



Theory-based evaluation: Options and choices for practitioners

Prof David Parsons, FRGS, FHEI, FAcSS
(Leeds Beckett University)

Social Research Association Workshop

© P&A Research and Consulting, 2021

INTRODUCTION

This resource pack provides take away resources which provide more detail and illustrations of some of the key content of the course. There are four resources:

- *Resource A: Case study of Plan Einstein Evaluation (making a case for TBE)*
- *Resource B: A layered approach to setting a theory base for an evaluation*
- *Resource C: Example of a synthesised ToC graphic from 'Launchpad'*
- *Resource D: Example of a theory of change*
- *Resource E: Some analytical methods to support theory based evaluation.*

Each resource supplements the content of the course.

We hope these resources will be useful when participants come to apply the knowledge and skills developed through the course in their own practice. Participants are welcome to share these resources with colleagues but any use should acknowledge the authorship and copyright of the resources.

Resource A: 'Making the Case for TBE'



The Utrecht Refugee Launchpad

The Utrecht Refugee Launchpad (URLP) was launched in Utrecht as 'Project Einstein' with funding from the EU Urban Innovative Actions Programme. One of four test bed projects across Europe aimed at large scale piloting of innovative approaches to asylum-seeker and refugee housing, it was based in Overvecht, a deprived district in the city of Utrecht. The initiative aimed to employ the principles of co-living and co-learning to create a more inclusive approach to the reception of asylum seekers, while simultaneously generating benefits for neighbourhood inhabitants in the local area.

Background

From late 2016, 38 local young people (Dutch and EU citizens) were expected to live in a building, known locally as *Plan Einstein*, alongside around 600 asylum seekers in the same site in an asylum seeker centre, with some shared facilities. All residents of the Centre (Plan Einstein) as well as people from the surrounding neighbourhood were to be encouraged to share activities, including social events, language courses in business English, classes in entrepreneurship and international business as well as expert coaching and opportunities to connect people to incubate new business ideas and develop relationships in the local business environment. In so doing the project aimed to:

- Equip participants with 'future-proof' skills that would be of benefit to them, whether in the Netherlands or elsewhere
- Help repair asylum seekers' 'broken narratives' and build stronger transition potential if and when asylum seekers were granted refugee status
- Halt the negative spiral created by the dominant 'exclusion and containment' approach to reception of asylum seekers and promote early integration of refugees.

The initiative also aimed to provide lessons about transferable practice for other urban neighbourhood approaches to asylum seeker and refugee integration in the Netherlands and more widely in Europe.

The evaluation

The EU funded projects was to run from late 2016 to Spring 2019. An evaluation was commissioned as an embedded part of the programme, led by COMPASS at the University of Oxford. A number of user stakeholders were involved in agreeing evaluation needs including the EU Urban Innovations Team, the municipal authorities in Utrecht, *Centraal Orgaan Opvang Azielzoekers* (the Dutch asylum seeker agency) – with different and often conflicting requirements.

There was an early ambition (for the municipal authorities) for a fully experimental method, with the EU being less prescriptive and valuing a more flexible approach. There was mixed enthusiasm for contrasting experiences both between the four EU funded projects and with impacts from more conventional ‘closed’ settlement centres. There were also different expectations of scope with different needs for covering the constituent participants – the asylum seekers, the young people co-residents, and the neighbourhood residents, with the municipal authorities keen to focus on the latter impacts, the COA on the asylum seekers with a tracking approach to extend to those securing refugee status, and the EU all three.

Resolution – making the case

Resolving the conflicting needs – and practical possibilities was compounded by the different location of evaluators, Project Einstein and Utrecht authorities, COA (Amsterdam) and the EU UIA (Brussels), different expectations of ‘robustness’ and future use and the impending start of the launch pad. Resolution involved a combination of bi-lateral and multi-lateral scoping and needs resolution discussions led by COMPASS and steered through external independent advice. The advice had early identified the likely better fit of a theory-based approach and a case was put forward based on:

Agreeing common ‘goal and scope’ requirements with all three ‘users’

- Setting out alternative ‘how to’ pathways for the evaluation – to include an RCT, QE and A TBE focus, with pro’s and con’s for each.
- Early dismissal of an RCT on the grounds of multi-factor influences on outcomes, ethical requirements and lack of a stable environment – likely volatile ‘recruitment’ and changing refugee mixes.
- Exploration of an adapted QE approach – subsequently abandoned when it was clear that appropriate in programme comparators were running to different timeframes (later starts).
- Further ‘making a case’ exploration of how an adapted form of TBE could fit needs.

This extended process saw users brought to the (inevitable) conclusions of the suitability of an adapted TBE approach, but with the municipal authorities a reluctant convert. In the event the case was made that:

- Ethical considerations ruled other possibilities
- ‘Start-up’ comparative data would not be available on a comprehensive and comparable basis – comprising a non-experimental approach
- The initiative operated in a highly volatile situation where there were likely to be multiple influences on outcomes for the different groups, and where many of those influences were only likely to emerge as the project progressed.
- Plan Einstein was highly experimental – and the evaluation needed to get under the skin not just of what worked (and what did not) but if the set up assumptions and

expectations had stood up, and the implications for changed thinking about what worked for any future such innovation.

- TBE offered a strong focus for qualitative evidence – this became important to the municipal authorities who were especially attracted by the prospect of ‘good news’ case studies which might be used for promotional purposes.

The case for TBE was consequently made progressively, building on reshaping initial expectations, and with all parties eventual enthusiast for the TBE approach.

The TBE approach

The theory-based evaluation followed a conventional approach to establishing and testing (and refining) a Theory of Change, clarified with partners to outline interim outcomes and a causal chain for achieving longer-term outcomes. In employing this approach it was keen to shed light less on a new and experimental project and the contexts in which it works, less seeking to show conclusively whether it ‘works’ but rather understanding ‘if, how or why does this work, for whom, in what circumstances?’ The aim was to better understand causal influences and their inter-relationships, against the original assumptions and expectations, so that it can be replicated or adapted elsewhere.

Following the initial ToC development, which also proved a valuable process for reflection within the delivery team, and for building the working relationships between the embedded researchers at Plan Einstein and COMPASS, they have used an adapted form of Contribution Analysis to underpin the TBE. This aims to understand if and how the project contributes to generating longer-term change among its beneficiaries, who include three groups: the asylum seekers, the local young people recruited from the neighbourhood, and residents of the wider neighbourhood itself.

The TBE has involved collating available COI data, supplemented by survey evidence from participants and a ‘living lab’ approach involving active relationships and staged experiential review with participants and multiple stakeholders. This was felt to be best placed against URLP partnership’s intention to innovate, be adaptable to the circumstances it finds, and evolve in content and process through collaborative innovation.

The TBE proved to be well placed subsequently to cope with unexpected changes in the content of the initiative, and in particular changes to:

- The COI placement strategy used at Plan Einstein
- Unexpected change of mix for young co-residents (students, non-local NEETs)
- Modified premises and common facilities
- Disruption to available common data form COI and others
- Changed political leadership (and direction) at the municipal authority.

The TBE approach has been sufficiently flexible to embrace these changes in an evaluation already welcomed as innovative and constructive by the UIA funders.

October 2019

Resource B: A layered approach to setting a *theory base* for an evaluation

This course sets out a 'layered approach' to building a theory platform for an evaluation. A theory platform may have value for almost any impact evaluation but is essential for a 'theory-based' evaluation. The layered approach provides for opportunities for different levels of intensity of theory setting and different potential for user organisations (eg those commissioning an evaluation or using it) to absorb the theory context, and TBE approaches. This brief guide introduces the different 'layer' options, why and how they are useful, who should do them, and how to get started in putting one together to help set up an impact evaluation. It also provides a few links to further guidance and sources of information.

Getting started

A theory platform is a tool which can be put together in different ways depending on the needs and starting point for an impact evaluation, its focus and needs and who it is to be done for (and used by). In all cases, its aim is to systematically set out, or surface, the assumptions underpinning why an intervention is to (or will) take place and how it is expected to work. This is an important starting point for any impact evaluation but in practice these assumptions and expectations are often unclear or insufficient.

How can a theory-base help an evaluation?

The starting point for most evaluations is where decision makers need to know if an intervention is doing what was expected of it, and these will (hopefully) be set out as evaluation objectives and/or goals for scope, timing and reporting. There may also be some background information on the intervention to say something about scale, focus, resourcing and perhaps targeted activities (and maybe planned achievements) but too often this will not give the evaluator enough (or any) detail about its underpinning rationale. The background information available to the evaluator on 'rationale' is too often light on detail such as:

- The nature of the intervention 'issue' (the problem or challenge which the intervention is set up to address)
- The context and environment – internal to the intervention and external – in which it to take place, and how this may affect aims and likely achievements
- What is understood from earlier research or any past attempts to tackle this issue
- What assumptions underpin how the intervention will work to deliver the necessary changes or impacts (ie not just the activities but how these are expected to be forged and inter-relate to make a difference)
- What sorts of changes or impacts are expected from those interactions – for who, what, how much and when.

Often it is not that these things are unknown or hidden; but they are too often assumed by policy makers, planners or others setting up interventions.

An impact evaluation is best placed to answer the evidence needs of decision makers when they can 'surface' these assumptions. Setting a theory base is a tool for getting these out into the open. With this, the evaluators have clearer understanding of what needs to be looked at to test if and how the intervention has worked against expectations of it. This is important because an intervention may 'fail' not because its activities or implementation are ineffective but because the expectations set for them (and which underpinned its roll-out) were ill-considered, inappropriate or unrealistic. An evaluation needs to distinguish between these if it is to provide reliable and appropriate evidence for decision makers.

Who does it?

A well planned intervention may already have given a lot of thought to 'rationale' and expectations and this will be shared with the evaluators as background information, perhaps as a logic chart (or chain) or even a fully worked programme theory (for the intervention) or theory of change (ToC). Far too often this is not the case, or the logic chart or ToC does not give enough detail for the evaluator to be confident they have an evaluation design which will look at the 'right' things. In these situations the evaluator may need to produce their own (or an enhanced) theory-base, working with the programme or policy team and key stakeholders to do so.

How should we go about it?

The layered approach to theory setting provides for more flexibility to use an approach which is fitted to particular needs and circumstances – and crucially to the 'users' ability to work with theory-based approaches and to engage appropriately with these. There are various ways in which the rationale, assumptions and expectations can be set out but common ways are:

- Using *available documentation* to set out, share and agree the intervention rationale
- Setting up a *logic chart* or chain, or modifying an existing one
- Establishing a *theory of change* (ToC) for the intervention.

These are not the only pathways but they are the most common; each is considered briefly below. The 'further information' section provides some further 'how to' practical guidance.

Available documentation: Although it may not have been shared (yet) with the evaluators those planning the intervention may have already given a lot of thought to its underpinning rationale. In this case, the evaluation may be able to get enough about assumptions and expectations from previous planning and other documentation (eg programme proposal, budget or business case scenario or project implementation document - PID). If these are available the evaluator will be looking for information on, for example:

- The background to the problem or challenge which the intervention will tackle
- Scene setting information about the nature of the problem or challenge (eg from any past policy or academic research or previous evaluations of any past initiatives)
- What assumptions have been made about the start-up of the intervention (budget, timeframe, partners and providers to be involved, etc)
- Assumptions also about the activities and deliverables needed for the intervention to work – and about expected scale, focus and participation, and how these inter-relate

- Expectations about any baseline, benchmarking or anticipated activity monitoring within the intervention
- Expectations also about the outcomes/impacts to be achieved, their likely nature and scale (what would 'success' look like) and how these will be measured.

Any information which is missing (or thin) might be filled in by scoping or inception discussions with programme or policy staff (pre design of the evaluation).

<i>Pro's</i>	<ul style="list-style-type: none"> • <i>Rapid and may fit well where a very fast start is needed</i> • <i>Simple – uses available/accessible evidence</i> • <i>Evidence sources will be understood and trusted by users</i>
<i>Con's</i>	<ul style="list-style-type: none"> • <i>Limited to evidence on contexts and those aspects of A and E that are part of the planning process</i> • <i>Likely to have major 'theory' gaps and to provide little in hidden assumptions; not suited to TBE for complex interventions</i> • <i>May say little about 'risks'.</i>

Logic charts: Where background documentation cannot meet the evaluator's needs for a theory base, they may need to put something together themselves; Logic charts are a common and trusted way to do this. We refer here to 'logic charts' but they go under many names (ie logic chains, logic models, intervention logic models) and they can be set out in different ways – as sequence or flow charts, as linked box charts, a structured narrative or some combination of these. There is no 'right' way to scope out a logic chart but a sound principle is to keep them as simple as possible to better communicate the 'theory'; multiple bullet points can work better than long descriptions.

While approaches to setting out logic charts differ, they all focus on setting out the main expected elements of the intervention and the sequence in which they will occur for:

- The 'issue': The challenge or problem the intervention is addressing
- The inputs needed to make the intervention happen
- The outputs - what will come from these inputs (NB. this may include any output or deliverables targets that have been set up for the intervention)
- The outcomes – the 'consequential changes' expected to occur in the short term (NB the logic chart might also split these into 'short term' and 'medium term').
- The impacts – the longer term consequential changes expected (NB after whatever lead time is needed for these longer term effects to take place).

Logic charts may also set out the activities to be conducted in the intervention to achieve the required outputs, although sometimes these can be subsumed into the 'outputs'. When put together, the logic chart needs to be shared appropriately with intervention partners and relevant stakeholders (and modified accordingly) to test if it is a fair and accurate reflection of the intervention assumptions and expectations. Annex 1 gives some more detail of what might be included in a logic chart, and the sequence in which they need to be articulated.

<i>Pro's</i>	<ul style="list-style-type: none"> • <i>Widely understood by policy people, planners and some stakeholders</i> • <i>Provides for simplified description of anticipated mechanisms and for multiple input-activity-output-outcome relationships</i> • <i>Strong on assumptions about sequencing of actions and intended outcomes.</i>
<i>Con's</i>	<ul style="list-style-type: none"> • <i>Not strong on assumptions beyond sequencing elements</i> • <i>Linear emphasis - may not show complexities in expected transformation processes</i> • <i>Says little or nothing about context and risks.</i>

Theory of change (ToC): A theory of change is a more refined approach to rationalising what the intervention is about and says (a lot) more about its underpinning assumptions and expectations. Well designed and tested, it is the gold standard for a theory-base and it provides everything that a good logic chart can ... and much more.

If the evaluation is using a more fully developed 'theory-based' approach, having a ToC will be an essential starting point to the method and design. Putting together a theory of change is not for those short on time or for the faint hearted, but nor are they rocket science. This guide can only touch on the how to' but more guidance can be sought from further sources of information (see below).

There are different ideas about what a ToC should look like although quite a lot of common ground on what they should cover. A common starting point is with a 'logic chart' type approach to sequence what is expected of the intervention inputs, activities, etc through to the long term vision on the change it expects to accomplish. This is usually through a conventional 'linear' logic chart or perhaps one set out as a flow chart with 'transformation' links between elements built into it.

It then goes further by adding more on the context of the problem or challenge being addressed and the intervention assumptions and expected mechanisms. This is usually by adding an accompanying narrative on:

- **Context:** A more fully worked statement of the nature of the 'problem' being addressed by the intervention – possibly using past evidence or available data to scope the key audience, the 'problem' and its distribution, its known effects, costs and consequences, the actions 'gap' why any past actions may not have worked, and what is known or expected of the causal influences (mechanisms) which underpin 'the problem' that the intervention seeks to address.
- **Expectations:** What consequential changes (outcomes and longer term impacts) are expected from the effective functioning of the interventions processes and transformation mechanisms; this usually distinguishes between early measureable outcomes, intermediate outcomes (those bridging between the early changes and those expected over a longer term, and longer term impacts – often occurring only after some lead time.
- **Assumptions:** The key assumptions that underpin the different steps planned in the intervention (as mapped in the logic chart), for example, what is needed for the different inputs to be secured, what is assumed in reaching the audience to 'recruit' or engage those individuals the intervention needs to get to, etc.
- **Risks:** This theory profile is usually complimented by a similar (narrative) assessment of risks to this working model operating as expected; the potential disruptions – internal and external - to its effectiveness and effects.

There are other ways to set out the 'theory' and no standardised approach to what's 'right'. A common framework is the Realist Evaluation use of – context – mechanisms – outcomes. This can work very well where the 'mechanisms' make it clear what respectively the tangible and intangible assumptions underpinning the cause-effect relationships of the intervention are thought to be and the expectations (outcomes in CMO) of the intervention are expected to be. 'Mechanisms' also need to set out coherent risks to the intervention working or working well. Using the CEAsR framework helps to make sure that none of these important aspects are missed; stakeholders may also find them easier to understand.

Like Logic models, ToCs need to be well tested – and refined – by appropriate consultation with stakeholders. This will take time but is always time well spent for the evaluation.

This may all seem a lot to achieve; it is. It also needs time at the start of an evaluation – and before starting with design – to produce a workable and credible theory of change. However, tackled this way it provides the necessary foundation for the evaluation to look not just at how the intervention is working, what it produces (outputs, outcomes and impacts) but also to critically assess these against the assumptions built into the intervention planning.

<i>Pro's</i>	<ul style="list-style-type: none"> • <i>Gold standard 'theory' platform for a TBE (well done and tested well)</i> • <i>Provides for much stronger content than other theory base options</i> • <i>Provides also for a full range of theory elements and their inter-relationships to set out a causal chain for the intervention</i>
<i>Con's</i>	<ul style="list-style-type: none"> • <i>Needs adequate resource to set out a well drafted 'first fit' initial ToC</i> • <i>Cannot be rushed – it is timely, staged process to iteratively develop a ToC</i> • <i>Users need to be confident in the ToC approach and committed to appropriate engagement with it</i> • <i>Needs extensive and well intentioned stakeholder engagement to review, refine and revise the ToC before and after the evaluation process.</i>

Further information

Some useful 'how to' sources and can be found for logic modelling in:

Better Evaluation web site; See guidance on programme and logic models in evaluation:
<https://www.betterevaluation.org/en/search/site/Logic%20Models>

Evaluation Support Guide 1.2: Developing a Logic Model from Evaluation Support Scotland:
<http://www.evaluationsupportscotland.org.uk/media/uploads/resources/supportguide1.2logicmodelsjul09.pdf>

Evaluation Toolbox – Programme Logic: A working guide to logic modelling for evaluators from an Australian team:
http://www.evaluationtoolbox.net.au/index.php?option=com_content&id=30

Magenta Book: Guidance on Evaluation. Published by HM Treasury this is cross-government guidance on effective evaluation practice:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

And for developing and applying theory of change:

Creating your Theory of Change: New Philanthropy Capitals Practical Guide:
<https://www.thinknpc.org/resource-hub/creating-your-theory-of-change-npcs-practical-guide/>

NESTA: Guidance on Developing a Theory of Change for Your Programme:
https://media.nesta.org.uk/documents/theory_of_change_guidance_for_applicants.pdf

Annex: Possible content of a logic chart or chain

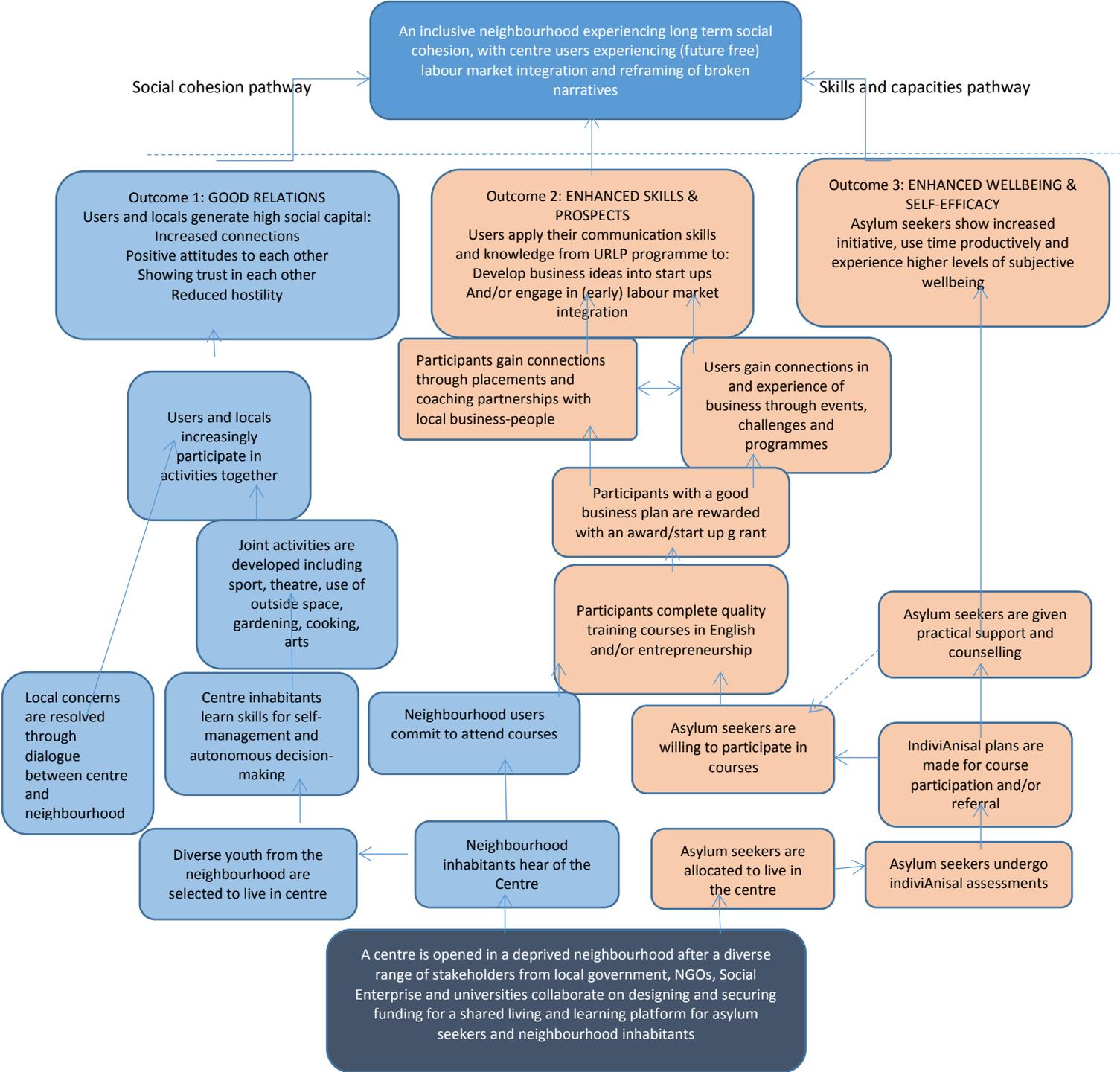
ISSUE: Intervention problem/challenge the intervention is expected to address	INPUTS: Inputs needed for the initiative or intervention to take place	ACTIVITIES: Activities and actions to take place as a result of the inputs	OUTPUTS: Deliverables resulting from the intervention activities	OUTCOMES: early 'consequential changes' expected from the outputs and deliverables	IMPACTS: Longer term consequential changes to meet long term vision for successfully addressing the intervention 'issue'
<ul style="list-style-type: none"> • Articulates the problem or challenge on which the intervention focuses • Sets out known dependencies, constraints and barriers to be tackled, etc – eg from past research. 	<ul style="list-style-type: none"> • Start/set up foundations for the intervention to take place • Includes resources (eg staff, time, budgets, etc) • Also provider engagement, partnerships, protocols and agreements needed for the proposed activities/quality to be set up. 	<ul style="list-style-type: none"> • All assumed component activities needed to underpin effective delivery of the intervention from the inputs • Sets out also activity inter-relationships and key delivery processes to underpin the actions including data/monitoring/QA 	<ul style="list-style-type: none"> • The key quantifiable deliverables resulting from (each) intervention activity • Sets out any targets for achievement(s) expected from the completion of the inputs and designated activity(s) (ie. what, by when, who by, where, what level, etc) 	<ul style="list-style-type: none"> • Intended <i>early</i> consequential changes and effects arising from the intervention (ie to be generated by the inputs+ activities + outputs) • May distinguish between 'short term' and 'medium term' outcomes • Includes direct and in-direct 	<ul style="list-style-type: none"> • The intended longer term (measureable) impact of the intervention • Impacts reflect how the intervention will successfully address the original issue/problem

				(knock on) outcomes	
<i>EXAMPLE: This might be a 'bullet point' synthesis on the problem or challenge, its scale and scope, known barriers to change and what (if any) past efforts to fix it are thought not to have worked.</i>	<i>EXAMPLE: The budget allocated for the intervention and what it does (and does not) fund; the interventions post-planning timescale and when it is expected to have achieved measureable change; etc.</i>	<i>EXAMPLE: The likely promotion/marketing or referral process to 'recruit' participants; actions to address the planned changes – training courses or new qualifications to be designed and run (by who), counselling or mentoring sessions to be run (and how counsels mentors are to be engaged); etc.</i>	<i>EXAMPLE: Achieved no. of participants (people, companies, etc) recruited; no of those expected to qualify or complete (after drop out); provider and/or practitioner no's trained to support programme; etc.</i>	<i>EXAMPLE: Raised participant awareness of ...; understanding of ...; increased levels of confidence for ...; etc (ie measureable 'soft' outcomes). Intermediary 'hard' outcomes could be included such as early behaviour changes among participants which indicate likely longer term impacts.</i>	<i>EXAMPLE: Indicators for sustained behaviour changes among participant's (ie contributing to the long term vision for change needed). EG - an unemployed young person achieving first job may be an intermediary outcome; the long term impact will be sustained working or employment for 6, 12 or months.</i>
-	-	NB All of the key (constituent) activities need be identified. These will usually need to be quantified (no's, frequency or range). This will set down any expected or contracted targets.	NB Focussing on key measureable achievement for (each) key deliverable. Don't forget to include both staged (eg interim) and/or final outputs.	NB Impacts can often need lead times to be realised; short and medium term outcomes may focus on early 'soft' 'impacts' (eg attitude change) and intermediary 'hard' outcomes.	NB. Impacts take account of lag/lead times to these being achieved and this may be tracked among participants post-intervention.

Resource C: Example of synthesised ToC for the Utrecht Refugee Launchpad

Theory of Change Graphic (summary)

The vision: To develop an inclusive approach to asylum seeker reception and integration, which beginning from day one connects newcomers with neighbourhood inhabitants through learning and living together, and encourages participation in a mutually supportive and cohesive neighbourhood, giving both constituents better opportunities for the future (See Resource A - with thanks to COMPASS).



Resource D: An example Theory of Change

Introduction

Following the *Review of Online Gambling* (2018) and 2019 national consultation, the Gambling Commission introduced a sector wide ban on the use of credit cards for gambling across Great Britain. The prohibition came into effect on 14 April 2020 and affects all 'land-based' and online gambling activities. To help with the roll out of the ban, and to underpin what will be a theory-based evaluation of it, the Commission has developed a *Theory of Change* (ToC) for the initiative¹.

The ToC sets out the circumstances surrounding the credit card ban, and is summarised below. This shows its aspirations and how it is expected to work, the aspirations for the ban and *explicit* and *implicit* assumptions which underpin it. It provides a *framework* for looking not just at how much and how well it is working (against expectations) but also if the founding assumptions and expectations were sound (and if they need to change). This 'framework' looks at each of:

- The **context** within which the ban will be introduced and the causal mechanisms which the ban needs to address
- The **expectations** of how the ban will bring about change and its outcomes
- The **assumptions** underpinning how it is expected to work and its effective implementation
- The likely disruptive **risks** which might circumvent or reduce the effectiveness of the ban.

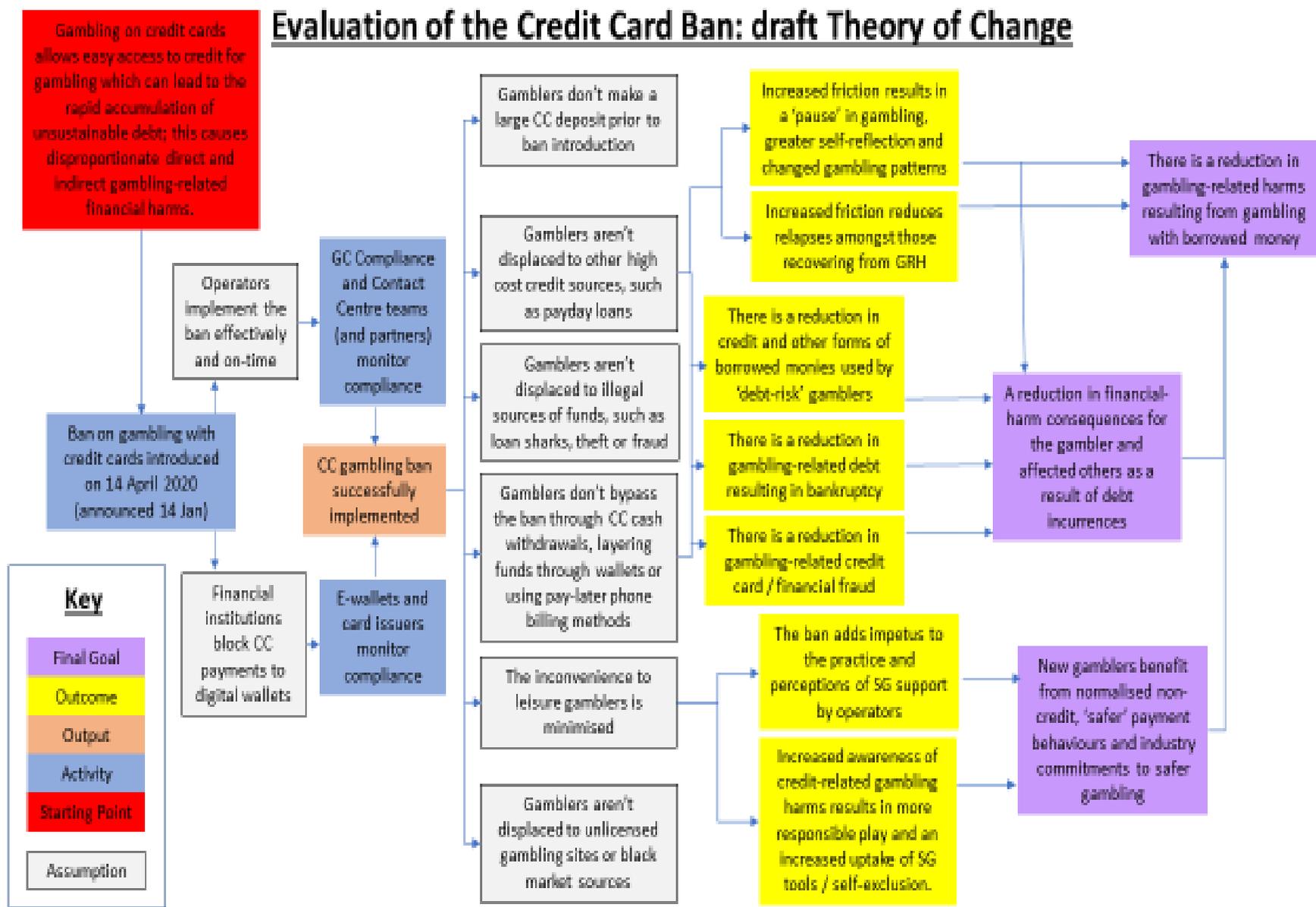
This CEAsR profile is the backbone of the 'theory of change'. The credit card ban aims to reduce gambling-related harms resulting from the use of borrowed monies. It is expected to constrain the ability of at-risk and problem gamblers to incur unsustainable levels of gambling debt by, for example, maximising borrowing up to the spending limits on their credit cards.

Evaluation of the ban and the Theory of Change

Evaluating the ban will need to look at changes after April 2020 in payment and borrowing practices for gambling activities, especially among at risk gamblers. It will also need to assess a range of other impacts including changing levels and nature of incurred debts, and set this within a complex environment of other influences on changed behaviours.

¹ The ToC has been developed with the Working Group over several weeks, informed by earlier research and the evidence from the national consultation and has also taken into account further comments from a number of stakeholders.

Evaluation of the Credit Card Ban: draft Theory of Change



This will need to go beyond unpicking a straightforward cause and effect and needs to be framed around the assumptions expected to shape the effectiveness of the ban. The ban may also have unintended consequences which are difficult to anticipate but which need to be taken into account to give a rounded view of its impacts and how these came about. The ToC is the starting point for this and the cornerstone of the subsequent evaluation.

To provide for this 'testable rationale', the ToC sets out the *Context – Mechanisms – Expectations - Assumptions* (CMEA) profile of the credit card ban. To this is added the known or likely risks to the effective working of the ban, such as at risk gamblers gravitating towards payday loans and overdrafts as new forms of credit. After the evaluation commences we expect emerging evidence will further refine the framework.

Context

Credit cards provide a convenient means of borrowing money to fund gambling and can facilitate high levels of gambling debt although research as yet can tell us little about the exact nature and direction of causation. Evidence does show a relationships between frequency and intensifying levels of credit card use and increased gambling problems.

There have been various credit related restrictions introduced outside the UK² although full prohibition on credit card use as a regulatory measure is unusual. Few borrowing related regulatory methods outside the UK have solid evaluation evidence to show effects but they do show that a credit card ban may have unintended effects such as substitution of credit card use with other forms of borrowing, such as overdrafts, unsecured high cost loans, or illegal activity to raise funds. In Britain there have been some restrictions on the use of credit cards for gambling which include, in particular:

- Established bans on the use of credits cards for some forms of gambling, notably for all gaming in casinos, bingo halls and arcades, and also for retail (direct over the counter) sales in licensed bookmakers (unless the customer holds an account with the bookmaker). These involve bans on the use of credit cards in particular types of gambling premises and do not cover all land-based gambling, or online gambling.
- Some credit card providers already ban or restrict the use of their cards for payment for gambling, including American Express but also some retailer credit cards (eg Asda credit cards which are issued by Creation Financial Services Ltd).
- Non regulatory and usually voluntary restrictions on credit card use for gambling activities such as the use of blocking software. This includes financial institutions, including the clearing banks, providing for customers to voluntarily add or remove gambling restrictions on their accounts to block transactions made to gambling related merchants. Such restrictions are applied individually to both credit and debit

² For example: financial institutions banning or limiting the use of credit cards for online gambling, voluntary blocking of card use for gambling transactions, limitations on the use of credit cards in high value gambling activity including casinos, restricting the number of payment options that can be used simultaneously and setting and enforcing maximum spend limits.

cards held in the customer's name (but not on behalf of any other party) and would include digital wallet transactions (eg Apple Pay, Google Pay or Samsung Pay). The restrictions are voluntary and can be removed by the customer at any time although most are subject to a cooling off period (e.g. 24 or 48 hours).

Against this background, the ban has been introduced as a 'levelling up' of some existing prohibitions and extending them to cover virtually all GB-based gambling.

More is known of the harms from problem gambling than the contribution of borrowing practices to causing these. Borrowing related harms are likely to be about incurring greater levels of debt than can be sustained through the financial resources of the player which in turn causes direct and indirect harms to gamblers and affected others (eg spouses, dependents, relatives, friends) resulting from the debt. While credit card use may only directly fund a small proportion of UK gambling activity (by value) it has disproportionate effects of harm for gamblers who are experiencing problems with their gambling behaviours. While the actual causal mechanisms are not absolutely clear, we can assume the harm (causal) mechanisms centre on:

- The ease by which credit cards can be obtained and used to gamble, including by those who may be at risk of using a large proportion of their credit facilities such that they incur unsustainable levels of debt.
- The large amounts of easily incurred credit that can be accessed through credit cards and the consequent potential levels of debt to be incurred.
- Providing a payment mechanism that facilitates continuous play without natural breaks.

Looked at in this way, the *cause* part of the cause-effect chain seems likely to involve:

- Easy access to payment for gambling through credit cards may allow gambling harms to persist or intensify.
- This easy access can cause direct and indirect harms by increasing the risk that unsafe gambling practices can put the gambler into levels of incurred debt beyond that which they can easily repay.
- This debt can accumulate faster than through other payment methods especially in quick-turn around and large volume gambling activities.

In terms of *effects*, harms are not limited to the direct consequences of unpaid incurred debt. Indirect effects include effects on affected others and are likely to include:

- Relationship harm
- Loss of employment and earnings
- Loss of agency including from credit agency and related restrictions
- Health related harms, primarily poor mental health
- Recourse to illegal activities including theft and fraud.

Another causal mechanism is that these and other harms are likely to be intensified where gamblers already have financial problems. This may include individuals who are experiencing problems with their gambling having already increased (optimized) their degree of borrowing from other sources to fund their gambling, including unsecured loans, pawning, and from illegal money lenders (loan sharks).

Expectations

The underlying expectation of the introduction of the ban is that it will achieve a step change in the reduction of harm from credit-based gambling, and especially for gambling related debt. There is no explicit expectation that it would reduce the overall levels of low-risk, moderate-risk or problem gambling according to the PGSI. In fact, some of those experiencing the highest levels of gambling-related harm may already not be able to hold credit cards due to persistently poor credit ratings, and as such the ban is unlikely to directly impact them.

Instead the expectation is that banning credit cards would have most effect on gamblers with impulsive or compulsive behaviours who systemically or occasionally resort to gambling through credit cards. Some will be at risk of generating levels of personal debt which may become unmanageable, or otherwise putting the individual into medium or long term gambling-related debt with consequent negative impacts. The ban may also prevent those currently experiencing lower levels of harm from similar impulsive credit card gambling which might otherwise lead them to experience greater levels of harm.

Reducing harm effects from debt in this way was expected to mean that more gambling activity was conducted affordably. More specifically, the regulatory ban was expected to:

- Cause a change in payment and gambling behaviours so that more gamblers gambled safely and were less likely to incur debt related harms.
- Increase the safety of payment behaviours for gambling, because of the 'frictions' which a credit card ban is expected to introduce into the process of individuals securing and using credit to fund gambling.
- Reduce credit card related financial fraud associated with some compulsive gamblers.
- See a net reduction in the overall levels of credit, and any other forms of borrowed monies, used by the cohort of gamblers who previously used credit cards to generate unsustainable debt.
- Have a preventative effect on new adult gamblers using credit to support unsafe gambling behaviours, especially among young adults. In particular, the ban would be normalised among those potential future gamblers who are about to turn 18, as they would not have experienced anything other than a ban.
- Reducing levels of recidivism among those recovering from gambling compulsion or other unsafe gambling practices, by reducing access to 'easy' credit which could otherwise facilitate recidivism.

In addition, it was expected:

- There would be measurable evidence of a reduced incidence of gamblers experiencing credit related gambling debt from unsafe gambling practices.
- There also would be measurable evidence of a reduced incidence of gamblers experiencing other credit related financial harms such as bankruptcy.
- More gamblers experiencing problems with their gambling could become aware of the harms from their gambling behaviours and register for single or multiple operator self-exclusion schemes (*NB. an indirect but nonetheless important expectation*)
- There could be a greater impetus to the practice and perceptions of safer gambling support by operators.

It was also expected that the ban would involve minimal disturbance to 'safe' gambling practices among recreational or leisure gamblers.

Assumptions

There are various 'assumptions' which appear to underpin the expected effectiveness of the ban. These can be loosely grouped into three separate but inter-related groups of working assumptions:

- **Start-up assumptions** related to operator and other preparations for the ban to be effective from 14 April
- **Technical**, mainly systems orientated assumptions
- **Assumptions** about changes to behaviour for gambling payment practices (ie needed to underpin the ban achieving its desired effects).

Start-up assumptions: It is assumed that operators within the regulatory remit in Britain will:

- Appropriately adapt and modify their systems to ban payment by credit cards.
- Organise and facilitate any dependencies on payment processors, so that the ban could be implemented for the start date.
- Organise and implement and any required changes to client terms and conditions (eg for accounts) and associated pre-ban communications with customers.
- Communicate the start of the ban, and how it will affect gambling payments, to all their customers in ways which minimise disruption to their customers
- Communications with customers about the credit card ban will also be safe and reliable i.e. they will not increase the potential for those at risk of harm seeking alternative forms of credit risks that would circumvent the ban.
- Operators will be motivated to ensure that the start-up requirements are in place and working, because of the combination of potential enforcement action and reputational harm.

Technical assumptions: Payment processors, including clearing banks, will be able to distinguish (and block) credit card payments for in-scope gambling related merchant category codes and will put in place any technical or capacity adjustments necessary to ensure compliance. This may require that:

- Gambling premises will be able to distinguish between credit card payments for gambling activity and the use of credit cards for non-gambling purchases such as food and beverage available on the same premises.
- Adjustments through merchant category codes will also be able to identify any credit card payments to society lottery operators (e.g. credit card payments for continuous payment authorities to participate in lottery prize draws) distinct from the use of credit cards for any non-gambling services provided by the society.
- Financial institutions providing credit cards will also be able to put in place, and communicate to credit card holders, the necessary changes to terms and conditions for consumers and within the required notice period (2 months as required by Payment Services Regulations)
- Adjustments will also be made by financial institutions to ensure that credit card payments into digital wallets are also blocked, including transactions by credit cards through intermediary processors such as Apple Pay, Google Pay or Samsung Pay.

Behavioural assumptions: Gamblers will experience a reduction in harm effects from credit-based gambling activity by:

- Gamblers currently using credit cards for all or some of their gambling activity will change their behaviour to use cash funds or other forms of payment not dependent on borrowing.
- Behaviour change among compulsive or impulsive gamblers currently using credit cards will be influenced by the increased *frictions* brought about by a credit card ban (ie removing their ability to use credit cards may help to introduce a 'pause' in the gambling experience)³.
- Gamblers using credit cards will not substitute other forms of payment for gambling which involve illegal sources of funds including through theft, loan sharks or fraud.
- Gamblers using credit cards will also not switch from regulated land-based or online operators to *black market* sources (such as online gambling sites which are not licensed to provide online gambling facilities in Britain⁴)
- Inconvenience caused by the credit card ban for leisure gamblers will be minimised with little or no disruption to their gambling activity.

³ That is, if the gambler is not able to use credit cards, but during gambling activity considers accessing other forms of borrowing to fund their gambling, the ban may afford a period of reflection before they arrange, access and secure alternative forms of credit to substitute for credit cards.

⁴ For example, non-licensed operators who nevertheless are able to advertise on Google etc and to otherwise make their facilities available to British users.

It is also assumed that the ban may have little or no effect on some highly compulsive and problem gamblers who already have substantial unsustainable debt (i.e. those who have had their credit cards blocked by their card issuers, or by gambling operators, or who cannot otherwise access credit cards due to a poor credit rating). It is recognised that some other problem and compulsive gamblers will always find a way to raise other funds.

The ban is chiefly aimed at those gamblers who are most likely to benefit from the frictions caused by the ban i.e. by having limited or no immediate access to alternative sources of credit, and where stopping a potential debt-incurring gambling action is easier for that gambler than finding a way to circumvent the lack of immediate credit. In particular, those experiencing lower or moderate levels of harm may be encouraged (by the friction, introduced by the credit card ban, in the journey of accessing and using borrowed funds to gamble) to reflect on their gambling expenditure before seeking other forms of funding their account with borrowed funds.

Disruptive risks

There is a lack of evidence from international sources on the processes affecting regulatory restrictions on gambling with credit, so we know little about what factors might constrain the effectiveness of the credit card ban – disruptive risks. However, in discussion we identified a series of potential disruptions (risks) to the working assumptions which included:

- Gamblers who had used credit cards finding sources of alternative, and possibly higher cost, credit.
- Gamblers could also anticipate the ban by using their credit cards to make a substantial deposit prior to the start of the ban. While this would not be a sustained risk it could have short term disruptive effects i.e. delaying or minimising the early effectiveness of a ban by skewing 'normal' behaviour in the first few months.
- Gamblers could 'displace' their direct use of credit cards by 'indirectly' using them to make monetary deposits into digital-wallets (e.g. where credit card funds are 'layered' from one digital wallet to another) or withdrawing cash from an ATM with a credit card and depositing that cash into their current account or onto a pre-paid card; and where the operator or e-wallet does not put in place mechanisms to identify where the payment originated from (eg to distinguish between payment by credit card from debit card or other 'cash' funds).
- Online gamblers in particular may switch to or intensify other forms of short-term credit which would substitute for credit cards, such as through their use of mobile phone payment methods which operate as a form of credit facility. These allow the consumer to make a payment such as a gambling deposit to an operator and charge that deposit to their mobile phone bill, and where the phone contract will be recorded as making the payment for gambling with the consumer paying the cost in arrears in their next (monthly) phone bill.
- The credit card ban could lead to an expansion of unregulated sources of gambling (e.g. black market) with adverse consequences for gambler safety and access to other harm minimisation initiatives such as self-exclusion.

The ability of the evaluation to isolate and identify impacts of the ban might also be complicated by other effects on gambling funding, such as limits to borrowing imposed across the board by financial institutions (or their regulators) or to caps on online spending.

There could also be longer term unintended consequences which we cannot anticipate, but which the design of the evaluation would need to be able to identify.

Where next for the ToC?

Put together the *Context – Expectations - Assumptions – Risks* (CEAsR) profile set out here, together with the implementation risks, makes up a preliminary ToC. As at July 2020 the start of the evaluation of the ban has been delayed due to the disruptive effects of COVID-19 on the land-based sector of gambling, but the ToC will be used to inform early planning work including baseline data collection. It remains a working model which will be refined as evidence emerges of the bans effectiveness and as changes to behaviours and any unintended consequences become clearer.

Resource E: Some analytical methods to support Theory-based Evaluation

What is it about; where does it fit?	Some pro's	Some Con's
Bayesian Updating		
<p>Uses <i>Bayes theorem</i> to help show the extent to which the TBE evidence supports contribution claims by assessing the probability of a contribution being valid. It can use different methods to estimate simple probabilities including empirical evidence, modelling or 'subjective probabilities' (via consensus). Bayesian Updating has been an established analytical feature for other professions and can be well-placed for assessing causality of outcomes from (mainly) qualitative contribution claims. It can help put some of the rigor back into evaluations which may lack quantification options, and is suited to evaluation of multi-activity programmes in complex settings. In the UK, DFID and others have used it in TBE contexts for impact evaluations.</p>	<ul style="list-style-type: none"> • Useful in multi-activity, complex settings where contribution claims cannot be directly observed and measured. • Can help evaluators test and build consensus for contribution from including stakeholders in assessing the strength of a causal contribution claim. 	<ul style="list-style-type: none"> • Needs experience and sensitive handling to be seen to work well. • Reliability of assessment depends on selected probabilities. • Not easy for some users to get to grips with (relies on formulas & probabilities). • Can be rigid and inflexible – testing parameters need to be anticipated early in methodology.
Contribution Analysis		
<p>CA has been around for some time but only recently gaining currency among evaluators. It focuses on assessing the likelihood of a contribution to an outcome or set of outcomes from the intervention itself. It provides for a</p>	<ul style="list-style-type: none"> • Highly flexible application – can be used in widely different intervention contexts and evidence mixes. 	<ul style="list-style-type: none"> • CA depends on the quality of the initial causal chain and the

<p>seven-step process for theory-development, testing and evidence based integration and review (some authors have condensed this to six steps).</p> <p>The approach is iterative and progressive process, mapping and refining a causal chain and adds additional evidence to that already available to look at how the contribution would have come about. It builds in allowance for other influencing factors and an embedded testing stage using 'knowledge others' (stakeholders) and can be used in diverse intervention settings.</p>	<ul style="list-style-type: none"> • Can be well fitted to evaluations where there is little scope for conventional counterfactual methods to assess contribution of an intervention. • Able to embed a theory of change, and use a progressive and iterative process to help critically review – or revise – it. • Well fitted to sensitive evaluations with multiple stakeholders who may hold varied of conflicting views about the effects, effectiveness or desirable outcomes of what is being evaluated. 	<p>enhancements from added evidence.</p> <ul style="list-style-type: none"> • Causal assessment stage is essentially subjective; lack of rigor may not give users used to quantification much confidence in the attribution. • Not well suited to evaluation contexts where there is a lot of variation in implementation, or changes over time.
<p>Contribution tracing</p>		
<p>Not a variant of CA, but rooted in hybrid of Process Tracing (see below) and Bayesian updating using mixed (quali-quant) methods. Uses participatory methods to establish outcomes traces and using probability-based validity of contribution claims.</p> <p>Unlike CA, CT is a rigorous method guided by explicit criteria for data collection and measuring confidence and probability assessment to quantify the level of confidence in a particular contribution claim. Like CA, it builds in consultation with 'critical friends' and relevant stakeholders.</p>	<ul style="list-style-type: none"> • CT is a focussed methods using only evidence likely to increase or decrease confidence in specific contribution claim. • Precise (guided) application contributes to the clarity and quality of the underpinning theory of change. • Confidence in the analysis is enhanced by appropriate use of 'critical friends' during the testing phase. 	<ul style="list-style-type: none"> • A rigorous method which needs systematic and careful handling especially of the undertaking schedule. • Not suited to short duration intervention with insufficient time for 'traces' to be realistically observed. • Needs considerable time and care to explore alternative explanations.

Most significant change

MSC is a highly participatory method with stakeholders embedded in the process of determining 'most significant' changes to be analysed. It involves the collection of 'change stories' and selection of the most significant of these stakeholder panels which review impacts and their determinants.

MSC is more commonly applied in monitoring actions but can be extended to TBE in complex intervention settings when outcomes are uncertain at the start of an evaluation, or where these vary widely across participants.

- Useful in multi-stakeholder contexts where stakeholders may not agree on which outcomes are important.
- Participatory depth can help build 'theory' understanding across stakeholders.

- Lacks quantification; may not be seen as credible to users who are used to numerical demonstrations of impact.
- Needs a constructive participatory context & well engaged stakeholders.
- Is time consuming and resource intensive - needs strong facilitation.

Process tracing

A structured method centred on individual cases of change (which can be used in multiples) to test if a causal-effect expectation explains the outcome being assessed. Evaluators 'trace' outcome and implications which would be expected if the causal chain (theory of change) being tested were true.

It uses intensive case study methods and can be combined with Bayesian updating (as above) to increase the rigour of causal claims. Uses various built in logical tests to help assess and demonstrate validity.

- Well suited where there is scope for case intensity and no available counterfactual.
- Well suited to embedded evaluation (*tracking forward* methods) or ex-post evaluation (*tracking back* methods).
- Better geared to small-n or segmented analyses involving single or small number (multiple) case situations.

- Very high intensity method – not well suited to interventions with variation in application or non-stable contexts.
- Needs high level of qualitative skills and used systematically and with rigour to prevent rater or inferential errors.
- Tracking back cases need a clear retrospective evidence trail and continuity of actors which may be challenged by organisation change or long time frames.

		<ul style="list-style-type: none"> Multiple process tracing cases risks complexity in explanation.
Qualitative Comparative Analysis		
<p>QCA is an established method which provides for systematic comparisons of outcome influences based on qualitative knowledge with some quantified testing to indicate reliability of assessments. It compares different aspects of an intervention effects set against contextual factors to identify various patterns and better understand the different characteristics (or combinations) linked to these.</p> <p>It is useful in complex settings where multiple influences need to be in place to achieve outcomes. Robustly applied it can identify success factors (and dis-enablers) and where these work in combination(s), and is well suited where there is expected to be considerable (eg geographic) variation in intervention effectiveness. It can work within a TBE by using the ToC to help anticipate factors of interest in transformation processes.</p>	<ul style="list-style-type: none"> Allows for both complex causation (combinations of factors) and multiple causes of an outcome to be accounted for usually in post hoc evaluation. QCA works best when data on all the cases of interest are available and the number of cases is neither too small nor too large, around ten to fifty cases. 	<ul style="list-style-type: none"> Not well suited where it needs larger numbers of cases for confidence in the analysis. Does not always provide for clear messages (eg which cases represent more 'success' or 'failure' than others). Not a participatory methods so may present challenges in building confidence in findings, and allowing for unobserved (alternative) explanations.

NB. This resource is adapted from the Magenta Guide, 2020 and includes supplementary material from the authors teaching materials and practical experience as an evaluation practitioner.