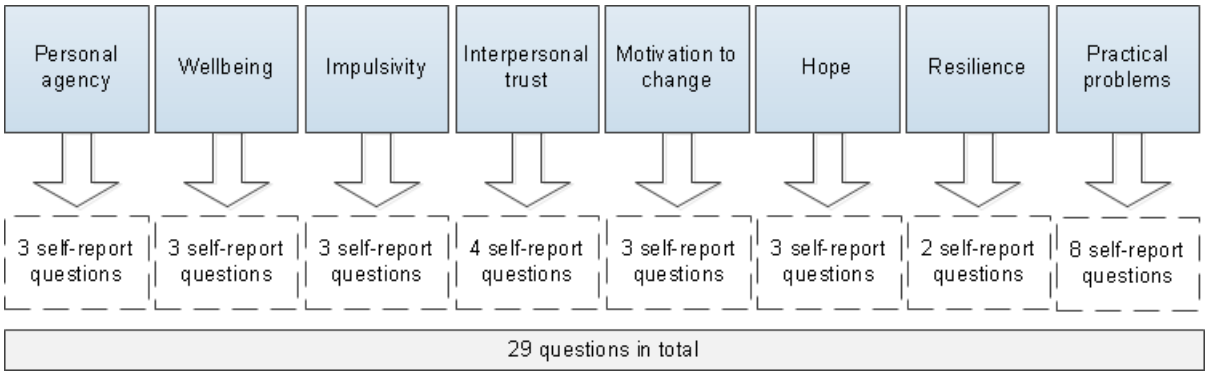
** Significant at p <.001, * significant at p <.05

The instrument after prison testing

As a result of the analysis described so far, the research team produced an instrument consisting of the 21 items which made up the seven dimensions with optimum factor solutions (i.e. sets of items) shown in Table 3.3, plus eight items measuring practical problems taken from CRIME-PICS II (29 items in total). The instrument was then piloted further in a community setting, as described in the next section. Its structure is summarised in Figure 3.2 below, and a full copy of the instrument as piloted is reproduced in Appendix A.

For six of the dimensions an increase in score indicates a positive change. The exception is impulsivity where, as is the case for practical problems, a reduction score indicates a positive change.

Figure 3.2: Overview of the IOMI



3.6 Stage 6: Piloting in the community with providers

This step of the testing sought to explore usability of the 29-item IOMI in a community setting and to investigate the ability of the instrument to measure change over time. To this end, a number of agencies which provided mentoring or arts projects in the community were recruited to pilot it. Six (four delivering mentoring projects and two delivering arts projects) agreed to do so. A description of these organisations is provided in Appendix E. One used the instrument for a period of six months, and the others for between two and five months. The plan was for the providers to ask programme participants to complete the instrument at the start of the intervention (T1) and again at the end (T2), in order to develop pairs of questionnaires which would allow the research team to explore change over time by undertaking a pre and post comparison. Participants filled in the instrument by hand, and programme staff collected the completed instruments for the research team.

Unfortunately, while all the providers successfully requested participants to complete IOMI at T1, two of the mentoring projects did not achieve any second administrations of the instrument at T2, and one produced only one. By contrast, two arts projects achieved 100 per cent follow-up rates (producing 10 T1s and 10 T2s, and 8 T1s and 8 T2s, respectively) and one mentoring project achieved 44 T2s out of a group of 49 clients who had completed T1s. In all, this left data from 63 pairs of instruments which had been completed by the same people at T1 and T2, and a further 44 T1 instruments with no matching T2.[23]
These results are in some ways very encouraging and in others disappointing. On the one hand, they suggest that, if an agency is highly committed to following-up its clients and asking them to complete a second questionnaire, it can achieve this in a high percentage of cases. On the other hand, the almost total lack of completed T2 instruments from three of the

[23] Three participants reported having received support in completing the questionnaire at T1, and four reported such support at T2.

four mentoring projects was at odds with the views expressed during the consultation that there was a desire among providers to measure their work. Although some providers co-operated fully and effectively throughout, securing the sustained involvement of others was more challenging than anticipated. Barriers to participation in the pilot included concerns about funding, which meant involvement in the pilot was not a priority; and a gap between strategic-level staff who signed up to the pilot and operational staff who felt they did not have the time to put it into practice (a finding echoed in previous studies, especially where small providers are concerned, for example, Meek et al., 2010; Mills et al., 2011). There was also variability between providers, as well as individual staff members, in the extent to which they thought outcome measurement was important. The experience of piloting the instrument raises questions about how to maximise support from providers in future for further testing and validation of the IOMI.

Limitations of the data from providers

Sample size: 63 pairs of instruments is sufficient for initial analysis, but cannot provide a basis for firm conclusions about the validity, reliability and sensitivity to change of the instrument.

Unbalanced contribution of different providers: Over two-thirds of all completed T1 and T2 pairs came from one provider, and three of the six providers produced only one T2 between them. This raises questions about the generalisability of the findings.

Time period between completion of T1 and T2 instruments: this ranged from two weeks to two months. Interventions with offenders do not always have fixed programme length and it is not always clear at what point they should be deemed to have been completed. There was probably some inconsistency between cases in how completion was defined in the mentoring project. Some T2s were deliberately brought forward at the end of the pilot period in order to maximise the number of cases available for analysis. This could have had some impact on the findings, as further change (negative or positive) could have taken place after the T2 instrument was completed.

Attrition: It could be argued that the results are affected by case attrition in that, across the six projects which agreed to take part in the pilot, 44 of the 107 individuals who completed T1 questionnaires did not complete a second one. Certainly, case attrition can potentially introduce bias into results. For example, those completing T2s may be more likely to demonstrate positive change than those who are not contactable, or who refuse to do so. However, in the pilot there was very little attrition of cases within the three projects which

made a real effort to obtain T2 completions. This considerably reduces the risk of this kind of bias in the analysis below.

Method of analysis

Completed pairs of instruments from the provider testing were subject to three forms of analysis:

- Descriptive analysis - age and gender of the sample.
- Internal consistency was once again calculated using Alpha coefficients for dimensions at T1 and T2. Sum scores were calculated for each dimension by summing responses to each item related to that dimension, creating a T1 and T2 sum score for each participant on each dimension.[24]
- A Mixed Factorial Multivariate Analysis of Variance (MANOVA) was conducted, with gender, age and project as 'between subject variables', and time (T1 and T2) as a 'within subject variable' across domains. This allowed for analysis of whether the instrument was sensitive to change at the multivariate level across domains and at the univariate level for each domain, as well as assessing the potential impact of age, gender and project.

Additionally, responses to the five questions about relationships with staff were reviewed.

Missing data

As is typical with self-report surveys administered in criminal justice settings, a small number of participants returned incomplete questionnaires, with one or more items lacking a response. In dealing with missing data, researchers can either exclude such participants from analysis altogether, or exclude them only from analysis of the specific variables where their data is missing. In order to maximise available data (given the small sample size) the latter option was pursued.[25]

[24] Cases with missing data were excluded.

[25] Analysis was conducted following the selection of the 'exclude cases pairwise' option for bivariate correlations and 'exclude cases analysis by analysis' for t-tests and analysis of variance, rather than the 'exclude cases listwise' option, However, in order to assess the potential impact of this process, analyses were repeated with the more conservative 'exclude cases listwise' option (reducing the sample subjected to analysis), and this process revealed the same results, confirming that this approach to dealing with missing data while maximising sample size did not adversely impact on the analytical process.

Results of analysis

Age and gender: Ages of participants ranged from 18 to 67 years, with an average of just under 33 years (see Table 3.5). Around 75% of the sample were male.[26]

Table 3.5: Respondents completing the IOMI at T1 and T2 – stage 6 piloting with providers

Project	Number of paired responses	Male respondents	Female respondents	Gender not reported	Average age (years)
Mentoring projects 1 and 4	45	25	7	13	31.8
Arts project 1	10	5	5	0	39.5
Arts project 2	8	8	0	0	30.5
Total	*63*	*38*	*12*	*13*	*32.9*

Internal consistency: As illustrated in Table 3.6, alpha coefficients were once again for the most part moderate to excellent, though as expected were lower in some cases and more variable compared to the results from the stage 5 prison testing. The alpha coefficients for the resilience subscale which comprises of only two items were particularly low. This is not surprising given that the sample was comparatively small (63 people).

Table 3.6: T1 and T2 reliabilities – stage 6 piloting with providers

Dimension	Alpha coefficient at T1	Alpha coefficient at T2
Resilience	.24	.47
Wellbeing	.83	.83
Agency / self-efficacy	.72	.76
Trust	.77	.75
Impulsivity	.65	.76
Hope	.74	.67
Motivation	.61	.72
Practical problems	.74	.59

MANOVA

Gender: There was a no significant multivariate effect of gender on domains overall or as an interaction with time, demonstrating that male and female scores across domains from T1 to T2 did not differ significantly. However, this should not necessarily be interpreted as an indication that gender will not have an impact on changes in each dimension across time, as there were only 12 women in the sample.

[26] 50% of respondents were aged under 30 years, and the remaining 50% aged 30-67 years. Gender was reported by 50 of the participants, and of these 38 were men and 12 were women.

Age differences: There was no significant multivariate effect of age (when participants were categorised as 30 years or under and over 30 years) overall, or over time across domains.

Project: There was no significant multivariate effect of project on domain scores overall or across time.

Domains: In terms of within subjects effects across time (T1 and T2), there was a significant multivariate effect of time across domains (F (8, 31) = 2.932, p <.05) indicating that when all domains are considered together, there was significant change from T1 to T2. Changes in scores for each dimension were also observed to be in a favourable direction (for all dimensions, an increase in score indicates a positive change, but for impulsivity and practical problems a reduced score indicates a positive change). Furthermore, as shown in Table 3.7, univariate analysis revealed that all changes for individual domains were statistically significant except for wellbeing, agency / self-efficacy and motivation.

Table 3.7: Summary of univariate MANOVA results – stage 6 piloting with providers

Dimension	Potential range	Time 1 mean score (SD)	Time 2 mean score (SD)	Mean change	Degrees of freedom	F
Resilience	(2-10)	5.60 (1.70)	6.30 (1.71)	.71	1, 38	6.435*
Wellbeing	(3-15)	9.58 (2.75)	10.29 (2.70)	.70	1, 38	2.592
Agency / self-efficacy	(3-15)	9.92 (2.26)	10.68 (2.37)	.76	1, 38	2.361
Trust	(4-20)	14.95 (3.00)	16.42 (2.65)	1.48	1, 38	9.808**
Impulsivity	(3-15)	10.71 (1.76)	9.08 (3.07)	-1.63	1, 38	11.99**
Hope	(3-15)	9.62 (3.00)	10.66 (2.80)	1.04	1, 38	5.460*
Motivation	(3-15)	12.35 (1.88)	13.08 (1.68)	.73	1, 38	3.900
Practical problems	(8-32)	17.73 (5.66)	14.86 (4.33)	-2.86	1, 38	8.535**

Note. * p <.05, ** p <.01 (n=63)

Staff relationship questions: A ceiling effect was observed in relation to these questions, with nearly all programme participants reporting very good relationships with staff. These results raise doubts about the value of these questions. However, further evidence would be needed before making a final decision about their removal from IOMI.

3.7 Stage 7: Test-retest with prisoners

The purpose of a test-retest exercise is to investigate the stability of responses to items in an instrument over time. This is typically assessed by administering the same instrument to the same population at two different time points. There is usually a relatively short period of time between each administration, with as few as possible significant events or interventions

having taken place in the intervening period. The length of time left between the test and re-test has to balance:

- The risk of generating recall, carry-over, or practice effects – a risk if the period is too short.
- The risk of generating differences in scores caused by real changes in the respondent's state, which can falsely indicate that the instrument is unreliable.

In selecting the period of time between the test and retest, a review was conducted to identify whether the pre-existing scales used in the development of IOMI (see Section 3.4) were designed for re-assessment at certain time periods. Taking this into account, it was decided to aim to retest all participants at a point between one and three weeks after their first completion of the questionnaire.[27] The support of a second prison, HMP Moorland, was secured to conduct the test re-test. Arrangements for distributing the instruments and information provided to prisoners were the same as in stage 5. Respondents were given seven days to complete the questionnaire in each round. This meant that the length of time between the test and re-test for any individual was between 8 and 21 days.[28]

Table 3.8 shows the mean scores for each dimension, the degree of change between scores, alpha coefficients, and the test re-test correlations for T1 and T2. MANOVA results indicated that the degree of change from T1 to T2 is very small: only one dimension (Trust) showed a significant change in mean sum score at the 0.05 level (F (1,126) = 5.90, p =.016). This suggests that the measure is stable over short periods of time for all domains other than trust.

[27] A targeted search of relevant bibliographic databases identified relevant information about temporal stability in relation to five out of the eight scales. This showed that the General Perceived Self-Efficacy Scale (Schwarzer and Jerusalem, 1995), the Brief Resilience Scale (Smith et al., 2008) and the UCLA loneliness scale (Russell, 1978) all have good temporal stability for a minimum period of 2 weeks and up to 3 months for adults (Cacioppo et al., 2007; Chiu and Hector, 2004; Smith et al., 2008). Temporal stability seems to be high for much longer time periods when these measures are administered to middle aged and older adults, where temporal stability can be sustained after 1 or 2 years (e.g. Cacioppo et al., 1996). In terms of the Adult State Hope Scale, no data appears to be available for temporal stability in adults, although for children it can be reliably achieved after 1 month (Snyder et al., 1997). On the other hand, temporal stability for the University of Rhode Island Change Assessment Scale (University of Rhode Island) has been deemed very weak (Carel et al., 2006).

[28] If respondents completed the T1 instrument on first day of distribution and the T2 questionnaire on the last day of distribution, this left 21 days between the test and retest. If respondents competed on the last day of T1 distribution and first day of T2 distribution, this left 8 days between the test and retest.

Table.3.8: Test re-test analysis results - stage 7

Scale	Mean score at Time 1 (SD)	Mean score at Time 2 (SD)	Alpha score Time 1	Alpha score Time 2	Pearson's Test-retest correlation
Resilience	6.78 (1.71)	6.85 (1.69)	.45	.55	.67**
Wellbeing	10.93 (2.44)	11.02 (2.25)	.80	.81	.79**
Agency / self-efficacy	11.19 (2.16)	11.07 (2.05)	.75	.76	.79**
Trust	14.79* (3.21)	15.15* (2.89)	.81	.82	.76**
Impulsivity	7.80 (2.98)	8.02 (3.17)	.81	.89	.79**
Hope	10.96 (2.76)	11.03 (2.70)	.79	.81	.76**
Motivation	12.10 (2.23)	11.99 (2.34)	.75	.81	.71**
Practical problems	14.94 (5.66)	14.53 (4.33)	.63	.66	.68**

* Significant change at 0.05 level. ** Significant correlation at 0.001 level.

This low level of change was found in a situation where T1 and T2 were relatively close together, and where most participants had not been subject to any significant interventions between these times. It can be contrasted with the very different results for the 63 cases analysed in stage 6, where the changes in mean scores before and after arts or mentoring interventions (and over generally longer periods than in the test-retest) were statistically significant in all but one of the dimensions, suggesting sensitivity to change. While further testing is necessary before confident conclusions can be drawn, these promising results suggest that the measure is stable over short periods of time for all the domains other than trust, while also being sensitive to change.

Importantly in terms of internal consistency, Table 3.8 also shows that the Alpha coefficients were generally very similar at T1 and T2. Furthermore, test re-test analysis revealed that pairs of scores at T1 and T2 on all six domains were significantly correlated (Pearson correlation coefficients) at the 0.001 level. This provides a further strong indication of the reliability of the instrument.

4. Developing the guidance, costing tool and data entry tool

4.1 Guidance document

A guidance document is included in the toolkit. The key aims of the guidance notes are to provide user organisations with:

- An understanding of the background to the design and development of the instrument and of the evidence used to underpin that design.
- Details concerning the reliability and validity of the instrument.
- Clear information about the key dimensions being measured, why these dimensions are important for assessing changes facilitated by the work that they deliver, and how they can be understood in relation to their own theories of change.
- Clear and accessible instructions for administration, data entry, scoring and reporting, to ensure consistency of use across multiple sites and organisations.
- Broad advice about how data generated by the instrument might be understood and how such data could be used as part of wider evaluation work.
- Some awareness of issues concerning use of the instrument with particular sub-groups (for example, young people, participants at women-only projects, etc.).
- Information about different ways of linking the tool to a range of other kinds of information or data to validate or challenge the picture emerging from offenders' self-reported data (for example, information from client management databases on appointments kept, time spent with practitioners, changes in accommodation or drug use, progress recorded in prison records, and practitioners' reflections on offenders' progress).

An important part of any further testing would be to assess the clarity and usability of the guidance document, from the perspective of providers.

4.2 Data entry tool

The toolkit includes a data entry and reporting tool for the IOMI in database format. The tool allows the user to enter responses onto an on-screen version of the instrument. The package:

- Converts each response into a raw score (for example, with 'strongly agree', 'agree', 'neither agree nor disagree', 'disagree', and 'strongly disagree' being scored as 5, 4, 3, 2, and 1 respectively for most questions, and reverse scored for some), and groups and sums these by dimension so that a raw score is recorded

for Agency / self-efficacy, Hope, Impulsivity/problem-solving, Motivation to change, Resilience, Interpersonal trust, and Wellbeing.

- Calculates a score which allows reports to be harmonised across dimensions.
- The reporting side of the package allows projects to generate reports on individual clients and to report on particular dimensions in aggregate.

The use of this data entry tool was not trialled with any of the providers testing the instrument in the field. It is important that the tool enables staff in provider organisations, who are not statisticians, to easily analyse and interpret the scores from the instrument. Currently, the tool is limited to producing individual-level results in the forms of graphs and tables. Any future versions of the data entry tool should enable users to construct offender cohorts, outcomes for which can be analysed at group level.

The data entry tool is available on request.[29]

4.3 Costing tool

One objective of this project was to develop and include in the toolkit a resource that providers could use to gather information about the costs of delivering their intervention. It has increasingly become a condition of funding that providers undertake some sort of assessment of the costs and benefits of the work that they deliver. However, undertaking cost-effectiveness analysis (CEA), cost-benefit analysis (CBA), value for money (VFM) assessment or social return on investment (SROI) analysis can be complicated, time-consuming and expensive to undertake, and is therefore a challenge for many providers.

Within the scope of this project, the objective was to develop a simple tool which could:

- Allow providers to estimate the costs of what they deliver.
- Allow providers to calculate unit costs in a number of different ways.
- Be used by providers to gather and present cost information in a transparent and consistent manner.

The availability of this kind of cost data can facilitate comparisons across different providers who use the tool. Such comparisons need to be made carefully, given that unit costs can vary widely across different user groups. This cost data can also provide information useful

[29] To access the data entry tool contact research@arcs-ltd.com for a link and login details.

for commissioners, who need to consider the appropriate balance between levels of resource input and expected outcomes.

Development and testing of the costing tool

The development and testing process was designed to ensure that the costing tool:

- Did not have any key gaps or omissions in the eyes of providers.
- Was sufficiently clear to be understood and used consistently by key provider staff.
- Would generate costing and unit cost summaries that will be of use to providers.
- Would generate the kind of information that will be relevant to other assessments that providers might wish to undertake as part of wider evaluation work.

A draft costing tool in a spreadsheet format was designed by members of the research team experienced in delivering CEAs and CBAs as part of the evaluation of a wide range of social programmes. The draft costing tool, along with guidance on its use, was sent to three providers delivering arts and mentoring programmes in the community. Site visits were conducted with these providers, during which the costing tool was discussed. Providers were asked to comment on the usability of the tool, whether the data requested by the tool was available for their project, and any potential difficulties in its use. The cost tool was revised in light of these discussions and sent again to three projects (two mentoring, and one arts) who were asked to complete the tool and comment again on any problems.

The costing tool

The final version of the tool is made up of four spreadsheets, with three of these being used by providers to input information, and one sheet being automatically populated by the programme. The three data entry sheets cover, respectively:

- Running costs (by year).
- Numbers of service users and their level of participation in a project or intervention (by year).
- Set-up costs.

The summary sheet uses totals from the other sheets to calculate weighted annual costs by year. It also calculates unit costs both by type of participant (in terms of how intensive their involvement with a project is), and by year, and it also calculates an overall unit cost (which is a weighted average).

The costing tool is published alongside this report on gov.uk. Instructions about how to complete the costing tool are included in Section 8 of the guidance document, which is also published on gov.uk.

5. Summary of findings and implications

This project has involved the production of a toolkit for providers of arts and mentoring interventions for offenders, which they can use in the delivery and assessment of their own work. The toolkit includes:

- The Intermediate Outcomes Measurement Instrument (IOMI).
- A document which describes the IOMI, and gives guidance on how it should be administered and how it could be used in a wider evaluation.
- A data entry tool to facilitate the collection and analysis of data from the IOMI.
- A tool for the collection and presentation of cost data, to assist organisations in measuring the cost of delivering their programmes, which can feed into wider cost-benefit or cost-effectiveness assessments.

The instrument is outlined in Appendix A of the Technical Appendices published alongside this report; the questionnaire is also available as a standalone document for download on gov.uk, along with the guidance document and costing tool. The data entry tool is available on request.[30]

The toolkit is theoretically informed and is anchored in careful reviews of evidence from research literature, as well as extensive consultation with providers of mentoring and arts programmes. It is emphasised that, although the results of preliminary analysis of data from piloting with samples of offenders were promising, an instrument of this kind inevitably requires extensive testing over longer periods of time (including the collection and analysis of reconviction data on those who have completed the questionnaires) before serious claims can be made for its validity and reliability or for its value as a proxy for reoffending.

5.1 Providers' interest in and capacity for measurement

Most of the providers of arts and mentoring projects contacted in the course of this research displayed a strong interest in more robust means of demonstrating the effectiveness of their work. It was widely admitted that there was an over-reliance, particularly in the voluntary and community sector, on anecdotal evidence or small numbers of case studies. Although valuable in terms of 'bringing to life' what providers do and its impact on some clients, it was recognised that such evidence is inadequate as a measure of performance in a world in which securing ongoing funding increasingly depends on evidencing improved outcomes. At

[30] To access the data entry tool contact research@arcs-ltd.com for a link and login details.

the same time, a certain amount of resistance was encountered, particularly among front-line staff and volunteers (rather than managers) to the idea of introducing 'measurement tools'. This was opposed in some cases on principle, but most commonly because of concerns about resources. It was made clear that many of the smaller providers simply do not have the capacity to collect and analyse significant amounts of data. So, while there was clearly an appetite for measurement, there was also a demand for simplicity and succinctness in any evaluation tool.

5.2 'User-friendliness' of the instrument

As outlined in Chapter 3, the results of piloting exercises (with prisoners and those involved in community arts and mentoring programmes) indicated that the items in the IOMI were readily understood and few were considered ambiguous. The instrument was almost unanimously agreed by staff and offenders who took part in these exercises to be user-friendly, interesting, and quick and easy to complete. However, the research team did encounter difficulties in securing the involvement of some providers in the piloting process, indicating that even with a user-friendly instrument there may be challenges in ensuring its take-up and use.

5.3 Reliability and validity

Data from piloting was used to undertake preliminary statistical testing of the reliability and validity of the instrument. Analysis based on calculating alpha coefficients for each key dimension indicates that it has a good level of internal consistency, although some alpha coefficients were low. As the instrument is theoretically based and is anchored in consultation with a wide range of experts and experienced providers, it also has strong face validity. Moreover, MANOVA analysis of changes in individuals' scores before and after experiencing rehabilitative interventions – based on a relatively low number of cases – showed that all changes were in the right direction. Changes on five out of the eight dimensions were statistically significant, suggesting the IOMI is sensitive to change.

In short, the indications at this early stage of the piloting and testing process are positive and encouraging. However, as outlined in the next section, there are several steps which should be undertaken before any firm claims can be made about the validity and reliability of the instrument, its ability to measure meaningful change, or its association with reoffending rates.

5.4 Future testing and piloting

It is suggested that the following tasks should be undertaken in order to validate the instrument and establish how useful it will be:

Sensitivity to change: Further testing is needed to explore the sensitivity of the instrument and of its individual dimensions to change over time, in both the long and short term.[31] Tests are also needed to assess its ability to capture change over different time periods. This could lead to a recommendation as to the optimum period of time between T1 and T2.

Floor and ceiling effects: It is important to establish whether specific dimensions are vulnerable to floor or ceiling effects (i.e. whether there are trends for responses to be consistently low or consistently high). A ceiling effect was observed during the initial testing in questions about relationships with staff, where offender responses were consistently high. This raised doubts about the value of these questions, although further evidence is needed before making a final decision on removing them from the instrument.

Comparison with results from other instruments and official data: To further establish validity of the IOMI, scores could be compared and cross-tabulated with results from other already validated tools which measure the same or similar constructs, or from official datasets held by criminal justice agencies (such as Offender Assessment System (OASys) scores).[32] This would help to address the possible problem with IOMI that it is based entirely on self-report data, which is sometimes considered less reliable than other sources of information.

Insufficient items in some dimensions: Two dimensions (resilience and agency / self-efficacy) had poor internal consistency, which on reflection may be improved by introducing additional new items for these dimensions (resilience has only two items at present), or by combining dimensions.

Associations with reoffending: Once the reliability of the instrument has been assured, a high priority will be to identify whether positive changes on the dimensions in the instrument, either singly or in combination, are associated with reductions in reoffending. In addition,

---

[31] It is likely that the dimensions are not equally subject to short-term change, although more evidence would be required before they could be ranked precisely in that respect. Impulsivity, for example, is a dimension that is not likely to change from day to day, and interpersonal trust might also not be subject to very short-term changes. By contrast, levels of hope and perhaps motivation could change more quickly.

[32] See National Offender Management Service (2015) for a compendium of research and analysis on OASys.

correlations between IOMI scores and those from assessment tools such as OASys could be investigated.

Further testing with other groups of offenders and other kinds of intervention: The testing and piloting conducted looked predominantly at the use of the IOMI instrument with adult male offenders in mainstream prisons and in voluntary community-based projects. While some preliminary piloting was undertaken with projects that were women-only or that worked with younger people, more testing would be needed to explore any variations in applicability of the dimensions and questions in the instrument by age or gender. Separate testing would also be needed regarding its suitable use and validity with other populations, such as high-risk offenders or substance misusers. Finally, it is possible that the toolkit could be used for measuring the impact of other types of interventions with offenders, beyond mentoring and arts programmes, but this too would need further investigation and testing.

5.5 Policy and practice implications

Unless further testing and analysis is conducted, any results based on the instrument should be treated with caution, and should not be used as a basis for decision-making. In particular, the instrument should not be used to compare the performance of different providers.

Potential value of the instrument

That having been said, over the longer term, the instrument has considerable potential as a tool for a number of policy and practice related purposes. These include:

Assistance to commissioners: Commissioners of rehabilitation services have a strong need for evidence about the effectiveness of different providers, and of different approaches to working with offenders. An instrument of this kind will not necessarily show that any one intervention has led directly to a reduction in reoffending. However, it has the potential to support other evidence that a particular agency or intervention has helped offenders to achieve positive changes which may contribute to reduced reoffending or eventual desistance from crime. Its usefulness in this respect will be greatly enhanced if further research were conducted to demonstrate a correlation between scores on the IOMI and actual reoffending rates. However, the instrument should not be used in crude fashion or on its own as a performance indicator by commissioners or others to compare the effectiveness of the different agencies that they are monitoring.

Self-evaluation: Independently of proving their worth to commissioners, many organisations (particularly in the VCSE sector) are keen to improve their services by finding out more about

the outcomes that their services are producing for clients. The IOMI potentially offers a simple to use, 'off-the-shelf' tool for this purpose. For those agencies which put high value on reducing reoffending, it could offer an early indication of effectiveness in this regard (rather than having to wait nearly two years for actual reoffending rates). For those who value other kinds of outcome, it could offer measures of changes in offenders closer to what they are directly setting out to help them achieve (such as changes in wellbeing or resilience). Evidence of such changes can also be fruitfully shared with offenders as part of a collaborative way of working.

Developing the evidence base: The instrument also potentially offers an extra resource for research into 'what works' in offender rehabilitation, especially into the effectiveness of small-scale and less mainstream interventions such as arts programmes. The rapid evidence assessments (REAs) conducted found limited reliable evidence about the impact of these programmes. The IOMI could be used to develop the evidence base, not only through use in small-scale evaluations, but more ambitiously by combining data from numerous small or medium sized providers who individually would have insufficient numbers of cases to allow robust quantitative analysis.

A large pool of data could be used to identify norms across different intervention types and different types of clients, eventually enabling services to compare their impact against these norms. Impact across different types of interventions could also be assessed. For example, a large scale study might find that the arts are particularly effective at increasing levels of hope, thereby allowing commissioners to employ a more nuanced approach to the commissioning of services for particular client groups.

A large national dataset could also allow for very detailed examination of relationships between specific dimensions, or of sensitivity to change over time, by dimension. In relation to contributions to desistance theory specifically, such a dataset could also allow for some detailed analysis of pathways, with multiple readings being assessed by group.

Use of the instrument at this stage

Bearing in mind the aforementioned caveats, providers are encouraged to use the IOMI (as in Appendix A), asking clients to complete it both before and after an intervention is delivered (or in a staged manner during lengthier interventions). The results cannot be claimed at this stage to constitute a reliable proxy for reduced reoffending, but the instrument has face validity and analysis to date suggests that it would not be unreasonable (as long as over-

claiming is avoided) to present IOMI scores alongside other indicators to help make a case that clients have undergone some change.

References

Allan, M., Grace, R. C., Rutherford, B. and Hudson, S. M. (2007) 'Psychometric assessment of dynamic risk factors for child molesters' *Sexual Abuse: A Journal of Research and Treatment,* 19, pp 347-367

Barnett, G., Wakeling, H., Mandeville-Norden, R. and Rakestrow, J. (2012) 'How Useful Are Psychometric Scores in Predicting Recidivism for Treated Sex Offenders?' *International Journal of Offender Therapy and Comparative Criminology,* 56(3), pp 420-446

Beech, A. and Ford, H. (2006) 'The relationship between risk, deviance, treatment outcome and sexual reconviction in a sample of child sexual abusers completing residential treatment for their offending' *Psychology, Crime & Law* 12, pp 685-701

Beech, A., Friendship, C., Erikson, M. and Hanson, R. K. (2002) 'The relationship between static and dynamic risk factors and reconviction in a sample of UK child abusers' *Sexual Abuse: A Journal of Research and Treatment,* 14, pp 155-167

Beech, A. R. (1998) 'A psychometric typology of child abusers' *International Journal of Offender Therapy and Comparative Criminology,* 42, pp 319-339

Bilby, C., Cauldfield, L. and Ridley, L. (2013) *Re-imagining futures Exploring arts interventions and the process of desistance*. Arts Alliance. http://www.artsevidence.org.uk/media/uploads/re-imagining-futures-research-report-final.pdf accessed 3 July 2018

Burnett, R. and Maruna, S. (2004) 'So Prison Works, Does It? The Criminal Careers of 130 Men Released from Prison under Home Secretary, Michael Howard' *Howard Journal* 33(4), pp 390-404

Burnett, R. and McNeil, F. (2005) 'The Place Of The Officer-Offender Relationship In Assisting Offenders To Desist From Crime' *Probation Journal,* 52(3), pp 221-242

Burrowes, N., Disley, E., Liddle, M., Maguire, M., Rubin, J., Taylor, J. and Wright, S. (2013) *Intermediate outcomes of arts projects: a rapid evidence assessment*. London: NOMS. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254450/Intermediate-outcomes-of-arts-projects.pdf accessed 3 July 2018

Cacioppo, J., Amaral, D., Blanchard, J., Cameron, J., Carter, C., Crews, D., Fiske, S., Heatherton, T., Johnson, M. K., Kozak, M., Levenson, R., Lord, C., Miller, E., Ochsner, K., Raichle, M., Shea, M., Taylor, S., Young, L. and Quinn, K. (2007) 'Progress and Implications for Mental Health' *Perspectives on Psychological Science,* 2, pp 99-123

Cacioppo, J. T., Petty, R. E., Feinstein, J. A. and Jarvis, W. B. G. (1996) 'Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition' *Psychological Bulletin,* 119, pp 197-253

Carel, J. C., Elie, C. and Ecosse, E. (2006) 'Self-esteem and social adjustment in young women with Turner syndrome—influence of pubertal management and sexuality: population-based cohort study' *Clin Endocrinol Metab* 91, pp 2972-2979

Chiu, F. T. and Hector, W. (2004) 'Validation of the Chinese general self-efficacy scale among individuals with schizophrenia in Hong Kong' *International Journal of Rehabilitation Research,* 27(2), pp 159-161

Clancy, A., Hudson, K., Maguire, M., Peake, R., Raynor, P., Vanstone, M. and Kynch, J. (2006) *Getting Out and Staying Out: Results of the Prisoner Resettlement Pathfinders*. Bristol: Policy Press

Cox, A. and Gelsthorpe, L. (2008) *Beats and Bars. Music in Prison: An evaluation*. Cambridge: Institute of Criminology

DiClemente, C. C., Schlundt, D. and Gemmell, L. (2004) 'Readiness and stages of change in addiction treatment' *American Journal on Addictions,* 13(2), pp 103-119

Farrall, S. and Calverley, A. (2006) *Understanding Desistance from Crime*. Milton Keynes: Open University Press

Fetzer Institute (n.d.) 'Self Report Measures for Love and Compassion Research: Loneliness and Interpersonal Problems'

Field, A. (2005) *Discovering Statistics using SPSS*. London: Sage.

Gordon, D., Levitas, R., Pantazis, C., Payne, S., Townsend, P., Bradshaw, J., Middleton, S., Bramley, G., Bridgwood, A., Maher, J. and Rowlands, O. (1999) *Millennium Poverty and Social Exclusion Survey. Questionnaire Appendix* http://www.bristol.ac.uk/poverty/pse/poverty_and_social_exclusion_online_appendix_%20final%20version.pdf accessed 3 July 2018

Hanson, R. K., & Bussière, M. T (1998) 'Predicting Relapse - a Meta-Analysis of Sexual Offender Recidivism Studies.' *Journal of Consulting & Clinical Psychology,* 66, pp 348-362

Hudson, S. M., Wales, D. S., Bakker, L. and Ward, T. (2002) 'Dynamic risk factors: The Kia Marama evaluation' *Sexual Abuse: A Journal of Research and Treatment,* 14, pp 103-119

Humphreys, C. and Stirling, T. (2008) 'Necessary but not sufficient: housing and the reduction of reoffending' Cardiff: Welsh Government

Huntley, F. L., Palmer, E. J. and Wakeling, H. C. (2012) 'Validation of an Adaptation of Levenson's Locus of Control Scale With Adult Male Incarcerated Sexual Offenders' *Sexual Abuse: A Journal of Research and Treatment,* 24(1), pp 46-63

Hutcheson, G. and Sofroniou, N. (1999) *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models*. London: Sage

Kazimirski, A., Ógáin, E., Van Vliet, A., Harries, E., Pritchard, D., Hunter, G., Jolliffe, D., Moretti, A., Turnbull, P., & Skrine, O. (yet to be published). Measuring Change in Offender Relationships: the development of a toolkit to measure the impact of relationship interventions with offenders. London: Ministry of Justice

Kline, P. (1999) *The Handbook of Psychological Testing*. London: Routledge

Laub, J. and Sampson, R. (2001) 'Understanding desistance from crime' *Crime and Justice* 21, pp 1-69

Lösel, F. (2012) 'Offender treatment and rehabilitation: What works?'. In Maguire, M., Morgan, R., Reiner, R. (ed) *The Oxford Handbook of Criminology*, pp 986-1061 Oxford: Oxford University Press

M&A Research (2013) *CRIME-PICS II manual*. M&A Research. http://www.crime-pics.co.uk/publications/cpicsmanual.pdf accessed 3 July 2018

M&A Research (n.d) *Conflict Resolution, Impulsivity and Aggression Questionnaire*. http://www.crime-pics.co.uk/criaq.html accessed 3 July 2018

Maguire, M., Holloway, K., Liddle, M., Gordon, F., Gray, P., Smith, A. and Wright, S. (2010) 'Evaluation of the Transitional Support Scheme' Cardiff: Welsh Assembly Government https://gov.wales/docs/caecd/research/101108-transitional-support-scheme-en.pdf accessed 3 July 2018

Maguire, M. and Nolan, J. (2007) 'Accommodation and related services for ex-prisoners'. In Hagley-Dickinson, A. and Hagley-Dickinson, L. (ed) *Prisoner Resettlement: Policy and Practice*, pp 144-172 Devon: Willan

Maguire, M. and Raynor, P. (2006) 'How the Resettlement of Prisoners Promotes Desistance from Crime: Or Does It?' *Criminology and Criminal Justice,* 6(1), pp 17-36

Maruna, S. (2001) 'Making good: The rhetoric of redemption'. In Maruna, S. (ed) *Making good: How ex-convicts reform and rebuild their lives.*, pp 85-108. Washington, DC US: American Psychological Association

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

Meek, R., Mills, A. and Gojovic, D. (2010) *The role of the third sector in work with offenders: the perceptions of criminal justice and third sector stakeholders*. Third Sector Research Centre Working Paper 34

Mills, A., Meek, R. and Gojkovic, D. (2011) 'Exploring the relationship between the voluntary sector and the state in criminal justice' *Voluntary Sector Review,* 2(2), pp 193-211

National Offender Management Service. (2015). A Compendium of Research and Analysis on the Offender Assessment System (OASys). London: Ministry of Justice

Østerås, N., Gulbrandsen, P., Garratt, A., Benth, J. S., Dahl, F. A., Natvig, B. and Brage, S. (2008) 'A randomised comparison of a four- and a five-point scale version of the Norwegian Function Assessment Scale' *Health and Quality of Life Outcomes,* 6(14)

Raynor, P. (1998) 'Attitudes, social problems and reconvictions in the 'STOP' probation experiment.' *Howard Journal of Criminal Justice,* 37, pp 1-15

Raynor, P., Ugwudike, P. and Vanstone, M. (2014) 'The impact of skills in probation work: A reconviction study' *Criminology and Criminal Justice*14 (2), 235-249.

Russell, D., Peplau, L. A. and Ferguson, M. L. (1978) 'Developing a measure of loneliness' *Journal of Personality Assessment,* 42, pp 290-294

Russell, D., Peplau, L. A.. & Ferguson, M. L. (1978) 'Developing a measure of loneliness' *Journal of Personality Assessment,* 42, pp 290-294

Sapouna, M., Bisset, C. and Conlong, A.-M. (2011) *What Works to Reduce Reoffending: A Summary of the Evidence Justice Analytical Services Scottish Government*. Available at http://www.scotland.gov.uk/Resource/0038/00385880.pdf accessed 3 July 2018

Schwarzer, R. and Jerusalem, M. (1995).Generalized Self-Efficacy scale. (GSE) http://userpage.fu-berlin.de/~health/engscal.htm accessed 3 July 2018

Shapland, J., Bottoms, A., Farrall, S., McNeill, F., Priede, C. and Robinson, G. (2012) 'The quality of probation supervision – A literature review: summary of key messages' https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/217435/quality-of-probation-supervision.pdf accessed 3 July 2018.

Sherman, L., D. Gottfredson, D. MacKenzie, J. Eck, P. Reuter and Bushway, S. (1997) *Preventing Crime: What Works, What Doesn't, What's Promising*, College Park, MD: Department of Criminology and Criminal Justice, University of Maryland, 1997

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P. and Bernard, J. (2008) 'The Brief Resilience Scale: Assessing the Ability to Bounce Back' *International Journal of Behavioral Medicine,* 15, pp 194-200

Snyder, C. R., Sympson, S. C., Ybasco, F. C., Borders, T. F., Babyak, M. A. and Higgins, R. L. (1997) 'Development and validation of the State Hope Scale' *Journal of Personal and Social Psychology,* 70(2), pp 321-335

Taylor, J., Burrowes, N., Disley, E., Liddle, M., Maguire, M., Rubin, J. and Wright, S. (2013) *Intermediate outcomes of mentoring interventions: a rapid evidence assessment*. London: NOMS https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254452/Intermediate-outcomes-of-mentoring-interventions.pdf accessed 3 July 2018

Wakeling, H., & Barnett, G. (2014). The Relationship Between Psychometric Test Scores and Reconviction in Sexual Offenders Undertaking Treatment. *Aggression and Violent Behaviour, 19*, 138-145

Weaver, B. and McNeill, F. (2010) 'Travelling hopefully: desistance research and probation practice'. In Brayford, J., Cowe, F., Deering, J. (ed) *What Else Works?: Creative Work with Offenders*, pp 36-60. Devon: Willan