

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



**Methodology for evaluating complex social interventions: Understanding partner
violence prevention for young people in Mexico City**

SHELLY MAKLEFF

**Thesis submitted in accordance with the requirements for the degree of
Doctor of Philosophy
of the
University of London**

June 2020

Department of Public Health, Environments and Society

Faculty of Public Health and Policy

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funding received from the Sexual Violence Research Initiative / World Bank Group, International Planned Parenthood Federation/Western Hemisphere Region through a donation from a private donor, and the Advancing Learning and Innovation on Gender Norms (ALIGN) Research Fund.

Statement of own work

I, Shelly Makleff, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

A solid black rectangular box redacting the signature of the author.

Date: 1 June 2020

Acknowledgments

This thesis would not have been possible without the support and efforts of many people. First, I am deeply grateful to my supervisory team. Cicely Marston has provided impeccable advice at every step along the way, from conceptualising the study, to joining me in Mexico for the pilot, through writing retreats and publications. She has consistently encouraged me to improve and refine my text, and move it to another level of abstraction. Working with Cicely over the last six years has made my PhD, for lack of a more sophisticated word, fun – and I hope our collaboration doesn't stop here. Mark Petticrew stepped in midway as my second supervisor and enriched the project with his reflections about complexity and evaluation, drawn from years of contributions to the field. His publications have influenced my thinking from the beginning of the PhD process, and I am deeply appreciative of the opportunities I've had over the last few years to discuss my project with him and learn from his thoughtful comments and guidance. In addition, I'd like to thank Antonio Gasparri for serving on my advisory committee.

This research is centred on an evaluation in Mexico, and I would like to acknowledge Jovita Garduño, Icela Zavala, Ivon Silva and Doroteo Mendoza, as well as their many dedicated colleagues at Mexfam, for the collaboration. Their insights and dedication ensured that this study was grounded in the reality and needs of Mexfam and the young people they serve. A special thanks to Jovita, who overcame earthquakes, torrential rains, and long commutes to build strong relationships with the study site and conduct the fieldwork for this study; and to the *Gente Joven* team for their willingness to complicate the already challenging process of providing comprehensive sexuality education by engaging in this study. And of course, my deepest gratitude to the staff and students at the Cetus 154 school in Mexico City for their warm welcome, enthusiasm throughout the study, and for entrusting me with their experiences and stories. My past and present colleagues at IPPF/WHR have also played an important role. Jimena Valades and Rebecca Koladycz are the reason I became involved in the study in Mexico in the first place, and they have been thought partners along the way. Marissa Billowitz, Mariana Cruz, Florencia Barindelli, and Susana Medina have brought their years of programmatic experience to the process of sensemaking and dissemination, and as co-authors.

I am extremely grateful for the funding received for this project from Mr. Stanley Eisenberg through his generous donation to IPPF/WHR, the ALIGN research fund in support of analysis and dissemination, and the Sexual Violence Research Initiative World Bank Group Development

Marketplace Award 2019 for writing, dissemination, and translation of evidence to practice. I'd also like to thank the anonymous reviewers of the articles that comprise the core of this thesis, whose comments invariably have improved the work.

Finally, I'm indebted to my family and friends, who have provided encouragement and distraction along the way. Ron was a surprisingly willing Gantt-chart buddy, and my parents have unsurprisingly been as supportive as ever. Grace Besharah has been a patient grammar guru, and Ale Garita and Meradith Leebrick opened their doors so I had two homes away from home in Mexico City. My friends who also happen to be researchers have graciously engaged in many impromptu conversations about this PhD over the last six year – in particular, Sarah Baum, Brianna Keefe-Oates, Rupal Sanghvi, Karen Bailey, Carolyn Birdsall, and Julia Bailey. Heidi Moseson gave me a crash course in list experiments, and Kemi Akinfaderin has shown endless patience and provided comic relief as I disappeared into the final months of writing. I'd also like to thank my Amsterdam community for forgiving all the last-minute cancellations while I was 'in the zone' with writing. Last but not least, GG has gracefully toed the line between supportive partner and provider of sage academic advice, while (usually) managing not to treat me as if I was one of his students. Thank you for everything, from encouraging me to embark on this journey to patiently talking things out over many meals and hikes. One step closer to retirement, one step closer to Tickle Me Elmo?

Abstract

Many evaluation approaches do not account for temporality and complexity. This thesis is a methodological examination of evaluation techniques based on a case study, which was designed with a dual purpose: first, to evaluate a sexuality education programme with a focus on intimate partner violence (IPV) prevention in Mexico City; and second, to provide empirical data for the thesis.

The case study was an evaluation with a longitudinal quasi-experimental design. Data collection methods were semi-structured in-depth interviews, focus group discussions, self-administered questionnaires, and observations of the intervention. I used thematic analysis to examine intervention effects.

The methodological exploration used a qualitative observational design based on the case study, exploring questions about the utility of qualitative longitudinal and complex adaptive systems approaches in evaluation and how qualitative and quantitative approaches to data collection compare. Evaluation data collected in Mexico served as raw data, and I wrote fieldnotes about the evaluation process. I used framework analysis, applied a complex systems approach, compared data collected through different methods, and identified barriers to high-quality data.

In the evaluation, we found evidence that the intervention contributed to changes in beliefs, intentions, and behaviours related to gender, sexuality, and IPV. The methodological analysis showed that repeat interviews illuminated how the intervention influenced relationship trajectories and provided contextualised data about lived experiences. A complex adaptive systems approach helped us examine system-disruptive elements of the intervention. Challenges to data collection included earthquake-related delays, social complexities, the shifting nature of relationship experiences, and variability in motivation to participate in the study. A reflexive discussion of such barriers to high quality data should inform interpretation of research findings.

I argue that evaluation methods should be designed to engage with unpredictability, interaction, temporality and change and should centre on building contextualised understanding of pathways to impact. Evaluations should engage stakeholders and beneficiaries to ensure relevant research questions and define what 'meaningful' evidence entails; this will facilitate utilisation of findings.

Table of Contents

STATEMENT OF OWN WORK	2
ACKNOWLEDGMENTS	3
ABSTRACT	5
TABLE OF ABBREVIATIONS	8
LIST OF TABLES AND FIGURES	9
CHAPTER 1: INTRODUCTION	10
1.1 BACKGROUND.....	10
1.1.1 <i>The politics of evidence</i>	10
1.1.2 <i>Methodology for the evaluation of complex and social interventions</i>	14
1.1.3 <i>Gender-transformative programming for the prevention of intimate partner violence</i>	19
1.2 AIM AND OBJECTIVES.....	24
1.3 RESEARCH QUESTIONS	25
1.4 STRUCTURE OF THE THESIS.....	25
1.5 THEORETICAL FRAMEWORKS AND CONCEPTUAL MODELS INFORMING THIS THESIS.....	27
1.5.1 <i>Complexity theory and complex adaptive systems</i>	27
1.5.2 <i>Social norms theory</i>	28
1.5.3 <i>The gender system</i>	29
1.5.4 <i>Social-ecological model for partner violence</i>	29
1.5.5 <i>The use of theoretical frameworks and models throughout this thesis</i>	30
CHAPTER 2: INTERVENTION DESCRIPTION AND SETTING	33
2.1 STUDY PARTNERS	33
2.2 COLLABORATION MODEL	34
2.3 DESCRIPTION OF THE INTERVENTION	35
2.4 INTERVENTION THEORY OF CHANGE	36
2.5 INTERVENTION SETTING.....	37
2.6 DISSEMINATION.....	38
CHAPTER 3: METHODS	40
3.1 STUDY DESIGN.....	40
3.1.1 <i>Disciplinary and epistemological perspective</i>	40
3.1.2 <i>Research approach</i>	41
3.1.3 <i>Thesis components</i>	45
3.1.4 <i>Study design for thesis component 1: Evaluation of IPV prevention intervention in Mexico</i>	46
3.1.5 <i>Study design for thesis component 2: Methodological exploration</i>	47
3.2 DATA COLLECTION.....	47
3.2.1 <i>Overview of data collection methods for thesis component 1: Evaluation of IPV prevention intervention</i>	47
3.2.2 <i>Data collection methods for thesis component 2: Methodological exploration</i>	56
3.3 DATA ANALYSIS	57
3.3.1 <i>Thematic analysis based on programmatic theory of change</i>	57
3.3.2 <i>Quantitative analysis of questionnaire data</i>	59
3.3.3 <i>Qualitative longitudinal analysis using framework method</i>	59
3.3.4 <i>Iterative ground approach to analysis for methodological exploration</i>	61
3.3.5 <i>Post-hoc analysis based on social complex adaptive systems concepts</i>	62
3.3.6 <i>Comparison of self-administered questionnaire and one-on-one interview data</i>	63
3.4 ETHICS.....	66

3.4.1	<i>Ethical considerations for thesis component 1: Evaluation of IPV prevention intervention in Mexico</i>	66
3.4.2	<i>Ethical considerations for thesis component 2: Methodological exploration</i>	70
3.5	REFLEXIVITY	71
3.5.1	<i>Motivation for this study</i>	71
3.5.2	<i>My role as researcher</i>	72
CHAPTER 4: PREVENTING INTIMATE PARTNER VIOLENCE AMONG YOUNG PEOPLE—A QUALITATIVE STUDY EXAMINING THE ROLE OF COMPREHENSIVE SEXUALITY EDUCATION		
4.1 INTRODUCTION		
4.2 ABSTRACT		
4.3 RESEARCH PAPER 1		
CHAPTER 5: EVALUATING COMPLEX INTERVENTIONS USING QUALITATIVE LONGITUDINAL RESEARCH: A CASE STUDY OF UNDERSTANDING PATHWAYS TO VIOLENCE PREVENTION		
5.1 INTRODUCTION		
5.2 ABSTRACT		
5.3 RESEARCH PAPER 2		
CHAPTER 6: APPLYING A SOCIAL COMPLEX ADAPTIVE SYSTEMS APPROACH WHEN EVALUATING IMPLEMENTATION OF A SCHOOL-BASED INTERVENTION FOR INTIMATE PARTNER VIOLENCE PREVENTION: A CASE STUDY IN MEXICO		
6.1 INTRODUCTION		
6.2 ABSTRACT		
6.3 RESEARCH PAPER 3		
CHAPTER 7: COLLECTING BETTER DATA ON SEXUALITY, RELATIONSHIPS AND VIOLENCE IN SCHOOLS: EMPIRICAL EVIDENCE FROM AN EVALUATION		
7.1 INTRODUCTION		
7.2 ABSTRACT		
7.3 RESEARCH PAPER 4		
CHAPTER 8: DISCUSSION		
8.1 SUMMARY OF RESULTS		
8.2 IMPLICATIONS FOR THE EVALUATION OF SOCIAL AND COMPLEX INTERVENTION		
8.2.1	<i>Data collection challenges when examining sexuality, relationship and violence</i>	179
8.2.2	<i>Complementary and alternative evaluation approaches</i>	180
8.2.3	<i>Co-production of evaluation</i>	183
8.3 CONCEPTUALISING ‘COMPLEXITY’ IN EVALUATION PRACTICE		
8.4 A TYPOLOGY OF CHANGE FOR ENGAGING COMPLEXITY IN EVALUATION		
8.5 THE ROLE OF THEORETICAL FRAMEWORKS IN THE THESIS		
8.6 CONTRIBUTION OF THE THESIS		
8.7 STRENGTHS AND LIMITATIONS OF THE STUDY		
8.8 CONCLUSION		
REFERENCES		
APPENDICES		

Table of abbreviations

CAS – Complex adaptive systems

CSE – Comprehensive sexuality education

EBM – Evidence-based medicine

GBV – Gender-based violence

IPPF/WHR – International Planned Parenthood Federation / Western Hemisphere Region

IPV – Intimate partner violence

LMIC – Low- and middle-income countries

LSHTM – London School of Hygiene and Tropical Medicine

Mexfam – Fundación Mexicana para la Planeación Familiar

MRC – Medical Research Council

NGO – Non-governmental organisation

RCT – Randomised controlled trials

SRH – Sexual and reproductive health

TOC – Theory of change

WHO – World Health Organization

List of tables and figures

List of tables

TABLE 1: RATIONALE FOR MIXING DATA COLLECTION METHODS AT DIFFERENT POINTS OF INTERFACE IN THE STUDY.....	43
TABLE 2: SUMMARY OF RESEARCH QUESTIONS FOR EACH THESIS COMPONENT AND CHAPTER	45
TABLE 3: SUMMARY OF DATA COLLECTION METHODS IN EVALUATION OF IPV PREVENTION INTERVENTION.....	48
TABLE 4: PURPOSE FOR EACH QUESTIONNAIRE ITEM	54
TABLE 5: TIMELINE OF DATA COLLECTION ACTIVITIES, BY METHOD AND GROUP	55
TABLE 6: QUESTIONNAIRE ITEMS ABOUT RELATIONSHIPS, SEXUALITY AND VIOLENCE	63
TABLE 7: CHARACTERISTICS AT BASELINE OF PARTICIPANTS IN MEXFAM'S COMPREHENSIVE SEXUALITY EDUCATION INTERVENTION.....	84
TABLE 8: AGE AND SEX OF INTERVIEW AND FOCUS GROUP DISCUSSION PARTICIPANTS, BY DATA COLLECTION METHOD.....	85
TABLE 9: TIMETABLE FOR LONGITUDINAL QUALITATIVE INTERVIEWS.....	105
TABLE 10: DEFINITIONS OF KEY COMPLEX ADAPTIVE SYSTEMS TERMS	127
TABLE 11: CHARACTERISTICS OF INTERVENTION GROUPS IN THE STUDY	131
TABLE 12: QUESTIONNAIRE ITEMS ABOUT RELATIONSHIPS, SEXUALITY AND VIOLENCE (DUPLICATE FROM CHAPTER 3).....	154
TABLE 13: DIFFERENCE IN PROPORTION OF COMPLETE RESPONSES BETWEEN MEN AND WOMEN, BY GROUP AND TIMEPOINT	162
TABLE 14: EXTENT OF CONSISTENCY BETWEEN QUESTIONNAIRE AND INTERVIEW DATA	167
TABLE 15: POTENTIAL REASONS FOR MISSING, INCOMPLETE AND INCONSISTENT DATA OBSERVED IN THIS STUDY	168
TABLE 16: CORE QUESTIONS THAT ENGAGE WITH COMPLEXITY IN EVALUATION	187
TABLE 17: TYPOLOGY OF CHANGE TO ENGAGE WITH COMPLEXITY IN EVALUATION	188

List of figures

FIGURE 1: REVISED CONCEPTUAL FRAMEWORK FOR PARTNER VIOLENCE (FROM HEISE, 2011).....	30
FIGURE 2: LEVEL OF INVOLVEMENT FOR EACH PARTNER ORGANISATION, BY STUDY PHASE	34
FIGURE 3: THEORY OF CHANGE DEVELOPED TO INFORM RESEARCH DESIGN.....	37
FIGURE 5: PROPORTION OF COMPLETE RESPONSES TO QUESTIONNAIRE ITEMS ABOUT SENSITIVE TOPICS, BY GROUP, TIMEPOINT AND GENDER	162
FIGURE 6: PROPORTION OF COMPLETE RESPONSES TO QUESTIONNAIRE ITEMS ABOUT SENSITIVE TOPICS, BY QUESTION AND GENDER	163
FIGURE 7: DIFFERENT FACETS OF COMPLEXITY ENCOUNTERED IN THE EVALUATION IN MEXICO.....	186

Chapter 1: Introduction

The aim of this thesis is to explore methods and approaches suitable for the evaluation of social and complex interventions, using as a case study the evaluation of an intimate partner violence (IPV) prevention programme implemented by a community-based organisation in Mexico. The case study was designed with a dual purpose in mind: first, to identify the mechanisms through which an intervention helps prevent IPV in Mexico; and second, to provide data for an exploration of methodology suitable for evaluating programmes that engage complex social issues in the context of international development and global health programmes.

In this chapter, I provide theoretical background about the politics of evidence, methodology to account for complexity in the evaluation of social interventions, and gender-transformative programming for the prevention of IPV. I then describe the aim and objectives, research questions, and structure of the thesis. Finally, I present the theoretic and conceptual frameworks that inform this work.

1.1 Background

This thesis lies at the intersection of three main subjects: the politics of evidence, methodology to account for complexity in the evaluation of social interventions, and gender-transformative programming for the prevention of IPV. Below I provide an overview of the background for each of these topics before describing their intersection as it relates to the particular study aims and objectives.

1.1.1 The politics of evidence

The choice of research method has long been influenced not only by scientific standards but also by politics and disciplinary or philosophical perspectives (Oakley, 1999). From the 1960s, methodological debates placed quantitative and qualitative methods in opposition with each other – the so called ‘paradigm wars.’ Beginning in the 1970s, the feminist movement, among other political and social forces, influenced a prioritisation of qualitative methods and distrust of quantitative methods for feminist research and in some branches of the social sciences, adding a gendered dimension to perceptions of methodological rigour (Oakley, 1999).

Over the last two decades, however, the divergence between quantitative and qualitative schools of thought has declined, bringing with it a growing use of ‘mixed methods’ research, which integrates qualitative and quantitative approaches to data collection, analysis and interpretation (Bryman, 2006b; Guest, 2013; Mertens and Tarsilla, 2015). While there are ongoing debates about how to integrate potentially incompatible research paradigms within mixed methods research (Hall, 2013; Shannon-Baker, 2016), research drawing on different methods and paradigms has become commonplace, and is increasingly a central approach to evaluation research (Pluye and Hong, 2014; Mertens and Tarsilla, 2015). More recently there has also been an expansion of synthesis approaches that incorporating qualitative and mixed methods studies in addition to the more traditional syntheses of only experimental or qualitative studies (Pluye and Hong, 2014). Different synthesis approaches that draw on a range of types of studies can be particularly relevant for learning about social and complex interventions (Petticrew *et al.*, 2013) – which may be more difficult to understand using only one type of data.

The adversarial debates that characterised the paradigm wars have largely ended (Sinkovics and Alfoldi, 2012), with broad consensus on the utility of a wide range of methods to be used to address different research questions (Stern *et al.*, 2012; Lambert, 2013). In evaluation research too, what can be referred to as the ‘first-generation’ impact evaluation question of ‘does it work?’ (White, 2014) has given way to a broader range of evaluation questions. Realist evaluation, for example, addresses ‘what works, for whom and under what circumstances’ (Pawson and Tilley, 1997) – questions that have increasingly been incorporated into evaluations (Adam *et al.*, 2012; Bonell *et al.*, 2012). Addressing this range of evaluative questions would necessitate a range of evaluation methods and approaches.

Despite the broad acceptance of mixed method approaches to address a range of evaluation and research questions, some researchers describe ways in which qualitative data may still be treated as less rigorous than quantitative data (Giddings and Grant, 2007; Howarth *et al.*, 2016). One author recently reflected that “the paradigm wars are alive and well in many of our disciplines, worldwide”, with “qualitative researchers [...] put on the defensive when it comes to their research practices” (Given, 2017) – though this may manifest in subtle ways. For example, reviewers of publications or proposals commonly apply the quality standards of quantitative research when reviewing qualitative publications or proposals, which can lead to the questioning

of sample sizes or lack of researcher objectivity, or may imply in their remarks that the main value of qualitative methods is to inform or support the superior quantitative methods (Given, 2017).

The prioritisation of certain types of evidence is similarly visible in the fields of international development and global health, where an emphasis is placed on demonstrating measurable programme impact (Ebrahim and Rangan, 2014) and using quantitative results to ensure accountability and provide evidence about 'what works' (Eyben, 2013). This 'results and evidence agenda' – which drives the production of evidence and results – is promoted by a broad range of organisations such as the World Health Organization and World Bank, as well as governments, non-governmental organisations (NGOs), and private philanthropy (Eyben, 2013). These actors are influential in determining how programme success should be defined and measured, often dictating from the top down what should be considered meaningful in NGO and development work (Eyben, 2013; Adams, Craig and Samen, 2016). The donor-driven reporting requirements and evidence demands that accompany this focus on quantitative results and evidence have been critiqued for their incompatibility with the work of many development and global health organisations (Eyben, 2013; Adams, Craig and Samen, 2016). Despite the frequent mismatch between donors' evidence requests and organisational capacity and needs, many organisations comply with donor demands due to their 'desperate desire to get the grant' (Eyben, 2013).

Organisations may design evaluations to meet donor demands rather than their own needs, risking a waste of limited NGO resources when the evaluation findings are not utilised (Liket, Rey-Garcia and Maas, 2014). NGOs generally are under more pressure to demonstrate 'upwards accountability' to the donor than 'downwards accountability' to the beneficiary or stakeholders (Ebrahim, 2003). One notable exception is an ongoing donor initiative, the Fund for Shared Insight, which funds projects focusing on mechanisms of accountability to programme beneficiaries with a conscious awareness of power relations and silenced versus privileged voices (*Fund for Shared Insight*, no date). However, this model does not represent the standard approach of most donors, who often expect grantees to demonstrate 'impact' without putting resources such as general support or substantial evaluation funds towards that demand (Ebrahim and Rangan, 2014). Small NGOs in particular may struggle to fulfil the demands to demonstrate certain forms of results and evidence; these forms of data collection and evidence generation are often unfeasible given organisational resources and staff capacity, incompatible with the needs and priorities of

programmatic work on the ground, or poorly suited to capturing the type of work that NGOs are carrying out (Adams, Craig and Samen, 2016).

A range of challenges faced by NGOs can contribute to low quality evaluation. Some of the barriers to high quality evaluative evidence include limited funding and human resources, the complexity and unpredictability of programmes, having insufficient technical expertise to use statistical methods, programmatic time constraints, and the pressure to show positive findings (Merchant-Vega, 2011; Roetman, 2011; Bamberger, Rugh and Mabry, 2012; Lakshman *et al.*, 2014). These factors, among others, have contributed to critiques of low quality evaluations in the field of international development – the so-called ‘evaluation gap’ (Savedoff, Levine and Birdsall, 2006). Poorly conducted evaluations have ‘no benefits, and can easily do harm to programmes and their stakeholders by creating inaccurate or misleading information’ (Braverman, 2012). For example, poor quality evaluation findings can result in the expansion of programmes that do not work, discontinuation of those that do work well, and lost opportunities for programmatic improvements and learning (Bamberger, 2007).

Some authors argue that the evidence-focused agenda promoted by donors, governments and agencies reflects a bias towards quantitative metrics and randomised controlled trials (RCT) as the most rigorous forms of evidence in global health and international development (Eyben, 2013; Adams, Craig and Samen, 2016). This aligns with the hierarchy of evidence adopted from evidence-based medicine (EBM), which similarly considers RCTs as the ‘gold standard’ in terms of research rigour (Petticrew and Roberts, 2003; Cowen and Cartwright, 2019). There has been a growing focus on ‘rigorous’ evaluation to fill the ‘evaluation gap’ in international development (Savedoff, Levine and Birdsall, 2006). However, certain methods have been privileged over others in this process. Advocates of experimental methods have seemingly co-opted the term ‘rigorous’ as being almost synonymous with RCTs, sometimes to the point of excluding other methods from consideration (Ravallion *et al.*, 2018; Stern *et al.*, 2012).

The primacy of experimental methods sometimes appears to distract from careful thought about the appropriateness of those methods to the research question. While many evaluation guidelines state that the research or evaluation question should precede and inform the choice of appropriate study design and data collection method (AusAID Office of Development Effectiveness (ODE), 2012; Craig, Dieppe, Macintyre, Michie, *et al.*, 2008; Stern *et al.*, 2012), in practice there

may be pressure to demonstrate the ‘impact’ of programmes (Ebrahim and Rangan, 2014), often using experimental methods (Petticrew, Chalabi and Jones, 2012). This pressure is premised on the inaccurate ‘assumption that the quality of evaluations can be judged solely on the basis of their design’ (Liket, Rey-Garcia and Maas, 2014), which risks that method choice will be based solely on the hierarchy of evidence. Instead, evaluation design should ensure ‘methodological aptness’ – selecting methods based on the premise that ‘different types of research question are best answered by different types of study’ (Petticrew and Roberts, 2003). For example, qualitative methods, while low on the hierarchy of evidence supported by evidence-based medicine, may well be the best suited and most rigorous option for answering questions about lived experiences, such as how people make health-related decisions or experience interventions; these are questions that RCTs and other quantitative methods are ill suited to answer (Popay, 1998; Petticrew and Roberts, 2003). Indeed, there are many circumstances in which experimental or statistical designs and methods are not the best method for evaluating social and complex interventions (Bamberger et al., 2012; Christ, 2014; Picciotto, 2014; Stern et al., 2012).

1.1.2 Methodology for the evaluation of complex and social interventions

Evaluators and donor agencies have begun to acknowledge that RCTs may be unsuitable for the majority of evaluation questions relevant to international development and global health (AusAID Office of Development Effectiveness (ODE), 2012; Bamberger et al., 2012; Stern et al., 2012). In the words of a proponent of this view, ‘the sudden popularity of experimental methods in development evaluation is counterintuitive. Whereas these methods are best suited to the assessment of simple and stable programmes, the development enterprise is mostly made up of complex, adaptable interventions implemented in volatile environments’ (Picciotto, 2012). Many critiques of experimental methods for evaluating complex interventions reflect the failure of this approach to answer questions about how an intervention works (Cartwright and Hardie, 2012). This is sometimes called the ‘black box problem’ of evaluation, referring to the ‘practice of viewing social programs primarily in terms of effects, with little attention paid to how those effects are produced’; this is in contrast to theory-based evaluation, which attempts ‘to unpack programmatic “black boxes” and explain how and why programs work (or fail to work) in different contexts and for different program stakeholders’ (Astbury and Leeuw, 2010). There is growing consensus that evaluations of social complex interventions – regardless of the methods being used – would benefit from theorising how interventions work and understanding which components of the

intervention contribute to change and in what situations and contexts (Bonell *et al.*, 2012; Silva *et al.*, 2014; Moore *et al.*, 2015; Howarth *et al.*, 2016).

Responding to these methodological challenges, efforts are underway to provide guidance on how to best evaluate social and complex interventions. In 2000, the United Kingdom Medical Research Council (MRC) published the Framework for the Development and Evaluation of RCTs for Complex Interventions to Improve Health, which has been extensively cited (Medical Research Council, 2000). In 2008 these guidelines were updated with a 'more flexible, less linear' process for developing and evaluating social and complex interventions; the revision provided successful worked examples of evaluations using a diversity of research methods (Craig, Dieppe, Macintyre, Michie, *et al.*, 2008). The revised MRC guidelines, as well as other publications and working papers addressing the evaluation of social and complex interventions, presented a broad set of research designs and analysis possibilities for situations where traditional RCTs might not be suitable. These include innovations in statistical modelling and use of statistical methods such as propensity scores, instrumental variables, interrupted time series studies, and regression dilution bias (Bonell *et al.*, 2011; Cousens *et al.*, 2011; Stern *et al.*, 2012). Others have proposed that RCTs could adopt a realist approach to engage with theory while determining the effects of complex interventions (Bonell *et al.*, 2012). This growing literature, which examines the potential of a broadening scope of methods for evaluation practice, is suggestive of a gradual shift in the hierarchy of evidence as it relates to impact evaluation.

However, the approaches examined in the methodological guidance documents and working papers described above rarely critique the 'results and evidence agenda' itself or the power imbalances implicit in how 'rigour' is defined. Of those who propose a broadening of evaluation methods, many maintain that experimental methods may be the 'gold standard' but acknowledge that they are not always appropriate or possible, and therefore consider statistical or quasi-experimental alternatives to RCTs (Bonell *et al.*, 2011; Cousens *et al.*, 2011; Stern *et al.*, 2012). What is considered to be rigorous or meaningful evidence is generally defined from the perspective of those who have the power to dictate what methods and approaches are prioritised and funded (Eyben, 2013).

With some exceptions – notably among feminist research approaches (Adams, Craig and Samen, 2016; Leung *et al.*, 2019), much of the literature exploring alternatives for evaluation and research

methods for complex interventions does not question whether other voices, such as beneficiaries or field-based partners, might value different forms of evidence and different conceptions of success. Reflecting on the challenges of demonstrating ‘success’ in global maternal and child health projects, Adams et al. suggest that experimental and statistical methods may not be suited to assess the types of work that is having impact on beneficiaries; accountability to partners may necessitate a shift away from the methods promoted through the current dominant perceptions of rigour.

We might need to shift how we talk about evidence in ways that are accountable to our partners, and in ways that are accountable to different kinds of familiar but too often invisible and unquantifiable things. This includes incorporating cultural knowledge in intervention designs [...]. Our goal [...] is not so much to argue against the RCT but to argue for the use of different kinds of models that count outcomes and think of evidence in different ways. (Adams, Craig and Samen, 2016)

This approach to defining alternative forms of meaningful evidence based on partners’ priorities is in line with feminist approaches to research (Leung *et al.*, 2019) and related critiques of colonial knowledge hierarchies that are based on and prioritise Eurocentric ideas while studying indigenous communities or programmes based in the Global South (Mohanty, 1988; Smith, 2012). Similarly, for research aiming to address social injustice, ‘there is a need to unsettle traditional concepts of what counts as research, as evidence, as legitimate inquiry’ (Denzin, 2019).

In addition to the power imbalances that may be perpetuated through donor-initiated demands for particular forms of evidence (Eyben, 2013), there are also barriers to the adoption of the experimental and quasi-experimental methods promoted in recent evaluation guidelines as well as novel statistical approaches to evaluation (Merchant-Vega, 2011). For example, the MRC guidelines for evaluation of complex interventions are unlikely to be feasible for small NGOs working on local scales, or for researchers working in under-resourced organisations in low- and middle-income countries – who may lack the resources to use such evaluation techniques (Lakshman *et al.*, 2014). In addition, programmes aiming to influence social change face particular challenges in measuring intangible and shifting changes in norms (Costenbader *et al.*, 2019). Yet evaluations of social change initiatives often use standard assessment tools or evaluation models to quantify such changes without questioning whether the assumptions of these tools – such as that of linearity – are applicable to processes of social change (Lacayo, Obregón and Singhal,

2008). One commonality within the debates about appropriate evaluation methodology and hierarchies of evidence is the acknowledgement that complexity should be addressed when evaluating social and complex interventions, regardless of field and topic. How to do so, however, remains a subject of ongoing discussion.

In evaluation and programme planning, complexity can be considered at various levels, including related to the intervention itself, the type of human behaviour or belief system that it aims to influence, and the context or setting in which it is implemented. First, complex interventions are commonly defined as having multiple interacting programme components (Campbell, 2000) – for example, working with individuals or families as well as intervening on a community-level through public awareness. Second, complexity can also reflect the topic that the intervention aims to influence. Interventions that aim to shift complex social phenomenon – as in the case of gender-based violence prevention programmes – are by nature complex, reflecting human behaviour and the influence of social norms (Zimmerman *et al.*, 2016). Finally, complexity can be conceived of as a characteristic of the system or setting in which an intervention takes place (Shiell, Hawe and Gold, 2008; Hawe, 2015). For instance, unstable political situations, election cycles, changes in local authorities and local and international organisational leadership, weather events, and shifting social norms are among the myriad contextual factors that can affect intervention implementation, often in unpredictable ways.

Contextual information is crucial to consider in evaluation because an intervention may work in one setting but not elsewhere (Craig, Dieppe, Macintyre, Mitchie, *et al.*, 2008) or may successfully influence its context, thereby making the programme mechanisms less relevant in the future, even in the same location (Moore and Evans, 2017). There is growing consensus that interventions and their context mutually interact and influence, and that programme theory needs to draw on deep local knowledge of contextual factors to help understand this (Howarth *et al.*, 2016; Moore and Evans, 2017). Despite the growing consensus that context must be taken into account when interpreting evaluation results, intervention components and their context are rarely reported on in detail (Howarth *et al.*, 2016), perhaps reflecting word count limits in journals or a lack of guidance on such reporting; proposed guidelines for describing interventions do not directly address reporting on context (Hoffmann *et al.*, 2014).

However, strategies have been proposed to ensure that context is taken into account when conducting evaluation of social and complex interventions. Evaluation research conducted in partnership with stakeholders, who have a nuanced understanding of the intervention in context, can benefit evaluation and research practice (Zimmerman *et al.*, 2016; Moore and Evans, 2017). Co-production centres evaluation decisions on what is important and feasible for all partners – including the programmatic teams and field-based organisations (Aniekwe *et al.*, 2012). This process can help ensure that evaluation findings are useful and utilised, and can influence changes in practice well before, and regardless of, publication (Jung, Harrow and Pharoah, 2012). By ensuring that all partners are at the table with equal footing, co-production may also contribute to the development of methodology that challenges knowledge hierarchies and ensures research questions are meaningful in local contexts (Durose *et al.*, 2012; Leung *et al.*, 2019).

In addition to co-production of research, additional strategies can help learn about intervention context and its interaction with an intervention – and how these contribute to complexity. These include process evaluation (Moore *et al.*, 2015), qualitative research, and mixed methods research (Howarth *et al.*, 2016). A complex adaptive systems approach (CAS) has also been widely discussed over the last decade as having potential for evaluation of social and complex interventions (Craig and Petticrew, 2013; Moore *et al.*, 2019). This approach can help examine the influence that an intervention has on a system, referring to the group setting in which an intervention takes place. A growing number of evaluations are identifying the intervention setting as a social system – for example, a hospital or school, where human interaction is a key aspect of complexity – and examine how and why an intervention influences the system (Datta and Petticrew, 2013; Thompson *et al.*, 2016).

Building on the notion of complexity as a feature of the particular system in which an intervention is implemented (Shiell, Hawe and Gold, 2008; Hawe, 2015), interventions can be conceptualised as events that aim to disrupt complex systems (Hawe, Shiell and Riley, 2009; Moore *et al.*, 2019). This helps bring the focus of evaluation away from individual-level change in behaviours or beliefs to instead examining system-level mechanisms of change that can have wider impact (Moore and Evans, 2017). A complex adaptive systems approach may be useful for such an analysis. However, despite growing interest in evaluating complex interventions using a complex systems approach, there remains limited guidance on how to do so and few examples of real-life evaluations that use this approach (Craig & Petticrew, 2013; Moore *et al.*, 2019; Smith & Petticrew, 2010).

Complex systems approaches to evaluation may hold particular promise as a strategy to account for complexity in social change interventions because social norms are by nature complex, and are important drivers of a range of harmful behaviours and health outcomes (Jewkes, Stern, & Ramsoomar, 2019; Mackie & Lejeune, 2009; Weber et al., 2019). For example, there is growing evidence from around the world that community-based interventions addressing gendered social norms and attitudes can help prevent IPV (Jewkes et al., 2019). These studies, however, rarely adopt a complex systems approach, despite its promise for the evaluation of social change interventions (Lacayo, Obregón and Singhal, 2008) that address complex social phenomena, such as IPV.

1.1.3 Gender-transformative programming for the prevention of intimate partner violence

Intimate partner violence (IPV), defined as physical, sexual or psychological violence perpetrated by an intimate partner, is a global human rights concern and has significant negative health outcomes (World Health Organization/London School of Hygiene and Tropical Medicine, 2010; The World Bank, 2013; Arango *et al.*, 2014; Ellsberg *et al.*, 2015). It is estimated that 30% of all women who have been in a relationship worldwide have experienced IPV (World Health Organization, 2013). IPV is 'a concrete manifestation of inequality between the sexes' (Garcia-Moreno *et al.*, 2005) and is considered a form of gender-based violence (GBV).

IPV prevention programmes focused on adolescents and young people are generally school- or community-based. Dating violence prevention programmes, more often in high-income countries, tend to include a focus on training bystanders to contribute to prevention by addressing the acceptability of violence and teaching how to address, diffuse and avoid potential conflict or violence (Coker *et al.*, 2017; Bush *et al.*, 2019; Gupta and Santhya, 2019; Muñoz-Fernández *et al.*, 2019; Miller *et al.*, 2020). Some programmes, generally in low- and middle-income, address both IPV and HIV (Pronyk *et al.*, 2006; Mathews *et al.*, 2016; Saul *et al.*, 2018), which are influenced by similar underlying factors.

There are significant investments in developing and evaluating GBV and IPV prevention interventions around the world. For example, the What Works to Prevent Violence Against Women and Girls Global Programme, funded by the UK Department for International Development, invested over £25 million in research and innovation in fifteen countries to learn about 'what works (and doesn't) to prevent violence' (Jewkes et al., 2019). This investment

addresses calls for GBV prevention interventions based on programme theory (Michau *et al.*, 2015), and there has been an increasing focus on developing culturally and contextually responsive theories of change for gender-based violence prevention in different settings (Moosa *et al.*, 2012; Eisenbruch, 2018). There is also a growing consensus that research is needed to build our understanding of how to transform the social norms that underlie and sanction violence (Heise, 2011; Jewkes, Flood, & Lang, 2015; Jewkes *et al.*, 2019).

Evidence does exist showing that interventions can prevent violence. A growing number of experimental and quasi-experimental evaluations of prevention interventions have reported effects that support IPV prevention, such as declines in self-reported victimisation of sexual or physical IPV, declines in self-reported perpetration of sexual or physical IPV, attitude shifts in favour of gender equality, decreased acceptability of physical IPV, or improvements in other sexual and reproductive health (SRH) outcomes interrelated with IPV (Abramsky *et al.*, 2014; Dunkle, Stern, Heise, & Chatterji, 2019; Ellsberg *et al.*, 2015; Jewkes *et al.*, 2019; Wagman *et al.*, 2015). The interventions found to have some detectable effects on IPV include group education interventions in low- and middle-income countries (LMIC) (Ellsberg *et al.*, 2015) and community-based interventions with a focus on shifting social norms (Jewkes *et al.*, 2019) – some of which also incorporate work with couples on relationship dynamics (Starmann *et al.*, 2017; McGhee *et al.*, 2019; McLean, Heise and Stern, 2019). A ‘gender-transformative’ approach is one central aspect of many of the interventions that have demonstrated changes in key behavioural and attitudinal outcomes supporting IPV prevention (Barker, Ricardo, & Nascimento, 2007; Fulu & Kerr-Wilson, 2015; Heise, 2011; Jewkes *et al.*, 2008; Pulerwitz, Barker, Segundo, & Nascimento, 2006; Pulerwitz *et al.*, 2010; Verma *et al.*, 2008). This approach addresses harmful gender norms by aiming to influence gendered power differentials and shift them to become more equal (Dworkin, Fleming and Colvin, 2015; Ruane-McAteer *et al.*, 2019).

Programmes to prevent IPV among young people is of particular importance because across countries and settings, IPV risk tends to be higher among younger women (Abramsky *et al.*, 2011; Stöckl *et al.*, 2014). An 81-country study estimated a 29.4% lifetime prevalence of IPV among 15-19 year old girls who have ever had a partner (World Health Organization, 2013). Yet despite the risks of IPV among younger women, there are ‘enormous gaps in our understanding of ways to reduce the experience of gender violence among adolescent [10-19 year old] girls’ (Blanc *et al.*, 2013).

A handful of reviews, systematic reviews and meta-analyses since 2014 have explored the effectiveness of IPV prevention programming with adolescents and young people in particular (De Koker *et al.*, 2014; De la Rue *et al.*, 2014; De La Rue *et al.*, 2017; Jennings *et al.*, 2017; Meinck *et al.*, 2019; Lee and Wong, 2020). Overall, these studies provide evidence that IPV prevention programming for young people can impact on knowledge or attitudes, while evidence on the impact on behaviours remains mixed. A recent systematic review and meta-analysis of teen dating violence prevention programmes in middle and high schools found evidence of effects on IPV perpetration, but not victimisation or bystander behaviours (Lee and Wong, 2020). A 2019 systematic review and meta-analysis of programmes focusing on IPV prevention among young people affected by HIV found reductions in exposure to IPV, but not perpetration (Meinck *et al.*, 2019). A systematic review of IPV preventions with young people, primarily in high-income countries, found more evidence for short-term than long-term effects, with mixed evidence overall for programme effectiveness (Jennings *et al.*, 2017). A 2017 meta-analysis (De La Rue *et al.*, 2017) and 2014 systematic review (De la Rue *et al.*, 2014) of school-based dating violence interventions did not find compelling evidence of effects on perpetration or victimisation. In contrast, a 2014 systematic review of RCTs of adolescent IPV prevention programmes did find some evidence of effectiveness in school- and community-based interventions on both perpetration and victimisation (De Koker *et al.*, 2014). A review of interventions focused on prevention of IPV among adolescents found some evidence of effectiveness of school-based programmes in high-income countries, and found evidence for the effects of community-based gender norms interventions more broadly (Lundgren and Amin, 2015a). Together, these studies highlight that IPV prevention with adolescents and young people is possible, and suggest that further research on school-based IPV prevention programming in low- and middle-income countries is needed.

A number of recent commentaries and editorials suggest that in order to attain broader behaviour change that translates into reductions in perpetration and victimisation for adolescents and young people, programmes need to focus on a range of interlinked outcomes and risk and protective factors across the social ecology. Prevention programming could be more effective if it moved beyond individual risk factors to a focus on social contexts more broadly and on the intersecting aspects of the lives of young people (De La Rue, 2019; Mannell *et al.*, 2019). In addition, IPV prevention programmes for youth would benefit from addressing interlinked behaviours and 'shared risk and protective factors in a non-siloed fashion' (Debnam and Temple, 2020).

While the growing evidence base provides a strong indication that IPV can be prevented, there remain weaknesses and gaps in the evidence. For example, some impact evaluations of IPV prevention programmes have shown no evidence of change in key behavioural outcomes, and some studies have inconclusive outcomes – sometimes because the duration or coverage of the programme was likely insufficient to result in changes in key outcomes (Arango *et al.*, 2014; Jewkes *et al.*, 2019). Others have shown incremental shifts in attitudes or behaviours but not a community-wide reduction in IPV rates (Gibbs *et al.*, 2019; McLean, Heise and Stern, 2019). Another limitation of the evidence base about IPV is that it is skewed towards high-income countries (Arango *et al.*, 2014; Wu *et al.*, 2020) and towards treatment and response interventions rather than prevention (Arango *et al.*, 2014; Ellsberg *et al.*, 2015). However, recent investments, such as those through the What Works Programme, have aimed to remedy these gaps by supporting the development and evaluation of prevention interventions in low- and middle-income countries, primarily in Asia and Africa (Jewkes *et al.*, 2019).

While it is a priority to learn about how IPV prevention works, it is also important to consider the many challenges inherent to researching relationships between intimate partners. Because it is difficult to directly observe behaviour within relationships, many studies rely on self-reported changes in knowledge, attitudes or behavioural intentions as a proxy for behaviour change (Lundgren and Amin, 2015b). However, such measures are limited because knowledge and attitudes do not necessarily translate into behaviour change (Kim and Hunter, 1993; Morrison, Ellsberg and Bott, 2007; De la Rue *et al.*, 2014; Lundgren and Amin, 2015b) and some behavioural change may precede – or occur without – a corresponding change in attitudes or norms (Jewkes *et al.*, 2019; McLean *et al.*, 2019; Pierotti, Lake, & Lewis, 2018). When behavioural outcomes are measured, research on relationship and sexual behaviours typically relies on self-reported data about perpetration, victimisation, and sexual risk behaviours (Arango *et al.*, 2014; Barker, Aguayo, & Correa, 2013; Fonner, Armstrong, Kennedy, O’Reilly, & Sweat, 2014; Jewkes *et al.*, 2008; Lundgren & Amin, 2015; Sarnquist *et al.*, 2014; Smith, White, & Holland, 2003). Yet studies rarely triangulate self-reported behavioural data, for example by seeking confirmation from partners (Barker, Ricardo and Nascimento, 2007; Lundgren and Amin, 2015b). One study that did triangulate behavioural outcomes, the Stepping Stones evaluation in South Africa, found a decline in self-reported perpetration among men but no accompanying decrease in self-reported victimisation among women (Jewkes *et al.*, 2008; Arango *et al.*, 2014). Similar discrepancies in the IMAGES study (Barker, Aguayo and Correa, 2013) and SHARE evaluation (Wagman *et al.*, 2015)

have also raised doubts about the reliability of self-reported behavioural data related to IPV. Similarly, a recent study in Australia found variation over time in women's self-reporting of lifetime IPV experiences (Loxton, Powers, Townsend, Harris, & Forder, 2019).

A significant challenge when researching IPV prevention efforts is the complexity of these interventions and the contexts in which they are implemented. Such interventions are complex in that they often intervene at multiple levels of the ecological model – including the individual, relationship, community and societal levels (Heise, Ellsberg, & Gottemoeller, 1999; Jewkes, 2017; Meinck et al., 2019; Andrew Morrison, Ellsberg, & Bott, 2004). To do so, programmes generally have multiple interacting parts – commonly considered to be a characteristic of complex interventions (Craig, Dieppe, Macintyre, Mitchie, *et al.*, 2008). IPV prevention interventions are also complex in because they are carried out in families, communities and schools – social settings in which interpersonal interaction is a central characteristic that needs to be accounted for in evaluation (Garcia-Moreno et al., 2014; Jewkes et al., 2019; Lundgren & Amin, 2015). Gender-transformative IPV prevention programming is complex in its engagement with gendered social norms, which vary by context and are among the drivers of IPV (Jewkes et al., 2019). Approaches to IPV or GBV prevention that engage social norms necessitate evaluation methods suited to detect hard-to-measure changes in these norms as well as in personal beliefs (Perrin *et al.*, 2019). In addition, evaluations may benefit from a focus on detecting moderate and incremental shifts in attitudes or behaviours, even if community-wide transformations in harmful social norms that underlie violence are not detected (Gibbs *et al.*, 2019; McLean, Heise and Stern, 2019). It is also a growing priority to explore evaluation methods that help understand the pathways to IPV prevention in different contexts (Fulu & Kerr-Wilson, 2015; Jewkes et al., 2015; Michau et al., 2015).

Because IPV is a social phenomenon and interventions addressing it are complex in their nature, this topic can serve as a case study to explore methodology and data collection methods suited to addressing complexity in evaluation. There is a need to identify and explore methods and approaches for the evaluation of social interventions that take complexity into account, are feasible for organisations operating on smaller scales, and are perceived as sufficiently 'rigorous' to satisfy the demands for high quality evidence from the perspectives of diverse evaluation stakeholders.

1.2 Aim and objectives

This thesis builds on the literature about these three intersecting themes – the politics of evidence, methodology to account for complexity in the evaluation of social interventions, and gender-transformative programming for the prevention of intimate partner violence – to explore evaluation approaches suitable for social interventions implemented in complex contexts. It builds on and engages the growing literature that argues for the broadening of evaluation methods to account for complexity and context. It does so by using as a case study an evaluation of a school-based comprehensive sexuality education programme with a focus on IPV prevention in Mexico City.

The aim of this study is to explore research approaches and data collection methods suitable for the evaluation of social interventions, using the case of IPV prevention programmes. The study will examine how differing approaches, methods, voices, and conceptions of rigour influence the research process and engage with complexity.

The research objectives are:

1. To examine research processes, methods, and co-production approaches of a case study evaluation in Mexico, and to reflect on their appropriateness for evaluating IPV prevention programmes in particular, and social interventions in general.
2. To conduct a co-produced evaluation study of a school-based IPV prevention programme in Mexico City in order to:
 - a. describe and document the intervention context and how it influences and interacts with the intervention in practice;
 - b. examine whether the intervention has any discernible effects on participants that may contribute to the prevention of IPV and to more equitable relationships; and
 - c. explore any gradual processes of change among participants and the aspects of the intervention that may be contributing to these.

1.3 Research questions

The research questions for the evaluation in Mexico City of a comprehensive sexuality education programme with a focus on IPV prevention are as follows:

Research question 1: What are the mechanisms through which the intervention can promote gender-equitable and non-violent relationships among young people?

Research question 2: How does the intervention appear to influence participants' relationship trajectories and interact with their personal context and circumstances?

The research questions for the methodological exploration, which uses the evaluation in Mexico as a case study, are as follows:

Research question 3: Which methodological approaches and data collection methods are appropriate and well-suited for the evaluation of social interventions in general, and IPV prevention programming in particular? Specifically, what are the advantages, disadvantages and challenges encountered when using qualitative longitudinal methods, complex adaptive systems, and quantitative survey approaches to account for complexity when evaluating an IPV prevention intervention?

1.4 Structure of the thesis

This is a research paper style thesis, organised around four research papers. Each of these papers contributes to the overall methodological exploration by examining the evaluation of the IPV prevention programme from different angles, applying a range of methodological approaches and considering their benefits and challenges.

In Chapter 1, I describe the importance of exploring appropriate methodological approaches for the evaluation of social and complex interventions and provide a rationale for using IPV prevention as a case study. In Chapter 2, I will introduce the intervention being evaluated – a school-based comprehensive sexuality education programme focused on IPV prevention in Mexico City. In Chapter 3, I will present the research approach and methods used for the evaluation in Mexico and the subsequent methodological exploration. In Chapters 4 through 7, I will draw on this foundation to present the results of the four interrelated analyses in four research papers, as described below:

- In Chapter 4, I begin by describing the gaps in the evidence about IPV prevention and the importance of examining whether comprehensive sexuality education may contribute to such efforts. I will then describe the context of IPV in Mexico and detail the school-based intervention that we evaluated, as well as the methods we used to learn about experiences in the intervention. I will then present the types of change that I identified among participants and the aspects of the course that appear to contribute to these.
- In Chapter 5, I first discuss the promise of qualitative longitudinal research as a strategy to take into account the shifting nature of people's experiences in interventions, and then describe why this research approach may be particularly useful for evaluating IPV prevention interventions. I will then present five case histories that show pathways through which the programme appears to influence relationship trajectories, and how personal context and circumstances influence participants' responses to the intervention. I will then examine participant perceptions of the repeat interview process to assess if this approach was feasible and acceptable. Finally, using the evaluation in Mexico City as a case study, I will consider the benefits and challenges of using qualitative longitudinal methods and the ways in which this approach could contribute to future evaluations.
- In Chapter 6, I first discuss the growing interest in a complex systems approach to evaluating social and complex interventions. I will describe why this approach may be particularly relevant when examining school-based interventions and those focused on IPV prevention. I then provide examples of how a social complex adaptive systems approach can shed light on the ways in which interpersonal interactions, group dynamics, the core messages of the course, and programme context influence intervention implementation and outcomes. I will then consider the system-disruptive potential of the evaluated intervention and reflect on the ways in which gender-transformative interventions may disrupt the social systems in which they are implemented. Finally, I will consider the potential of applying a social complex adaptive systems approach in future evaluations, particularly those focused on social change.
- In Chapter 7, I will describe the challenges collecting sexual health data from young people in schools. I will then explore the quality of data collected in the evaluation in Mexico City, comparing the self-administered, structured questionnaire data with the semi-structured

one-to-one interview data and identifying barriers to collecting high-quality data in practice. Finally, based on these findings, I will consider which methods are appropriate for the evaluation of social and complex sexual health programmes and reflect on how data collection challenges should be reported.

In Chapter 8, I provide a summary of the results of these four analyses to orient the reader, and then discuss the implications of these, present a framework of the different facets of complexity that can influence intervention evaluation in practice, and consider evaluation questions that can be used to ensure that evaluation design takes the diverse aspects of complexity into account.

1.5 Theoretical frameworks and conceptual models informing this thesis

This thesis draws on a range of theoretical frameworks and conceptual models. In particular, it is informed by complexity theory and complex adaptive systems approaches; social norms theory; the concept of the gender system; and the social-ecological model for partner violence. Each is described below, followed by a general discussion of the justification for choosing these particular frameworks and models for this study. In section 1.5.5, I show the chapter(s) where each of these comes into play throughout the thesis.

1.5.1 Complexity theory and complex adaptive systems

Complex adaptive systems (CAS) are studied in different fields in the natural as well as social sciences. Plsek and Greenhalgh define a CAS as ‘a collection of individual agents with freedom to act in ways that are not always totally predictable, and whose actions are interconnected so that one agent’s actions changes the context for other agents’ (Plsek & Greenhalgh, 2001). Examples of CAS include ‘the stock market, a colony of ants, the human body immune system, and just about any collection of humans’ (Plsek, 1997).

Social complex adaptive systems is a term that differentiates social systems such as communities, hospitals, health systems and schools – in which people are the ‘interacting agents’ – from the complex systems often examined in research from the natural world (Keshavarz *et al.*, 2010). Keshavarz *et. al.*, in a study of school-wide health promotion programming, argue that schools should be conceptualised as social complex adaptive systems, defined as follows:

At its core, a complex adaptive system comprises a population of diverse rules-based agents, located in multi-level and inter-connected systems in a network

shape. A system is characterised by the behaviour of individual agents. Agents in complex adaptive systems are often numerous, dynamic, autonomous, highly interactive, learning and adaptive. Agents of complex adaptive systems act in ways that are based on a combination of their knowledge, experience, feedback from the environment, local values and formal system rules. These change over time leading to continuously changing interactions and adaptations that are often novel and are hard to predict, especially in social systems. Agents in complex adaptive systems interact with and adapt to each other and the system within the network. Complex adaptive systems are open systems with fuzzy boundaries and also highly context dependent in terms of time, history, and space including location and proximity. Complex adaptive systems also have distributed control. Consequently, complexity that is not necessarily a characteristic of individual agents, emerges at system level. (Keshavarz *et al.*, 2010)

Hawe, Shiell and Riley argue that interventions should be conceptualised as ‘events in systems that either leave a lasting footprint or wash out’ (Hawe, Shiell and Riley, 2009). An intervention can be designed to disrupt the system in which it is implemented. This design should be based on understanding the particular phenomenon being intervened on, in context. With this framing, understanding the context in which an intervention is implemented becomes a central aspect of evaluation (Hawe, Shiell and Riley, 2009; Moore *et al.*, 2019).

1.5.2 Social norms theory

Social norms influence human behaviour. Social norms are sometimes conceptualised as patterns in behaviour or as collective attitudes, but in the field of global health they are more commonly considered to be beliefs about what others in a community do and think (Cislaghi and Heise, 2018b). Cialdini, Kallgren and Reno consider both descriptive norms (‘the perception of what most people do’) and injunctive norms (‘the perception of what most people approve or disapprove’) as factors that influence individual behaviours and beliefs (Cialdini, Kallgren and Reno, 1991).

Mackie and colleagues reviewed sixteen definitions of social norms from different fields and authors and found three elements in common (Mackie *et al.*, 2015). First, norms are created by social expectations. Second, norms are held in relation to a reference group (or groups). And third, norms are maintained by social influence. Importantly, social norms do not always correspond with individual attitudes, as social norms are ‘externally motivated’ while individual attitudes are ‘internally motivated’ (Mackie *et al.*, 2015). However, while attitudes and social norms differ, they are interrelated, because attitudes are socially constructed and influenced by social norms (Learning Collaborative to Advance Normative Change, 2019).

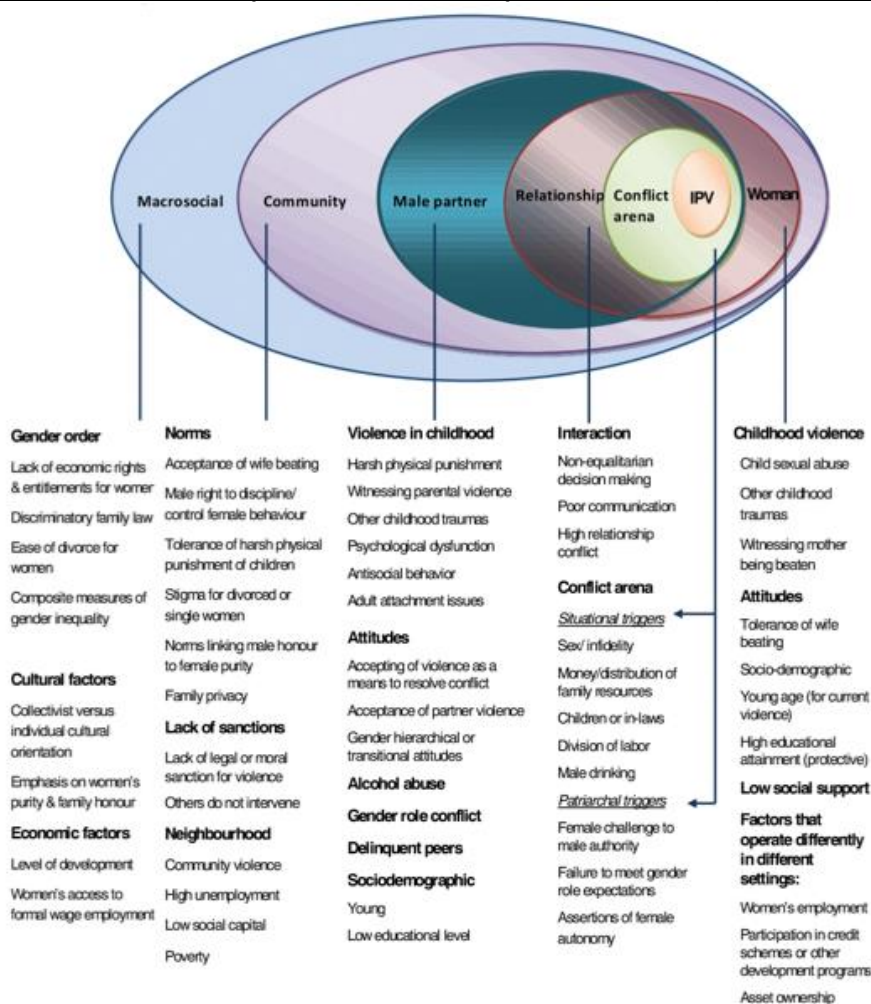
1.5.3 The gender system

The gender system reflects what is expected of women and of men in a particular setting (Hirdman, 1991; Ridgeway and Correll, 2004). Hirdman referred to the 'gender system' as the different tasks and roles that are assigned to men and women in a particular context (Hirdman, 1991), and Ridgeway and Smith-Lovin define it in terms of the 'processes that both define males and females as different in socially significant ways and justify inequality on the basis of that difference' (Ridgeway and Smith-Lovin, 1999). Gender systems are held in place by gendered social norms (Heise *et al.*, 2019). 'Gender-transformative' programmes, which aim to shift social norms to become more equitable (Dworkin, Fleming and Colvin, 2015), are in essence designed to change the gender system in the different contexts in which they are implemented.

1.5.4 Social-ecological model for partner violence

The social-ecological model for violence against women, originally proposed (Heise, 1998) and subsequently adapted by Heise (Heise, 2011), presents interrelated factors that contribute to partner violence at the individual, relationship, community, and societal levels. The 2011 version of the ecological model (Figure 1) differentiates between community-level social norms and the 'gender order' more broadly in society. Social norms are among multiple factors that influence behaviour and contribute to IPV – or to its prevention. Therefore interventions should not rely on norms-focused strategies in isolation (Cislaghi and Heise, 2018b) but rather are more likely to be effective if they intervene at multiple levels of the social ecology (Michau *et al.*, 2015).

Figure 1: Revised conceptual framework for partner violence (from Heise, 2011)



1.5.5 The use of theoretical frameworks and models throughout this thesis

Here, I provide a rationale for choosing the particular theoretical frameworks and conceptual models for this study and describe where in the thesis each is applied and its purpose. First, I describe two cross-cutting frameworks and theories that played a role throughout the thesis: social norms theory and the social-ecological model for partner violence. Then, I describe the frameworks and theories that emerged as the study progressed and thesis developed: complexity theory and complex adaptive systems, as well as the gender system. For each, I provide a rationale for inclusion in this thesis and describe where and how they were applied in the thesis.

Social norms theory

Social norms theory informed the development of the gender-transformative intervention that is at the centre of this thesis. As such, social norms theory is cross-cutting in its influence throughout the thesis. In Chapter 4, I examine whether intervening on social norms related to sexuality and gender may contribute to IPV prevention. In Chapter 5, I describe a qualitative longitudinal approach to evaluation as a strategy to examine processes of change and gradual shifts in beliefs and behaviours interrelated to social norms. In Chapter 6, I examine the potential of the intervention in terms of its ability to disrupt social norms, such as forms of acceptable commentary and behaviour, among and in each intervention group. In Chapter 7, I consider how social norms can influence participant responses when collecting data about sexuality and relationships. In Chapter 8, I discuss the challenges of measuring complex lived experiences influenced by social norms, and consider different approaches through which evaluations can consider and taking these into account.

Social-ecological model for partner violence

The social-ecological model has influenced thinking on IPV prevention programming for decades, and informed the conceptualisation of the intervention. As such, the model is cross-cutting in its influence throughout the thesis. In Chapters 4 and 5, I consider the mechanisms through which the intervention can influence not only individual participants, but also their relationships and partners. I also explore any indirect effects on peers and families, at school and at home. In Chapters 6 and 7, I consider the influence of events and interactions at the classroom and school levels as well as the broader influence of localised factors such as environment or climate as well as norms in the community. These different levels of analysis and inquiry are informed by the social-ecological model's articulation of the multiple levels through which an intervention can influence on IPV prevention.

Complexity theory and complex adaptive systems

Complexity theory generally, and complex adaptive systems in particular, emerged as this thesis developed as an important lens through which to view the evaluation of complex, social, and public health interventions. These concepts help create a structure through which to examine complexity and social interactions in practice, and increasingly feature in the academic literature about evaluation, despite a dearth of empirical examples of the use of these concepts in evaluations of social interventions such as IPV prevention programmes. The thesis draws on

complexity concepts primarily in chapters 6 – 8. In Chapter 6, I introduce social complex adaptive systems as an analytical framework through which to re-examine qualitative data collected in the evaluation process. In Chapter 7, I consider how different evaluation methods can address complex concepts, such as relationships, sexuality and violence, in dynamic contexts. In Chapter 8, I draw on both of these chapters to map all of the ways in which complexity influenced the evaluation in Mexico and propose a series of questions and considerations that can be taken into account in evaluations to engage with complexity.

The gender system

The ‘gender order’ features in the social-ecological model for partner violence, and the gender system is tied to and influenced by gendered social norms (Heise *et al.*, 2019). As I began to engage with complex adaptive systems and the concept of interventions as system-disruptors, I began to consider whether the concept of a ‘gender system’ might provide a useful lens through which to examine how gender-transformative programmes influence on participants and help shift social norms. In Chapter 6, I explore system disruption as a concept useful for understanding the influence of gender-transformative programming. Specifically, I consider how rules of engagement that govern interaction in implementation groups receiving the intervention might shift. I examine some of the key elements of the intervention that seem to shift beliefs about gender and violence, and provide examples of how the intervention influenced and aimed to reshape the ‘gender system’ in groups of intervention participants. In Chapter 8, I explore this further.

Chapter 2: Intervention description and setting

The case study explored in this thesis is a co-produced evaluation of an intervention that was implemented in Mexico City. In this chapter, I will introduce the study partners and their collaboration model, describe the intervention, present the programmatic theory of change, and provide detail about the study setting. I will then discuss the translation of evidence to practice.

2.1 Study partners

International Planned Parenthood Federation is a global non-governmental organisation focusing on sexual and reproductive health and rights. The Western Hemisphere Regional Office (IPPF/WHR) works with local affiliates across Latin America and the Caribbean. IPPF/WHR was responsible for fundraising and coordinating between the programmatic and research components in this project, which was funded by a private donor. IPPF/WHR staff also collaborated in the research process and on dissemination of findings. I was a long-term staff member at IPPF/WHR, and for two years I was concurrently the IPPF/WHR staff member responsible for coordinating and implementing the research project while also serving as the co-Principal Investigator through my role as a PhD student at the London School of Hygiene and Tropical Medicine (LSHTM). I discuss this dual role further in the reflexivity section (3.5.2).

Fundación Mexicana para la Planeación Familiar, A.C. (Mexfam) is the IPPF/WHR affiliate in Mexico. The community-based organisation provides sexual and reproductive health services through its network of clinics as well as community programmes throughout the country. Mexfam's youth programme, *Gente Joven* ('Young People'), aims to reduce unplanned pregnancies, sexually transmitted infection and HIV by providing comprehensive sexuality education (CSE) and offering youth-friendly health services. Mexfam's CSE programme was revised in 2016 to incorporate a focus on IPV prevention. This updated intervention with a focus on IPV was evaluated as part of this thesis. Mexfam was responsible for implementing the intervention, collaborating on the evaluation, and translating evidence into practice to improve their programming.

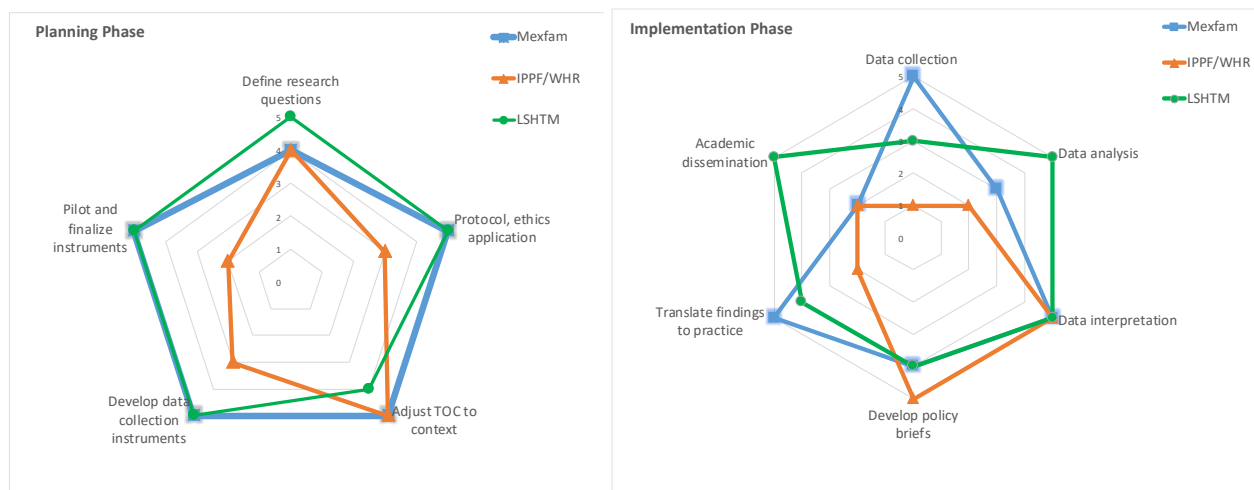
LSHTM was the research partner for the evaluation of the CSE programme with an IPV prevention focus. There were two co-Principal Investigators for the evaluation: myself, as PhD candidate, and Professor Cicely Marston, in her role as my PhD supervisor and advisor to IPPF/WHR on the

evaluation in Mexico. My role was to lead on all aspects of research design, implementation, analysis and dissemination, with support from Professor Marston.

2.2 Collaboration model

The evaluation of Mexfam’s comprehensive sexuality education programme was co-produced with the partner organisations. A topic of growing interest in the international development sector, co-production entails full partnership between implementers and researchers to ensure the research is feasible and relevant to all partners (Aniekwe *et al.*, 2012). Co-producing research ‘is increasingly seen as one of the most effective ways to ensure research impact,’ as it supports timely research utilisation (Jung, Harrow and Pharoah, 2012) and harnesses local expertise to ensure research questions are locally meaningful (Durose *et al.*, 2012). An important aspect of research collaboration is ensuring that roles and responsibilities are clearly defined so that partners can contribute at a different level of intensity based on the needs and their capacity on the particular action or topic. At the beginning of the study, I facilitated a discussion to define expectations for the roles and responsibilities across the different project phases (Figure 2).

Figure 2: Level of involvement for each partner organisation, by study phase



I was responsible for study design, conception of the papers, data analysis, and writing, as well as managing the partnership between the three organisations. I went to Mexico at least two times per year. While there, I facilitated collaborative study design, theory of change development, and data interpretation meetings with Mexfam and IPPF/WHR staff. I also participated in fieldwork. The research assistant, Jovita Garduño, was responsible for managing the relationship with the

school, recruiting participants, and for data collection. At the beginning of the study, the research assistant and I held meetings with school staff to inform them about the study, conducted informational sessions with parents to obtain consent, and then implemented the baseline questionnaires, including the informed assent process with the students. During the semester, we both observed intervention sessions and wrote field notes. The research assistant conducted all of the in-depth interviews and repeat interviews – I observed some of these with participant permission. At the end of the semester, we together implemented the endline questionnaires and co-facilitated focus groups with students, teachers and the Mexfam staff health educators.

I also provided feedback on Mexfam’s programmatic work, in particular with relation to the translation of the research findings into practice. For example, Mexfam developed a Manual for the Strengthening Programme for the Successful Implementation of Comprehensive Sexuality Education that is already being used at their sites throughout Mexico to provide ongoing training for health educators and to inform programmatic improvements. These materials will be shared with other service-delivery organisations that are members of the IPPF/WHR network to inform programmes elsewhere, particularly in Latin America.

2.3 Description of the intervention

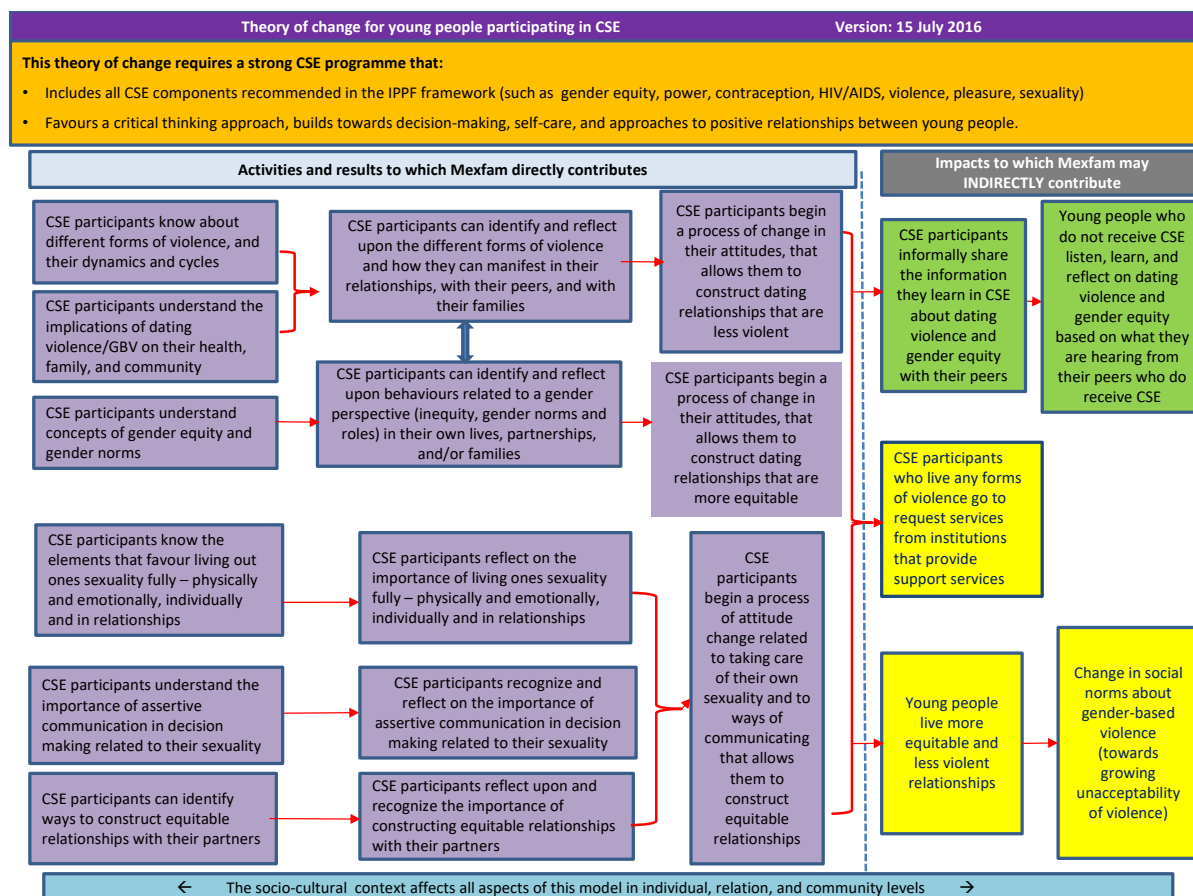
Mexfam developed the intervention that was evaluated in this study. The comprehensive sexuality education programme with a focus on IPV prevention was centred on a manual-based curriculum drawing from international standards for sexuality education (United Nations Educational Scientific and Cultural Organization, 2018). The course comprised ten two-hour sessions delivered by trained health educators under the age of 30 who were staff at Mexfam. The course was delivered over one semester in mixed-gender classroom groups of approximately 20 young people. Course topics included sexual diversity, gender, equitable relationships, IPV, pregnancy prevention, sexually transmitted infections, and other sexual and reproductive health topics. Both harmful and favourable attributes of relationships were explored. IPV was presented as a range of behaviours that could be perpetrated or experienced by both men and women. Types of violent behaviours discussed in the course including excessive jealousy, possessive behaviour over a partner such as monitoring of social media or cell phones, and other behaviours that cause emotional, physical or sexual harm.

The intervention used a 'gender-transformative' approach (Dworkin, Fleming and Colvin, 2015). Specifically, it was designed to facilitate critical reflection about gender norms and violence, among other topics. It aimed to shift individual attitudes and group norms related to gender, sexuality and violence through participatory methods that reinforced key intervention messages. Course activities were designed to build knowledge, develop assertive communication and interpersonal skills, educate participants about the right of young people to receive healthcare, inform them about where they could seek such services, and encourage access to clinical and support services in cases of IPV or other sexual or reproductive health needs.

2.4 Intervention theory of change

At the beginning of the project, I facilitated a two-day meeting in Mexico to develop a programmatic theory of change (TOC) for Mexfam's comprehensive sexuality education programme and hypothesise how it might reduce and prevent IPV. Participants in the meeting included Mexfam health educators, youth representatives from the *Gente Joven* programme, and Mexfam and IPPF/WHR staff members specialising in research and evaluation as well as community programming and sexuality education. We identified pathways through which Mexfam's comprehensive sexuality education programme might contribute to IPV prevention. This served as a starting point for ongoing conversations within the research team that contributed to ongoing refinements of the TOC (Figure 3). In collaboration with the Mexfam team, I developed research questions to explore the pathways set out in the TOC (Appendix 1) as part of the evaluation. Throughout the research process, I used the ongoing data analysis process to inform iterative refinements of the TOC (Appendix 2, final version of TOC shown in Spanish).

Figure 3: Theory of change developed to inform research design



2.5 Intervention setting

The intervention and evaluation were implemented at a state-run technical secondary school in the southern part of Mexico City, in a commercial zone of the Tlalpan area. The school delivers vocational training on topics including automotive mechanics and food preparation. It primarily serves students from lower-middle income families who live throughout Mexico City, sometimes travelling up to two hours on public transportation to school (Gómez Espinoza, 2006). Mexico City, the capital city of Mexico, is among the largest cities in the world. The metropolitan area has over 21 million inhabitants, nine million of whom live within the designated city boundaries. The city is situated in a seismic zone, and was impacted by two significant earthquakes in September 2017 (Chavez *et al.*, 2017), during the research period.

Mexico City has more liberal social policies than the rest of Mexico, for example with decriminalised abortion in the first 12 weeks of pregnancy since 2007 (Becker and Olavarrieta,

2013) and legal same-sex marriage since 2009 (Unzelman, 2011). Despite this progressive legal context in terms of sexual and reproductive rights in Mexico City, cultural values and social norms in Mexico generally support traditional gender roles (Aguilar Montes de Oca *et al.*, 2013). Culturally, there is a focus on family obligations, with men often viewed as the fiscal provider and leader while women are expected to provide for the emotional and physical needs of the family (Castillo *et al.*, 2010). In addition, social norms perpetuate beliefs that women should be virginal and pure (Espinosa-Hernández, Vasilenko and Bámaca-Colbert, 2016) while men should be powerful and dominant (Parker, 1996) and sexually virile (Szasz, 1998).

There are high levels of interpersonal violence in Mexico (Imbusch, Misse and Carrión, 2011), with spikes in homicide rates related to organised crime and gang activity (UNODC, 2019). According to recent estimates, 43.9% of women in Mexico (ages 15 and older) have ever reported experiencing intimate partner violence (Instituto Nacional de Estadística y Geografía, 2018b). In this context, prevention of intimate partner violence, including for young people, is an urgent priority.

2.6 Dissemination

With funding through the 2019 Development Marketplace Award for Innovations in Gender-based Violence, I contributed to three co-authored policy briefs that are available in both English and Spanish. In addition, in 2019 I presented one conference presentations at the Sexual Violence Research Initiative Forum in Cape Town, and a colleague from Mexfam presented on my behalf at the World Association for Sexual Health conference in Mexico City. IPPF/WHR and Mexfam have also published a series of blogs hosted on the Sexual Violence Research Initiative website based on the evaluation experience. In addition, I was the lead author on a briefing paper and blog post about the study process and findings on the Advancing Learning and Innovation on Gender Norms (ALIGN) website. All publications and other dissemination materials are written in both English and Spanish when possible, to ensure access to information among the study partners in Mexico and other IPPF/WHR partner organisations in Latin America.

I was the lead author on four peer-reviewed publications (listed below) and was responsible for study design, conception of the papers, analysis and writing, with significant contributions from Cicely Marston in her role as my PhD supervisor. Because of the co-produced nature of the evaluation, selected IPPF/WHR and Mexfam staff are co-authors based on their involvement in study design, data collection, interpretation of findings, or critical revisions of the papers, as

appropriate in accordance with the International Committee of Medical Journal Editors (ICJME) authorship guidelines. All publications are currently or will be made available through open access journal articles or using repositories that comply with copyright requirements.

Paper 1 (Chapter 4) was published in 2019 in *Sexuality Research and Social Policy*:

Makleff S, Garduño J, Zavala RI, Barindelli F, Valades J, Billowitz M, Silva Márquez VI, Marston C. "Preventing intimate partner violence among young people – a qualitative study examining the role of comprehensive sexuality education." *Sexuality Research and Social Policy* (2019). <https://doi.org/10.1007/s13178-019-00389-x>

Paper 2 (Chapter 5) was submitted to *Qualitative Health Research* in May 2020:

Makleff S, Garduño J, Zavala RI, Valades J, Barindelli F, Cruz M, Marston C. "Evaluating complex interventions using qualitative longitudinal research: A case study of understanding pathways to violence prevention." (Revision responding to peer review is currently under review).

Paper 3 (Chapter 6) has been published in 2020 in *Health Policy and Planning*, accompanied by a podcast I recorded that is available on the journal website:

Makleff S, Billowitz M, Garduño J, Cruz M, Silva Marquez VI, Marston C. "Applying a social complex adaptive systems approach when evaluating implementation of a school-based intervention for intimate partner violence prevention: A case study in Mexico." *Health Policy and Planning* (2020). <https://doi.org/10.1093/heapol/czaa067>

Paper 4 (Chapter 7) will be finalised and send to a journal in late 2020:

Makleff S, Garduño J, Silva Marquez VI, Medina S, Barindelli F, Marston C. "Collecting better data on sexuality, relationships and violence in schools: empirical evidence from an evaluation."

Chapter 3: Methods

In this chapter, I present the research approach of the thesis as a whole and then describe the two components of the thesis: first, the evaluation of an IPV prevention intervention in Mexico; and second, a methodological exploration drawing on the evaluation in Mexico as a case study to consider methods and approaches that account for complexity. For each, I describe the study design, data collection and analysis. I then discuss ethical consideration and reflexivity of the researcher.

3.1 Study design

3.1.1 Disciplinary and epistemological perspective

This thesis is located at the intersection of the fields of public health and international development. My approach is multidisciplinary, “draw[ing] on knowledge from different disciplines” in ways that interact to contribute to the overall work (Choi and Pak, 2006). This multidisciplinary occurs at the level of the research team, which included programmatic staff and practitioners working closely with evaluators and researchers. It also reflects the literature that informed the study in the planning, implementation, analysis and dissemination stages. My work is informed by academic work from a range of disciplines in the social sciences: social psychology, organisational effectiveness, sociology, gender and sexuality studies, and development economics. In addition, I drew on historical literature about Mexico and its educational system as part of my focus on understanding the context in which this study took place.

I draw on constructivist epistemology, which considers that there is no one ‘truth’ and that reality is constructed in the mind, with the possibility for many possible constructions and explanations (Guba and Lincoln, 1989). I draw on Denzin’s argument that mixed methods research based on constructivist paradigms has potential to contribute to research addressing social justice aims (Denzin, 2012). Mertens and Tarsilla build on this, arguing that “mixed methods that begin in the constructivist paradigm afford the opportunity to assess the interpretive, contextual level of experience where meaning is created and provides a roadmap to address social justice” (Mertens and Tarsilla, 2015).

While the commonly used ‘convergence’ approach to triangulation, ‘assumes that the datasets produce the same picture of reality’ and corresponds to a positivist approach, there are

applications of triangulation that are compatible with constructivist approaches (Nightingale, 2009; Denzin, 2012). In this study, I adopt a 'complementarity' approach to triangulation, which 'seeks to produce a fuller picture of the research questions by combining information from different methods or different observers. In this kind of analysis, quantitative and qualitative datasets can be combined together. The results are not expected to be the same, but rather to make sense in relation to each other and to help create a fuller picture of the research problem by creating more complete information about a topic' (Nightingale, 2009). Similarly, Denzin proposes that triangulation should not be fixed to any particular research paradigm, but rather can serve as a form of methodological bricolage that can contribute to research with social justice aims (Denzin, 2012).

3.1.2 Research approach

3.1.2.1 Researcher as bricoleur

This thesis draws on the notion of bricolage as 'a critical, multi-perspectival, multi-theoretical and multi-methodological approach to inquiry' (Rogers, 2012). I engage Denzin and Lincoln's conception of the researcher as a 'narrative bricoleur' by drawing from multiple voices to engage different perspectives and construct a narrative of experiences being studied (Denzin and Lincoln, 2005). Kincheloe builds on earlier work by Denzin and Lincoln and describes a bricoleur as avoiding 'monolithic knowledge' and 'one-dimensionality' in research as in order to acknowledge and engage with the complexity of lived experiences (Kincheloe, 2005). A bricoleur will 'understand that complexity sets the stage for the need for the bricolage, the necessity of new ways to understand the complications of social, cultural, psychological, and educational life' (Kincheloe, 2005).

I adopt a bricolage approach to research, engaging with multiple perspectives, voices and methods to build a story of the intervention being evaluated. This approach – and particularly an approach to bricolage that mixes different methods of data collection – allows for practice-based knowledge and lived experiences of participants to be prioritised in analysis as forms of meaningful evidence. This is a way to 'privilege the lived experiences of those most affected by the intervention' (Hesse-Biber, 2012), drawing from feminist research core values of collaboration, participation and accountability (COFEM, 2018; Leung *et al.*, 2019).

I engage with another facet of bricolage in this thesis in terms of designing the methodological exploration (thesis component 2). The idea for the methodological exploration came about in dialogue between my own experience conducting evaluation in an NGO context and Professor Marston's methodological expertise and interests. The conceptualisation and evolving design of the methodological exploration was in this way 'co-produced', bringing together different experiences and perspectives; this can be seen as a 'bricolage' approach as well.

3.1.2.2 A 'qualitatively driven' mixed methods approach

In this thesis, I explore narratives about young people's lived experiences in their social and intimate relationships, examine how these are influenced by an intervention, and consider which evaluation methods can account for and explore personal context and lived experiences. My approach to this study is inspired by Jennifer Mason's 'creative' and 'qualitatively driven' mixed methods approach, which is based on a constructivist epistemology (Mason, 2006). Mason argues this approach is particularly useful for research about lived experiences because it can engage with the multi-dimensionality of lived experiences. She suggests that having 'a creative palette of methods of data generation' can help researchers think beyond the typical research questions and approaches central to their field of study, and in this way can encourage innovative research to address questions of how and why social phenomena occur in context.

For this study, I adopted a qualitatively driven mixed methods approach for two main reasons. First, to gather complementary and diverse forms of data that would shed light on different aspects of intervention functioning and of complex experiences related to relationships and sexuality. Second, to allow for an exploration and comparison between different data collection methods, to serve as a basis for the methodological exploration.

3.1.2.3 Rationale for using a mixed-methods approach at different research stages

There are calls to be explicit about the rationale for combining quantitative and qualitative data collection methods in a study, while also acknowledging that the purpose for doing so may shift as the study advances (Bryman, 2006; Guest, 2013; Schoonenboom, Johnson, & Froehlich, 2018). Some researchers have advocated for the rationale for method choice to be articulated not for the study as a whole but within a study (Schoonenboom, Johnson and Froehlich, 2018) or at 'the *point of interface* between two data sets' [emphasis in original] (Guest, 2013). Below, I describe the rationale for mixing different data collection methods and their resulting data at three points of

interface of the study: data collection, analysis and the interpretation stage. I draw from an influential classification system of the rationale for mixed methods research (Greene, Caracelli and Graham, 1989) and a more recent classification that builds on the former (Bryman, 2006a). The various rationale used for mixing data collection methods in this thesis are summarised in Table 1.

Table 1: Rationale for mixing data collection methods at different points of interface in the study

Point of interface	Rationale
Data collection	<ul style="list-style-type: none"> • ‘Different research questions’ rationale (Bryman, 2006) • ‘Development’ (Greene, Caracelli and Graham, 1989) • ‘Sampling’ rationale (Bryman, 2006) • ‘Utility’ rationale (Bryman, 2006)
Analysis	<ul style="list-style-type: none"> • ‘Initiation’ rationale (Greene, Caracelli and Graham, 1989) • ‘Expansion’ rationale (Greene, Caracelli and Graham, 1989)
Interpretation	<ul style="list-style-type: none"> • ‘Completeness’ rationale (Bryman, 2006) • ‘Complementarity’ rationale (Greene, Caracelli and Graham, 1989) • ‘Triangulation’ rationale (Greene, Caracelli and Graham, 1989)

The Mexfam team was closely involved in the process of study design for the evaluation of their comprehensive sexuality education programme. I will refer to decisions made about methods of data collection in the collective voice (‘we’) to reflect their role in this process.

At the *data collection* stage, we decided to include a range of methods because we anticipated challenges in gathering meaningful data regarding relationship, violence and sexual health experiences among intervention participants. We opted for diversity in data collection methods to help overcome the many documented challenges to measuring IPV and other sexual health behaviours (Fenton *et al.*, 2001; Ghanem *et al.*, 2005; McCallum and Peterson, 2012; Young *et al.*, 2016; Loxton *et al.*, 2019).

Each data collection method had a different purpose (detailed below in Table 3). For example, we designed the survey questionnaire to learn about the experiences, attitudes and characteristics of a larger sample of participants in a cost-effective manner, despite our suspicion that a quantitative survey approach might not elicit meaningful data on sensitive topics. In addition, the

programmatic team at Mexfam wanted to include knowledge and satisfaction questions— these data were collected through the questionnaire and provided to the programmatic team at Mexfam for internal use, but were not analysed as part of the study. We complemented the structured questionnaire with longitudinal qualitative interviews that would allow us to learn in detail about the lived experiences of a small subsample of participants. In addition, we incorporated focus groups with students, teachers and health educators into the study to learn about knowledge, beliefs and social norms about gender and IPV in the school context from different perspectives. This corresponds to the ‘different research questions’ rationale (Bryman, 2006). We also used the baseline questionnaire data to develop a purposive sample for in-depth interviews; this aligns with the ‘development’ (Greene, Caracelli and Graham, 1989) or ‘sampling’ (Bryman, 2006) rationale. Finally, we designed the study with an eye to ensuring that different audiences, such as practitioners, donors and policy makers, would find the results useful and to use methods that were considered rigorous and meaningful by different audiences; this aligns with the ‘utility’ rationale, which has particular relevance for applied research (Bryman, 2006). This rationale led us to include quantitative data collection methods to complement the qualitative methods that predominate in this study, in part because qualitative data alone may be perceived as less rigorous (Giddings and Grant, 2007; Given, 2017).

At the *analysis* stage, I compared individual participant responses from different data sources and explored convergence and divergence. I approached the analysis expecting some divergence between individual responses because participants might feel uncomfortable disclosing personal information about their relationship history, sexual experience, sexual orientation, or experiences of violence. This reflects the ‘initiation’ rationale, which ‘seeks the discovery of paradox and contradiction’ (Greene, Caracelli and Graham, 1989), based on the expectation that information from different data sources for each individual participant might not align. In addition, I used the different types of data that were collected in the evaluation in Mexico to inform an analysis of the challenges and benefits of different data collection methods for learning about sensitive or taboo topics such as sexuality, relationships and violence in this setting. This reflects the ‘expansion’ rationale, which can ‘extend the breadth and range of enquiry’ (Greene, Caracelli and Graham, 1989), in this case extending to also include an exploration of evaluation approaches and data collection methods.

At the *interpretation* stage of the evaluation, the rationale was to bring together different perspectives and strands of data to gain a more complex and nuanced understanding of whether and how the intervention ‘worked’ to prevent partner violence. This aligns with the ‘completeness’ rationale that allows ‘a more comprehensive account of the area of inquiry’ (Bryman, 2006a) and the ‘complementarity’ rationale that ‘seeks elaboration, enhancement, illustration’ (Greene, Caracelli and Graham, 1989). I was guided by the principle that if all three sources confirmed a certain event, the evidence supporting it would likely be stronger because it provides a way to bring together different perspectives and constructed meanings of collective experiences. This aligns with the ‘triangulation’ rationale, which seeks corroboration of the results from different sources (Greene, Caracelli and Graham, 1989). I employed a ‘complementarity’ approach to triangulation, which can align with the constructivist approach that underlies this study. In a complementarity approach to triangulation, findings from different data collection methods should together help create a fuller picture’ of phenomenon or concept being studied (Nightingale, 2009).

3.1.3 Thesis components

There are two components of this thesis, each of which reflects different research questions. First, the evaluation of an IPV prevention intervention in Mexico examining research questions 1 and 2; and second, a methodological exploration, which uses the evaluation in Mexico as a case study, examining research question 3. Each research question is examined in multiple thesis chapters (Table 2).

Table 2: Summary of research questions for each thesis component and chapter

Thesis components	Research question	Chapters
Component 1: Evaluation of IPV prevention intervention	1: What are the mechanisms through which the intervention can promote gender-equitable and non-violent relationships among young people?	4, 5
Component 1: Evaluation of IPV	2: How does the intervention appear to influence participants’ relationship trajectories and interact with their personal context and circumstances?	4, 5

prevention intervention		
Component 2: Methodological exploration	<p>3: Which methodological approaches and data collection methods are appropriate and well-suited for the evaluation of social interventions in general, and IPV prevention programming in particular?</p> <p>Specifically, what are the advantages, disadvantages and challenges encountered when using qualitative longitudinal methods, complex adaptive systems, and quantitative survey approaches to account for complexity when evaluating an IPV prevention intervention?</p>	5, 6, 7

3.1.4 Study design for thesis component 1: Evaluation of IPV prevention intervention in Mexico

Together with Mexfam and IPPF/WHR, I led a quasi-experimental longitudinal evaluation of an intimate partner violence (IPV) prevention programme in Mexico City in 2017 and 2018. There were two phases, both of which took place at the same school. First, from March to June 2017, we conducted a pilot study to accompany the pilot implementation of the intervention. This phase focused on exploring intervention acceptability and feasibility and developing an understanding of the intervention context. Mexfam piloted their comprehensive sexuality education course with one classroom of third-year students (approximately 45 participants with a median age of 16), and I piloted all the data collection instruments with support from the research assistant based in Mexico. Based on the pilot experience, I refined the data collection instruments before the research phase began, and revised versions were shared with the ethics committees that approved the study (detailed in section 3.4). The second phase, from September to December 2017, involved a quasi-experiment longitudinal study with an intervention and comparison group of first-year students (median age of 14). We selected first year students because they were in their first semester at the school so were not exposed to the pilot process, and also because they were unlikely to interact in any school-based activities with the older students who had participated in the pilot during the prior academic year.

We selected an intervention and comparison group from the daytime (*matutino*) and evening (*vespertino*) programmes. These two programmes use the same school facilities but operate at different hours with different students and teachers. This separation ensured minimal contamination between the intervention and comparison groups. We used a coin toss to assign the daytime programme as the intervention and afternoon programme as the comparison group. The school staff assigned three classrooms of first-year students in the daytime programme as the intervention group and three classrooms of first-year students in the evening programme as the comparison group, based on scheduling and logistical considerations. Each classroom had approximately 40 students, so the intervention and comparison groups each comprised approximately 120 students. The comparison group received the intervention in 2018, after the study ended.

3.1.5 Study design for thesis component 2: Methodological exploration

The methodological exploration used a qualitative observational design. The evaluation in Mexico (thesis component 1) served as a case study to explore the methodological questions about the methods and approaches best suited for the evaluation of social and complex interventions (thesis component 2).

3.2 Data collection

3.2.1 Overview of data collection methods for thesis component 1: Evaluation of IPV prevention intervention

We selected data collection methods to address the two research questions for thesis component 1 (Table 2), as well as additional questions reflecting the programmatic theory of change (listed in Appendix 1). For the intervention group, the data collection methods comprised self-administered (pen-and-paper) baseline and endline questionnaires, direct observation of intervention sessions, separate focus groups with teachers and health educators, separate focus groups with male and female students who participated in the intervention, one-time in-depth interviews at the end of the semester, and repeat interviews throughout the intervention with a small group of 'case study' participants. For the comparison group, the data collection methods comprised self-administered (pen-and-paper) baseline and endline questionnaires and separate focus group discussions with male and female students. We conducted all data collection in Spanish. I present the sample,

implementation timeline, and purpose for each of the different forms of data collection we used in the evaluation in Table 3.

Table 3: Summary of data collection methods in evaluation of IPV prevention intervention

Method	Sample	Participants	Timing	Purpose
Intervention group				
Baseline questionnaire	All intervention students with informed assent and parental consent were invited	124 respondents	Aug 2017	<ul style="list-style-type: none"> • To understand the context across the full study sample and learn about the experiences of young people with relationships and IPV. • To assess baseline knowledge and attitudes related to violence and gender. • To inform sampling frame for interviews and focus group discussions. • To enable integration of qualitative and quantitative data from multiple timepoints for individual participants, in order to explore in more depth participant attitudes, intentions and behaviours and learn about the effects of the intervention.
Endline questionnaire	All intervention students with informed assent and parental consent were invited	95 respondents	Dec 2017	<ul style="list-style-type: none"> • To learn about satisfaction and self-reported level of participation. • To examine any changes from baseline to endline in knowledge or attitudes among individual participants and the group average. • To enable analysis comparing intervention and comparison groups.
Focus group with students	Volunteers from those who completed the baseline questionnaire	24 participants total in three focus groups	Dec 2017 and Mar 2018	<ul style="list-style-type: none"> • To learn about knowledge, beliefs and social norms about gender and IPV in the group. • To learn about any IPV in the school context and how young people respond to and describe it.

				<ul style="list-style-type: none"> • To examine participant perceptions of the influence of the course on beliefs about gender and IPV among the group. • To assess which aspects of the course were most influential on participants.
Focus group with health educators	All Mexfam health educators who implemented CSE for the study (including pilot)	Five participants in two focus groups (one in pilot, one research phase)	Jun 2017 (pilot) and Dec 2017 (study)	<ul style="list-style-type: none"> • To learn about health educator perspectives about the implementation process and curriculum strengths and weaknesses. • To learn about health educator observations of changes they observed in participants and which elements of the course may have triggered those.
Focus group with teachers	All teachers of the classrooms who received the CSE intervention (including pilot)	Five participants in two focus groups (one in pilot, one research phase)	June 2017 (pilot) and Dec 2017 (study)	<ul style="list-style-type: none"> • To explore perceptions of implementation and of how the intervention worked in the school context. • To learn if teachers had observed any changes in participants that may be related to the course.
Repeat interviews ('case studies')	Purposive sample based on baseline questionnaire data to ensure diversity of characteristics and experiences	Nine participants: five women, four men. Each completed two to four interviews, for a total of 33 repeat interviews.	Every 4-6 weeks during course; with final interview in Mar 2018	<ul style="list-style-type: none"> • To assess possible influence of the course on participants' beliefs and behaviours related to IPV prevention and gender norms. • To learn about how the course shifted participants' beliefs, relationship trajectories, and pathways to IPV prevention, with a focus on gradual processes of change and 'tipping points' or triggers. • To learn about participant relationship and sexual history, experiences and perceptions of IPV, and family and social context. • To learn about perceptions of the intervention and how participants felt their classmates responded.

				<ul style="list-style-type: none"> • To learn about experiences participating in the questionnaires and interviews. • To assess which aspects of the course seemed influential on IPV prevention. • To assess if the course influenced on care-seeking behaviour for IPV or other sexual health needs.
One-time in-depth interviews	Purposive sample based on baseline questionnaire data to ensure diversity of characteristics and experiences	20 (10 women, 10 men)	Mar to Apr 2018	<ul style="list-style-type: none"> • To assess possible influence of the course on participants' beliefs and behaviours related to IPV prevention and gender norms. • To learn about participant relationship and sexual history, experiences and perceptions of IPV, and family and social context. • To learn about perceptions of the intervention and how participants felt their classmates responded. • To learn about experiences participating in the questionnaires and interviews. • To assess which aspects of the course seemed influential on IPV prevention. • To assess if the course influenced on care-seeking behaviour for IPV or other sexual health needs.
Observation	The research assistant observed sessions weekly; I also observed occasional sessions when in Mexico.		Sep to Dec 2017	<ul style="list-style-type: none"> • To observe implementation and contextualise the data gathered through other data collection methods. • To observe the body language of the group during the intervention and interactions among participants.
Comparison group				
Baseline questionnaire (pen-and-paper)	All comparison students with informed assent and	116 respondents	Sep to Oct 2017	<ul style="list-style-type: none"> • To understand the context across the full study sample and learn about the experiences of young people with relationships and IPV.

	parental consent			<ul style="list-style-type: none"> • To assess baseline knowledge and attitudes related to violence and gender.
Endline questionnaire	All intervention students with informed assent and parental consent were invited	57 respondents	Dec 2017	<ul style="list-style-type: none"> • To enable analysis comparing intervention and comparison groups.
Focus group discussions	Volunteers	17 participants total in 3 focus groups	Dec 2017 and Apr 2018	<ul style="list-style-type: none"> • To learn about gender norms and knowledge and beliefs about IPV, and whether these differ between intervention and comparison groups. • To learn about any IPV in the school context and how young people respond to it.

3.2.1.1 Design of data collection instruments

The initial program theory of change was the basis for developing all of the data collection instruments. We mapped each assumption and outcome from the theory of change to questions in interview guides, focus group discussion guides, and the questionnaires. I drafted all of the data collection instruments based on a review of the literature, with a focus on addressing the assumptions and hypotheses in the programmatic theory of change. I refined each instrument iteratively in collaboration with study partners, and each instrument was piloted and refined with the participation of several youth volunteers at Mexfam. Each instrument was then piloted with a classroom during the pilot phase of the study. I subsequently refined each based on the pilot experience, and implemented the final versions of each instrument in the full study. More detail about the development of key data collection instruments follows.

In-depth and repeat interviews with students

The design of the one-time and repeat interview guides focused on eliciting stories about the lived experiences of the participants and their friends and assessing the possible

influence of the course on these. Rather than asking directly about attitudes about gender norms of violence, we sought to hear about their experiences, and through these narratives we would interpret information about their attitudes or beliefs and the social norms in their community. If participants had no personal experiences of relationships or violence to draw from, or did not want to share these experiences, we included more generic questions that could be asked in those cases. We used similar questions in the two interview guides, but gave more time to prompt on personal experiences and family and social context in the repeat interviews, and asked about recent sessions of the course in each of the repeat interviews.

Focus group discussion guides with students

Unlike the in-depth interviews, where we focused on participants' lived experiences, in the focus group discussions with young people we focused on how participants spoke about relationships and what they believed others thought and did in terms of gender, sexuality and relationships. We sought to learn about participants' perceptions of the types of relationships and manifestations of intimate partner violence among students at the school, as well as any influence of social norms on these. One way we did this was by presenting relationship-related scenarios that reflect real-life circumstances that young people might encounter, and asking what they thought others their age might do and should do in these situations. This approach draws from the use of vignettes to measure social norms in order to provide a non-threatening way for participants to share beliefs (Blum *et al.*, 2019). We considered using the detailed vignettes from the intervention curriculum as discussion points in the focus group discussions, but ultimately used shorter questions such as 'Imagine this situation: Another student at the school is in a violent relationship and asked a range of questions including what they believed the student should do, and differences in what is acceptable based on the type of relationship. The focus group discussion guide for students is available in Appendix 3.

Focus group discussion guides with teachers and health educators

The focus group discussion guides with teachers and health educators focused on learning about perceptions of the implementation process and observations of any shifts they noticed among the students in the class. The guide focused on their observations and perceptions of the trajectories of participants through the intervention and which elements of the intervention appeared to influence on participants.

Survey questionnaire

The survey questionnaire was developed to describe the sample with respect to self-reported experience of relationships, sexuality, and partner violence, gender-equitable attitudes, attitudes about the acceptability of violence, and knowledge about where to seek support for sexual health needs and in cases of partner violence.

Questionnaire development drew on a review of the following instruments that have been used for similar research. The Gender Equitable Men (GEM) Scale, designed for evaluating interventions that 'promote gender equity and reduce HIV/STI and violence risk' (Pulerwitz and Barker, 2008), has already been adapted to and used with men and women in Mexico (Barker, Aguayo and Correa, 2013). The WHO multi-country study (Garcia-Moreno *et al.*, 2006) measures physical and sexual violence and controlling behaviours by an intimate partner. The Sexual Relationship Power Scale measures power within intimate partnerships (Pulerwitz, Gortmaker and DeJong, 2000). Other studies identified items related to violence against women and girls (Mūkoma *et al.*, 2009; Rijdsdijk *et al.*, 2011; Holden, Bell and Schauerhammer, 2015). I also reviewed self-efficacy scales related to sexual health (Koch, Colaco and Porter, 2011) and relationships (Lopez, Morua and Rice, 2007).

Drawing from the sources noted above, I identified potential items that aligned with the research questions or addressed aspects of the program theory of change that we aimed to explore. I then reviewed the list of items that might be relevant to the study, adapted them as needed to the context of this study, and translated the most relevant items to Spanish. I iteratively refined drafts of the survey based on conversations with the research team. Mexfam requested to include knowledge items and questions that they could use for program monitoring in the questionnaire. As such, I included questions in the instrument that were for the use of the study partners, and would not be incorporated into the data analysis for the evaluation. Refined drafts were shared with the *Gente Joven* educators for feedback. We then tested the questionnaire items with small groups of young people, to ensure comprehensibility and relevance. Once we had a full draft of the questionnaire, we implemented it during the pilot phase of the study. In this pilot phase, we also conducted focus groups to learn about how young people at the school talked about their relationships, sexual behaviours, and violence. We used some of the phrasing we heard from the

focus groups to refine the question wording after the pilot phase, and made other revisions based on the pilot data.

The final version of the questionnaire implemented after these adaptations (Appendix 4) has 35 questions. The purpose for each item is described below (Table 4).

Table 4: Purpose for each questionnaire item

Question	Purpose
1-5	Knowledge questions included for Mexfam’s internal purposes. Intended to assess any change knowledge about contraception from baseline to endline. This can provide evidence for TOC assumption about participants paying attention to the intervention contents.
6-11	Knowledge questions included for Mexfam’s internal purposes. Intended to assess any change knowledge about sexually transmitted infections including HIV/AIDS from baseline to endline. This can provide evidence for TOC assumption about participants paying attention to the intervention contents.
14-16	To conduct a list experiment (Blair, Imai and Lyall, 2014; Moseson <i>et al.</i> , 2017) in order to attempt to calculate a population estimate for the class of how many have experienced dating violence.
12-13, 17-21	To gather data about participants’ relationship and sexual experience from their own perspective in order to understand the lived experience of the students; to inform the selection of a purposive sample for the qualitative methods that represents a range of relationship and sexual experience.
22	To have a sense at baseline of attitudes about gender equity among participants, and to assess change in agreement with gender equitable statements (or disagreement with gender inequitable statements) from baseline to endline.
23-25	To understand the lived experience of the students in relation to IPV at school/in their lives in general
26-27	To understand participants’ attitudes about IPV and the acceptability of violence, and see if these attitudes change from baseline to endline.

28-29	To see if there is change in knowledge about different types of IPV from baseline to endline.
31	To gather sociodemographic information (religiosity).
30, 32_pre	To understand where young people go to seek information and support in relation to IPV.
32_post	To assess the dose of each student to the intervention.
33-35	To assess the satisfaction of the student with the CSE course.

3.2.1.2 Data collection process and timeline

We began baseline data collection in the intervention group in August 2017 and in the comparison group in September 2017. However, two significant earthquakes in Central Mexico in September resulted in one day of school closure after the first earthquake and nearly three weeks after the second one. This delayed our data collection timelines (Table 5). The questionnaire implementation was particularly impacted by the school closures. The baseline data collection for the comparison group was not yet complete at the time of the school closures; we finalised this data collection after the school reopened, nearly six weeks after the beginning of the semester. The earthquake was traumatising for many students, and they appeared distracted and nervous after their return to school. This may have influenced responses to the questionnaire in ways that we can not ascertain. The delays and circumstances of the baseline data collection for the comparison group led us to question the comparability of the data with the intervention group. I discussed the data quality risks with members of the research team and decided that data quality for the comparison group questionnaires were compromised and that quantitative analysis would be unlikely to yield meaningful findings.

Table 5: Timeline of data collection activities, by method and group

August 2017 – May 2018	Aug	Sep ¹	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
INTERVENTION GROUP										
Baseline questionnaire	X									
CSE course		X	X	X	X					
Observation of CSE sessions		X	X	X	X					
Repeat interviews		X	X	X	X		X	X		

¹ Significant earthquakes on September 7 and 19 resulted in cancelled school days on September 8th and from September 20 until October 8, 2017.

Endline questionnaire					X					
Focus groups – teachers and health educators					X					
Focus groups – students					X		X	X		
In-depth interviews							X	X	X	
COMPARISON GROUP										
Baseline questionnaire		X	X							
Endline questionnaire					X					
Focus groups – students					X		X	X		
CSE course							X	X	X	X

After the school reopened, the school administration announced a two-week extension of the semester. This would allow Mexfam to provide the intervention in full and ensure sufficient time for endline data. However, three weeks before the end of the semester the school administration reversed this decision and planned to end the semester one week earlier than expected. This had implications for the study, as the endline questionnaire had to be rescheduled into the now compressed last week of the semester. We were able to implement the endline questionnaire during the examination period and immediately before the two-month Christmas holidays. The school setting was chaotic and absentee rates were higher than usual. In the intervention group, 95 of the 124 baseline respondents in the intervention group (77%) completed the endline questionnaire. In the comparison group, endline data collection was more challenging, with less than half of respondents completing the questionnaire (57 of 116 participants; 49% response rate). The delays and their impact on data quality are explored further in Chapter 7.

3.2.2 Data collection methods for thesis component 2: Methodological exploration

To examine the methodological questions addressed in this thesis, I wrote field notes about my observations throughout the evaluation process. I also engaged in conversations with the research team about their experience with the research process. I recorded these conversations, as well as my own thoughts and experiences, in analytic memos (Saldaña, 2009). The different forms of data collected in the evaluation of the IPV prevention intervention in Mexico also served as raw data for the methodological exploration.

3.3 Data analysis

3.3.1 Thematic analysis based on programmatic theory of change

I analysed the qualitative data using thematic analysis with a combination of deductive and inductive approaches (Fereday and Muir-Cochrane, 2006). For the deductive analysis I used a 'theoretical' thematic analysis, which is 'driven by the researcher's theoretical or analytic interest' (Braun and Clarke, 2006). Specifically, I created a 'start list' of codes (Miles and Huberman, 1994) based on the hypotheses expressed in the theory of change and the broader research questions of the study. For the inductive analysis, I created new codes as I reviewed the transcripts in order to capture emerging concepts 'without trying to fit it into a pre-existing coding frame' (Braun and Clarke, 2006). This combination of approaches to thematic analysis allowed me to generate both data-driven and theory-driven categories for analysis (Gläser and Laudel, 2013).

The first step of thematic analysis was familiarisation with the data (Braun and Clarke, 2006). All interviews and focus groups were professionally transcribed in the original Spanish, and I conducted the coding, analysis and interpretation in Spanish. I imported interviews into a qualitative data management software (NVIVO Server 11.0) as soon as they were transcribed and reviewed them within two weeks. I wrote memos about each interview to summarise emerging themes. I also listened to the full audio recording of most of the interviews. Because the language used by the students in the study was informal, often grammatically incorrect and used local colloquialisms, I involved the research team members based at Mexfam to assist with interpretation of any phrases that were difficult for me to understand.

The next step was to index all the transcripts using the 'start list' of codes as well as inductively developed new codes that I identified as I reviewed the transcripts. As I became more familiar with the qualitative data, I adopted a 'progressive focusing' approach, which entailed an iterative process to adjust the codes and adapt the data collection process to better fit with the evolving focus of the study (Stake, 1981; Sinkovics and Alfoldi, 2012; Schutt and Chambliss, 2013). For example, I provided feedback to the interviewer on refinements to the interview guide to ensure that she asked questions that focused on focal themes of the study. I also identified follow-up questions for each repeat interview that the interviewer would incorporate into the next interview, a benefit of qualitative longitudinal research noted by others (Vincent, 2013; Burke *et al.*, 2019).

Once the majority of interviews were transcribed, imported into NVIVO and indexed, I then reviewed the memos I had written for each interview to identify themes across participants. I exported from NVIVO all the codes linked to each main theme and combined them into one document, to integrate data from different sources (interviews, focus groups) and participants (teachers, health educators, students). I then reviewed the excerpts within each theme and began an iterative process of reorganising the main themes and the codes assigned to each. At this stage, I began a process of analytic memo-writing to facilitate the analysis process (Saldaña, 2009), moving from summarising the excerpts within each theme to interpretation of the data. I periodically shared the memos with my doctoral supervisor, who was also co-Principal Investigator on the evaluation study in Mexico, for input and discussion. In addition, I presented the emerging themes and analysis to colleagues from Mexfam and IPPF/WHR at multiple points in the process and facilitated discussions in which we collectively analysed and discussed the data. These conversations led me to return to the transcripts to reconsider the findings and then update the thematic narratives accordingly.

I then explored whether there was evidence to support the different hypotheses articulated in the theory of change. For each hypothesis and assumption in the theory of change, I consolidated illustrative excerpts from the data that provided evidence about that topic. I then examined the thematic summaries I had already written to identify any phenomena or themes highlighted in the summaries that had not yet been linked to the theory of change. I used these to update the theory of change, adding new types of outcomes that were not part of the original TOC. Based on the emerging findings, I also changed the order of the outcomes and removed hypotheses for which there was no evidence. I then facilitated an in-person meeting with Mexfam and IPPF/WHR staff in June 2018 to interpret the data and refine the theory of change into proposed pathways of change that were based on the findings. The four pathways were: (1) communicating about relationships, sexuality and violence; (2) taking protective and preventative actions to promote equitable and less violent relationships; (3) accessing violence-related and sexual and reproductive health services; and (4) shifting attitudes and behaviours related to gender, sexuality and violence. This final TOC was substantially more detailed than the original version and was grounded in the evidence that emerged from the study (Appendix 2, shown in Spanish).

The results of this thematic analysis are reported in Chapter 4.

3.3.2 Quantitative analysis of questionnaire data

Using baseline questionnaire data from the intervention group, which was completely collected before the earthquakes in September 2017, I used descriptive statistics (frequencies, percentages and means) in STATA to report on the sociodemographic information and self-reported experiences and beliefs of participants. These are reported in Chapter 4 (Table 7).

I conducted preliminary analyses using endline data from the intervention group, presenting the percentages of those reporting on their satisfaction with the course and number of sessions attended. I then conducted paired baseline-endline comparisons between groups of knowledge items (questions 28-30) about violence using a Wilcoxon matched-pairs signed-ranks test. For question 28, I first constructed a variable to calculate the number of types of behaviours from the list that participants correctly identified as forms of IPV, and then compared the mean score between baseline and endline. I conducted similar analyses using a Wilcoxon matched-pairs signed-ranks test to compare between responses at baseline and endline in the items examining attitudes about violence (question 27) and attitudes about gender (question 22). However, due to the limitations of questionnaire data quality described elsewhere (Chapter 7), these analyses, along with other preliminary explorations of the questionnaire findings, were discussed internally with the study partners but are not reported on in this thesis.

We included questions 14-16 in an attempt to use a list experiment to increase reporting of experiences of IPV. In list experiments, sensitive or stigmatised behaviours and experiences are not asked directly of the participant, which has been found to increase reporting of such behaviours or experiences (Blair, Imai and Lyall, 2014; Moseson *et al.*, 2017). However, many of these questions were left blank, and we were unable to conduct meaning analysis of these data, which are not reported further in this thesis.

3.3.3 Qualitative longitudinal analysis using framework method

Building on the thematic analysis, I then focused on the 33 repeat interviews with nine 'case study' participants. I conducted a qualitative longitudinal analysis focused on one pathway from the theory of change: taking protective and preventative actions to promote equitable and less violent relationships. Specifically, I wanted to understand how the intervention might influence participants' relationship trajectories. In order to manage the relatively large quantity of data, I used the Framework Method; this approach uses a matrix-based format to organise data, which is

useful for case-based research because 'the views of each research participant remain connected to other aspects of their account within the matrix so that the context of the individual's views is not lost' (Gale *et al.*, 2013).

I followed the five steps of framework analysis: familiarisation; identifying a thematic framework; indexing; charting; and mapping and interpretation (Ritchie and Spencer, 1994). I had already completed the familiarisation step in the prior analysis phase, as described above. For the second step, I used the updated TOC as the theoretical framework. For the third step, I again reviewed all the transcripts and indexed excerpts relevant to the focus of the analysis on relationship trajectories. For the charting phase, I used a 'time-ordered, sequential' matrix-based approach tailored to longitudinal qualitative analysis (Grossoehme and Lipstein, 2016). In Excel, I created one matrix per participant (example in Appendix 5) with one column for each interview and one row for each key theme: comfort discussing sexuality and relationships; seeking information and care; gender norms; sexual diversity; relationships and how to behave; violence; jealousy; intervention group dynamics; and observations about course impact on others. These were identified based on the revised TOC and on themes that had emerged in the prior analysis.

I then created a consolidated matrix focused on change by theme. This matrix had one column per participant and one row per theme. In each cell, I summarised the change, or lack thereof, within each theme per participant (example in Appendix 6). This consolidated change matrix facilitated the final stage of framework analysis: mapping and interpretation. In this phase, I wrote analytic memos to reflect on patterns, similarities, or differences between cases and identified factors that appeared to influence the participant relationship trajectories or their experiences in the intervention. I then shared the summary matrix and analytic memos with members of the research team for group discussion and interpretation, which aided in refining the analysis. I selected five of the nine 'case study' participants who described experiences that were commonly discussed in the repeat interviews, and I iteratively wrote a case history for each (Thomson, 2007) to demonstrate how the intervention influenced their relationship trajectory in relation to IPV prevention.

These case histories are presented in the results section of Chapter 5.

3.3.4 Iterative ground approach to analysis for methodological exploration

I used an iterative, grounded approach to define the analysis strategy for the methodological exploration. At the outset of the study, I set out potential questions that could guide the analysis for the methodological exploration:

- How useful might quantitative data be in evaluations of this type of intervention?
- Why do certain types of data take priority over others? What are the external and internal pressures that influence method choice? To what extent is method choice related to the perceived rigour of the research?
- Is it useful include a comparison group in a qualitative research design? How does it influence the credibility of findings?
- What additional data are gained from the longitudinal case study approach in examining individual change in response to an intervention? Does this approach allow for the attribution of change to an intervention?
- How can qualitative research be used to attribute results to an intervention, and will this approach be considered less rigorous by key stakeholders?
- How does co-production of evaluation research influence research question development, research design, method choice, utilisation of findings, and perceptions of the rigour of the research?
- To what extent are co-produced programme evaluation findings considered relevant and rigorous to local implementing partners as well as the broader academic and practitioner community?

I planned to iteratively define the most relevant questions and related analyses in response to the evolving experience in the field and nature of the data that we gathered in the evaluation in Mexico.

As the evaluation advanced, I began to refine the analytic focus for the methodological exploration. I reviewed field notes and email communications from the evaluation process. I then became familiarised with the data from each data collection method: focus groups with students, focus groups with teachers, focus groups with health educators, one-time in-depth interviews, longitudinal qualitative interviews, and the baseline and endline questionnaires. I also reflected on

the unique circumstances and context of this evaluation. For example, disruptions such as earthquakes and flooding delayed the research and intervention timelines, and led me to exclude much of the quantitative data from analysis. I wrote analytic memos (Saldaña, 2009) to develop my ideas for analyses that appeared relevant to the particularities of the evaluation experience in Mexico.

Based on this iterative process of defining an analysis strategy, I decided to focus the methodological exploration on three evaluation approaches. First, I examined the challenges and benefits of longitudinal qualitative research for evaluation (Chapter 5) because of the potential for repeat interviews to shed light on participant experiences throughout the intervention and help us consider how shifting timelines and other day-to-day events influenced participant responses to the intervention (Saldaña, 2003). Second, I chose to explore the use of a complex adaptive systems approach for evaluation (Chapter 6) to address the influence of unpredictable events on evaluation and intervention processes. Finally, in response to the low quality of the quantitative questionnaire data, I focused on the challenges encountered in the evaluation in Mexico (Chapter 7). These analyses are summarised below and described further in the relevant chapters.

3.3.5 Post-hoc analysis based on social complex adaptive systems concepts

This analysis focused on the complexity of the intervention process and examined the diversity of experiences in and outcomes of the intervention. The analysis drew from two data sources that included rich detail about intervention processes. First, the focus groups with the facilitators of the course; and second, the 33 repeat interviews that took place throughout and after the intervention.

I conducted a post hoc analysis of the qualitative data from these two sources, examining it against the attributes of social complex adaptive systems identified by Keshavarz et al. as particularly relevant to schools: diverse and dynamic agents, interaction, unpredictability, emergence, and context dependency (Keshavarz *et al.*, 2010). I compiled excerpts from the transcripts that provide examples for how each of these complexity concepts appear to have occurred in the implementation of the intervention and how they may have influenced the course outcomes. I paid particular attention to narratives from different sources – facilitators and students – about the same event, allowing comparison from different perspectives. I also aimed to identify ways in which the evaluated comprehensive sexuality education programme with a

gender-transformative approach was potentially disrupting the system in which it was implemented (Moore *et al.*, 2019), for example by shifting the formal and informal rules that influence individual and group behaviour among the intervention participants.

3.3.6 Comparison of self-administered questionnaire and one-on-one interview data

I iteratively developed an analysis strategy based on a phased exploration of the data. Because of the data collection challenges experienced with the questionnaire, I conducted descriptive statistics to assess the quality of the questionnaire data in both the intervention and comparison groups. I compared levels of incomplete data (missing and ‘prefer not to say’) between different questions, and between men and women, intervention and comparison group participants, and repeat interview participants and one-time interviewees, at baseline and endline. I focused this analysis on the ten questionnaire items that addressed participant experiences with relationships, sexuality and violence (Table 6).

Table 6: Questionnaire items about relationships, sexuality and violence

	Item	Response categories	Comments about item wording
1	Have you ever gone out with someone, as in a dating relationship? It could be with a boyfriend, girlfriend, friends with benefits, etc.	Yes No I would rather not say	These question were worded based on some of the most common ways that young people at the school talked about different types of relationships in focus group discussions.
2	Are you currently going out with someone, as in a dating relationship? It could be with a boyfriend, girlfriend, friends with benefits, etc.	Yes No I would rather not say	

3	Have you ever had sexual contact? For example, touching or being touched on the genitals, anal sex, vaginal sex or oral sex?	Yes No I would rather not say	This question was worded based on testing of the item with small groups of young people.
4	Have you ever experienced dating violence in your own relationship (past or present)?	Yes No I would rather not say	We used this question to identify participants who had ever experienced IPV, based on their own definition at the time of the survey. This helped us identify a sample for qualitative data collection with some type of IPV history. This question is not intended to measure prevalence.
5	Do you think the majority of your female friends have already had sex?	Yes No I would rather not say	These questions were worded based on testing of the item with small groups of young people. We tested various wording options for this question, but ultimately failed to include non-binary gender identities.
6	Do you think the majority of your male friends have already had sex?	Yes No I would rather not say	
7	As far as you know, has there been any type of dating violence at your school? It could be among	Yes No I would rather not say	We used this question to learn about whether participants had observed this form of violence to be common in their school, based on their own definition at

	boyfriend/girlfriend, friends with benefits, etc.		the time of the survey. This question is not intended to measure prevalence.
8	With whom would you say that you prefer to have a dating relationship? It could be serious, casual, friends with benefits, etc.	<p>With someone of the same sex as me</p> <p>With someone of the opposite sex as me</p> <p>With people of both sexes</p> <p>I don't know</p> <p>I would rather not say</p>	We tested various wording options for this question, but ultimately failed to include non-binary gender identities.
9	Have you ever felt agreed to have sexual contact with someone even though you didn't want to?	<p>Yes</p> <p>No</p> <p>I would rather not say</p>	This question aimed to identify any experience of sexual coercion, to allow sampling of these participants for interviews.
10	Have you ever heard of, seen or experienced sexual harassment in your school? For example, whistles, sexual advances, inappropriate looks, unwanted touching, kissing, pressure to have sex through threats, etc.	<p>Yes</p> <p>No</p> <p>I would rather not say</p>	We used this question to learn about whether participants had observed or experienced this form of violence to be common in their school, based on their own definition at the time of the survey. This question is not intended to measure prevalence.

I also compared responses to the seven question that were asked directly in both in-depth interviews and self-administered questionnaire (Table 6). In this comparison, I identified the

extent of the consistency between data sources for each of the 28 participants who had completed the baseline and endline questionnaires and at least one in-depth interview. This participant-level analysis resulted in 196 comparisons – seven questions for each of the 28 participants. I categorised each comparison as consistent, partially consistent, inconsistent, or insufficient data. I approached the comparison of questionnaire and interview data acknowledging that conflicts between data points are not necessarily an indicator of a problem in the data, but may instead reflect that different methods of data collection result in data that express different aspects or attributes of a complex topic (Slonim-Nevo and Nevo, 2009).

I used descriptive statistics to report the proportion of complete questionnaire responses overall. I then compared this proportion of complete responses between intervention and comparison groups using a two-sample test of proportions to test for any difference in proportion of complete responses at baseline and endline overall and by gender (as shown in Chapter 7, Figure 5 and Table 13).

I then conducted a thematic analysis (Braun and Clarke, 2006) to examine excerpts from the interview transcripts that reflected participant experiences completing the questionnaire and their perceptions of how their classmates responded. I also reviewed field notes and emails about the implementation conditions for the questionnaire and wrote analytic memos (Saldaña, 2009) reflecting on these to identify barriers to data collection.

3.4 Ethics

The study was approved by LSHTM Research Ethics Committee in the United Kingdom (LSHTM Ethics Ref: 11864) and *Comité de Bioética y Ciencia Para la Investigación, Centro de Investigación Clínica Acelerada (CICA)* in Mexico. Approval letters are presented in Appendix 7.

3.4.1 Ethical considerations for thesis component 1: Evaluation of IPV prevention intervention in Mexico

3.4.1.1 Working on IPV with minors

This research addressed issues of violence and relationships among 14- to 17-year-olds. This is a vulnerable population, and we prioritised protecting participants and their wellbeing throughout the research process. We followed ethical guidelines for research of this nature (Jewkes, Dartnall and Sikweyiya, 2012; Devries *et al.*, 2015) and were in contact with Mexfam programmatic and

clinical staff to ensure that all procedures complied with best practices for the context, and that participants' needs were met and rights were respected.

Because of the focus of this study on IPV and lived experiences, some participants disclosed experiences of family or relationship violence. Mexico City has a protocol for the identification, referral and care in cases when violence occurs within the school system. The protocol applies for school staff, but not for those working for non-governmental organisations such as Mexfam. Mexfam therefore developed its own protocol, in line with Mexican law, for disclosures of violence by research participants. If the participant wanted the Mexfam team to accompany them to formally report the case of violence, the team would do so. If the participant did not want to report the case, the team respected this preference. If participants approached a Mexfam health educator or research team member and disclosed either violence or a related health concern, they were referred for appropriate services. If those services were available at Mexfam, the participant was referred for free services, covered by the project budget. If Mexfam did not provide the needed services, the participant was referred for free services in the public sector.

3.4.1.2 Research team training and safety

All research team members completed ethics training, reviewed Mexfam and IPPF's child protection policies, and were trained on confidentiality, safety and protocols for the disclosure of violence. Research team members in Mexico met monthly, more often if needed, to debrief and discuss issues and concerns that emerged throughout the study. Trained counsellors at Mexfam were available to provide free counselling to research team members as needed.

3.4.1.3 Informed consent

The director of the participating school signed a memorandum of understanding with Mexfam to consent to participate in the research. The document served as institutional consent for the school's participation in the research. The school leadership was responsible for notifying school staff and teachers about the details of the research, and ensuring they understood that their individual participation was voluntary.

According to Mexican law, young people under 17 years of age must provide informed assent to participate in research and the consent of parents is also required. We held in-person information sessions with parents before the study began to inform them about the study objectives and to

answer their questions. Parents were able to withdraw the student for whom they were responsible by text, phone, or in writing. In case the parents were unable to attend the in-person meetings, they received the information sheet and consent form from the teacher. The decision to take a passive consent approach was made in order to include as many participants as possible from the classroom, to increase the quality of the study and the likelihood that the findings will be relevant to a diversity of youth.

Only students with parent/guardian consent were eligible to participate in the study. We implemented an informed assent process for all the students in the intervention and comparison classrooms who chose to participate in the questionnaires. The assent materials were designed in text and format to be comprehensible to young people. The research assistant read the materials out loud to each class, answered questions, and gathered signed assent forms. For the in-depth interviews and focus group discussions, we implemented informed assent processes using materials adapted to each data collection method before collecting data.

Teachers and health educators – all of whom were over 18 years old – provided written informed consent to participate in focus group discussions. They were asked to read the informed consent materials, given the opportunity to ask any questions they might have, and then signed the consent form.

3.4.1.4 Confidentiality

All research information from participants is confidential, and only the research team has access to the data, which is securely stored. The research team used unique identifiers rather than full names for field notes, interview transcripts, and audio recordings. We used pseudonyms for analysis and when disseminating results. We used the unique identifier to link each students' baseline questionnaire with their endline questionnaire and any interviews or focus groups they participated in. This identifier is the only way a student can be identified; the master list that matches a student to a code is stored in a separate locked cabinet in Mexfam. In order to protect confidentiality, identifying information was not stored with research data.

The interviewer did not ask questions about names or other identifying information during the audio-recorded interviews or focus groups. If identifying information was mentioned by the

participant during the audio recordings, the transcription used only a first initial. Any identifying information was removed for analysis and publication.

3.4.1.5 Risks to participants

Participants discussed sensitive topics during this study, which may have made them feel uncomfortable or distressed. During the informed consent and assent processes, each participant was told they could discontinue the interview, focus group discussion, or questionnaire at any time; this message was repeated at the beginning of each interview and at any difficult or emotional moments during interviews. Participants were informed in the informed assent process of potential implications of disclosing personal information about sexual activity or experiences of violence. The agreement signed with the participating school ensured that the school would support Mexfam in implementing the research in the most ethical way possible and safeguarding the participants.

3.4.1.6 Benefits for participants

The information gathered through this research will support the development of strategies to support young people in relation to their health and relationships. Participants may not gain personally from participating but will be helping other young people by informing programmatic improvement. This is an opportunity for participants to make sure their voices are heard and to have the chance to share their experiences.

3.4.1.7 Compensation for participation

Small gifts in the form of a gift card (the equivalent of less than 10 GBP) were given to participants to compensate them for their time in the interviews and focus group discussions. In the case of focus group discussions, snacks and beverages were offered to participants. Students were not given any compensation for participation in the baseline or endline questionnaires as these took place during class time.

3.4.1.8 Protection of data

Data protection is of paramount importance. Data security was ensured through standard practice for document storage in compliance with relevant LSHTM data protection and management policies. All research documents are stored on the LSHTM and IPPF/WHR secure servers. We used unique identifiers codes and pseudonyms rather than full names for interview notes, field notes,

interview transcripts, and audio recordings. Field notes, meeting minutes, transcripts, and recordings were only shared within the research team. In accordance with LSHTM's retention requirements, primary research data will be retained in secure long-term storage for 10 years following study completion.

3.4.2 Ethical considerations for thesis component 2: Methodological exploration

3.4.2.1 Risks to participating organisations

The methodological exploration included observation of a co-produced evaluation as it was carried out. This brought ethical challenges in that the work of the partner organisations – local and regional NGOs – became data for analysis and critique. NGOs, particularly those working in sexual and reproductive health, are often the targets of religious or government opposition, and there is a risk that information from this study, taken out of context, could create reputational harm or otherwise be a risk for the partner organisations.

3.4.2.2 Benefits for participating organisations

This study and the resulting publications and dissemination material can give visibility to the work of the partner organisations and position them as research-producing organisations. Staff have had opportunities to gain experience in research processes through their participation in this study, which also can benefit the institutions.

3.4.2.3 Institutional consent

I discussed the project, including the risks and benefits, with senior management at Mexfam and IPPF/WHR, who consented to participate. We agreed, in line with the commitments of the co-produced study, that the institutions would be able to review publications and other dissemination materials before they were submitted to ensure that there was no reputational risk.

3.4.2.4 Protection of data

All the field notes, emails, and other research data from this study are only accessible to the research team, and data are securely stored. Data security was ensured through standard practice for document storage in compliance with relevant LSHTM data protection and management policies. All research documents are stored on the LSHTM and IPPF/WHR secure servers. In

accordance with LSHTM's retention requirements, primary research data will be retained in secure long-term storage for 10 years following study completion.

3.5 Reflexivity

3.5.1 Motivation for this study

I was motivated to begin my doctoral studies to address some of the challenges and tensions I experienced in over a decade working on evaluation in a sexual and reproductive health international NGO. The politics of evidence was a backdrop to this work but was normalised to the point where it became business as usual. In the context of limited resources that is common for NGOs, evaluation can often be seen as instrumental – one of its important functions is to demonstrate the success or promise of programmes as a way to generate more funding. Because of this, we (myself and other staff working on evaluation) sometimes made methodological decisions to generate the types of evidence that would resonate with donors. For example, having quantitative figures and statistics was a strategy to appear more 'rigorous', to 'prove' that our programmes or projects had an impact, or to generate information that would translate well into an annual report or other institutional communications materials. The selected methods were not always best suited for the evaluation question, we often lacked resources to conduct evaluations well, and the requests or demands of donors frequently dictated what we measured and how we defined success – even if this was entirely unsuitable to the programme or context in which we were working. This situation also led us to sometimes collect data that were not a priority from the perspective of the service or programme, were unrealistic to gather, or had little meaning in context. While there were some opportunities to monitor and evaluate programmes rigorously – mainly when evaluation staff time and resources were included in the budget – we often conducted 'quick and dirty' evaluations that allowed us to gather some useful information to inform programming and report to donors while working within short-term timeframes and with insufficient funds.

Complexity affected all of the NGO's programmes, which focused on sensitive sexual health topics in often restrictive legal and social environments in a context of shifting political leadership, resistance from religious groups, and unpredictable weather events and disasters such as earthquakes. Yet in our approaches to monitoring and evaluation, there was no formal approach to address such complexity – while we were engaged in implementing and evaluating 'social

complex interventions,' we did not use those terms. We footnoted exceptions when there were dips or peaks in the data – in essence, explaining them away – and proceeded to work within the standard evaluation toolkit that most of us were versed in from public health, social science or international relations training.

My experience is by no means unique to the particular NGO where I worked. Studies signal that similar issues with evaluation are widespread in the NGO and international development sector (Laybourn, no date; Merchant-Vega, 2011; Liket, Rey-Garcia and Maas, 2014). This reflects the politics of funding and evidence (Eyben, 2013), the training that many evaluators receive, the tension between the urgency of direct service provision and the resources needed to conduct evaluation, and the disconnect between academic conceptualisations of evaluation research and much evaluation practice on the ground.

After years working in this context, I felt compelled to question the pressure – from not only donors, but also institutionalised within NGOs policies and practices – to use certain methods and answer questions that donors, governments, and organisational management had decided were priorities. I also observed that the methods in the NGO evaluation toolkit were unable to adequately address the inherent complexities of the diverse contexts in which we were working. This thesis presents my efforts to grapple with these tensions by attempting to conduct and learn from an NGO evaluation done 'the right way' – that is, with full participation by all partners throughout the evaluation process, particularly when defining what should be measured and how; a lack of donor pressure influencing the aims and methods of the evaluation; sufficient time spent to ensure high quality data; generating data that is not only useful locally for the project itself but also can inform other similar programmes; and using the study experience to engage with the broader debates about methodology for evaluation of social and complex interventions in ways that can inform future evaluation practice.

3.5.2 My role as researcher

I agree that 'no research is neutral' (Charmaz, 2006), and approached the research process reflexively to examine my own preconceptions and those of the research team and participants. I bring to this research my own experiences and knowledge from a career in sexual and reproductive health in Latin America and from my prior collaboration with Mexfam, as well as my personal beliefs about gender equality and IPV.

My role in this project entailed working at the crux of many of the tensions at the core of co-produced research. I played two different roles in this study. As the co-Principal Investigator working towards my PhD, I was tasked with meeting academic standards and planning for publishable research. As a long-term staff member, then consultant, at IPPF/WHR responsible for the partnership with Mexfam (IPPF/WHR's affiliate in Mexico), my role was to ensure the research was feasible and useful for all involved organisations. My role as IPPF/WHR staff and as the budget-holder for the research for the first two years of the project put me in a position of relative power with respect to decision-making. This tension also reflected my position representing an international NGO, also risking a situation in which the organisation would exert power through funding streams. In addition, as a white woman from the United States leading a Mexican research team, I was also aware of potential power imbalances and perceptions of colonial presence. I made all efforts, through close communication both in-person and remotely, to build a culture of transparency, trust and equal decision-making between all research partners and developing collective processes for analysis, interpretation and decision-making. To reflect the shifting roles that I played in this project, I alternate between the use of 'I' and 'we' – the former when describing work that was more specific to the PhD thesis, and the latter when referring to the co-produced evaluation in Mexico with Mexfam and IPPF/WHR.

There are other power imbalances that I was also cognisant of in the implementation of this project. The local research team was predominantly highly educated, earning a reasonable wage by local standards, and trained in sexual and reproductive health and rights. There were differences in belief systems with some of the teachers at the school who were resistant to violence-prevention training, and in age and socio-economic background between the research team and research participants, who were 14- to 17-year-old students in a public-school setting from lower-middle income families. We engaged in reflexive practice in debriefing sessions to discuss related issues and how they could be addressed.

Chapter 4: Preventing Intimate Partner Violence Among Young People—a Qualitative Study Examining the Role of Comprehensive Sexuality Education

4.1 Introduction

In this chapter, I will present the evaluation of a comprehensive sexuality education programme in Mexico City that focuses on IPV prevention. This evaluation is the first component of the thesis; it serves as the case study for the methodological exploration that is the second thesis component. This chapter introduces the intervention being evaluated, which aims to promote gender-equitable and non-violent relationships among young people in Mexico City. This analysis presented here addresses the first research question of the thesis: What are the mechanisms through which the intervention can promote gender-equitable and non-violent relationships among young people? This chapter provides the foundation for the methodological explorations detailed in Chapters 5, 6 and 7.

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1400132	Title	Ms
First Name(s)	Shelly		
Surname/Family Name	Makleff		
Thesis Title	Methodology for evaluating complex social interventions: understanding partner violence prevention for young people in Mexico City		
Primary Supervisor	Prof Cicely Marston		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Sexuality Research and Social Policy		
When was the work published?	2019		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item.
----------------------	-----------------

SECTION D – Multi-authored work

<p>For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)</p>	<p>I am the first author of this paper, and was responsible for study design, conception of the paper, data analysis, and writing, and also participated in data collection. Cicely Marston supported on study design, conception of the paper, analysis and writing. Because of the co-produced nature of the evaluation case study, key members of the partner organizations are co-authors based on their involvement in study design, data collection, or interpretation of findings, as well as critical revisions of the paper.</p>
---	---

SECTION E

Student Signature	
Date	1 June 2020

Supervisor Signature	
Date	1/6/20

Wednesday, April 29, 2020 at 14:39:13 Central European Summer Time

Subject: Re: Query for Sexuality Research and Social Policy (13178)
Date: Tuesday, April 28, 2020 at 10:25:25 PM Central European Summer Time
From: Melissa Delgado
To: Shelly Makleff

*** This message originated outside LSHTM ***

Dear Dr. Shelley Makleff,

Your article is Open Access.

Thank you.

Best Regards,
Sassa

Melissa T. Delgado
Production Coordinator
Journals Production

Springer Nature
T +63 (035) 532-9912 loc. 23618
M +63 9167525199

melissa.delgado@springer.com
www.springernature.com

Springer Nature is one of the world's leading global research, educational and professional publishers, created in May 2015 through the combination of Nature Publishing Group, Palgrave Macmillan, Mac

From: Shelly Makleff <Shelly.Makleff@lshtm.ac.uk>
Sent: Tuesday, April 28, 2020 5:28 PM
To: Melissa Delgado
Subject: Query for Sexuality Research and Social Policy (13178)

Dear Sir/Madam,

In 2019 I published an open access article with you. I will be submitting the article as part of my PhD and need to show evidence that I retained the copyright for the work. I'd like to request a response that confirms that this article is open access and I have retained copyright.

The article is:

Makleff S, Garduño J, Zavala RI, Barindelli F, Valades J, Billowitz M, Silva Márquez VI, Marston C. "Preventing intimate partner violence among young people – a qualitative study examining the role of comprehensive sexuality education." *Sexuality Research and Social Policy* (2019).
<https://doi.org/10.1007/s13178-019-00389-x>

Sincerely,
Shelly

--

Shelly Makleff, MPH
PhD Candidate at London School of Hygiene and Tropical Medicine

Page 1 of 2

4.2 Abstract

Comprehensive sexuality education may help prevent intimate partner violence, but few evaluations of sexuality education courses have measured this. Here we explore how such a course that encourages critical reflection about gendered social norms might help prevent partner violence among young people in Mexico. We conducted a longitudinal quasi-experimental study at a state-run technical secondary school in Mexico City, with data collection including in-depth interviews and focus groups with students, teachers, and health educators. We found that the course supported both prevention of and response to partner violence among young people. The data suggest the course promoted critical reflection that appeared to lead to changes in beliefs, intentions, and behaviours related to gender, sexuality, and violence. We identify four elements of the course that seem crucial to preventing partner violence. First, encouraging participants' reflection about romantic relationships, which helped them question whether jealousy and possessive behaviours are signs of love; second, helping them develop skills to communicate about sexuality, inequitable relationships, and reproductive health; third, encouraging care-seeking behaviour; and fourth, addressing norms around gender and sexuality, for example demystifying and decreasing discrimination towards sexually diverse populations. The findings reinforce the importance of schools for violence prevention and have implications for educational policy regarding sexuality education. The results suggest that this promising and relatively short-term intervention should be considered as a school-based strategy to prevent and respond to partner violence.

4.3 Research Paper 1

Background

Comprehensive sexuality education may help prevent intimate partner violence among young people by addressing inequitable relationships and the harmful gender norms that perpetuate violence. Despite this possibility, few evaluations of sexuality education have examined whether and how these programmes can prevent or mitigate intimate partner violence, which can be defined as ‘any behaviour within an intimate relationship that causes physical, psychological or sexual harm to those in the relationship’ (Krug *et al.*, 2002). Instead of measuring outcomes related to violence, studies evaluating sexuality education typically document reductions in HIV, sexually transmitted infection and unintended pregnancy rates (Kirby, Laris and Rolleri, 2006; Kirby, 2008; Fonner *et al.*, 2014).

It is important to identify which aspects of interventions are most likely to help prevent intimate partner violence. Sexuality education has long been conceptualised as tackling violence and gender equality in addition to providing information about health and sexuality (*Report of the International Conference on Population and Development, Cairo, 5-13 September 1994*, 1995). Drawing from this framework, what is termed an ‘empowerment approach’ to sexuality education incorporates content about gender and power to promote equitable relationships (Haberland and Rogow, 2015; UNFPA, 2015; United Nations Educational Scientific and Cultural Organization, 2018). According to a comprehensive review of 22 studies, sexuality education and HIV prevention programmes that address the topics of gender and power dynamics within intimate partnerships are five times more likely to reduce rates of sexually transmitted infection and unintended pregnancy than programmes that exclude these topics (Haberland, 2015). The same review concludes that it is important to incorporate gender and power as themes in sexuality education in order to address harmful gender norms and ‘increase the chances that young people will have relationships characterised by equality, respect and nonviolence.’ A qualitative study in Cambodia and Uganda similarly found that comprehensive sexuality education holds promise to prevent violence against women and girls by promoting gender equitable attitudes, life skills and changes in community norms (Holden, Bell and Schauerhammer, 2015).

Gender norms and other social norms influence the perpetration of violence (Instituto Nacional de Salud Pública, 2015; Jewkes, Flood and Lang, 2015). For instance, the IMAGES study in Brazil, Chile

and Mexico found that perpetration of partner violence is associated with behaviours and beliefs corresponding with dominant expressions of masculinity (Barker, Aguayo and Correa, 2013). Norms also reflect social expectations in terms of gendered behaviours, including sexual behaviours and preferences – for instance defining socially acceptable ways to behave in relationships or express one’s sexuality (Marston, 2005; Pulerwitz and Barker, 2008). Gender norms are of particular relevance for young people, who are in the process of gender socialisation (Blum, Mmari and Moreau, 2017; Chandra-Mouli *et al.*, 2017). Social norms related to gender and sexuality may even be reinforced in school settings, which are often heteronormative (Youdell, 2005; Toomey, McGuire and Russell, 2012); some of these norms may be a source of not only partner violence but also interpersonal violence in schools, such as bullying based on gender identity or sexual orientation.

Across diverse contexts and types of interventions, a ‘gender-transformative’ approach – aiming to shift social norms, particularly harmful gender norms – has been shown to be a central component of programmes that reduce partner violence (Jewkes *et al.*, 2008; Verma *et al.*, 2008; Rottach, Schuler and Hardee, 2009; Pulerwitz *et al.*, 2010; Heise, 2011; Ricardo, Eads and Barker, 2011; Ellsberg *et al.*, 2015; Fulu and Kerr-Wilson, 2015). A promising strategy for interventions that aim to shift social norms is to work with a group of individuals who will become agents of change to influence their community (Cislaghi and Heise, 2018a). This approach is used by *Hombres Unidos Contra La Violencia Familiar* (Men United Against Family Violence), a violence prevention programme working with Hispanic men in the United States that demonstrated promising results such as participants rethinking their ideas about violence and engaging in conversations in their community about partner violence (Nelson *et al.*, 2010). The SASA! programme in Uganda also takes this approach, and the programme has contributed to more egalitarian relationship dynamics as well as being associated with reductions in reported intimate partner violence (Kyegombe *et al.*, 2014; Starmann *et al.*, 2017). Evaluations of both these programmes also identified facilitator-led group discussion as a key strategy to encourage reflection about violence-related social norms (Nelson *et al.*, 2010; Kyegombe *et al.*, 2014). These strategies to shift social norms, as well as other elements that are considered central to prevention of partner violence, are also included in international standards for comprehensive sexuality education (United Nations Educational Scientific and Cultural Organization, 2018). However, these guidelines are not always implemented (Montgomery and Knerr, 2018) and so the effects of well-implemented sexuality education programmes on violence prevention remain unclear. Whether sexuality education that

adopts an empowerment approach and addresses harmful gender norms can prevent intimate partner violence in different settings – and how it might do this – requires further investigation.

In addition to the scarcity of research evaluating how sexuality education may address partner violence, there are further gaps in the literature about prevention. For example, there is little research in low- and middle-income countries about the ‘primary prevention’ of partner violence – intervening to prevent violence before it takes place (Arango *et al.*, 2014; Ellsberg *et al.*, 2015). We also lack evidence about effective strategies that prevent partner violence among adolescent girls (Blanc *et al.*, 2013), who may be at elevated risk of these types of violence (Morrison, Ellsberg and Bott, 2007; Abramsky *et al.*, 2011; World Health Organization, 2013). The growing literature on violence prevention calls for more research in low- and middle-income countries that examines not only programme effectiveness but also how the process of change happens (Kyegombe *et al.*, 2014; Fulu and Kerr-Wilson, 2015; Starmann *et al.*, 2017). This longitudinal study responds to these gaps in evidence by examining the mechanisms through which a comprehensive sexuality education intervention in Mexico may support the prevention of intimate partner violence among young people.

Intimate partner violence is common in Mexico: According to recent estimates, 44% of women aged 15 years and older in Mexico report at least one incident of partner violence in their lifetime (Instituto Nacional de Estadística y Geografía, 2018b). In a different study, over half of students surveyed at the Mexican National Polytechnic Institute reported ever experiencing romantic jealousy in a relationship (Tronco Rosas and Ocaña López, 2012). In addition, 10% of women and 13% of men reported having exerted controlling behaviours, such as monitoring a partner’s cell phone, email or social media, more than once in a relationship (Tronco Rosas and Ocaña López, 2012). The researchers note that these behaviours are often perceived as displays of caring but may indicate or lead to relationship violence (Tronco Rosas and Ocaña López, 2012).

Nearly 39% of the Mexican population is younger than 20 years old (Instituto Nacional de Estadística y Geografía, 2018a) and the country has high school-attendance rates (98% in primary and 79% in secondary school) (United Nations Children’s Fund, 2015). Despite some opposition in Mexico, school-based sexuality education in Mexico has potential for substantial reach (Chandra-Mouli *et al.*, 2018) and could contribute to violence prevention. In 2016, Fundación Mexicana para la Planeación Familiar (Mexfam), a non-governmental organisation that provides clinical and

community-based health care services in Mexico, revised its comprehensive sexuality education curriculum to include content explicitly aimed at preventing intimate partner violence. This work built on prior versions of the programme that had been found to improve communication about sexuality among young people and influence gender norms (Marston, 2004).

The updated comprehensive sexuality education course comprises 20 hours of curriculum delivered over a semester in weekly sessions by young people (aged 30 or younger) who are staff health educators for Mexfam's *Gente Joven* ('Young People') programme. The course uses a gender-transformative approach, tackles gender and power dynamics as cross-cutting themes, and includes a comprehensive set of topics including sexuality, intimate partner violence, unintended pregnancy and relationships. The course employs participatory techniques and encourages critical reflection on violence and gendered social norms. Students are given information on where and how to seek support for sexual and reproductive health and violence and are told about their right to seek health services. For brevity, in this paper we will refer to the comprehensive sexuality education course with a violence prevention component run by Mexfam as 'the course', and unless stated otherwise, all mention of violence refers to intimate partner violence.

Mexfam partnered with International Planned Parenthood Federation/Western Hemisphere Region and the London School of Hygiene and Tropical Medicine to pilot, implement and evaluate their updated comprehensive sexuality education course using mixed methods of data collection. This article presents participant experiences as well as how the course appears to support the process of prevention and response to intimate partner violence.

Methods

In 2017 and 2018, we conducted a longitudinal quasi-experimental study in Mexico City. The analysis presented in this article reflects primarily qualitative data focusing on the intervention group. Baseline sociodemographic and descriptive statistics for the intervention group are presented in Table 7; further quantitative analysis is not yet complete and is not presented here. The study took place at a state-run technical secondary school located in a commercial zone of the Tlalpan area in the southern part of the city. The school operates morning and afternoon sessions, each with different students and teachers, delivering vocational training in subjects such as automotive mechanics and food preparation. Students are primarily from lower-middle income families from Mexico City (Gómez Espinoza, 2006), and travel to the school from different parts of

the city. We first conducted a pilot in one classroom of the afternoon session and then used a coin toss to assign the school's morning programme to receive the course the following semester, with three classrooms participating.

We aimed to explore the ways the course may contribute to preventing partner violence, in line with the call for studies that examine the mechanisms of violence prevention (Starmann *et al.*, 2017). In the planning phase, programme and research staff collaboratively developed a series of hypotheses about how the programme might affect participants, for example learning to identify different types of partner violence, shifting attitudes about violence and gender norms, sharing information about violence, gender or sexuality with peers, seeking support and services if they experience violence, and ultimately experiencing less violent or more equitable relationships. We grouped these hypotheses into what we termed a 'theory of change' (Silva *et al.*, 2014; Breuer *et al.*, 2016), which informed the data collection methods and data analysis.

Students in the intervention classrooms were told by school officials that they were expected to participate in the comprehensive sexuality education course during a weekly tutoring session, but that taking part in the study was optional. Students who were 14 to 17 years old, provided informed assent and obtained parental consent were eligible to participate in the study. Of the 185 students receiving the course in the pilot and intervention semesters, 157 (85%) were eligible and agreed to participate in this study. They were asked to complete a baseline questionnaire in which they self-reported age, sex, relationship status, sexual history and experience of intimate partner violence. We used these baseline responses from students receiving the intervention to purposively select a subsample for qualitative data collection that was heterogeneous with regard to the reported characteristics. In addition, some students who were not originally sampled approached the research team and asked to participate in interviews or focus groups. If eligible for the study, they were invited to participate even if they had not completed a baseline questionnaire. In total, 47 students (30% of students receiving the intervention) comprised the subsample participating in interviews and focus groups. Table 7 presents baseline sociodemographic and descriptive data for this qualitative subsample and the intervention group as a whole.

In addition to students, we also invited all teachers assigned to the intervention classrooms and all Mexfam health educators providing sexuality education to these classes to participate in focus

group discussions, after obtaining written informed consent. We gave all focus group and interview participants a gift card as compensation for their time and offered them subsidised services through Mexfam’s network of clinics. This study was approved by the London School of Hygiene and Tropical Medicine Research Ethics Committee in the UK and *Centro de Investigación Clínica Acelerada (CICA)* in Mexico.

Table 7: Characteristics at baseline of participants in Mexfam’s comprehensive sexuality education intervention

	Interview and focus group participants ^a (n=39) ^b	All baseline questionnaire respondents (n=124) ^c
Sex		
Male	59%	55%
Female	41%	45%
Age		
14	26%	27%
15	67%	60%
16	8%	12%
17	0%	1%
Sexual orientation		
Gay	3%	5%
Bisexual	18%	10%
Heterosexual	72%	76%
Prefer not to say	8%	7%
Don’t know/no response	0%	2%
Ever had relationship		
Yes	72%	70%
No	26%	28%
No response	3%	2%
Ever had sexual contact		
Yes	41%	38%
No	33%	37%
Prefer not to say	23%	23%
No response	3%	2%
Ever had sex without wanting to		
Yes	10%	10%
No	72%	75%
Prefer not to say	15%	14%
No response	3%	2%
Ever experienced partner violence		
Yes	13%	13%
No	72%	80%
Prefer not to say	15%	7%
No response	0%	1%

Note. Figures are rounded and so may not add up to 100.

^a These are a subsample of the baseline questionnaire respondents.

^b There were 47 interview and focus group participants; eight did not complete the baseline questionnaire.

^c Data from the pilot semester are excluded from the table because we adapted the questionnaire after pilot implementation.

Qualitative data collection methods with course participants were as follows: We used *observation* throughout the semester to better understand how students interact and engage during the course. We used *case studies* to explore participant trajectories (Grossoehme and Lipstein, 2016) during the intervention, conducting up to four interviews with each of nine course participants (five female, four male) approximately monthly throughout and up to three months after the intervention. After the intervention ended, we selected an additional 10 male and 10 female course participants for *one-time in-depth interviews* and conducted three *focus group discussions*, two with young men (n=18) and one with young women (n=6). Table 8 summarises the characteristics of interview and focus group participants.

Table 8: Age and sex of interview and focus group discussion participants, by data collection method

Data collection method	Conducted	Number of participants			Participant age
	Total	Women	Men	Total	Median (range)
Repeat interviews - students	33	5	4	9	15 (14-17)
One-time interview - students	20	10	10	20	15 (14-17)
Total interviews	53	15	14	29	
Focus groups – students ^a	3 ^b	6	18 ^b	24	15 (14-16)
Focus groups - health educators	2	4	1	5	26 (23-29)
Focus groups - teachers	2	5	0	5	No data
Total focus groups	7	15	19	34	

^a Six of these students (4 men, 2 women) also participated in interviews.

^b We were aiming to include an approximately even number of men and women in the sample, so we invited more men to the focus groups based on our difficulties recruiting them for individual interviews. Unexpectedly all the invited men participated, so we conducted two focus groups with men and only one with women, resulting in more men than women in the focus groups.

We also conducted separate focus group discussions with five teachers and five health educators after the pilot semester and after the full intervention (participant characteristics detailed in Table 7). All five teachers were women; we did not collect age data. The health educators had a mean age of 26.4 years, ranging from 23 to 29 years old. Four were women and one was a man. Two of the health educators participated in both focus groups.

Mexfam staff and local professional transcriptionists carried out verbatim transcription of audio-recorded interviews and focus groups. Two team members (JG and SM) listened to all of the interviews and spot-checked transcription quality. Two team members (SM and CM) led data analysis using the original transcripts in Spanish. We created a ‘start list’ of codes based on the

research questions and hypotheses (Miles and Huberman, 1994). We then reviewed and indexed all transcripts according to these codes. We created new codes throughout the process to capture emerging concepts, allowed for a 'combination of data-driven and theory-driven strategies of category creation' (Gläser and Laudel, 2013). We used a 'progressive focusing' approach, iteratively adjusting the codes and data collection processes as we became familiarised with the data and refined our focus of inquiry (Stake, 1981; Sinkovics and Alfoldi, 2012; Schutt and Chambliss, 2013). Once most transcripts were indexed, we began subsequent analysis by developing code summaries (Miles and Huberman, 1994) to compile and integrate data from different sources and begin to draw conclusions pertaining to each of our hypotheses. We used analytic memo-writing to aid the analysis process (Saldaña, 2009) and examined longitudinal data to explore 'epiphanies', 'tipping points', and the 'unfolding' of gradual change (Farrall, 1996). One team member (SM) translated the final quotations selected for this paper in discussion with the other authors (native speakers of Spanish and English) to capture nuance, using '...' to designate omitted text. We periodically conducted meetings to discuss emerging findings. We refer to participants using pseudonyms.

Results

We found that Mexfam's comprehensive sexuality education course influenced participants in the following principal areas which we will discuss in turn below: critical reflection about social norms; shifts in attitudes and knowledge about violence and gender; increased communication about relationships, sexuality and violence; taking protective and preventative actions related to violence, relationships and sexual and reproductive health; and changing norms around gender and sexuality.

Critical Reflection and Attitude Shifts Related to Violence and Relationships

Students, health educators and teachers told us that during the course many participants reconsidered their perceptions of jealous and controlling behaviours that occur in relationships, such as monitoring a partner's social media. They said the course encouraged reflection and debate about the types of violence that can occur in relationships as well as the positive attributes of relationships. Some participants described reflecting on their own behaviour and experiences as a result of using the *Violentómetro* tool (Tronco Rosas, 2012), which was presented in the course, to analyse whether they were experiencing violence in their own relationships. The tool,

developed based on formative research in Mexico (Tronco Rosas and Ocaña López, 2012), visually depicts manifestations of partner violence ranging from subtle to severe, including types of psychological, physical and sexual violence.

Students and health educators frequently mentioned that the group discussions in the course played a role in the process of reflecting on participants' perceptions of jealousy and controlling behaviours in relationships.

There is a lot of 'close your Facebook, we'll only use messenger', or 'block that boy [on your social media]' or 'don't dress that way' or 'why are you talking to him' ... or many [such] things, right? And, well, we thought that this was fine, they do it because they love me.... And then [the facilitator] made it very clear to us that this is not because they love you, but actually because they are a possessive person. That is, if they really loved you, they would accept you as you are. (Laura, young woman, 16)

Some participants said that before the course they didn't think jealousy in a relationship was problematic but during the course they realised that it was. Julián said he now believes jealousy 'is bad, because if you have trust in your partner, why are you going to be jealous over them?' but that before the course he thought it was a way to show love: 'I used to say ... that if they weren't jealous then they didn't love you, things like that' (Julián, young man, 15). Julián and other participants said they now considered jealousy to be a negative attribute in a relationship, but this view was not universal among the interviewees. 'Jealousy can be good because, well, it is a way to express that you care about the person. But if it becomes excessive it can be dangerous' (Vicente, young man, 15).

Alberto said that before the course many classmates believed that jealousy was a sign of love, but that even while they were starting to change their views, a television programme showing at the same time as the course was promoting the opposite message.

There was a moment when the penny dropped for them [i.e. they realised that jealousy was bad], but afterwards the famous [television] programme *Enamorándonos* [*Loving each other*] began ... and the famous sexologist said, 'if there wasn't any jealousy, there wasn't any love.' And – I think that put them [my classmates] back to where they had been before the course. (Alberto, young man, 15)

Some participants mentioned that the course highlighted not only aspects of relationships that should be avoided, such as manifestations of violence, but also positive attributes of relationships.

'We used an *amorómetro* [love thermometer] that started with trust, respect, avoiding jealousy. And [learned that] by avoiding jealousy most of your problems will stop. Because most problems come from jealousy' (Young man, focus group).

One teacher said she observed students questioning and reflecting on their ideas about relationships and love during the course.

For them, at this age, is it really love? Or is it a simple preference, or is it simply affection? [These questions] often confuse [them]. So, for them ... to be able to reflect ..., to engage again [with these ideas] and reflect about their interpersonal relationships, such as friends with benefits, or friend-boyfriend, or friend or acquaintance – that is what they experienced [in the course]. (Teacher)

Critical Reflection About Gender, Sexuality and Relationships

Some participants said the course encouraged them to reflect on their beliefs about gender and sexuality, for example during facilitated activities in which the group engaged with and questioned prevailing social norms, such as those related to gender equity. 'We were debating and [pause] came to the conclusion that [pause] they [women and men] are equal, that women can do what men can do, and men can help women and women can help men' (Alberto, young man, 15). Others said that hearing what classmates said during the group discussion helped them reflect on their own individual beliefs.

One of the things my classmate said stayed with me. He said that the man has to work and the woman [should stay] in the house.... It made me, like, think.... I think that a woman doesn't need to always be at home... um, as if it were a prison.... I think you need to give freedom to both people in a relationship. (Marco, young man, 15)

Some participants said they changed their personal attitudes about gender and sexuality during the course, and several said that the course helped them accept their own sexuality and feel more confident talking about it.

Before, I was not sure of myself.... And ... well, it turns out that ... various lessons during the course ... helped me reflect and ... realise ... whether I was [bisexual] or not, that I was born this way, and this is who I am. (Karina, young woman, 15)

Several participants also said the course helped them develop more self-respect and confidence in their ability to make the right decisions about relationships and sexuality. For example, one participant told us the course helped her reflect and come to the conclusion that she was not

ready to start having sex. One teacher said that she thought the course helped prepare young people for their interpersonal and romantic relationships.

From what I observed during the course sessions, it seems to me that for the participants it was a watershed moment, it created a different vision for their own lives, their family life, their relationship with school, and friends, and above all to help them rethink – as young people – the sexual and emotional direction of their lives. (Teacher)

Increased Communication About Relationships and Sexuality

Students, teachers, and health educators said that participants became more comfortable talking about relationships and sexuality as the course progressed.

Before the course, it made us ... a bit embarrassed to talk about [sexual and reproductive health]. But afterwards we understood, with the course, that it was, like, very natural to talk about it. It's like any other thing, and so I now feel fine talking about it. (Gerardo, young man, 15)

Several students also said that when the Mexfam health educators shared their own personal experiences during the sessions it helped the course participants to open up about their own lives.

For some, an important part of the course was hearing the other participants' views. For some participants, this process helped them identify supportive peers. For example, one focus group participant said it was valuable to know what her classmates thought about the course topics, and another participant added that this helped them know who they could trust. 'I know that all the women in my class ... think the same as me, and I know that if anything ever, well, happens to me, I know I can talk to any of them about it' (Young woman, focus group).

Protective and Preventative Actions Related to Violence and Health

Participants in this study told us that they engaged in a range of direct actions to mitigate or respond to violence in their own relationships and those around them. Many participants said they shared the information learned in the course about relationships and violence with their friends and family members, and some said they intervened in violence around them during or after the course.

[Her boyfriend] told her that without him she was nothing, that she would never find anyone better than him.... She told us, well, that she wanted to leave him but ... that he, well, wouldn't allow it. And the other [friend] was just ... she was sad because she didn't want to be with her boyfriend anymore ... and he told her

that if she left him, he would kill himself. So, we told her that no, that she should leave him, that [pause] she should tell his mother, someone who can take care of him.... So, because of the course I already knew how I could help. (Judith, young woman, 14)

A handful of participants said they talked with their partners about different types of violence versus positive aspects of relationships.

When they told us ... about what is love and what is not love.... I told him [my boyfriend] '... they told us that jealousy is bad', and he replied, 'that's right, because it means a lack of trust', and in this way, we sometimes talked about [the course contents]. (Silvia, young woman, 14)

One participant said the course prompted him to talk with his girlfriend about his dislike of her controlling behaviour in their relationship. Others said that they noticed changes in their classmates' relationships. For example, one participant told us about a male classmate who at first wouldn't let his girlfriend talk to other men, but then 'his way of thinking changed and he relaxed about it' once the course began to address the topic of relationship violence (Marco, young man, 15). Another participant said a classmate had disclosed to the class that she had spoken to her boyfriend about the violent behaviours in their relationship and believed the relationship had improved as a result. Several individuals also told us that classmates had left controlling or violent relationships during or because of the course.

[The course] left them with a clear idea of what was really going on in their relationship, so they decided [to leave], saying 'it's true, it's not that he loves me. This [being possessive over me] is a type of violence.' (Laura, young woman, 16)

Students and health educators told us that course participants approached members of the Mexfam team for advice and support related to relationships and partner violence, either for themselves or to get advice on how to support friends who were experiencing violence.

There was one classmate, [the topic of] relationship violence made a big impression on her.... Because I think her boyfriend used to be, well, jealous. He used to ask for her cell phone and things like that. And, well, I think that she approached one of the girls [health educators]. And she also asked for help. (Gerardo, young man, 15)

Several Mexfam health educators said that course participants came to them to ask for information or support related to relationship violence, as in the case of one young woman who said she was scared to leave a jealous boyfriend.

A young woman ... approached me and said, 'I just got back together with my boyfriend, but he is very jealous.' And I told her about her options [to address her situation], and she told me 'I'm going to use one of those options ... but I'm scared.' (Health educator)

The course helped participants learn that it was possible for them and other young people to seek support at Mexfam or at other health centres, and to ask for help from the health educators.

[The health educators] are people you can trust. As time passed, well, they gave me confidence ... that if at any moment something happens and I need something, or want to know something ... well, I can ask them for help, it won't be a problem. At the beginning I felt a bit embarrassed [talking to them], but afterwards, no, I would feel relaxed. (Miguel, young man, 15)

One health educator said that as part of the course, they deliberately reinforced reflection about young people's right to services by repeatedly extending invitations to health services and reminding participants that it was their right to access these services. Another health educator said that they observed participants begin talking during the course about their right to healthcare and speculated that before the course they would not have considered it their right to access that healthcare.

The health educators said that participants contacted them in various ways, both during and after the course. Some approached them in person, often towards the end of the semester once they had developed trust in the health educators, and others contacted them by phone or WhatsApp. When the course ended, Mexfam organised a health fair at the school during which they provided free health services to a number of course participants in the organisation's mobile health unit. The Mexfam team said that course participants generally approached their mobile health unit rather than those of other organisations offering services at the health fair, and suggested that this may relate to the trust the students had developed in the health educators during the course.

Changing Norms around Gender and Sexuality

Participants described how they or their classmates changed their beliefs and behaviours relating to gender and sexuality. Several students and health educators said that participants shifted the way they spoke about these topics during the course.

[The course helped] a classmate, because he ... used to make this type of comment [disparaging women], and after this I also tried to explain to him that you have to respect women.... He no longer makes that sort of comment. (Alejandro, young man, 16)

Some participants said the course taught them to engage in dialogue or communicate assertively about sexuality or gender. When asked whether he spoke with peers about the topics discussed in the course, one young man said he engaged in conversations about sexual diversity by asking questions of a classmate who identified as gay. Another participant told us:

I had a classmate who used to say that gay people disgusted him.... And then after the course he started assimilating things and then he didn't think in that way anymore.... [Beforehand] I used to hang out with him, but he didn't know I was bisexual [...]. Afterwards ... he asked me if I was gay, and I said that 'no, I'm bisexual'. And he said to me 'that's fine, you're all right.'... After that, he changed his way of thinking. (Gerardo, young man, 15)

Students and health educators told us that a few course participants approached the Mexfam team to talk about how harmful gender norms affected their lives and to seek support. One young man who identified as homosexual told us he enjoyed wearing makeup, and that his family tried to prevent him from doing so. He said he wanted to bring his mother to Mexfam so she would learn to accept his gender expression.

I said [to my mother], it isn't fair that you criticise me, because you are completely interfering with the person I am..... I told my mother, if you want, I'll invite you to Mexfam, so that ... they can tell you that it isn't ok for you to interfere with the way I am. (Gilberto, young man, 17)

Judith also asked for help from the course facilitators so that her mother would begin to understand her and accept her sexual orientation. Similarly, a health educator said that after the last session of the course, a young woman asked for information about how to work with her mother to accept her sexuality.

Another health educator said that a course participant, who was in her first year of secondary school, approached them to discuss the gender beliefs in her family, specifically their resistance to her attending school because she was a woman.

She said to me, '... I think that it is economic violence.' I asked her why. 'It's that they don't give me [money] for food ..., for transportation.... I already talked to them and they told me that I don't have the right [to study] because I'm a woman. And I should get home and take care of my family.' (Health educator)

Discussion

Our study adds to the growing literature that suggests that changes in behaviours and beliefs that support the prevention of partner violence can be achieved in programmatic timeframes. We found that comprehensive sexuality education may help prevent intimate partner violence among young people, both in terms of prevention – addressing the harmful gender norms that underlie inequitable relationships as well as other risk factors for violence – and response – preparing young people to address and mitigate such violence if it happens. Our findings show how comprehensive sexuality education appears to help young people take a critical approach to social norms and respond to them accordingly. Students, teachers and health educators credited Mexfam’s course with influencing a range of attitudes and practices compatible with the objectives of gender-transformative programming and violence prevention efforts (Dworkin, Fleming and Colvin, 2015).

This study also responds to the call for research that examines the process of violence prevention. The course appears to promote critical reflection that helped change beliefs, intentions and behaviours related to gender, sexuality and violence. Reflection has also been shown elsewhere to support attitude and behaviour change in related areas (Nelson *et al.*, 2010; Kyegombe *et al.*, 2014). Aspects of Mexfam’s course that likely contributed to this process include the use of content relevant to participants’ lives and activities such as discussing vignettes designed to promote critical engagement with social norms. Group discussions promoting open dialogue between participants and facilitators provide a space to share experiences and beliefs, debate about contradictions among these, and begin to create individual and collective narratives that help them engage with and resolve dilemmas related to the course topics. The facilitators play a crucial role in ensuring the space is safe for what is sometimes very sensitive discussion, and it is clearly important for them to be adequately trained and supported if they are to be effective, a consideration noted in international guidelines for comprehensive sexuality education (United Nations Educational Scientific and Cultural Organization, 2018). This may be of relevance for programmes expecting teachers to deliver a comprehensive sexuality education curriculum with a gender-transformative approach, as they should not only be comfortable with course topics but also prepared to engage participants in critical reflection processes.

Young people are likely to encounter messages and information coming from credible sources that counter the teachings of a sexuality education course, as in the case of the television show asserting that jealousy is a sign of love; a short term intervention may not be able to counter these

contradictory societal messages, but can create space for reflection and assertive communication about different – for example, religious, cultural or scientific – understandings of love. By promoting critical reflection, comprehensive sexuality education may directly contribute to shifts in social norms. For example, facilitated group discussions questioning dominant ideas about gender equity and violence created opportunities for course participants to reconsider their beliefs, renegotiate norms within their group and later diffuse these shifts in ideas and norms within their community – a strategy for social norms change also reported elsewhere (Miller and Prentice, 2016; Cislighi and Heise, 2018a). We would suggest that this is a way that comprehensive sexuality education supports the prevention of intimate partner violence.

We identify four key elements of the course that seem crucial to supporting violence prevention: First, encouraging participants' reflection about romantic relationships, which helped them question whether jealousy and possessive behaviours are signs of love; second, helping them develop skills to communicate about sexuality, the characteristics of inequitable relationships and reproductive health; third, encouraging care-seeking behaviour; and fourth, addressing norms around gender and sexuality, for example demystifying and decreasing discrimination towards sexually diverse populations.

Mexfam's comprehensive sexuality education course helped participants consider a range of narratives of how love can be expressed, for example by rethinking jealousy and possessive behaviours as unwanted practices. This reconceptualisation, which was encouraged in the course, contradicts the mainstream construction of jealousy and controlling behaviours as a part of romantic love in Mexico (Flecha, Puigvert and Redondo, 2005; Tronco Rosas, 2012; Tronco Rosas and Ocaña López, 2012; Ruiz, 2015). The *Violentómetro* (Tronco Rosas, 2012; Tronco Rosas and Ocaña López, 2012) helped participants to identify and reflect on these and other subtle or less subtle forms of relationship violence. Building on the acceptability and usefulness of the *Violentómetro*, Mexfam used a similar format to showcase positive and equitable relationship behaviours that could replace unwanted or violent practices.

Developing communication skills was also important. Course activities appeared to help participants overcome their embarrassment and become more confident talking about sexual and reproductive health topics in a mixed-gender environment during the course. This also appears to have prepared them to communicate about these topics outside the course with family, peers and

partners. The group discussions during the course likely helped participants to give and seek advice about sexuality, relationships and violence within a supportive network of peers. By sharing information learned in the course with others, participants may well have created indirect effects in their social network that we have not been able to measure. Comprehensive sexuality education programmes seeking to shift gender norms could explicitly support participants to be agents of change in their communities, as in other violence prevention programmes (Nelson *et al.*, 2010; Kyegombe *et al.*, 2014; Starmann *et al.*, 2017) and other community programming seeking to shift social norms (Cislaghi, 2018).

The course encouraged participants to seek professional advice and support regarding sexual and reproductive health and relationships, with some students approaching Mexfam for information and referrals and a smaller number reporting that they or their peers accessed health services. This suggests that the course may have helped address some of the barriers to sexual and reproductive health and partner violence services commonly encountered by young people (Mejía *et al.*, 2010; Santhya and Jejeebhoy, 2015), and may relate to the health educators' work emphasising the right of young people to receive services, building trust with the students, providing frequent referral to trusted providers and being accessible by phone, social media and text message applications such as WhatsApp. After the course ended, some participants continued to contact the health educators through these avenues, and Mexfam sustained contact with students by providing information and services on school grounds using mobile health units during a school-wide health fair.

The course addressed social norms related to gender and sexuality, for example by demystifying sexual diversity and tackling homophobic discrimination. It may be that participants who are grappling with their own sexuality or gender identity were particularly motivated to engage with the course, but in any case, the positive impact reported by gay, lesbian, bisexual and questioning participants in the course highlights the importance of avoiding heteronormativity and addressing sexual diversity in sexuality education programmes. Engaging families and teachers, which was not done as part of the evaluated intervention, might improve parent-child communication and avenues for support within the school context, which could have particular benefits for those young people who find it challenging to communicate about their sexuality, are at risk of discrimination related to their sexual orientation or gender identity or are particularly vulnerable to intimate partner violence.

One of this study's strengths is the close collaboration between the research and programmatic teams throughout the research process, which helped ensure that study findings would be programmatically relevant and put into practice. The study also had limitations. Intervention participants interacted repeatedly with the research team and provided generally positive feedback about the Mexfam's course. It is possible that negative feedback was not given out of politeness or for some other reason. Also, study participants were volunteers and so may have been more inclined to talk about relationships, violence and sexuality than their non-volunteer peers. Interviews were supplemented with frequent observations of the classrooms and the course, which helped us put the interviews in context in the analysis.

Direct measurement of partner violence would be useful to assess the effects of the programme in the medium term, but this brings major conceptual and methodological challenges, which is why for this study we examined the process of violence prevention rather than attempting to measure it directly. As such, this study cannot quantify the effectiveness of the intervention, but rather provides an in-depth exploration of the program's influence on the factors hypothesised in the theory of change to contribute to the process of violence prevention. This study followed participants for up to three months after the end of the intervention. Longer-term follow-up research could assess whether the shifts in beliefs and behaviours experienced by course participants are sustained over the medium- to long-term, and to what extent the course may contribute to shifting norms not only among course participants, but also within their families and communities. The current findings reflect a comprehensive sexuality education programme that intervened only at the individual level rather than at wider social or cultural levels (Marston and King, 2006). It would be useful to assess whether similar interventions that systematically work not only with students, but also with teachers and families, result in intensified programme effects.

In conclusion, this paper highlights some mechanisms through which comprehensive sexuality education programming with a gender-transformative approach appears to have supported prevention of and response to intimate partner violence among young people in Mexico City within programmatic timeframes. The findings, which have implications for educational policy, reinforce the importance of schools both as settings for violence and for its prevention. In Mexico, where educational institutions may resist incorporating comprehensive sexuality education, these findings help demonstrate the importance of systematically implementing this type of intervention. The results suggest that this promising and relatively short-term comprehensive

sexuality education programme has potential for scalability within Mexico's educational curricula as a strategy to prevent and respond to partner violence and potentially reduce homophobic discrimination or other forms of interpersonal violence common in school settings. We identified programmatic elements that appear most likely to trigger change among participants in Mexico City, which should be tested elsewhere to examine whether or not they can have an impact on beliefs and practices related to intimate partner violence in other settings.

Chapter 5: Evaluating complex interventions using qualitative longitudinal research: A case study of understanding pathways to violence prevention

5.1 Introduction

The analysis presented in this chapter builds on Chapter 4, where I showed different types of change experienced by participants in a comprehensive sexuality education intervention focused on IPV prevention. In Chapter 4, I considered various mechanisms through which the intervention appears to contribute to IPV prevention using a cross-sectional analysis of the types of change observed across the study sample. In this chapter, I move to longitudinal analysis, examining how a small group of participants experienced the intervention over time. I do so by using a qualitative longitudinal approach to examine how the intervention influenced relationship trajectories. This analysis also serves as a case study to explore the ways in which repeat interviews can be useful for evaluation studies, particularly when evaluating programmes addressing social complex issues such as IPV.

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1400132	Title	Ms
First Name(s)	Shelly		
Surname/Family Name	Makleff		
Thesis Title	Methodology for evaluating complex social interventions: understanding partner violence prevention for young people in Mexico City		
Primary Supervisor	Prof Cicely Marston		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Qualitative Health Research
Please list the paper's authors in the intended authorship order:	Shelly Makleff, Jovita Garduño, Icela Zavala, Jimena Valades, Florencia Barindelli, Mariana Cruz, Cicely Marston

Stage of publication	Submitted
----------------------	-----------

SECTION D – Multi-authored work

<p>For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)</p>	<p>I am the first author of this paper, and was responsible for study design, conception of the paper, data analysis, and writing, and also participated in data collection. Cicely Marston supported on study design, conception of the paper, analysis and writing. Because of the co-produced nature of the evaluation case study, key members of the partner organizations are co-authors based on their involvement in study design, data collection, or interpretation of findings, as well as critical revisions of the paper.</p>
---	---

SECTION E

Student Signature	[Redacted Signature]
Date	29 / 5 / 2020

Supervisor Signature	[Redacted Signature]
Date	29 / 5 / 20

5.2 Abstract

Qualitative longitudinal research may help understand people's changing experiences during interventions: dynamics which are often overlooked in evaluations. We explore the role of qualitative longitudinal research using a case study: the evaluation of a partner violence prevention programme with young people. In this evaluation, we conducted 33 repeat interviews every one to two months with nine 'case study' participants, and analysed participant trajectories. We found that participants' relationship-related beliefs and intentions changed, promoting self-reflection that in turn helped alter their relationship dynamics. Our qualitative longitudinal approach allowed us to detect and track specific examples of change, identify influential elements of the programme, and gather contextualised data about participants' lived experiences. Qualitative longitudinal research provided evidence of gradual shifts on the pathways to violence prevention. Long term effects of violence prevention interventions are very hard to measure directly. We argue that a qualitative longitudinal approach provides a way to measure subtle changes that can serve as proxies for longer term impacts.

5.3 Research paper 2

Introduction

Evaluation research often overlooks the shifting nature of people's experiences in interventions over time. Complex interventions 'may change with time, for pragmatic reasons, change of external factors, or as a result of learning during the evaluation' (Craig, Dieppe, Macintyre, Michie, *et al.*, 2008). Evaluations could benefit from examining such variation in order to better understand participants' experiences in and responses to an intervention. However, there is little concrete guidance on how real world evaluations can do so (Craig, Dieppe, Macintyre, Michie, *et al.*, 2008; Moore *et al.*, 2015) – and such investigation is routinely omitted from evaluations altogether. To improve evaluations, we need worked examples and recommendations about how temporal change can be appropriately taken into consideration. In this paper, we examine the case of interventions designed to prevent or mitigate intimate partner violence (IPV). Such interventions aim to contribute to social change, and measuring their effects is challenging; capturing participant experiences repeatedly during and after an intervention may help to understand pathways to impact.

Qualitative longitudinal methods show promise for including temporal aspects of participant experience in evaluation research, and may be particularly well-suited for evaluating IPV prevention interventions. First, these methods are fitting to examine social policies or programmes in their real-life settings (Lewis, 2007; Thomson, 2007; Calman, Brunton and Molassiotis, 2013; Thomson and McLeod, 2015). Because IPV-related interventions are by their nature social – they address interpersonal relationships and are generally delivered to groups such as schools, families or communities – this is particularly relevant. Second, qualitative longitudinal research can be used to examine how and why change happens in relation to socio-cultural context (Holland, Thomson and Henderson, 2006). Interventions addressing IPV engage with social norms, which vary by context, and there is a growing consensus that research is needed to build our understanding of how to transform the social norms that underlie and sanction violence (Temmerman, 2014; Jewkes, Stern and Ramsoomar, 2019). Third, recent impact evaluations of IPV prevention programmes based on randomised controlled trials have suggested that while programmes may not achieve community-wide reductions in IPV, moderate and incremental shifts in attitudes or behaviours still occur (Gibbs *et al.*, 2019; McLean, Heise and Stern, 2019); qualitative longitudinal methods can help detect such shifts. Finally, by focusing on temporality

and change (Saldaña, 2003), this approach can respond to calls for research exploring the pathways to IPV prevention (Fulu and Kerr-Wilson, 2015; Jewkes, Flood and Lang, 2015; Michau *et al.*, 2015). Indeed, researchers who have evaluated IPV and school-based violence prevention programmes without collecting longitudinal qualitative data have tentatively suggested that the approach might have helped elucidate processes of change (Kyegombe *et al.*, 2014, 2017; Starman *et al.*, 2017).

Recent evaluations of IPV prevention interventions in Mozambique, Nepal, Rwanda and South Africa have included qualitative longitudinal methods as part of larger evaluations in order to better understand how an intervention may contribute to change or to learn about implementation processes (Burke *et al.*, 2019; McGhee *et al.*, 2019; Stern and Heise, 2019; Hatcher *et al.*, 2020). Authors of these studies described advantages of the qualitative longitudinal approach such as improving disclosure of sensitive issues (Stern and Heise, 2019), allowing for detailed follow-up to aid in interpretation of data (Burke *et al.*, 2019), and the ability to assess whether changes were sustained over the evaluation timeframe (McGhee *et al.*, 2019). Interview frequency varied: the study in Nepal conducted two interviews four months apart (McGhee *et al.*, 2019) while the other three waited one year between interviews. The Rwanda study conducted the first interview before the intervention (Stern and Heise, 2019); the rest started interviews during the intervention (McGhee *et al.*, 2019; Hatcher *et al.*, 2020) or after it ended (Burke *et al.*, 2019). None of the studies gathered qualitative data at shorter intervals during the intervention, limiting their ability to detect gradual or subtle shifts on the pathway to change.

Repeat interviews conducted at shorter intervals may be beneficial when working with young people and adolescents, who are at an age marked by rapid cognitive, physical and psychological changes (UNAIDS, 2004; Kågesten *et al.*, 2016; Blum, Mmari and Moreau, 2017; Mmari *et al.*, 2017). At this age, experiences and beliefs related to gender, sexuality and relationships evolve quickly (Price *et al.*, 2016; Blum, Mmari and Moreau, 2017), and research addressing IPV prevention with young people may necessitate an approach that accounts for short-duration relationships (Giordano *et al.*, 2010). This is in contrast to studies focused on cohabitating adult couples, which often examine shifts in relationship dynamics in long-term relationships, sharing of household tasks, and economic decision making (Starman *et al.*, 2017; McGhee *et al.*, 2019; McLean, Heise and Stern, 2019; Stern and Heise, 2019) – concepts of less relevance for young unmarried participants.

This analysis explores data from repeat interviews conducted with young people during and after their participation in a comprehensive sexuality education programme in Mexico City that aimed to prevent IPV and encourage non-violent and equitable relationships. We explore how incorporating this methodological approach in evaluations can help explore gradual shifts and the changing nature of participant experiences over time. The programme was designed and delivered by the non-profit organisation Fundación Mexicana Para la Planeación Familiar (Mexfam). The 20-hour curriculum involved topics including relationships, IPV, sexual diversity, sexually transmitted infections and contraception. Gender was a cross-cutting theme throughout the curriculum, which took a gender-transformative approach, i.e. aimed to shift harmful gender norms (Dworkin, Fleming and Colvin, 2015; Dworkin and Barker, 2019). The course was delivered by paid Mexfam health educators under 30 years of age, to groups of approximately 20 secondary school students between 14 and 17 years of age. Sessions were to take place in classrooms over one semester, in 10 two-hour sessions.

In this paper, we explore the contribution of qualitative longitudinal methods to the evaluation of the programme in Mexico; we address four questions: First, can a qualitative longitudinal approach help detect intervention effects? Second, can it identify pathways to IPV prevention? Third, are repeat interviews feasible and acceptable to participants? Fourth, in what ways could qualitative longitudinal methods contribute to future evaluations?

Methods

Study design

The evaluation in Mexico employed a longitudinal quasi-experimental design and was conducted at one school in Mexico City in 2017 and 2018. The study was co-produced by the International Planned Parenthood Federation/Western Hemisphere Region, the London School of Hygiene and Tropical Medicine, and Mexfam. More details about the course activities, study setting, research design, data collection and evaluation findings are reported elsewhere (Chapter 4).

Data collection

We conducted 33 repeat interviews every one to two months with nine IPV intervention participants. To identify suitable participants, we randomly selected from 87 intervention group participants who had reported in a baseline questionnaire any past experience of sexual contact,

romantic relationships, or IPV. We sought 10 such participants, with equal distribution by gender, to ensure diverse experiences while maintaining a manageable quantity of data. Nine of the 15 young people we approached in the first two months of the study agreed to participate. Seven participated in four interviews, one in three interviews, and one in two interviews – a total of 33 interviews. The nine ‘case study’ participants were distributed between four different implementation groups of approximately 20 participants each. At baseline, five identified as women/girls and four as men/boys; two were 14 years old, five were 15, one was 16 and one was 17. At baseline, all nine reported ever having a relationship, six ever having sexual contact, three having experienced IPV, and two having experienced sexual contact against their will.

Interviews took place every one to two months, over a six-month period, both during and after intervention (Table 9). The interviewer was a Mexfam staff member (JG), and interviews took place in a private space at the school or at Mexfam’s headquarters nearby. They were audio recorded with permission and transcribed in the original Spanish by professional transcriptionists and Mexfam staff members. All transcriptions were quality checked by the research team. After each interview, the interviewer wrote field notes about the experience.

Table 9: Timetable for longitudinal qualitative interviews

	Month 1 ²	Month 2	Month 3	Month 4	Month 5 (School holiday)	Month 6	Total
Participants							
Laura							4
Gilberto							3
Elena							4
Julián							4
Karina							4
Gerardo							4
Lizbeth							4
Israel							4
Beatriz							2
Total	2	8	6	8	0	9	33
CSE course							

² In September 2017 there were nearly three weeks of earthquake-related school closures while waiting for building inspections; the first interviews were delayed for many participants.

This paper examines the data from the repeat interviews. However, we also collected data for the evaluation from focus group discussions, one-time in-depth interviews, direct observation, and self-administered questionnaires before and after the intervention, as described elsewhere (Chapter 4).

The interview guide drew from the programme theory of change and was designed to explore shifts in participants' perceptions, expectations, intentions and behaviours pertaining to violence and relationships. In the final interview (two months after the intervention ended), we included questions about experiences with the repeat interviews and any changes that they considered may have been influenced by the intervention. We had intended to conduct the first interview with each participant before the start of the intervention, but delays in recruitment followed by earthquake-related school closures meant that they began after the start of the intervention.

At the beginning of the semester, parents and guardians were informed about the study and given the option to withdraw the student. Students provided written consent to participate in the research and received a gift card of 200 Mexican pesos (approximately US \$10, equivalent to one month of cell phone service with data) after each interview to compensate for their time. They were also offered free counselling and sexual and reproductive health services at Mexfam clinics and at a school-based health fair. We obtained ethical approval in Mexico (*Centro de Investigación Clínica Acelerada*) and the UK (the London School of Hygiene and Tropical Medicine Research Ethics Committee).

Analysis

SM reviewed interview transcripts as they became available and shared a list of follow-up questions with JG to incorporate into the subsequent interview – in this respect the data collection and analysis followed a 'progressive focusing' approach, iteratively adjusting data collection and analysis processes as we became familiarised with the data and refined our focus of inquiry based on earlier findings (Stake, 1981; Sinkovics and Alfoldi, 2012; Schutt and Chambliss, 2013).

We defined the themes for analysis from the programme theory of change (Figure 3 in Section 2.4). We analysed data from the 33 repeat interviews following the five steps of framework analysis: familiarisation; identifying a thematic framework; indexing; charting; and mapping and interpretation (Ritchie and Spencer, 1994). We used time-ordered, sequential matrices to

summarise change by theme for each case (Grossoehme and Lipstein, 2016), and then identified similarities and differences between cases. We aimed to build a 'thick' description (Geertz, 1973) of participant experiences in the intervention and identify any gradual shifts in what we refer to as their 'relationship trajectories.' By this, we mean participants' evolving and shifting beliefs, expectations, intentions and experiences related to romantic relationships. We developed exploratory case histories (Thomson, 2007) for five participants whose narratives contained common patterns from the data.

All authors worked with the quotations in the original Spanish and then four authors (CM, FB, JV, SM) translated them into English for presentation here. All interviewee names are pseudonyms to protect confidentiality.

Analysis for the methodological exploration in this paper also involved review of field notes written by the interviewer (JG), review of interview excerpts addressing participant perceptions of repeat interviews, and meetings between the first and second authors to document, reflect on and analyse the repeat interview experience. SM wrote analytic memos (Saldaña, 2009) to organise themes relevant to the methodological aspects of this study, and shared these with co-authors for input.

Results

The interview narratives suggested that the course influenced participants in relation to IPV prevention in two keys ways: first, shifting relationship-related beliefs and intentions; and second, promoting critical self-reflection that helped participants identify harmful aspects of their own relationships and alter their relationship dynamics. We present details of five case histories that illustrate typical ways in which the intervention appeared to influence relationship trajectories.

Influencing relationship-related beliefs and intentions

The course appeared to affect participant beliefs and intentions. These shifts were commonly related to ideas about jealousy, possessive behaviours, and the types of behaviour that are forms of IPV. The cases of Gerardo, Elena and Julián illustrate this:

Case 1: Gerardo (young man, 15 years old, 4 interviews)

Gerardo's narratives indicate a shift from talking about controlling and possessive behaviours in a relationship in broadly negative terms to referring to them as forms of violence.

In the first interview, Gerardo talked about his parents' negative views of IPV: he said his mother had accompanied a friend to report a physically violent husband to the authorities, and that his father told him never to hit a woman. He told us he was proud that his past relationships were not violent. 'I've never had, like, these modern relationships you could say, where there is violence, where if they don't hit you it's because they don't love you, and things like that. So, I feel like, proud to say that all my relationships have been like that... um, healthy.' In this interview he already spoke about jealousy and controlling behaviour as something to be avoided. For example, he said that 'if you start to ask to check your partner's phone, it's because there is no trust. And so that relationship will never survive.'

In the second interview, Gerardo spoke again about avoiding jealousy: 'I do not like being jealous because I think that trusting your partner is essential. So, I... if I see there is something going on that I don't like, instead of being jealous I try to talk things over.' While he already expressed negative views about possessive behaviour in his first interview, he nonetheless told us that he had put into practice what he had learned in the course. He said the course had helped him realise he did not need to agree to certain demands from a partner, such as 'not to accept that they want to check my phone constantly for no reason, or not to accept that they hit me, or things like that.' He also said that, based on what he learned in the course, he told a friend that she should not allow her boyfriend to monitor her phone.

Her boyfriend wanted to check her cell phone, and well, I told her that I had just remembered the topic [in the course] and [...] he didn't have a reason to check her phone. Because he is not even her parent to be able to do that.

From the third interview onwards, he started referring to monitoring of cell phones or social media as not only something negative, but as forms of violence. For example, in the fourth interview he said, 'It's like bullying a person, looking in their personal things. I do think this is a form of violence.'

Case 2: Elena (young woman, 14 years old, 4 interviews)

Elena's interviews over time suggest that the course encouraged her to expand her definition of IPV beyond the physical to include possessive and controlling behaviours, and to rethink her beliefs about acceptable ways of behaving in a relationship.

In the first interview, Elena told us that trust was an important aspect of a relationship and that she disapproved of IPV, which she said included physical and psychological violence, yelling or being controlling. In this same interview, she mentioned that her parents had a 'very normal' relationship because they did not hit each other. 'Between them it is good [...]. When they fight, it's more because I stay in the house [...] They only scream at each other, it is not that they hit each other or stuff like that.'

In the second interview, Elena told us that she learned in the course about IPV, how it manifests, and types of 'good' or 'bad' behaviour in relationships. She said she learned that IPV 'can start with jealousy and end with murder.' Elena also told us that her father acted possessive and controlling over her mother, for example monitoring her phone and forbidding her from doing many things – but she did not specify what she thought about this behaviour.

In the third interview, Elena said the course helped her learn how one should behave in relationships. She told us, 'before [...] I didn't pay attention to that [acceptable behaviour]. [...] But now that I see what's ok and what's not [...]. I'm beginning to take that more seriously.'

In the final interview, Elena – for the first time in our conversations – referred to jealousy and controlling behaviour as potential forms of violence 'if it gets to an obsessive point.' She also talked about her own feelings of jealousy in relationships with both men and women. She said it was important not to act on that jealousy by being controlling over a partner.

Well yes, I am a jealous person. But I can tell you that I am not one of those people who would tell you, 'Yes, you made me jealous, and so I am not going to let you do that, and I am going to forbid you from seeing that other person.'

The course seems to have encouraged her to reflect on how she wanted to act in her relationships. She appeared to begin to recognise her own feelings of jealousy, and stated her intentions to avoid acting on these by controlling a partner. Her narratives may also reflect a departure from her initial characterisation of her parent's relationship – in which they screamed at each other and her father controlled her mother's actions – as 'normal.'

Case 3: Julián (young man, 15 years old, 4 interviews)

Julián's narratives suggest that he became more convinced over time that jealousy in a relationship is negative. He described the jealousy-related information in the course as important and potentially influencing these beliefs.

In the first interview, Julián told us that his first girlfriend had been jealous and controlling, although he said that at the time, he had not realised it was a problem. He talked about different levels of jealousy, saying 'Yes, obviously, as a human, yes, yes I have felt jealousy. But it did not go further than saying "who is it?" or something like that. Never a single blow, or breakup, or shouting, or anything. There is jealousy and there is *jealousy*.'

In the second interview, he spoke again about this past relationship, saying it happened because he hadn't known what love should be like. He said the course activities helped him think differently, 'because you reflect [on the types of violence] and you say, "ah, I had this [happen to me]" or "I did that".'

In the third interview, Julián told us that learning about jealousy was an important aspect of the course for him. He mentioned twice that he had changed his beliefs, saying that before the course he had thought that jealousy was a form of love, but now he perceived it negatively. For example, he said:

[Before the course] I would say, if there is jealousy there is love. Or that if they are not jealous, they don't love you, and [things] like that. But I think that was my ideology – that you always need to have jealousy because it's a form of protection, to know that you love someone. [...] [Now I think], well, that it is bad, because if you trust your partner, why are you going to be jealous?

In the final interview, two months later, he again spoke of jealousy as something negative. However, this time he told us he always had thought jealousy was bad – a divergence from his narrative in prior interviews.

Interviewer: What did you think about jealousy before the course?

Julián: Well, that it is bad, because you are not sure about yourself, like, you think that the person is your property and you think that someone is going to take them away. Like no, you don't know how to trust in him... in your partner.

Interviewer: And during the course, did your perception change with regards to jealousy, or you continue to think the same way?

Julián: No. I still think the same.

Influencing relationship trajectories

The course appeared to affect participants' beliefs about relationships and influence their ongoing relationships at the time of the course. The cases of Laura and Beatriz illustrate this:

Case 4: Laura (young women, 16 years old, 4 interviews)

Over time, Laura's narratives describe a changing relationship trajectory in which she identified that the jealousy and controlling behaviour in her relationship were escalating, then gradually constructed a more positive relationship through assertive communication.

In the first interview, Laura already described having strong feelings against IPV because of the partner violence she observed between her parents and among other family members at home. She also talked about the harms of romantic jealousy and said she had intervened in a controlling relationship at school, saying to a classmate, 'he [your boyfriend] does not have rights over you, he is not your owner or anything like that.' We do not know if this exchange took place before the course began.

In the first interview, Laura also said that the aspect of her current relationship that she liked least was 'that he is very jealous.' In the second interview, she told us that her boyfriend had not always been jealous, but that things started to change a few months ago, after they had had sex for the first (and only) time. Since then, he had monitored her social media and phone, questioned her about her whereabouts, stopped her from talking to other men, intimidated people who spoke with her, spied on her, and told her how to dress. She said she did not like this behaviour and had told him he could not control her: 'If my dad doesn't forbid me from doing things, why would you?' She described the situation as 'a fight that never ends' and said she had considered ending the relationship, but every time she tried her boyfriend cried and convinced her to stay together. She said she did not want to have sex again, because she was concerned his controlling behaviour would further escalate.

When that happened [having sex], [...] it wasn't suddenly, like the next day, right? But yes, I don't know, maybe a week later he was like, 'I don't want you to see you like this, I don't want to see you with him' and things like that. [...] That's why I'm saying, for it [sex] to happen again – well, I don't think so.

In the third interview, Laura said she did not want to talk about her relationship. In the fourth and final interview, she told us her relationship had changed and that she no longer experienced harmful behaviours in her relationship. 'In my personal life... well, I have not been exposed to those situations... Before, I was, but not now.' She said she was fighting less with her boyfriend, and that her priority was now school rather than the relationship. She also described the relationship as something that might not last forever. She said she told her boyfriend, 'just because we are dating doesn't mean we are going to be together forever' and that 'we must also focus on things that really matter to us. That is, I am not saying the relationship doesn't matter, but at least for me the priority is to be in school and, to do well in school.'

When we asked if the course helped improve her relationship, she said it did by promoting communication, and because '[the facilitators] said that a relationship is between two people, no? That you have to have trust, that being jealous does not... does not benefit the relationship.' Laura said she attended all the intervention sessions and paid particular attention to the IPV activities. She told us the course helped participants understand that possessive behaviour is a type of IPV and that she felt more confident about what to do if she experienced it in the future.

Laura's teacher, when asked about any changes she had observed in the students during the course, mentioned a student who had been in a controlling relationship, but had lately become more confident and involved in her schoolwork – in part because of the course. While we can not be certain, we believe she was referring to Laura based on the description and detail of the situation.

She is a very committed girl, very studious... she has very clear goals for herself. However, she was taken by this infatuation, right? Then [...] she was very controlled by her boyfriend, very controlled. We realised that her boyfriend was violent towards her [...]. But now I notice her being more confident, happier. I am talking about something subjective [...] but her appearance is that she is back to that girl from the first semester, who came [to school] with enthusiasm. [...]. And especially this course, I think, helped her understand many things. (Teacher)

These narratives demonstrate shifts in how Laura – and seemingly her teacher – perceived and described her relationship trajectory. At first, she spoke negatively about controlling behaviour in

other people's relationships but did not refer to this in her own relationship as violence. In later interviews, she identified escalating controlling behaviour in her own relationship and also reflected on the course contents about IPV. By the final interview, she talked about her relationship as something that was less of a priority in her life, and suggested that she was no longer experiencing harmful behaviour in her relationship.

Case 5: Beatriz (young woman, 14 years old, 2 interviews)

The interviews with Beatriz suggest that she engaged in critical reflection and learning that influenced her relationship over time. She communicated with her partner about types of partner violence, identified her own excessive romantic jealousy, and expressed her intent to manage it and to avoid IPV in the future.

We interviewed Beatriz for the first time over halfway through the course and again two months after the course ended. Both interviews took place after the IPV-related session, which she said she had not paid attention to – possibly having missed much of it by leaving early. In the first interview, Beatriz described a broad definition of IPV that included jealousy and possessive behaviour, and talked about having intervened in partner violence she observed in her neighbourhood. She told us that what she liked least about her boyfriend was his jealousy and possessive behaviour, such as telling her what to wear or monitoring her social media. While the definition of IPV that she shared with us in this interview included these types of behaviour, she did not refer to her boyfriend's behaviour as violent.

In the second interview, Beatriz talked about a refresher course provided by Mexfam after the end of the semester, saying it had prepared her to prevent IPV in her future relationships. She also talked about the importance of learning the different types of escalating violence in a relationship, and said she told her boyfriend about these phases of violence.

Everything was explained to us step by step, all the phases [of violence] that led to, to the last thing, which was murder. And what it is, what should not be done, how to realise that the relationship was already really bad, and how to try to break up for the sake of your sanity.

In this interview, Beatriz also said the course helped her see 'that it is not good to be jealous of a person for any little thing' – she said she realised that she herself was becoming more jealous in her current relationship but was trying to control it.

I kind of knew who his girlfriends had been [...] and it bothered me that he talked to them. [...] I became even more jealous. But [...] from the last time [I felt jealous] I said [to myself], 'No, what are you doing. I mean, when did you become like this?' And I started to think things out a little more.

She told us she did not want to be violent in future relationships. 'To see something as small as ... a little push, a bite, one can get to the point of killing your partner if this goes on. And no, oh no, I don't want to be like that in the future.'

Participant descriptions of their experiences in the repeat interviews

Overall, the nine 'case study' participants had positive comments about the repeat interviews. Some said that the interviews helped reinforce the course contents. For example, Elena said she liked that the interviews made her remember what was discussed in course. Julián and Gerardo noted that while all of the participants in the course were given the opportunity to reflect, the repeat interviews reinforced the core topics and may have provided additional opportunity to reflect.

For some, the repeat interviews allowed them discuss topics they did not talk about with anyone else. For example, Julián said the interviews were a place to unburden himself, and Israel said the interviews 'helped me like [...] to let out what I had. [...] Clear my head, to... let off steam and then talk to someone, right?' The interview transcripts show how participants introduced topics that were troubling them – such as arguments with siblings, their sexual orientation, violence at home, or social conflicts – and discussed them at length over multiple interviews.

At the start, not all participants were comfortable sharing their feelings and personal information. While some shared sensitive information in the first interview – including sexual assault, suicide attempts, and family and relationship violence, others did not disclose these types of experiences until later interviews. Lizbeth and Karina each mentioned feeling nervous or uncomfortable about the idea of sharing personal information in their first interview. 'When you and I started to have these interviews, it was like, it was like, "oh, how scary, and what if she tells someone else [about what we discussed]"' (Lizbeth). Both of these participants said they ultimately decided to share sensitive information as they felt that the interviewer was open-minded and would not judge them.

Repeat interviews during and after the intervention were feasible and acceptable – and may also have reinforced some of the course messages. The qualitative longitudinal research process appears to have created opportunities for participants to introduce topics that were important to them, to have allowed trust to develop, encouraging disclosures that a one-time interview with the same participant might have missed.

Discussion

Using qualitative longitudinal research for an evaluation study allowed us to explore pathways to change, identify and learn about gradual and subtle shifts along these pathways, consider what aspects of the intervention appeared to contribute to change among participants, examine how context influenced participant responses to the intervention, and explore apparent inconsistencies in participant narratives. We found repeat interviews particularly useful to account for the weekly frequency with which participants were exposed to intervention messages and for research with young people, to engage with their rapidly evolving experiences with relationships, sexuality and violence.

The repeat interviews we conducted in Mexico suggest a number of pathways to IPV prevention among participants in a school-based comprehensive sexuality education programme. We observed a range of behaviours on these pathways: participants shifted their understandings of jealousy, possessive behaviour, and other forms of IPV; reconsidered their beliefs about acceptable and healthy behaviours in a relationship; communicated course learnings about IPV to their intimate partners; and identified and managed the harmful behaviours in their own relationship to construct more positive relationship dynamics. All of these changes can plausibly lead to less violent relationships in the future.

Recent studies illustrate two key factors that highlight the importance of detecting gradual and subtle change on pathways to IPV prevention. First, that prevention programmes sometimes lead to attitudinal or behaviour shifts without attaining transformational or system-level change; and second, that attitudinal change can happen without behaviour change, and vice versa (Pierotti, Lake and Lewis, 2018; Gibbs *et al.*, 2019; Jewkes, Stern and Ramsoomar, 2019; McLean, Heise and Stern, 2019). This may imply a growing emphasis on detecting incremental attitudinal and behavioural shifts and exploring the relationship between them. In contrast, system-level transformations, such as shifting unequal power relations between men and women or reducing

community-wide levels of IPV, often require long timeframes to be achieved and can be more challenging to influence – and to measure – than individual-level change. As such, it may be expedient to focus evaluation efforts on identifying gradual change processes and suggestive pathways to change, rather than investing in measurement of transformations that may not emerge within research timeframes.

Repeat interviews during and after an intervention helped illuminate aspects of the intervention that seem to contribute to shifts among participants. Based on aspects of the intervention that were mentioned during multiple interviews by different participants, we found that two messages in the course seemed particularly influential on relationship trajectories. First, the message that excessive jealousy and possessive behaviours could be forms of IPV; and second, that these behaviours as well as other less severe forms of physical aggression could be on the pathway to escalating physical violence. Participants seemed to use the course activities on these topics as a lens through which to reflect on their own past and present relationships. Our findings suggest first that beliefs about jealousy and possessive behaviour may be amenable to change, and second, that it is feasible to intervene to tackle excessive jealousy and controlling behaviour as part of IPV prevention interventions. These results have much in common with those of a study of the Indashyikirwa programme in Rwanda, which found that the realisation among participants that ‘minor’ forms of IPV can escalate to more serious forms of physical violence may have helped reduce tolerance to IPV (Stern and Niyibizi, 2018). The authors note the importance of raising awareness among participants about ‘less visible and scrutinised forms of violence, such as reportedly ‘minor’ forms of physical violence like slapping, emotional abuse, and sexual coercion in relationships’ (Stern and Niyibizi, 2018). This similarity in findings from different contexts and populations – unmarried adolescent students in Mexico and cohabitating adult couples in Rwanda – suggest that these intervention messages may well be useful and relevant elsewhere.

Repeat interviews conducted frequently may be particularly useful for studies examining phenomena and experiences that are complex and change over time in subtle ways, such as sexuality, relationships and violence. Many of our participants had multiple relationships and were experimenting sexually or questioning their sexuality over the data collection period (Chapter 4). Transition to adulthood is a time of rapid change (UNAIDS, 2004; Kågesten *et al.*, 2016; Blum, Mmari and Moreau, 2017; Mmari *et al.*, 2017), and frequent data collection may be particularly useful for this age group to capture evolving relationship experiences and reflections (Giordano *et*

al., 2010; Price *et al.*, 2016). Multiple interviews can help develop trust and rapport that is essential for research on sensitive topics, and may be particularly useful where the behaviours of interest are stigmatised or illegal.

Qualitative longitudinal research can illuminate the process and role of critical reflection in social change interventions. Critical reflection is an important component of interventions contributing to social norms change (Jewkes, Stern and Ramsoomar, 2019). The intervention we evaluated was designed to encourage reflection about beliefs and norms related to gender and violence. We found evidence that participants did engage in self-reflection about their past and present relationships in light of the information that was shared in the course, and that this influenced their beliefs about what is acceptable behaviour and how they hoped to experience and behave in their relationships in the future. These reflections appeared to be able to influence relationship trajectories even when only one partner was in the intervention, as in the case of Laura, or when a participant was not in a relationship at the time of the intervention, as in the cases of Gerardo, Julián and Elena. This is promising sign for interventions aiming to prevention IPV among young unmarried participants who may not yet have had many personal experiences to draw on.

While critical reflection often began in the classroom – such as in response to course activities or messages – our data suggest it then continued during the interview process. From the perspective of the participants, both intervention and evaluation activities may be experienced and interpreted as the ‘intervention.’ Our findings suggest that repeat interviews may have encouraged reflection about intervention messages beyond what would be anticipated through course participation alone. It is therefore possible that the repeat interview process contributed to quicker or larger shifts than we would have observed among other participants. Other researchers have pointed out that research processes, particularly longitudinal ones, can stimulate reflection (Oakley, 2016b) or improve recall among participants (Oakley, 2016a). Repeat interviews may improve our ability to detect and learn about the nature of participant experience in an intervention due to a potentially accelerated response. In reinforcing the critical reflection component, however, they may create a false impression of what the intervention has achieved and what similar interventions might achieve without the accompanying interviews.

Qualitative longitudinal research can contribute to highly contextualised data by creating multiple opportunities for participants to share information about their lives that they find to be important.

Topics raised by our interviewees were not always directly relevant to our central research questions, but they provided insight into participants' lived experiences and their family and social contexts, and helped us understand how these may have influenced their experiences of the intervention. Providing space for participants to lead the conversations to what they care about may help mitigate power imbalances between the interviewer and the interviewee (Collins, 1998; Vincent, 2013; Oakley, 2016b). Analysing case histories from repeat interview data (Thomson, 2007; Henderson *et al.*, 2012) allowed us to link participants' individual circumstances and context with their experiences in and responses to the intervention; this helped explain variability in how different participants experienced the intervention (Chapter 6). In their narratives, participants reflected on their family situations in relation to the course contents and described discussing the intervention with family members, suggesting that violence-related beliefs and experiences at home and participant trajectories through the intervention influenced each other.

Repeat interviews during and after the intervention allowed us to track shifts in how participants talked about key course topics as they went through intervention. However, these changes sometimes made interpretation more challenging. For example, Gerardo and Elena both began to refer to excessive jealousy or possessive behaviours as forms of violence in the final two interviews (at the end of the intervention and two months later). This may suggest a shift in their beliefs about IPV that was compatible with the aims of the intervention. However, as the intervention progressed and these messages were presented and revisited, participants may have wanted to present themselves to the interviewer in ways they believed were consistent with the course messages and therefore more desirable. In other words, it is important to be critical in interpreting the data given that genuine shifts in beliefs or behaviours (what the intervention hopes to achieve), may look, at interview, very similar to changes in how the less naïve participant wishes to present themselves in the interview – which may or may not be accompanied by deeper changes in attitude.

Inconsistencies in narratives can also arise in repeat interviews, and bring interpretation challenges. For example, in one interview Julián said the course helped him realise that jealousy in a relationship was 'bad', but two months later he said he always thought that jealousy was bad, even before the intervention. It is difficult to interpret this apparent inconsistency. Perhaps with the passing of time Julián assumed he had always thought of jealousy in negative terms, or maybe – because jealous behaviour was in conflict with the lessons of the intervention – he did not want

to acknowledge that he had ever thought otherwise (or both). This raises questions of how to reconcile divergent or inconsistent narratives that may well emerge over time during longitudinal data collection. Other qualitative longitudinal studies have similarly described challenges interpreting data when participants change their narratives about a particular topic (Calman, Brunton and Molassiotis, 2013), reinterpret prior interactions or occurrences (Lewis, 2007), or appear unaware that their perspectives may have shifted (Grossoehme and Lipstein, 2016). Oakley suggests that such 'discrepancies' between time points relate to participants having time to reflect on and process their experiences (Oakley, 2016a).

While inconsistent narratives may be common in qualitative longitudinal research, repeat interviews may provide some advantages when it comes to interpreting these. Researchers can review transcripts and audio recordings between interviews and suggest follow-up questions and prompts for the next interview; this can help identify unclear points or areas where detail is lacking, and provides opportunities to probe on any apparent conflicts in narratives (Vincent, 2013; Burke *et al.*, 2019). Every additional interview allows some scope to follow up on apparent contradictions and provides further opportunities to explore these, as sometimes contradictions are not obvious during the interview itself.

This study has limitations. First, we were unable to begin the repeat interview process before the intervention began because of delays related to flooding and earthquakes. Because of this, we missed the opportunity to create a qualitative baseline. However, we designed the interview guides to explore shifts that might have been influenced by the intervention. Second, participants may have chosen to overstate their favourable responses to the intervention because the interviewer worked for the organisation implementing the programme. Third, this analysis examined individual changes but did not explore broader shifts in social norms (e.g. changes in families, or in the wider context). We explore the potential of the intervention to shift norms elsewhere (Chapter 4, Chapter 6). Fourth, our findings highlight suggestive pathways to IPV prevention, but we cannot know whether these changes will ultimately prevent or reduce IPV, or whether any changes will be sustained over time. Fifth, the repeat interview participants may differ from others in the study in terms of their willingness to share sensitive information from early on in the research process. Finally, while we included a substantial number of interviews in this analysis, the number of participants is somewhat small. However, the added benefits of increasing the number of participants might not offset the attendant costs in terms of the labour

and resources required for logistics, transcription, and analysis. We gave careful consideration to the frequency of interviews and number of participants in order to keep a manageable amount of data while also gathering data at intervals suitable to learn about gradual shifts over time during the intervention.

Conclusion

Our qualitative longitudinal approach allowed us to detect processes of change among participants, identify aspects of the course that seemed to influence these, and gather contextualised data that helped us understand how participants' lives influenced their experiences in the intervention, and vice versa. We found that repeat interviews provided meaningful empirical evidence of how the intervention influenced participants – in this case, their relationships trajectories and pathways to IPV prevention, which can be measured in the short to medium term. This evidence can, we would argue, serve as a proxy for IPV prevention over the longer term where measurement is very difficult. This method can be used alongside other complementary approaches that are designed to engage a larger number of participants to jointly build a more complete picture of intervention mechanisms and effects over time.

Chapter 6: Applying a social complex adaptive systems approach when evaluating implementation of a school-based intervention for intimate partner violence prevention: A case study in Mexico

6.1 Introduction

In the previous chapter, I established that individuals have different trajectories through and responses to a comprehensive sexuality education intervention in Mexico that seeks to prevent IPV. I showed that their trajectories reflect their personal history, family context, and the circumstances of their relationships. In this chapter, I complement the individual-level analyses from Chapters 4 and 5 with an examination of collective or group-level experiences. I do so by adopting a complex adaptive systems approach to examine intervention processes. This links to the overall methodological exploration by considering how evaluation approaches can take complexity into account.

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1400132	Title	Ms
First Name(s)	Shelly		
Surname/Family Name	Makleff		
Thesis Title	Methodology for evaluating complex social interventions: understanding partner violence prevention for young people in Mexico City		
Primary Supervisor	Prof Cicely Marston		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Health Policy and Planning
Please list the paper's authors in the intended authorship order:	Shelly Makleff, Marissa Billowitz, Jovita Garduño, Mariana Cruz, Vanessa Ivon Silva Marquez, Cicely Marston

Stage of publication	In press
----------------------	----------

SECTION D – Multi-authored work

<p>For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)</p>	<p>I am the first author of this paper, and was responsible for study design, conception of the paper, data analysis, and writing, and also participated in data collection. Cicely Marston supported on study design, conception of the paper, analysis and writing. Because of the co-produced nature of the evaluation case study, key members of the partner organizations are co-authors based on their involvement in study design, data collection, or interpretation of findings, as well as critical revisions of the paper.</p>
---	---

SECTION E

Student Signature	[Redacted]
Date	29 / 5 / 2020

Supervisor Signature	[Redacted]
Date	29 / 5 / 20

6.2 Abstract

Despite calls for evaluation practice to take a complex systems approach, there are few examples of how to incorporate complexity into real-life evaluations. This paper presents the case for using a complex systems approach to evaluate a school-based intervention aiming to prevent intimate partner violence (IPV).

We conducted a post hoc analysis of qualitative evaluation data to examine the intervention as a potential system disruptor. We analysed data in relation to complexity concepts particularly relevant to schools: *diverse and dynamic agents, interaction, unpredictability, emergence, and context dependency*. The data – two focus groups with facilitators and 33 repeat interviews with 14 to 17-year-old students – came from an evaluation of a comprehensive sexuality education intervention in Mexico City, which serves as a case study for this analysis.

The findings demonstrate an application of complex adaptive systems concepts to qualitative evaluation data. We provide examples of how this approach can shed light on the ways in which interpersonal interactions, group dynamics, the core messages of the course, and context influenced the implementation and outcomes of this intervention. This gender-transformative intervention appeared to disrupt pervasive gender norms and reshape beliefs about how to engage in relationships.

An intervention comprises multiple dynamic and interacting elements, all of which are unlikely to be consistent across implementation settings. Applying complexity concepts to our analysis added value by helping reframe implementation-related data to focus on how the ‘social’ aspects of complexity influenced the intervention. Without examining both individual and group processes, evaluations may miss key insights about how the intervention generates change, for whom, and how it interacts with its context. A social complex adaptive systems approach is well-suited to the evaluation of gender-transformative interventions, and can help identify how such interventions disrupt the complex social systems in which they are implemented to address intractable societal problems.

6.3 Research paper 3

Introduction

Despite calls over the last decade for evaluation practice to take a complex systems approach and move beyond the individual to explore macro-level effects, there are few examples of how to incorporate the concept of complexity into real-life evaluations (Craig & Petticrew, 2013; Moore et al., 2019; Smith & Petticrew, 2010). Interest in evaluating complex and social interventions has grown steadily (Craig and Petticrew, 2013), yet ‘the literature [...] is thick with descriptions of complex, challenging interventions, but thin on practical advice on how these should be dealt with’ (Datta and Petticrew, 2013). Reviews have found that clinical and health promotion interventions, and hospitals and schools, are commonly the subject and sites of research discussing complexity (Datta and Petticrew, 2013; Thompson *et al.*, 2016). However, applications of complexity theory for interventions that address complex phenomena driven by underlying social norms, for example intimate partner violence (IPV), are rare. One study in New Zealand applied complexity theory to conceptualise the health care system response to IPV, but did not mention the additional complexities of social norms or gender (Gear, Eppel and Koziol-McLain, 2018). Taking these into account is important, as social norms are important drivers of the harmful global phenomenon of IPV (Jewkes, Stern and Ramsoomar, 2019) and gender itself is a complex social system that defines what we expect of women and of men in any given society (Hirdman, 1991; Heise *et al.*, 2019). There is a growing evidence base examining community-based interventions that address gendered social norms as a part of IPV prevention efforts (Jewkes, Stern and Ramsoomar, 2019); however, these studies rarely adopt a complex systems approach and few are carried out in schools. Here we consider the case for using a complex adaptive systems framework to evaluate school-based IPV prevention interventions.

Complexity, often considered in the evaluation literature to be an attribute of an intervention, can alternatively be conceived of as a characteristic of the system or setting in which an intervention takes place (Shiell, Hawe and Gold, 2008; Hawe, 2015). Building on this, Moore et al. conceptualise interventions as events that aim to disrupt complex systems (Moore *et al.*, 2019). This moves the focus of evaluative research away from individual behavioural change (Westhorp, 2012) to instead examining how a complex system – such as a hospital, school or community – responds to an intervention over time (Moore *et al.*, 2019). Evaluation conducted with a complex systems approach can seek to understand how an intervention – an event in a system – ‘begins to gain

traction within its context over time' to 'either leave a lasting footprint or wash out' (Hawe, Shiell and Riley, 2009). Context is a necessary focus when implementing or evaluating interventions with a complex systems perspective, as the 'effects of any intervention are influenced strongly by the starting points of the system they attempt to disrupt' (Moore *et al.*, 2019). In other words, the type and extent of change that an intervention creates in a system reflects local or particular characteristics.

Beyond the properties of the system in which an intervention is implemented, additional complexities may reflect the types of outcomes being addressed. IPV is a complex social phenomenon, and prevention programming may be more effective when it intervenes at multiple levels beyond the individual, for example, the relationship, community and societal levels (Heise, Ellsberg and Gottemoeller, 1999; Morrison, Ellsberg and Bott, 2004; Meinck *et al.*, 2019). IPV prevention interventions often use a 'gender-transformative' approach, which aims to shift gendered power differentials to become more equal (Dworkin, Fleming and Colvin, 2015; Jewkes, Flood and Lang, 2015; Michau *et al.*, 2015; Jewkes, Stern and Ramsoomar, 2019; Ruane-McAteer *et al.*, 2019). Such shifts, which are central to gender-transformative programmes, are also complex (Walters, 2004). IPV prevention interventions are usually implemented in group settings, such as communities, schools or families (Garcia-Moreno *et al.*, 2014; Lundgren and Amin, 2015b; Jewkes, Stern and Ramsoomar, 2019). We adopt the conceptualisation of schools – and other group settings – as *social* complex adaptive systems (Keshavarz *et al.*, 2010), which comprise 'a collection of individual agents with freedom to act in ways that are not always totally predictable, and whose actions are interconnected so that one agent's actions changes the context for other agents' (Plsek and Greenhalgh, 2001). Given the complexities inherent to IPV programming, a social complex adaptive systems approach may be well-suited for evaluating prevention interventions.

We present an illustrative case study applying a complex adaptive systems approach to an evaluation of a school-based comprehensive sexuality education intervention with a focus on preventing IPV. The study objectives are to consider whether complex adaptive systems concepts appear relevant for the evaluation; how this approach helps understand the intervention; and what it may add beyond a traditional evaluation perspective. We also aim to identify 'disruptive' elements of the intervention and examine how group dynamics and social context can influence participant experiences and intervention outcomes.

Methods

Theoretical perspective

Schools have many of the key characteristics of *social* complex adaptive systems – they comprise a population of diverse and changing individuals (or ‘agents’) such as students and teachers, who interact in often unpredictable ways that are linked to context, leading to non-linear and emergent outcomes (Keshavarz *et al.*, 2010). Therefore, a complex systems lens is appropriate for research about school-based interventions (Keshavarz *et al.*, 2010; Moore *et al.*, 2019). Keshavarz *et al.* identify key attributes of complex adaptive systems as particularly relevant to schools: *diverse and dynamic agents, interaction, unpredictability, emergence, and context dependency* (Keshavarz *et al.*, 2010). We present definitions of these terms used for this analysis in Table 10.

Table 10: Definitions of key complex adaptive systems terms

Complex adaptive systems terminology	Definition for analysis
Complex adaptive systems	‘At its core, a <i>complex adaptive system</i> comprises a population of diverse rules-based agents, located in multi-level and inter-connected systems in a network shape. A system is characterised by the behaviour of individual agents. Agents in complex adaptive systems are often numerous, dynamic, autonomous, highly interactive, learning and adaptive. Agents of complex adaptive systems act in ways that are based on a combination of their knowledge, experience, feedback from the environment, local values and formal system rules. These change over time leading to continuously changing interactions and adaptations that are often novel and are hard to predict, especially in social systems. Agents in complex adaptive systems interact with and adapt to each other and the system within the network. Complex adaptive systems are open systems with fuzzy boundaries and also highly context dependent in terms of time, history, and space including location and proximity. Complex adaptive systems also have distributed control. Consequently, complexity that is not

	necessarily a characteristic of individual agents, emerges at system level.’ (Keshavarz et al., 2010).
Diverse and dynamic agents	Agents ‘act in ways that are based on a combination of their knowledge, experience, feedback from the environment, local values and formal system rules’ (Keshavarz et al., 2010). For this analysis, we define <i>diverse and dynamic agents</i> as intervention participants, facilitators and teachers.
Interaction	For this study, we consider <i>interaction</i> to be the interpersonal and group dynamics among these agents as well as between them and their family members or peers.
Unpredictability	<i>Unpredictability</i> is explained as continuous changes and adaptations in the system that may be ‘novel and are hard to predict, especially in social systems’ (Keshavarz et al., 2010).
Emergence	We use the term <i>emergence</i> to reflect the unpredictable changes or outcomes in the system resulting from the ‘interplay of the many factors indicated above over time’ (Keshavarz et al., 2010).
Context dependency	<i>Context dependency</i> suggests that individual agents will behave differently depending on the unique context within each system; and that ‘different contexts create different influences on the way in which agents can function and on the complexity of introducing change’ (Keshavarz et al., 2010). This concept is similar to that of <i>path-dependence</i> , which considers systems as ‘sensitive to their initial conditions, so that the same force might affect seemingly similar organizations [systems] differently based on their histories’ (Lindberg & Schneider, 2013).

Study design

We conducted a post hoc analysis and examined existing qualitative data from a case study – a school-based evaluation – against key attributes of complex adaptive systems. We designed this case study with a dual purpose: first, as an evaluation of a specific intervention in Mexico, and second, as a methodological exploration to apply and test different research methods, approaches

and techniques during the course of the evaluation itself. The complex adaptive systems analysis presented in this paper responds to this second purpose.

We conducted the evaluation study in one school in the south of Mexico City in 2017 and 2018. The aim was to learn about participant experiences in the intervention and explore whether and how it contributed to IPV prevention. Three partner organisations collaborated on study design and implementation: Fundación Mexicana para la Planeación Familiar, A.C. (Mexfam) – a Mexican non-governmental organisation providing community-based health promotion as well as clinical services; International Planned Parenthood Federation/Western Hemisphere Region (IPPF/WHR) – an international non-governmental organisation; and London School of Hygiene and Tropical Medicine (LSHTM). The first author was affiliated with the latter two at the time of the study, and coordinated the collaboration.

The evaluation employed a longitudinal quasi-experimental design with an intervention and comparison group. Mixed methods of data collection were in-depth interviews, repeat (longitudinal) interviews, self-administered questionnaires, focus groups and observation. We collected data from students ages 14 to 17, teachers, and Mexfam facilitators. Participants received a gift card for each interview or focus group as compensation for their time and were offered subsidised services at Mexfam clinics. We obtained ethical approval in Mexico from *Centro de Investigación Clínica Acelerada* (CICA) and in the UK from the LSHTM Research Ethics Committee. We have reported elsewhere on the intervention content and context, participant characteristics, and evaluation findings – which suggest that the course promoted critical reflection that appeared to lead to changes in beliefs, intentions, and behaviours related to gender, sexuality, and relationships that supported the prevention of and response to IPV among young people (Chapters 4 and 5).

The evaluated programme comprised 10 two-hour sessions delivered over one semester to mixed-gender groups of approximately 20 young people, who remained together for the duration of the course (the ‘intervention group’). The intervention was implemented with two groups of students in their third year of secondary school (*preparatoria* in Mexico) during the pilot phase and six groups in their first year of secondary school during the full implementation; there was a possibility of exchange between members of different intervention groups in their other courses. Each group was assigned a young (under 30) facilitator who was staff at Mexfam, where they were

trained on the comprehensive sexuality education curriculum to ensure consistency in implementation while allowing for flexibility to adapt to emerging situations. The manual-based curriculum, designed based on international standards (United Nations Educational Scientific and Cultural Organization, 2018), comprised participatory activities, each building on prior sessions, to reinforce key messages about course topics. These included sexuality, gender, equitable relationships, IPV, and other sexual and reproductive health topics. Core messages included the importance of self-respect, promoting tolerance of difference, and considerations of power. IPV was presented as a range of behaviours that anyone in a relationship could perpetrate or experience, including excessive jealousy or control over a partner and other behaviours that cause emotional, physical or sexual harm. Activities were designed for factual learning, to encourage access to health services, to stimulate debate, reflection and group discussion, and to build interpersonal, communication and relationship skills. The intervention was gender-transformative in its aim to generate critical reflection about gendered social norms and shift individual attitudes and group norms related to gender, sexuality and violence.

Data collection

Focus groups with course facilitators were conducted at Mexfam's offices and co-facilitated by the research coordinator (JG) and the first author (SM) – both women. The objectives were to learn about intervention processes (challenges, group dynamics, activities) and facilitators' perceptions of the course effects. All Mexfam facilitators who implemented the curriculum as part of the study, as well as the manager of the comprehensive sexuality education programme, were invited to participate. We conducted one focus group in June 2017 with the two facilitators (one woman, one man) who implemented the intervention pilot. We held another in December 2017 with the programme manager (female) and all four facilitators (three women, one man) who implemented the course during the full study. Two facilitators attended both focus groups. Participants had a mean age of 26.4 years (range: 23 to 29).

Nine students (five women and four men) with a mean age of 15.1 years participated in repeat semi-structured interviews over a six-month period, during and after the intervention. We conducted 33 interviews in total; seven participants completed four interviews, one completed three, and another completed two. The research coordinator (JG) conducted these either in a private space at the school or at Mexfam's offices. Interviews included questions about

experiences in and perceptions of the intervention, group dynamics, and course effects. The sampling strategy and methodology of the repeat interviews are detailed in Chapter 5. Table 11 presents information about the facilitators and participants in each implementation group. We use pseudonyms to protect participant confidentiality.

Table 11: Characteristics of intervention groups in the study

Group	Timeframe	Facilitator	Gender balance	Repeat interview participants	Summary of dynamics and events in each group
1	January – June 2017 (pilot)	Paola (F) with support from Orlando (M)	60 % women 40 % men	Laura (F) Gilberto (M)	Developed trust over time Learned to engage in respectful debate
2	January – June 2017 (pilot)	Regina (F) with assistance from Orlando (M)	80 % women 20 % men	None	Women in the group debated whether a female classmate was experiencing subtle forms of IPV. The classmate and her boyfriend were there and denied that their relationship was violent.
3	August – December 2017	Patricia (F), then switched to Orlando (M)	55 % women 45 % men	None	Earthquake during session Change in facilitator after the earthquake was hard to adapt to for participants
4	August – December 2017	Berenice (F)	55 % women 45 % men	Beatriz (F) Elena (F) Israel (M) Julián (M)	Earthquake during session Conflict resolution session related to verbal aggression among participants

					<ul style="list-style-type: none"> - Participant yelled at facilitator and hit classroom wall - Some participants treated intervention as joke, distracted others - Respectful dialogue regarding sexual diversity - Male participants unwilling to participate in IPV-related activities
5	August – December 2017	Tania (F)	55 % women 45 % men	Karina (F) Lizbeth (F)	<ul style="list-style-type: none"> - Negative comment about abortion triggered strong reaction among female participants and facilitator - Some treated intervention as joke, distracted others - Improvements in group dynamics over time - Women took course more seriously than men
6	August – December 2017	Berenice (F)	55 % women 45 % men	Gerardo (M)	<ul style="list-style-type: none"> - Some treated intervention as joke, distracted others - Active engagement in activities, particular when group leader was present
7	August – December 2017	Regina (F)	55 % women 45 % men	None	<ul style="list-style-type: none"> - Two men made aggressive comments about women; women in the class appeared to participate less as a result

8	August – December 2017	Orlando (M)	55 % women 45 % men	None	Women more interested in the IPV-related contents than men
---	------------------------------	----------------	---------------------------	------	--

Analysis

Others have proposed that a qualitative approach is particularly fitting when applying complexity theory (Gear, Eppel and Koziol-McClain, 2018; Gomersall, 2018). We reviewed qualitative evaluation data – transcripts from two focus groups with course facilitators and 33 repeat interviews – to identify examples of key elements of complex adaptive systems: interactions between diverse and dynamic agents, unpredictability and emergent outcomes, and context dependency. We particularly sought examples where facilitators and students described the same events, allowing comparison of different perspectives. We also reviewed the transcripts for evidence of the intervention acting as disruptive to a system (Moore *et al.*, 2019). Specifically, the analysis aimed to identify ways in which this comprehensive sexuality education programme with a gender-transformative approach was potentially ‘disruptive’ to the formal and informal rules, such as social norms, that govern the system – in this case, the intervention group and its participants.

Results

The findings presented here demonstrate an application of complex adaptive systems concepts to qualitative evaluation data. We provide examples of how this approach can shed light on the ways in which interpersonal interactions, group dynamics, the core messages of the course, and context influenced the implementation and outcomes of this intervention.

Interactions between diverse and dynamic agents

The comprehensive sexuality education course promoted interaction in the intervention group setting. For the purposes of this analysis we defined these groups, rather than the school as a whole, as the system being examined. We applied the concept of ‘diverse and dynamic agents’ from complexity theory to mean the approximately 20 participants and one facilitator engaged together in each mixed-gender intervention group. There were eight such groups in the study, two in the pilot and six in the full study. There were slightly more women than men assigned to each group, with some variation (Table 11). We examined narratives from facilitators and students to

identify examples of how interactions and group dynamics influenced experiences in different groups. They described both positive and negative interactions. Positive experiences they mentioned included respectful dialogue and learning about the beliefs and experiences of other participants, while negative ones ranged from verbally aggressive behaviour to classmates not paying attention or interrupting class.

One facilitator, Berenice, mentioned an episode of physical aggression in Group 4. Specifically, she said that a female participant yelled at her and hit the wall before leaving in the middle of a session. Some participants also mentioned this event. In addition, Berenice and various students described bullying in the form of repeated teasing or mockery. Beatriz and Elena, students in this group, each mentioned that they had been bullied by classmates during the course. Beatriz spoke about her experience:

Sometimes my group bothers me a lot because there is a group of girls [...]. Not just with me, with many of my classmates, they have hurt them with the things they say. [...] They discriminate against me because of the colour of my skin. [...] There are also kids in my class that they make feel bad, and they lower people's self-esteem.

This example indicates that violence was exerted and experienced by both women and men, and that different types of violence were perpetrated during the intervention sessions. Veteran teachers in the school also told us this group of students was particularly unmanageable, and the facilitator (Berenice) noted Group 4 as an outlier in terms of group dynamics.

The difference was very noticeable. That is, not all classrooms have the same levels of violence, but in this group [...] the violence was very marked. [...] [In one activity] the group started to [verbally] attack each other, and it was one corner of the room against the other [...]. The students became more and more out of control.

Berenice (the facilitator) also told us that after two participants (one woman and one man) escalated their verbally aggressive behaviour towards each other, she organised an activity – a spontaneous session that was not part of the regular curriculum – in which they had to directly communicate about their conflicts. Several students in Group 4 also talked about a conflict resolution exercise that addressed the fighting in the group. According to Beatriz, who participated in this activity, this session helped the two participants improve their relationship. 'They started talking and became friends, they even go out together, it helped them a lot. [...] Yes, [the course]

influenced them, a lot. [...] [To have a] better relationship.’ Julián, also in this group, did not mention this particular event but told us that while at first the participants didn’t get along, they started to become more open because of the course, show more respect in class, and pay more attention.

Berenice similarly mentioned improvements in the behaviour of the two participants who were in conflict, as well as in the group dynamics overall. However, she still expressed concerns about the ongoing aggression and violence in Group 4.

I’m a little worried because I gained their trust [...] – not completely, not 100%. But I’m, like, worried that now in the next semester, [...] they’ll change classes, there will be more violence. Some girls commented that they were going to change schools because they didn’t like the atmosphere, I mean straight up they said, listen, it’s very violent here and I don’t want to be here. Maybe not [violence in romantic] relationships, but yes, other types of violence.

Other, less severe negative interactions were also mentioned by participants from multiple groups. For example, Beatriz and Elena in Group 4, Lizbeth in Group 5, and Gerardo in Group 6 all mentioned that some of their classmates treated the intervention as a joke or didn’t pay attention during the sessions, creating a distraction for those students who did want to participate.

Participants in various groups also described positive aspects of group dynamics or improvements in interactions over time. For example, two participants from Group 1 (Laura and Gilberto) and their facilitator (Paola) each told us the group learned to accept each other’s differences and debate respectfully. As Gilberto said, ‘[Paola] built trust, a warmth between everyone that was very nice, so much that we became like a little family again.’ Karina similarly talked about the openness and sharing that developed over time in Group 5. Israel described Group 4 as respectful and open regarding sexual orientation, saying he valued the opportunity to hear what others had to say.

What I noticed is that there is a lot of, like, diversity. It’s more that they don’t follow too much, um, this rule that boys like girls and girls like boys [...] [A] girl [...] said [to her friend], seriously, ‘Don’t you like Pedrito? [...] Or do you like Ana?’ And I noticed that and turned around to look. [...] They do respect these types of things, and it’s, well, it’s something nice. [...] In this class I’m like, twice as interested in seeing what the others will say, what we will talk about.

Using a complex adaptive systems approach brought attention to both positive and negative interpersonal interactions and how they influenced not only individual experiences but also the group-level dynamics in different implementation groups.

Unpredictability and emergent outcomes

A complex adaptive systems approach anticipates continuous change in ways that are unpredictable and influence the system's collective properties over time. We analysed the data to find reference to such unpredictable events or circumstances, and present here three examples and their influence on collective experiences in the intervention.

The first example relates to how aggressive commentary in class can influence the group dynamic. In one case, the Group 5 facilitator (Tania) told us that a young man in class said was fine for a woman to die from an unsafe abortion, upsetting her as well as several young women in the class. Karina, a student in this group, said:

I don't remember [the comment]. I think it was about the responsibility of having a baby, I don't know, that the women had to stay home to take care of the baby, something like that, but [it] sounded very... bad, from my way of thinking.

Multiple participants told us that this comment was a significant event in Group 5, triggering a debate. Karina and Lizbeth each spoke about how female participants challenged the young man's comment. They also said they observed a gradual shift in the types of comments made by this young man afterwards. For example, several months after the intervention ended, Lizbeth said she thought he was 'no longer *machista* [male chauvinist]' and had stopped making such comments.

Tania (the Group 5 facilitator) described this event, noting that young women spoke out and also saying she perceived changes in this young man over time.

In the first session [...] about unplanned pregnancy [...] [one participant] made it be known that he had no problem with unsafe abortion [...]. He said 'well, she should die' [...]. And so, the girls got angry. [...] And he said it again, like, very seriously [...]. And from there I confronted him, and [...] we started to question what he had said. And at the end of the session [...] it seems he realised that his comment was really aggressive. [...] Since then, he started to have more measured participation, also calmer [...]. This change in, um, attitude and thoughts doesn't happen in one hour, but [...] he started with an extreme comment and then after that his comments became very, very mellow. And, like, throughout the sessions [...] he ended up being one of those who participated the most, like when [the conversation] required reflection.

This provides an example of how female participants and the facilitator responded to an aggressive and gendered comment about abortion and appeared to disrupt a pattern of behaviour – in this case, gender-discriminatory commentary – to potentially shift the rules in the group about the acceptability of this type of comment.

In Group 7, the facilitator (Regina) told us about two men in the group repeatedly making aggressive comments that reinforced harmful gender norms, such as blaming women for not getting contraception, wearing short skirts, or having too many boyfriends. She said that despite her attempts to address the situation the men continued to make such comments; over time women appeared to participate less when these men were present.

A second example of a circumstance that influenced the collective experience in the intervention had to do with the willingness of participants to openly share their personal experiences with the group. Some participants noted that gay or bisexual participants in their group spoke openly about their experiences during the intervention sessions. For example, Julián (Group 4) said he had never had a gay friend before and that through the course he initiated conversations with an openly gay classmate to ask questions about sexuality. He also said the course helped participants learn about and accept sexual diversity, ‘to open their minds, you could say.’ Israel, also in Group 4, identified himself as gay. He described these conversations from his own perspective, describing a male classmate who asked him many interesting questions about the lesbian, gay and bisexual community. Israel said that many of his classmates seemed genuinely interested in the topic of sexual diversity. It seems that the intervention methodology created a space in which participants could express themselves freely about sexual orientation and influence the beliefs and social norms in the group around sexual diversity.

A third unpredictable event that influenced the collective experience was a magnitude 7.1 earthquake that hit central Mexico on September 19, 2017. This earthquake took place a few weeks after the intervention began, during concurrent sessions of the intervention attended by Groups 3 and 4. Participants, facilitators and teachers reported that the event was traumatic for many of them. Afterwards, Mexico City schools were closed until the buildings could be inspected. Participants, facilitators, and teachers did not know when the school would reopen. Ultimately, the semester restarted after three weeks and the intervention continued in a slightly condensed form. The Group 3 facilitator (Patricia) left her position at Mexfam after the earthquake and was

replaced by Orlando. Orlando and some participants told us that it was difficult for the group to readjust to a new facilitator in the middle of the course. We do not know the direct influence of the earthquake on intervention outcomes, but the facilitators told us that after weeks of school closures they had to rebuild group trust and address the emotions and trauma related to the earthquake. This further delayed implementation, reducing the hours available for the curriculum.

Context dependency

A complex adaptive systems approach expects context dependency, meaning that different participants (sub-systems) and intervention groups (systems) will respond differently to the intervention based on their initial conditions and context. We found a striking example of this, in which there were different responses to IPV-focused activities in two intervention groups that had the same facilitator. According to this facilitator, Berenice, the young men in Group 4 were unwilling to engage in the topic of IPV.

I noted that the girls were much more interested than the boys [...]. The boys [...] told me, 'I'm not interested at all in violence' [...]. And the girls were the ones who went and started to debate with the boys about why it was important for them to learn about this, no? And the boys said 'it's that really, I'm not interested, and you can tell me a thousand and one arguments, and no.' And in the last sessions I noticed more attendance by the girls than the boys.

Group 6 experienced the same curriculum as Group 4, also facilitated by Berenice, yet responded differently to the IPV-related activities. In Group 6, an influential young man was enthusiastic about the intervention and seemed to encourage classmates to engage actively, particularly on the topic of IPV. Berenice said:

The head of the group went on vacation, and he is a born leader, because when he went on vacation many of them stopped coming. And, when [he] came back, they came back, no? [...]. And when we started to talk about violence, the group leader was the one who began to participate the most, the one who showed most interest, and [following his lead], all the rest seemed much more interested in what we had to say. There was a moment when I didn't even talk. That is, I was just listening, and they were sharing. And, well, it was a circle of trust, they all started to share their ideas.

Another example of a possible gender difference in responses to the course was mentioned by the Group 8 facilitator, Orlando, who told us that women took the course more seriously than men. Similarly, Lizbeth, a participant in Group 5, told us that the women in her group took the course more seriously than the men:

The men take these things as a joke, that's why sometimes there are unwanted pregnancies because they don't take it seriously. [Instead of saying,] 'well, if my girlfriend doesn't know something, I'll pay attention so I can talk about it [with her],' it's like 'Hahaha, condoms, oh hahaha, this and that.' [...] I feel like the girls are taking it more seriously.

These narratives may indicate a gendered difference in how participants respond to the course, depending on the dominant gender norms and other power dynamics within each of the groups.

System-disruptive elements of the intervention

The comprehensive sexuality education programme being evaluated was designed with a gender-transformative approach, seeking to reshape pervasive gender norms within a system. Therefore, we examined the focus group and repeat interview data to assess whether there was evidence that the intervention disrupted norms within the intervention group or influenced gender- and IPV-related beliefs among participants.

Multiple participants described their perception that the intervention introduced new ideas that countered the status quo in terms of common beliefs and norms about relationships and IPV. Key intervention messages mentioned by participants were: (1) respecting oneself and one's own needs in a relationship, (2) how to behave in a relationship and the many types of IPV that can occur and escalate, and (3) the value of expressing yourself freely and accepting diversity, particularly regarding gender and sexuality.

The first core message about self-respect was mentioned by several women; none of the men directly mentioned this message, though several did mention learning to accept or respect their own preferences and needs in a relationship. The concept of self-respect seems to have influenced perceptions of how one should behave and expect to be treated in a relationship. Reflecting on this message, Lizbeth (Group 5) told us:

The most important part was this, to love myself. And that if I say something, it will be respected, because I'm saying it [...]. Maybe others don't respect it, but I myself will respect my own decision. [...] If someone else don't respect me, it's enough that I respect my decision.

Lizbeth said the course taught her to stand by her own decisions, changed her expectations of how her boyfriend should treat her, and helped her behave differently within that relationship. For example, she said:

It used to be like, 'it doesn't matter if you go out with another girl, we'll have sex anyway.' And now it is like, 'you want to be with someone else? Then leave.' [...] My character has become more... I do something because I said it.

The second core message, about the types of behaviour that are considered acceptable in a relationship, seems to have shifted perceptions among male and female participants. For example, Israel (Group 4) said that the course helped 'differentiate between what [type of behaviour] is love and what is not love.' In another example, Laura (Group 1) contrasted what she learned about relationships in the course to the beliefs and norms in her family, demonstrating the potentially disruptive influence of the course on her own understanding of acceptable behaviour in relationships despite what she learned at home.

In my mother's family, violence is that someone hits you. But they don't know that violence can also be that someone insults you, doesn't give you money [...]. Because I'm sure that none of my aunts, nor my grandmother, are aware that violence is to pinch them, to push them, to be verbally insulted, to not receive money to buy food or to survive, or that type of thing. That is what I liked the most [about the course].

The third core message described by participants was the importance of respecting diversity and expressing yourself freely, particularly regarding gender and sexuality. This was mentioned particularly by participants who identified as gay or bisexual. Gerardo (Group 6) said the course taught 'that you don't need to comply with stereotypes, the labels that people place on you. Like, if you are a man you have to like women and if you are a woman you have to like men. So, well, I feel that it [the course] gives you a basis to respect people's rights.' Israel (Group 4) similarly talked about the message that people should act freely and be themselves. He described an activity that caught his attention in the first intervention session, where the facilitator (Berenice) began to ask why the school uniform didn't include trousers for women or skirts for men, and why men shouldn't wear pink and women shouldn't wear blue.

I understood that the objective of the class was that we can express ourselves just as we are and don't need to be guided by what people say about 'boys should play with that, and girls should play that'.

Gilberto and Laura (Group 1) each talked about the process of learning to respect differences in opinion and debate respectfully. For example, Gilberto described a debate about social norms in the group:

We were all debating our points of view, arguing amongst ourselves but respecting each other. So, we always were able to understand, and always respected everyone, and more than anything, sharing our opinion. And at best these did differ [...] and we made comparisons about social stereotypes that we have from childhood, the strong social stereotypes about how to be a woman, how to be a man.

These core messages, which were part of a gender-transformative approach, positioned the intervention to influence and shift the gendered social norms and related beliefs that dictate how people should behave towards others, engage in relationships, and otherwise express themselves.

Discussion

The conceptualisation of an intervention as a disruptive event in a complex system (Moore et al., 2019) may be particularly well-suited for evaluating gender-transformative programmes. Such interventions aim to reshape – or disrupt – gender norms within local contexts. Applying the lens of system disruption to qualitative evaluation data helped us identify the key aspects of a comprehensive sexuality education programme in Mexico that appeared to disrupt the formal and informal rules that govern individual and group behaviours and beliefs, particularly those related to gender, relationships and IPV. The specific messages in Mexfam’s intervention about self-respect, how to treat others, and acceptance of diversity may have appeal beyond this particular context because they are broadly applicable regardless of location, gender, sexual orientation or relationship status. These messages seek to disrupt the rules of interaction in the system (the implementation group), including the gender system in that group – that is, what is expected of women and of men (Hirdman, 1991; Heise *et al.*, 2019). Because gender norms uphold and underlie gender systems (Heise *et al.*, 2019), gender-transformative programmes aiming to shift these norms are ultimately attempting to reshape the gender system in a community to be more equitable. By employing a complex systems approach, we were able to connect examples of individual-level changes or actions in the group setting to potential system-level shifts in the intervention group – as in the case of female participants speaking out against a classmate’s gender-discriminatory comment about abortion to influence what was considered appropriate to say in the group.

Our analysis shows that without examining both individual and group processes, evaluations may miss key insights about how the intervention generates change, for whom, and how it interacts with its context. Complementing the individual analyses so common in evaluation practice with a

complex adaptive systems approach can help ‘see the wood as well as the trees’ when conducting public health evaluation, as advocated by Smith and Petticrew (Smith & Petticrew, 2010). This combination of approaches allowed us to better understand the variability we observed in implementation environments, actors and outcomes.

A systems approach to evaluation acknowledges that social interventions are unpredictable and emergent. As noted in the UK Medical Research Council guidance for evaluating complex interventions, fidelity in implementing these types of interventions ‘is not straightforward’ (Craig, Dieppe, Macintyre, Michie, *et al.*, 2008). Therefore, it may not be reasonable to expect standardised implementation processes and consistent outcomes. In our study, we identified two similar groups of students who received the same intervention from the same facilitator, but their responses varied dramatically. Similarly, a vignette-based activity was implemented in all implementation groups, but only in one group did it seem to trigger a heated debate about gender norms and eventual shift in the conduct of some participants. These examples highlight the potential for variable responses to an intervention, even when implemented concurrently and in the same setting. In addition to the ways in which social interactions can lead to variation, a range of external factors can also contribute to variability in implementation. In our study, an earthquake impeded our ability to implement the intervention as planned. Such experiences raise questions about whether fidelity is a relevant concept (Hawe, Shiell and Riley, 2004, 2009) and challenge the concept of faithful implementation when implementing social interventions embedded in complex systems – which are context dependent, behave unpredictably and have emergent results that develop over time.

A complex adaptive systems approach to analysis helped us assess how unanticipated events such as the earthquake or particularly aggressive group dynamics seemed to influence individual participants as well as collective experiences. The experience in Mexico suggests that with well-trained and supported facilitators and a flexible curriculum, implementing teams can cope with unpredictable events and suboptimal implementation environments to constructively address implementation challenges and contribute to IPV prevention efforts. Flexibility to respond to challenges and adapt to local needs is relevant when implementing comprehensive sexuality education; a study of these programmes in four low- and middle-income countries highlighted the need for ‘mechanisms for feedback on implementation hurdles’ and adaptation of curricula for different contexts (Keogh *et al.*, 2018).

A systems analysis can help organisations prepare for different types of unpredictable occurrences and outcomes that may emerge during an intervention (Peters, 2014), in part by putting a spotlight on some of the most challenging aspects of implementation. For example, the intervention in Mexico was promoting equity, respect, and non-violent relationships while also, unintentionally and unwillingly, serving as a space in which bullying and aggression were perpetrated. Bullying is a common form of school violence and can entail physical, psychological or sexual forms of aggression (United Nations Educational Scientific and Cultural Organization, 2019). This highlights a tension inherent in implementing school-based violence-prevention interventions: because of the ongoing violence that can permeate schools, it may not be possible to implement school-based interventions in truly violence-free spaces. This raises questions about how interventions seeking to prevent IPV or other forms of violence can reduce the potential for harm and avert normalised forms of interpersonal violence – a concern also noted in a study of an IPV prevention programme in South Africa (Hatcher *et al.*, 2020). As good practice, training in classroom management and conflict resolution should equip facilitators to address bullying, aggressive behaviour, and other forms of harmful group dynamics. In the intervention presented in this paper, facilitators were trained in these topics and did respond directly to aggressive behaviour in the course. Despite this, interpersonal conflicts had a negative influence on the experiences of other participants. Although Mexfam routinely trains its facilitators in classroom management, based on this experience the organisation is developing a set of tools to systematically support their staff in addressing any harmful or aggressive behaviour that occurs during intervention implementation. Similar tools may be useful to support facilitators of school-based programmes addressing a range of types of violence in different contexts.

Despite this preparation and the programme's gender-transformative approach, and although the evaluation results suggested that the intervention in Mexico had positive effects on both young men and young women (Chapter 4), we observed some differences in responses to the intervention by gender. Overall, women appeared to take the course more seriously than men, though many men did engage actively during the course – in particular in one group with a socially influential male participant who encouraged classmates to participate. Several facilitators encountered varying degrees of resistance to the course among male participants, especially related to the topic of IPV. These findings have programmatic implications, highlighting the importance of working with both men and women while also identifying further strategies that effectively engage men in IPV prevention programming – also noted by other researchers (Peacock

and Barker, 2014; Dworkin and Barker, 2019). Other IPV prevention programmes have similarly found that male participants sometimes rebelled against or resisted intervention messages (McGeeney, 2015; Pierotti, Lake and Lewis, 2018). A study in South Africa described ways in which male facilitators of an IPV prevention intervention negotiated and engaged their own masculinities – sometimes limiting the potential of the programme to contribute to transformational change (Gibbs *et al.*, 2019). These examples are in line with the finding from a recent systematic review that gender-based violence interventions appear more effective at reducing exposure to IPV among young women than at reducing exposure or perpetration among young men (Meinck *et al.*, 2019).

This study had a number of limitations. First, the study was not designed with a complex adaptive systems perspective specifically in mind, and the analysis was conducted post hoc to complement individual-level analyses published elsewhere. However, the study design did include methodological exploration to take into account dynamic and contextual aspects of the intervention, which inspired the complex adaptive systems approach presented here. Second, the few programme activities that might influence school-level change, such as teacher training and a health fair, were implemented after the study ended to avoid contamination of the evaluation. Changes at the school or school system level would be relevant to examine from a systems perspective in future studies. Third, we included in this paper only a subset of complexity concepts, chosen for their relevance to these particular data. Because this paper aimed to explore the relevance and potential added value of a complex adaptive systems approach to evaluation, our intention was not to conduct a systematic assessment of all of the data from the evaluation in Mexico using all possible complexity-related concepts. In addition, complexity theory and complex adaptive systems are inconsistently defined and applied across studies (Walton, 2014; Thompson *et al.*, 2016), exacerbating challenges in defining an appropriate set of concepts that should be included in such an analysis. Future evaluation studies could contribute to refining the set of complexity concepts most relevant to intervention evaluation and consider adopting a systems approach to complement other analyses. This may be particularly relevant when evaluating gender-transformative programming and interventions that address complex social issues such as IPV. Finally, it would be interesting to engage a complexity approach to examine how system-level changes influenced by the intervention are sustained – or not – over time, but we do not have the long-term data needed to conduct this analysis.

Conclusion

An intervention is a composite of multiple dynamic and interacting elements beyond the (somewhat static) curriculum, including each participant's background and experiences, facilitator characteristics, group dynamics, and environmental or contextual factors – all of which are unlikely to be consistent across implementation settings. Applying complexity concepts to our analysis added value by helping us reframe implementation-related data to focus on how the 'social' aspects of complexity, particularly interactions among participants and facilitators, influenced the intervention. A complex adaptive systems approach also sheds light on some of the variation in experiences and outcomes across individuals and groups. A system-level focus is a useful complement to individual-level analyses, which may fall short when examining complex and norms-based outcomes such as IPV. A social complex adaptive systems approach is well-suited to the evaluation of gender-transformative interventions, and can help identify how such interventions disrupt the complex social systems in which they are implemented to address intractable societal problems.

Chapter 7: Collecting better data on sexuality, relationships and violence in schools: empirical evidence from an evaluation

7.1 Introduction

In the previous chapter, I described various interpersonal interactions and unpredictable occurrences that disrupted a comprehensive sexuality education intervention with a violence prevention focus in Mexico City. I examined the influence of such emergent and unpredictable events using a complex adaptive systems approach. In this chapter, I build on this exploration of complexity by considering which evaluation methods are best suited for evaluation studies that address complex concepts, such as relationships, sexuality and violence, in dynamic contexts like schools.

Here, I will explore the influence of the implementation conditions and data collection challenges experienced in practice in the evaluation in Mexico. I will compare data from self-administered questionnaires and one-on-one in-depth interviews and report on the quality of the data. Based on this empirical example, I will reflect on the challenges and limitations when collecting sensitive information from young people about their lived experiences as part of programme evaluations. This analysis contributes to the methodological focus of the thesis. Specifically, I explore appropriate and well-suited approaches for the evaluation of social and complex interventions by critiquing, and considering complementary methods for collecting data about shifting complex concepts related to sexuality, relationships and violence in dynamic social settings.

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1400132	Title	Ms
First Name(s)	Shelly		
Surname/Family Name	Makleff		
Thesis Title	Methodology for evaluating complex social interventions: understanding partner violence prevention for young people in Mexico City		
Primary Supervisor	Prof Cicely Marston		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Social Science and Medicine
Please list the paper's authors in the intended authorship order:	Shelly Makleff, Jovita Garduño, Vanessa Ivon Silva Marquez, Susana Medina, Florencia Barindelli, Cicely Marston

Stage of publication	Not yet submitted
----------------------	-------------------

SECTION D – Multi-authored work

<p>For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)</p>	<p>I am the first author of this paper, and was responsible for study design, conception of the paper, data analysis, and writing, and also participated in data collection. Cicely Marston supported on study design, conception of the paper, analysis and writing. Because of the co-produced nature of the evaluation case study, key members of the partner organizations are co-authors based on their involvement in study design, data collection, or interpretation of findings, as well as critical revisions of the paper.</p>
---	---

SECTION E

Student Signature	
Date	29 / 5 / 2020

Supervisor Signature	
Date	29/5/20

7.2 Abstract

Data quality issues such as inconsistencies and missing data are common when collecting information about young people's sexual health. Improvements are needed to ensure high quality data to inform effective programming for young people's sexual health and relationships. This paper provides an illustrative example of the challenges to collecting school-based data about sexuality and relationships as part of an evaluation, and exploring how these affect data quality for different methods.

We use empirical data from a school-based programme evaluation in Mexico City as a case study. The evaluation sought data on sexuality and relationships using self-administered questionnaires, observation, focus groups, and in-depth interviews. We explore the advantages and disadvantages of different data collection methods in practice, comparing data from questionnaires and interviews to examine data quality and identify challenges to collecting and integrating methods from different sources. We conduct descriptive statistics, review field notes, and conduct thematic analysis, drawing on research team and participant perspectives.

Data collection was influenced by the dynamic study environment, including earthquakes and extreme weather, social aspects of the data collection setting, and the sensitive nature of the data we collected. Most questionnaire responses were complete, with variability by group, timepoint, gender, and topic. Women had a higher proportion of complete responses than men at most timepoints. Intervention group men were the only group to increase their proportion of complete responses from baseline to endline. Items about sexual activity had the lowest proportions of complete responses. Participants varied in their willingness to share personal information through different data collection methods. Data from different sources and timepoints sometimes appeared contradictory, creating challenges with interpretation.

Our experience highlights the difficulties of anticipating or correcting for the many interacting real-world challenges of school-based sexual health research. It is vital that results for all types of studies – quantitative and qualitative – are accompanied by a reflexive discussion of data collection conditions and challenges that might impact on data quality. This allows for an assessment of data quality before using potentially meaningless results to inform decision-making.

7.3 Research Paper 4

Introduction

Missing and inconsistent data are common problems when researching young people's sexual health (Young *et al.*, 2016). Complete and high-quality data are needed to inform effective programming. Barriers to collecting accurate and meaningful sexual health information from young people include the limitations of self-reporting on sensitive behaviours (Fenton *et al.*, 2001; McCallum and Peterson, 2012) and the challenges of collecting data in schools, where many health interventions for young people are implemented (Aarons *et al.*, 2001; Aventin *et al.*, 2016). Improvements are needed to ensure high quality data for the evaluation of programmes addressing young people's sexual health and relationships. To inform improvements, we examine different techniques for collecting school-based data about young people's sexuality, relationships and partner violence in the context of a range of data collection challenges, using empirical data from a school-based evaluation in Mexico.

Many adolescent sexual health studies are conducted in schools, which provide efficient access to a large population of young people – particularly in countries with high levels of school attendance (Testa and Coleman, 2006). Schools can provide a relatively controlled environment in which to collect data, and school-based surveys typically have high response rates (Testa and Coleman, 2006). However, schools are complex systems operating in dynamic contexts (Hawe, Shiell and Riley, 2009; Keshavarz *et al.*, 2010), bringing a range of data collection and recruitment challenges. Weather events, earthquakes, school closures, students reassigned to different classrooms, or changes in school staffing can hinder data collection (Aarons *et al.*, 2001; Wong *et al.*, 2014; Naveed *et al.*, 2017; Baytiyeh, 2018). Parents or staff may resist sexual health programming (Chandra-Mouli *et al.*, 2018; Keogh *et al.*, 2018) or research on sexuality or violence, making it difficult to recruit schools at all (Aventin *et al.*, 2016). Even once schools agree to participate, data quality can be compromised by school dropout rates (Sosa-Rubi *et al.*, 2016), absenteeism, scheduling issues, classroom management challenges (Aarons *et al.*, 2001; Aventin *et al.*, 2016), a lack of private spaces for data collection (Milnes and Kendal, 2012), and difficulties retaining participants over time (Henderson *et al.*, 2010). Researchers have also observed that students may be uncomfortable answering sexual health questions in the presence of their teachers (Aarons *et al.*, 2001), or may opt out of participating in such studies altogether (Henderson *et al.*, 2010; Aventin *et al.*, 2016).

Regardless of the setting, gathering data about young people's sexual practices and relationships is influenced by the sensitive nature of adolescent sexuality (Mensch, Hewett and Erulkar, 2003). Young people may find questions about sex inappropriate or invasive (Young *et al.*, 2016). This may reflect social norms that stigmatise adolescent sexuality, especially for young women (Bhugra *et al.*, 2007; Mejía *et al.*, 2010; Smith *et al.*, 2016), and can influence participants' willingness to disclose their own sexual health experiences (Makleff *et al.*, 2019). Self-reported sexual health data can be unreliable for a range of reasons including lack of recall or the desire of participants to respond in socially acceptable ways (Fenton *et al.*, 2001; Ghanem *et al.*, 2005; McCallum and Peterson, 2012). Different strategies have been used in attempts to correct for such threats to the quality of self-reported sexual health data. However, studies examining the effects of the mode of data collection, question wording, incentives, fieldworker training, and matching interviewers and interviewees by gender have been inconclusive or shown only marginal improvements (Catania *et al.*, 1996; Collins *et al.*, 2000; McCallum and Peterson, 2012).

Mixed methods and longitudinal approaches can help address some of the challenges to collecting meaningful and accurate data about adolescent sexual health and relationships. Because every data collection method has different strengths and weaknesses, the use of multiple complementary methods in one study can help mitigate some of the challenges to obtaining high quality data (Pluye and Hong, 2014). Mixed methods studies draw on different types of data to consider various aspects of the same concept or experience (Creamer, 2018), thereby providing insight to different dimensions of lived experiences. In addition, collecting data longitudinally can help understand shifting experiences with sexuality and relationships (Stern and Heise, 2019) – and how interventions influence these (Chapter 5). This may be particularly useful when conducting research with adolescents, who are at an age characterised by rapid shifts in beliefs and experiences related to gender, sexuality, and relationships (Price *et al.*, 2016; Blum, Mmari and Moreau, 2017). While longitudinal and mixed methods approaches can help learn about multifaceted and evolving sexual health and relationship experiences, these approaches can also create challenges in interpretation by recording potential conflicts or inconsistencies between data points (Moffatt *et al.*, 2006; Slonim-Nevo and Nevo, 2009; Hesse-Biber, 2012; Creamer, 2018).

This analysis considers different approaches to gathering data about young people's sexuality and relationships. We draw on a case study experience of a school-based evaluation in Mexico in

which we experienced challenges to data collection, including two earthquakes and extreme weather events. The aim of this analysis is to provide an illustrative example of the challenges to collecting high-quality data in practice, explore how these challenges and the study environment impacted on data collection and data quality for different data collection methods, and consider the benefits and challenges of collecting and interpreting data from different time points and methods.

Methods

We use empirical data from a programme evaluation as a case study; the evaluation was a quasi-experimental longitudinal exploration of a school-based comprehensive sexuality education (CSE) programme with a focus on intimate partner violence (IPV) prevention. It was conducted in Mexico City in 2017 and 2018 and was collaboratively designed and implemented by the International Planned Parenthood Federation/Western Hemisphere Region (IPPF/WHR), Fundación Mexicana para la Planeación Familiar (Mexfam), and the London School of Hygiene and Tropical Medicine (LSHTM).

We have described further details of the evaluation elsewhere, along with its findings and details about the intervention (Chapters 4, 5, 6). The intervention we evaluated was developed by the non-governmental organisation Mexfam and took place in one secondary school (*preparatoria*) in the south of Mexico City with students between 14 and 17 years of age. The intervention comprised 10 sessions of approximately two hours each and was facilitated by trained Mexfam staff health educators. The course used participatory methods designed to generate reflection about gendered social norms, and included sessions examining relationship dynamics and types of IPV as well as other sexual and reproductive health topics such as pregnancy prevention and sexually transmitted infections. We piloted the intervention and the data collection instruments from March to June 2017, then implemented the full study from August to December 2017 with six classroom groups of young people – three intervention classes who received the course during the study, and three comparison classes who received it after the evaluation ended.

Data collection methods for the evaluation in Mexico City comprised self-administered (pen-and-paper) baseline and endline questionnaires, observation of intervention sessions, focus group discussions with teachers, students, and health educators, and in-depth interviews with students (one-time and repeat). We selected this range of data collection methods to generate evidence

that would be useful and meaningful to different audiences, such as practitioners, donors and policy makers. We included quantitative data collection methods (questionnaires) to complement the qualitative methods (interviews, observation, focus groups) that predominate in this study, in part because qualitative data alone are sometimes perceived as less rigorous by some audiences (Popay, 1998; Giddings and Grant, 2007; Howarth *et al.*, 2016; Given, 2017).

Instrument development drew on a review of instruments that have been used for similar research (Pulerwitz, Gortmaker and DeJong, 2000; Garcia-Moreno *et al.*, 2006; Lopez, Morua and Rice, 2007; Pulerwitz and Barker, 2008; Mükoma *et al.*, 2009; Koch, Colaco and Porter, 2011; Rijdsijk *et al.*, 2011; Barker, Aguayo and Correa, 2013; Holden, Bell and Schauerhammer, 2015). We adapted and refined items drawing on focus group discussions to learn about how young people at the school talked about their relationships, sexual behaviours, and violence. We tested individual items and then piloted the full instruments.

The work was conducted in Spanish. SM and JG held information sessions about the study with participants' parents, obtained parental consent and participant assent, implemented the baseline and endline questionnaires, and observed sessions of the sexuality education course. JG recruited participants, conducted interviews, and co-facilitated focus groups with SM or with Mexfam staff. We made observations about the evaluation process, recording these in field notes (Saldaña, 2009) describing our experiences using different data collection methods and approaches.

Participants in the interviews and focus groups, which took place outside of class time, were offered a gift card for use at a discount department store. Questionnaire respondents were not offered compensation as this took place during class. All study participants were offered free health services at Mexfam. The research team was versed in local laws around disclosure of violence and followed ethical guidelines for research of this nature (Jewkes, Dartnall and Sikweyiya, 2012; Devries *et al.*, 2015). All study participants gave informed consent to participate in the study, and we obtained parental consent. The study was approved by the *Centro de Investigación Clínica Acelerada*, Mexico, and the LSHTM Research Ethics Committee, UK. We use pseudonyms to protect participant confidentiality.

Data analysis

This analysis focuses on our experiences with two data collection methods in practice. First, baseline and endline pen-and-paper survey questionnaires, which we used to describe the full study sample in terms of sociodemographic information and self-reported experiences and beliefs. Second, in-depth interviews (repeat and one-off), which we conducted with a subset of participants to learn about their lived experiences, with a focus on their dating relationships, and to consider which parts of the intervention seem to influence participants' evolving experiences, understandings and beliefs about sexuality, relationships and violence. We selected these two data sources because they were prominent in our study, encompassing the vast majority of study participants.

We used an iterative, grounded approach to defining our analysis strategy to explore the advantages and disadvantages of different data collection methods. The analysis focused on questions about participants' experiences with relationships, sexuality and violence. These are sensitive topics for which we anticipated challenges gathering accurate information. Seven such questions were asked in both the interview guide and questionnaires (1-7 in Table 12); three additional questions were asked directly only in the questionnaires, but sometimes emerged as topics in the interviews (8-10 in Table 12).

Table 12: Questionnaire items about relationships, sexuality and violence (duplicate from Chapter 3)

	Item	Response categories	Comments about item wording
1	Have you ever gone out with someone, as in a dating relationship? It could be with a boyfriend, girlfriend, friends with benefits, etc.	Yes No I would rather not say	These question were worded based on some of the most common ways that young people at the school talked about different types of relationships in focus group discussions.
2	Are you currently going out with someone, as in a dating relationship? It could be with a boyfriend, girlfriend, friends with benefits, etc.	Yes No I would rather not say	
3	Have you ever had sexual contact? For example, touching or being touched on the genitals, anal sex, vaginal sex or oral sex?	Yes No I would rather not say	This question was worded based on testing of the item with small groups of young people.
4	Have you ever experienced dating violence in your own	Yes No I would rather not say	We used this question to identify participants who had ever experienced IPV, based on their

	relationship (past or present)?		own definition at the time of the survey. This helped us identify a sample for qualitative data collection with some type of IPV history. This question is not intended to measure prevalence.
5	Do you think the majority of your female friends have already had sex?	Yes No I would rather not say	These questions were worded based on testing of the item with small groups of young people. We tested various wording options for this question, but ultimately failed to include non-binary gender identities.
6	Do you think the majority of your male friends have already had sex?	Yes No I would rather not say	
7	As far as you know, has there been any type of dating violence at your school? It could be among boyfriend/girlfriend, friends with benefits, etc.	Yes No I would rather not say	We used this question to learn about whether participants had observed this form of violence to be common in their school, based on their own definition at the time of the survey. This question is not intended to measure prevalence.
8	With whom would you say that you prefer to have a dating relationship? It could be serious, casual, friends with benefits, etc.	With someone of the same sex as me With someone of the opposite sex as me With people of both sexes I don't know I would rather not say	We tested various wording options for this question, but ultimately failed to include non-binary gender identities.
9	Have you ever felt agreed to have sexual contact with someone even though you didn't want to?	Yes No I would rather not say	This question aimed to identify any experience of sexual coercion, to allow sampling of these participants for interviews.
10	Have you ever heard of, seen or experienced sexual harassment in your school? For example, whistles, sexual advances, inappropriate looks, unwanted touching, kissing, pressure to have sex through threats, etc.	Yes No I would rather not say	We used this question to learn about whether participants had observed or experienced this form of violence to be common in their school, based on their own definition at the time of the survey. This question is not intended to measure prevalence.

The analysis proceeded iteratively based on emerging findings. We reviewed data from all sources to assess potential data quality issues, and reviewed documents (field notes and email

communications) regarding data collection challenges to attempt to identify reasons for inconsistent, incomplete or inaccurate data. We also conducted a complementary thematic analysis (Braun and Clarke, 2006) of in-depth interview data from 19 participants to gather information about their experiences with the data collection process and perceptions of challenges to data quality.

We assessed the extent to which questionnaire responses were incomplete ('prefer not to say' or left blank) and complete (any response category other than 'prefer not to say') for ten questionnaire items about relationships, sexuality and violence (Table 12), drawing on 392 questionnaires (124 baseline - intervention group, 116 baseline - comparison group, 95 endline - intervention group, and 57 endline - comparison group). We used descriptive statistics to assess the proportion of complete questionnaire responses overall, then compared this between intervention and comparison groups at baseline and endline, men and women at baseline and endline, and between questionnaire items. We used a two-sample test of proportions in STATA 14.0 to assess if the proportion of complete responses was different between groups.

For 28 participants who had at least three data sources (baseline questionnaire, endline questionnaire, at least one in-depth interview), we compared their responses for seven questions (1-7 in Table 12). We assigned each of the 196 comparisons (28 participants, 7 questions each) to one of the following categories: (1) consistent – all three data points concur; (2) partially consistent – two data points concur, but a third is left blank or 'prefer not to say'; (3) inconsistent – at least two data points do not coincide; or (4) insufficient data – only one data point available. Among these 196 comparisons, we then compared the proportions of consistent responses between men and women, using a two-sample test of proportions in STATA 14.0 to test for differences.

When interpreting the data, we took into consideration contextual information about implementation conditions and participant experiences with data collection. This helped us assess whether inconsistencies or incomplete data might reflect problems with the method in practice – for example, poorly written questions, not asking a question in the interview, or implementation challenges (Rao and Woolcock, 2003; Pluye *et al.*, 2009; Slonim-Nevo and Nevo, 2009; Creamer, 2018). We consider inconsistencies or conflicts in the data using a 'complementarity' approach to triangulation, in which findings from different data collection methods are 'not expected to be the

same, but rather to make sense in relation to each other and to help create a fuller picture of the research problem' (Nightingale, 2009). This approach acknowledges that conflicts between data points can reflect diverse aspects of the concepts being studied (Slonim-Nevo and Nevo, 2009) or different moments in evolving life experiences, rather than assuming that there was only one 'correct' answer (Moffatt *et al.*, 2006; Creamer, 2018).

Results

Data collection was influenced by the dynamic study environment, including earthquakes and extreme weather, social aspects of the data collection setting, and the sensitive nature of the data we collected. These factors influenced recruitment and participation rates in the study.

The majority of questionnaire responses were complete, however there was variability by group, timepoint, gender, and topic. Women had a higher proportion of complete responses than men at most timepoints; the intervention group had more complete responses at endline than at baseline, reflecting an increase in complete responses among men. Items about sexual activity had lower proportions of complete responses than other questions. Participants described different potential reasons for inaccurate or incomplete data during the study, including concerns disclosing sensitive sexual health information, distractions during data collection, or a lack of interest or willingness to answer questions honestly. Participants varied in their willingness to share personal information through different data collection methods. Data from different sources and timepoints sometimes appeared contradictory or inconsistent, creating challenges with interpretation.

I. Illustrative example of challenges to collecting high-quality data in practice

During the evaluation, we encountered a range of challenges to collecting high-quality data. These included: (1) delays to data collection; (2) interpersonal interactions during data collection; (3) participant concerns about privacy and confidentiality; and (4) participant recruitment and willingness to participate. We describe each of these below.

Delays to data collection

Data collection in the evaluation in Mexico was disrupted by two earthquakes, torrential rains, and flooding. These led to delays to research activities (detailed in Appendix 8). In the first weeks of

the study, several scheduled information sessions for parents and guardians were delayed or cancelled due to heavy rains and flooding, the evacuation of the school during an earthquake drill, and a day of school closure for building inspections after the earthquake on September 7, 2017. In response, baseline questionnaire implementation for two of the three comparison classrooms was delayed by a week. Recruitment for the repeat interviews was also delayed several weeks, and the interview process began later than anticipated (Chapter 5).

On September 19, 2017, the school was closed again due to another significant earthquake, pending building inspection. The school reopened more than three weeks later. As a result, baseline questionnaire implementation for the third comparison classroom was delayed by five weeks. The intervention, which was suspended on September 19, 2017, resumed more than three weeks later in a slightly condensed format to account for the lost weeks of the semester. Because of the delays, the final session took place on the last week of the school year – one week later than anticipated. The endline questionnaire, which could not be implemented until the intervention was completed, was delayed to the final days of the semester, during the examination period and immediately before the Christmas holiday – a time at which rates of absenteeism were higher than usual. Because the school holidays began immediately after the intervention ended, the in-depth interviews scheduled for the end of the semester were delayed by two months to the beginning of the following semester.

Interpersonal interactions during data collection

In all groups, we encountered challenges related to interpersonal interactions while the questionnaire was being implemented. For example, we observed that students would stand outside the classroom window and gesture to those inside to come out, causing a distraction. We also saw many participants talking to each other as they responded to the questionnaire. In some cases, they appeared to answer questions together with a friend or groups of friends, or were looking at each other's responses. Such interactions seemed more common when the questionnaire was implemented in larger groups of students. For example, due to a scheduling error, two of the three intervention classrooms – nearly 100 students – completed the baseline questionnaire concurrently in an auditorium. Students were sitting near each other, and we observed extensive interaction between participants as they completed the survey, including unruly behaviour that disrupted the whole group. In the smaller groups, we observed

interpersonal interactions during questionnaire completion, but this caused less disruption than in the larger group.

The in-depth interviews were one-on-one, and therefore less susceptible to interruption due to interpersonal interactions. Some participants asked to complete the interviews with a friend; we were unable to do so as this violated the study protocol and could have influenced the nature of the interview and what was disclosed.

Participant concerns about privacy and confidentiality

The research team observed that students appeared concerned about privacy and confidentiality during the questionnaire implementation. For example, several young women asked out loud during implementation if we were really going to keep their data safe. The presence of teachers also may have concerned participants. In two of the comparison groups, a teacher was in the room while the students were completing the questionnaire. Students appeared uncomfortable and may have thought their teacher would have access to their responses. In one classroom, we observed that the teacher tried to increase participation rates by telling students they had to complete Mexfam's questionnaire before receiving their final examination scores. Although the research team immediately contradicted this and reiterated that participation was strictly voluntary, students may have felt coerced, potentially affecting their responses.

In addition, our field notes record challenges maintaining privacy when conducting in-depth interviews. The space that the school initially provided for the interviews was an office with several glass windows. School staff, including teachers, could see into interview room, and the interviewer and interviewees were concerned that the interview could be overheard. Interviewees appeared uncomfortable sharing private information in this context. In addition, first aid supplies were stored in this space, and on several occasions the school nurse entered unexpectedly to use the space, which ended the interview early. Based on these experiences, we requested a more private space for the in-depth interviews, far away from both classrooms and staff offices. Once this was granted, there were fewer disruptions, though cleaning staff occasionally entered the room, leading the interviewer to pause the conversation for a few minutes.

Recruitment and participation rates

Despite the delays, disruptions, and privacy challenges described above, the vast majority of students agreed to complete the questionnaire. Nevertheless, in one of the comparison classes a large group refused to participate in the baseline questionnaire. In this group, our fieldnotes record how one young man had initially refused to participate; after his refusal half of the group followed suit, either withdrawing from the study or leaving the questionnaire blank.

There were lower response rates in both the intervention and comparison groups at endline compared with baseline. More than three-quarters of the intervention group completed the endline questionnaire (95 of 116 participants; 77% response rate), while under half of the comparison group completed the endline questionnaire (57 of 116 participants; 49% response rate). This may reflect differences between the two groups in endline questionnaire implementation conditions and in the relationship between the research team and participants. The intervention group completed their endline questionnaires during the final intervention session, with most of the students present. The research team had observed a number of intervention sessions, so the students in the intervention group were familiar with the team that implemented the questionnaire. In contrast, we did not have regular sessions with the comparison group throughout the semester, and therefore had to ask their teachers for permission to enter the comparison classrooms. Many of the comparison group students were absent, some left the room when we arrived, some said they did not remember completing the baseline questionnaire, and most did not remember meeting the research team at the beginning of the semester.

We were able to conduct the anticipated number of endline in-depth interviews at the beginning of the following semester. However, recruiting for and scheduling these interviews was challenging, in part because some students had changed their phone numbers or moved to a different school during the semester break. One student who changed schools came to the Mexfam offices for the final interview, as former students were no longer allowed on school grounds.

II. Assessing data quality

In this section, we consider how the challenges encountered in the evaluation may have affected data quality. We describe the extent of missing and incomplete data in the questionnaire, and consider differences in complete data by timepoint (baseline and endline), group (intervention and comparison), gender (women and men), and question topic. The analysis of complete versus

missing and incomplete responses to the questionnaire is limited to the ten items about experiences of relationships, sexuality and violence. Because of the semi-structured nature of the in-depth interview guide, we do not attempt to quantify data quality for the interviews; we will consider the nature and quality of the qualitative data in the subsequent section.

Extent of complete data in questionnaire responses

Of the 3,920 questionnaire responses in the analysis (392 questionnaires with 10 items analysed), 86% of questionnaire responses were complete, 12% were 'prefer not to say,' and 2% were left blank (not shown). While most of the responses were complete, there were differences in the proportion of complete responses by group, timepoint, gender, and topic.

In the intervention group, there was a higher proportion of complete responses at endline than baseline (93% vs. 87%; $p < .001$). In contrast, in the comparison group there is no evidence of a difference in the proportion of complete responses between baseline and endline (79% vs. 79%) [not shown]. In the intervention group, for 9 of the 10 questions there was an increase in the proportion of complete responses between baseline and endline. This reflects an increase in the proportion of complete responses among male respondents in the intervention group from 84% at baseline to 93% at endline ($p < .001$), while complete responses among women stayed nearly constant (92% vs 93%; no evidence of a difference in proportions). [Figure 5].

Women provided complete responses more frequently than men in three of the four groups and timepoints [Table 13]. Specifically, women had a significantly higher proportion of complete responses than men in the comparison group at baseline (84% vs 73%; $p < .001$), in the comparison group at endline (87% vs 70%; $p < .001$), and in the intervention group at baseline (92% vs 84%; $p < .001$). It was only in the intervention group at endline that there was no difference in the proportion of complete responses by men and by women [Table 13].

Figure 4: Proportion of complete responses to questionnaire items about sensitive topics, by group, timepoint and gender

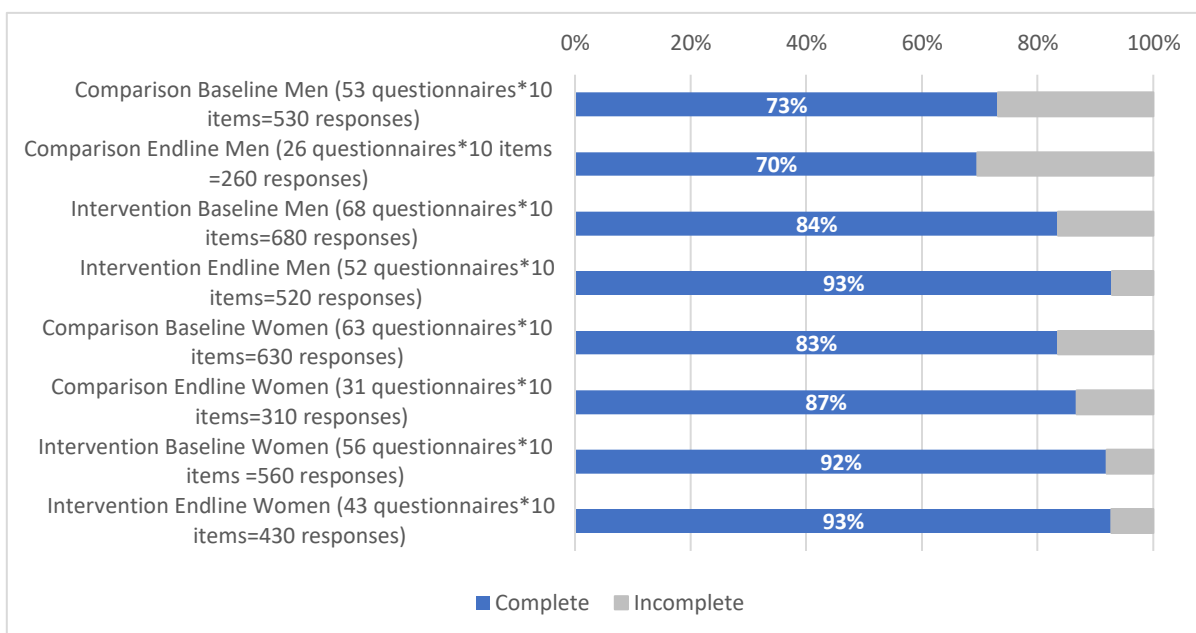


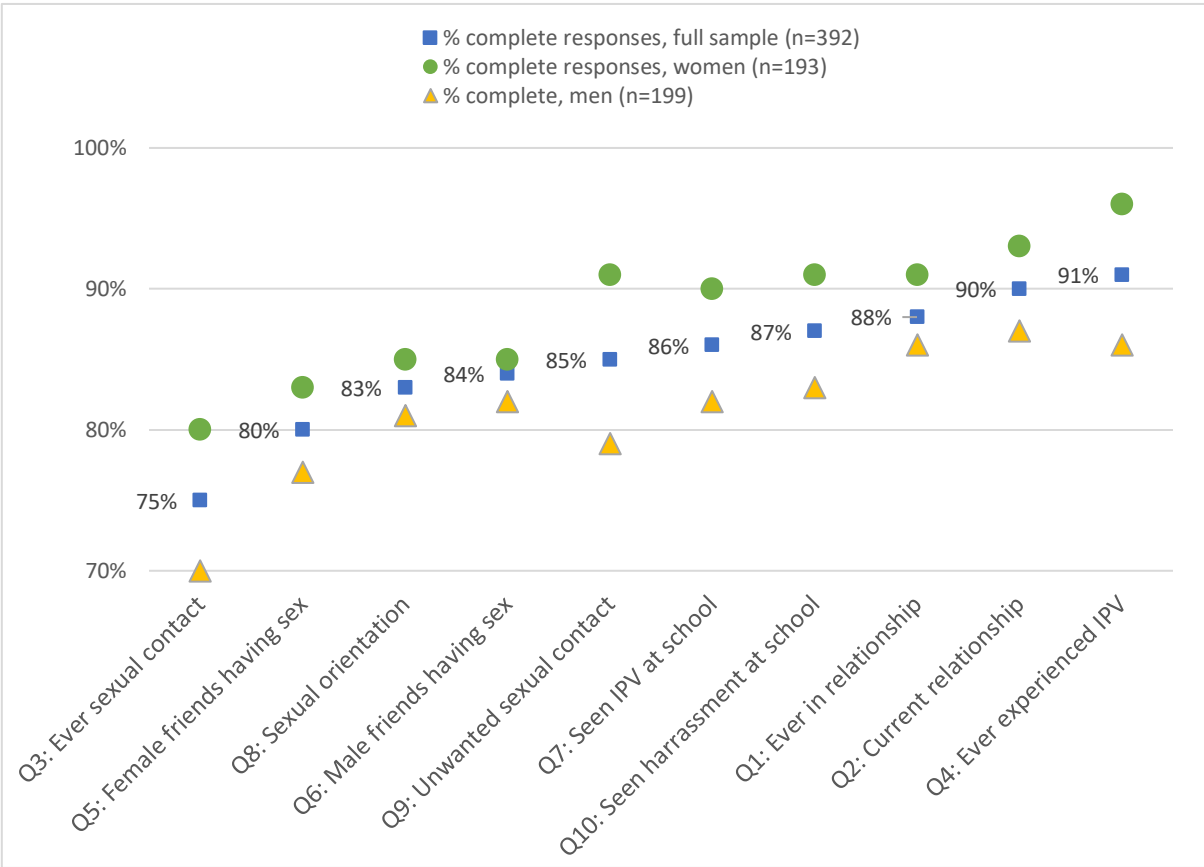
Table 13: Difference in proportion of complete responses between men and women, by group and timepoint

	Proportion of complete responses	p-value*
Comparison Baseline Men (53 questionnaires*10 items=530 responses)	73%	<.001
Comparison Baseline Women (63 questionnaires*10 items=630 responses)	84%	
Comparison Endline Men (26 questionnaires*10 items =260 responses)	70%	<.001
Comparison Endline Women (31 questionnaires*10 items=310 responses)	87%	
Intervention Baseline Men (68 questionnaires*10 items=680 responses)	84%	<.001
Intervention Baseline Women (56 questionnaires*10 items =560 responses)	92%	
Intervention Endline Men (52 questionnaires*10 items=520 responses)	93%	0.9554
Intervention Endline Women (43 questionnaires*10 items=430 responses)	93%	

*Calculated using two sample test of proportions

Questions related to sexuality and sexual activity had fewer complete responses than those about other topics. The questions with the lowest proportion of complete responses were about ever having had sexual contact (75% complete), whether female friends were having sex (80% complete), sexual orientation (83%), whether male friends were having sex (84% complete), or if they ever had unwanted sexual contact (85% complete). For every item, women had higher proportions of complete responses than men. [Figure 6]

Figure 5: Proportion of complete responses to questionnaire items about sensitive topics, by question and gender



Participant observations about data quality

We asked a subset of participants about the data collection process and how they thought their classmates responded. Overall, participants described mixed views in terms of whether their classmates answered questionnaire items honestly. Of the 19 participants who discussed this, four (Vicente, Elena, Israel, Beatriz) said they thought their classmates were generally honest in their

responses, and four (Cecilia, Alejandro, Mauricio, Judith) that their classmates were not honest. Others had more nuanced responses. Gloria said she thought participants might be honest about some questions, but not about intimate ones such as those about sexual contact. Gerardo said he was more honest when responding to the endline than the baseline questionnaire, and thought his classmates may have answered similarly. Seven of the 19 respondents (Marco, Miguel, Julián, Gabriel, Jazmin, Victor, Karina) observed that while some classmates probably answered honestly, others seemed to treat the questionnaire as a joke.

Participants described a range of potential reasons for incomplete or inaccurate responses to the questionnaire. Two participants (Julián and Elena) said that they or others wanted to answer items quickly in order to leave and socialise. Five young women (Karina, Cecilia, Jazmin, Gloria and Judith) said that participants might feel embarrassed, fear being judged, or worry about lack of confidentiality – particularly in relation to questions about sexual contact. As Jazmin said, ‘maybe [they didn’t answer accurately] because it made them embarrassed or they doubted that the information would be confidential.’ Five young men (Miguel, Vicente, Alejandro, Gerardo, Alberto) talked about finding the survey boring, not paying attention, or using it to joke around. Alberto commented about his classmates:

They didn’t like that it [the questionnaire] seems like a school test, because there were a lot of questions and I think they got bored. Once they saw there were boxes to fill out, they just did it randomly... or others answered ‘yes’ [or] ‘no’ randomly. (Alberto)

In contrast, some participants talked about reasons to answer the questionnaire accurately and completely. A few said that students may have been motivated to respond accurately to demonstrate their knowledge, show off that they were in a relationship, or because it was an important topic. As Israel said, ‘this is a topic that can personally help them, they know they can take care of themselves or benefit. I think they were sincere.’ Nine participants (Julián, Marco, Jazmin, Beatriz, Israel, Victor, Alejandro, Lizbeth) also described about the questionnaire as a tool to help build knowledge, for example reaffirming what they learned or helping them see their progress. As Victor said, ‘At first it [the baseline questionnaire] is to see what you know, right? To see the information that you have about these topics. And at the end, it is so you can see how much your knowledge improved.’

III. Comparing data gathered through interviews and questionnaires at different timepoints

The questionnaires were designed to describe the entire group of participants and help assess whether there were shifts in how participants reported their experiences and definitions of key concepts over time. The in-depth interviews were designed to provide insight to the evolving lived experiences of participants, and how the intervention might influence on their understandings and experiences. In this section, we compare self-reported data from three data sources: baseline questionnaires, endline questionnaires, and in-depth interviews.

Exploring differences in the nature of data collected in interviews and questionnaires

Willingness to share information appeared to vary by data collection method, timing and participant. Gerardo, a repeat interview participant who was recruited at the beginning of the semester, appeared willing to share personal information through both questionnaires and interviews from the beginning of the study. Gerardo self-reported that he was bisexual in both the baseline and endline questionnaires, and he also disclosed his sexual orientation voluntarily in an in-depth interview without being asked. ‘... I’m going to tell you this, I’m bisexual’. In contrast, Alberto did not share personal information in the baseline and endline questionnaires, responding ‘prefer not to say’ to many questions, but shared detailed information about his relationship in an in-depth interview two months after the semester ended.

‘I didn’t like it because she [the person I went out with] was very controlling. [...] And it’s that she told me, “without jealousy there is no love.” That was what she said. But me, I think that is the saying that bothers me the most in a relationship, because love is trust, supposedly. And well, no, jealousy doesn’t show trust.’

In some cases, data collected through different methods and at different timepoints seemed contradictory, creating challenging interpreting data – especially about concepts that the programme intervened on. For example, questions about IPV were worded so that the participant would define ‘partner violence’ based on their own understand of the concept at the time. In the baseline and ending questionnaires, Angélica reported that she had not observed IPV at school, whereas in the interview she described observing a situation of IPV and her reaction to it. ‘I saw it, yes [...]. We were in class and they were fighting, and she told him she didn't want [to be with him] anymore, and he yanked her.’ Similarly, in the baseline questionnaire Angélica self-reported no

past experiences of IPV, while at endline she responded that she had experienced it. In the interview, two months after the endline questionnaire, she shared detailed information about her experience of IPV before the intervention, in a past relationship.

‘I had a boyfriend [in secondary school] who forbade me, forbade me from dressing like I wanted and from talking to people... well, from boys. [...] He told me that without me he couldn’t live, and so... and that he was going to kill himself [if I left him].’

The differences in Angélica’s responses are difficult to interpret because the intervention aimed to shift participant understandings of the types of behaviours that comprise IPV. We do not know if her definition of IPV had shifted as a result of the intervention, which might have influenced the difference in her responses over time, or if there are other reasons for the discrepancy in her responses. For example, she may have become comfortable reporting on sensitive behaviours in the endline survey because she had grown to trust the implementation and research team over the course of the semester.

For some students, responses to questions about their own experiences varied at different timepoints of data collection, likely reflecting that that participants’ dating and relationships experiences were developing rapidly throughout the intervention. In the baseline questionnaire, Beatriz reported that she did not have a boyfriend, while in the endline questionnaire she reported that she did. In the interview that took place two months after the endline questionnaire, she said ‘Yes, yes I have a boyfriend, right now I do have a boyfriend’, and described that her relationship status had changed multiple times over the research period. Data collected through in-depth interviews allowed us to move beyond the limitations of close-ended responses to learn about the shifting and evolving experiences of participants, but also creating challenges in interpreting differences between data points.

Interview responses were often quite detailed and provided contextual information as well as illuminating the ways in which the intervention seemed to influenced participants – which the questionnaires were not designed to do. For example, Erika described how the course led her to reflect on her own relationship and whether she was experiencing violence

‘Yes, yes it [the course] made me think a lot about the relationship [...]. I don’t want to get to the point of blows. [...] Because when he, when he stops studying and I stay here at

school, he starts to be insecure, and actually a week ago, I don't remember if on Monday or Friday, he bit me and left my lip very, very swollen.'

Quantifying the consistency between data collected from different sources

Approximately half (53%) of the 196 comparisons had consistent responses from all three data sources. The remaining half (47%) had some form of missing, inconsistent, or incomplete data – 13% had two consistent responses but a third one that was missing or incomplete; 32% had inconsistent responses; and 3% had insufficient data for comparison (Table 14). There was a higher proportion of consistent responses among women than men (56% vs 48%), and among repeat interviewees than one-time interviewees (56% vs 51%); however, these differences are not statistically significant.

Table 14: Extent of consistency between questionnaire and interview data

	Consistent	Partially consistent	Inconsistent	Insufficient data	Total comparisons per row
	n (% of row total)	n (% of row total)	n (% of row total)	n (% of row total)	n (% of row total)
Women	59 (56%)	8 (8%)	37 (35%)	1 (1%)	105 (100%)
Men	44 (48%)	17 (19%)	25 (27%)	5 (5%)	91 (100%)
Total occurrences	103 (53%)	25 (13%)	62 (32%)	6 (3%)	196 (100%)

IV. Overview of reasons for missing, incomplete and inconsistent data

Data collection appeared to be influenced by diverse factors: the dynamic study environment, social setting in which questionnaire completed, timing of data collection, evolving nature of the concepts being measured, focus on sensitive topics, variation in individual interest or motivation to participate, and instrument design or mode of data collection. In Table 15, we summarise examples of each of these.

Table 15: Potential reasons for missing, incomplete and inconsistent data observed in this study

Dynamic study environment	<ul style="list-style-type: none"> • Natural hazards (earthquakes, flooding) caused school closures and schedule changes for data collection
Social setting in which questionnaire completed	<ul style="list-style-type: none"> • Participants interacted, i.e. completing questionnaire in consultation with friend(s) • Distractions during questionnaire implementation • Mass refusal to participate or opting out with classmates • Teacher in the room made participants uncomfortable • Coercion by teacher to complete questionnaire • Insufficient privacy in crowded room
Timing of data collection	<ul style="list-style-type: none"> • By endline, comparison group didn't remember research team or having completed baseline questionnaire • Seasonality of absenteeism (before holidays, examination period)
Evolving nature of concepts being measured	<ul style="list-style-type: none"> • Relationship experiences and life experiences shift over time, so data collection at different time points may elicit different responses • How participants understood and defined violence-related concepts shifted as they went through the intervention, which may have influenced how they responded to the questions over time
Sensitive topics	<ul style="list-style-type: none"> • Confidentiality concerns and fear of responses being leaked to teachers or parents – especially among women • Embarrassment or discomfort responding to sensitive questions
Variation in individual interest or motivation to participate	<ul style="list-style-type: none"> • Variability in interest to participate in the study and answer honestly; some participants took the questionnaire as a joke, others appeared to take it seriously • Women seemed more willing to respond than men, and had more complete responses • Some participants appeared particularly willing to share personal information from the beginning of study

	<ul style="list-style-type: none"> • As the intervention proceeded, intervention group participants may have become more interested in study topics, more comfortable talking about them, or more willing to complete the questionnaire
Instrument design and mode of data collection	<ul style="list-style-type: none"> • Some questionnaire items were not asked directly in interviews • Challenging to interpret ‘prefer not to say’ responses

Discussion

Reflexive contextualisation of findings based on data collection conditions and challenges is crucial to assess the utility and accuracy of data; without this, programmatic decision-making may be based on misleading or meaningless results. Our case study evaluation in Mexico illustrates the myriad challenges to collecting high quality data about relationships and sexuality from young people in schools – including earthquakes and floods, social interactions, and embarrassment. Instrument design and data analysis processes aim, with varied success, to account for a range of biases and other challenges. However, it is difficult to design away all of the barriers to high quality data, yet papers presenting study findings rarely detail these types of crucial conditions that might compromise data quality (Wells *et al.*, 2012). Reflexivity, which involves ‘continual evaluation of [...] the research process’ (Finlay, 2002), can help researchers engage with the unpredictability and ‘messiness’ of qualitative research fieldwork and data analysis (Billo and Hiemstra, 2013; Naveed *et al.*, 2017; Green and Thorogood, 2018; Sakata *et al.*, 2019) – but the concept is not usually associated with quantitative studies (Walker, Read and Priest, 2013). This may contribute to a lack of reflexive reporting of data collection conditions and challenges when authors present quantitative results. Regardless of the type of study, data collection challenges and conditions should be reported on systematically alongside results, necessitating “a cultural change [...] in which the open and honest reporting of such issues is seen as an indicator of study strength and researcher integrity” (Wells *et al.*, 2012). This can inform sensemaking and interpretation – particularly when studies address sensitive topics such as sexual health or are operating in challenging data collection environments such as schools.

Despite the challenges we encountered, we were able to adjust our research plans. In-depth interviews were adaptable, as they are conducted out of classroom hours and individually negotiated with participants. Interviews also created opportunities to ask participants about the

influence of the earthquakes on their intervention experiences, providing important context to inform our interpretation of data. In contrast, the process of adjusting questionnaire data collection proved far more challenging, as we had to negotiate for time to implement the questionnaire within a school schedule abbreviated by school closures. Ultimately, we were able to implement the endline questionnaire concurrently with school examinations and immediately before a two-months school holiday – negatively influencing data quality and participation rates for this method.

The data collection challenges in our study compromised the data quality of the questionnaires, leading us to exclude most of the quantitative data from analysis – thereby raising the relative importance of the qualitative data. Because the strength of evidence of stand-alone qualitative findings is still sometimes questioned (Giddings and Grant, 2007; Howarth *et al.*, 2016; Given, 2017), the exclusion of our quantitative results may negatively influence perceptions of the rigour of our findings. Statistical and experimental methods and metrics are often prioritised for demonstrating programme results and success in the fields of international development (Eyben, 2013) and global health (Adams, Craig and Samen, 2016) – even though qualitative methods (Popay, 1998) or a ‘qualitatively driven’ approach to mixed methods (Mason, 2006) may be better suited for answering questions about how programmes work in context.

While the school closures influenced all types of data collected in this study, we found variability in data quality by group, gender, timepoint, and topic that may not be fully explained by the earthquake-related delays. For example, the intervention group had more complete data and higher response rates to the questionnaire than the comparison group, particularly at endline. There are several possible explanations. First, the intervention group had experience discussing sensitive sexual health topics during the comprehensive sexuality education course (Chapter 4) – which the comparison group did not have at the time of the study. Second, the intervention group’s reportedly positive experience with Mexfam over the semester likely increased their willingness to complete the endline questionnaire – perhaps to show appreciation for the intervention. Third, the intervention group may have found the study topics to be relevant because of their recent experience with the programme. In line with our findings, others too have found that control group participants may respond to surveys differently because they feel resentment for not receiving a service (Bickman and Reich, 2009). Control groups may have lower response rates than intervention groups, especially as a study progresses (Oakley *et al.*, 2003). In

studies where the comparison group receives no intervention at all, it is important to consider the possibility of divergence over time between intervention and comparison groups in terms of their levels of motivation to participate in a study and respond honestly or completely to questions.

Men in this study seemed more likely than women to respond to questionnaire items at random, leave incomplete responses, or take the questionnaire as a joke. Men in the intervention group, however, were the only group of participants that significantly increased the proportion of complete responses to the questionnaire between baseline and endline – bringing the men and women in the intervention group to equal levels of complete questionnaire responses at endline. Something about male participants' experience in the intervention seems to have increased their interest or willingness to participate in the study, a difference we did not observe among women.

Women tended to describe discomfort responding to private questions and express concerns about confidentiality for both the interviews and questionnaires – concerns which were rarely brought up by men in this study. This suggests that women in our study may have been considering whether to respond honestly to questions about relationships, sexuality and violence in light of perceived risks. Other studies have similarly found that female participants are more likely to participate in research than their male counterparts (Collins *et al.*, 2000; Henderson *et al.*, 2010). This raises questions about how to account for gendered differences in data quality and response rates, and whether this phenomenon might influence data quality in others studies too.

The differences we observed in the level of complete data and in motivation to participate between groups and by gender highlight some of the challenges of using analyses that average quantitative data across participants. Averages can conceal differences in how an intervention influences different subgroups (Petticrew *et al.*, 2012) as well as variability in data quality – for example, how seriously different participants responded to the questions and whether they intended to answer honestly. Variability in data quality, such as the honesty of responses and willingness to participate, means that grouped quantitative analyses may combine accurate data with purposefully inaccurate or random responses – potentially obscuring meaningful responses with meaningless ones. As such, interpretation of group-level data frequencies may have limitations without contextual information about data quality and its variability.

Questionnaire items about sexual contact yielded the lowest overall response rates in the study, likely reflecting local social norms that stigmatise premarital sex among young people (Mejía *et al.*,

2010) and celebrate virginity and purity among women (Espinosa-Hernández, Vasilenko and Bámaca-Colbert, 2016). The research team observed that participants were more likely to respond completely and in detail to questions about sensitive topics during in-depth interviews – perhaps because these were carried out individually, in a space that ensured privacy, allowed time to build trust, and were less prone to disruption. This suggests that qualitative research has an important role to play in gathering meaningful data to inform evaluation research about sensitive topics.

We used complementary methods of data collection, and gathered information at multiple points over time, to mitigate some of the challenges of collecting data about sensitive sexual health topics. These diverse data sources and timepoints, however, also led to apparent contradictions between data sources. Apparent inconsistencies can provide insight to multifaceted and evolving lived experiences over time, and do not necessarily indicate poor quality data; they do, however, create challenges when interpreting data (Slonim-Nevo and Nevo, 2009; Creamer, 2018). Because beliefs, intentions and experiences evolve and shift over time, responses to questions about these may vary depending on when they are asked (Marston, 2005; Loxton *et al.*, 2019; Stern and Heise, 2019). Creating further measurement challenges, interventions often shift participants' beliefs and understandings of key concepts (Craig, Dieppe, Macintyre, Mitchie, *et al.*, 2008), potentially influencing their interpretation of the questions being asked, interest in responding, or perceptions about socially desirable responses. When conducting research with young people, data collection and interpretation should account for the rapidly evolving lived experiences, including short-duration relationships (Giordano *et al.*, 2010). Interpretation of data about adolescent sexuality and relationships should also take into account that what a participant reports at any given moment might provide meaningful information about their beliefs or experiences at that timepoint, but reflects one moment in a non-linear trajectory of rapidly evolving perceptions, understandings and experiences.

This study had a number of limitations. First, the design of the survey exacerbated some of the challenges with interpretation of data. The intervention was designed to influence participants' definitions or understanding of IPV, and we designed the questions to reflect a participants own definition of concepts such as IPV. As such, a change in their responses around IPV might reflect shifting definitions because of the intervention, or might be explained by actual changes in their lived experiences, such as new experiences of violence. Second, we asked questions at different points in time, increasing the likelihood that responses would differ due to changing circumstances

in participants' lives. Third, we considered but lacked the resources to use digital or computer-assisted data collection tools, which reduce inconsistencies and missing data in sexual health research in some low- and middle-income country contexts (Jaspan *et al.*, 2007; Bernabe-Ortiz *et al.*, 2008). However, digital data collection would not have remedied most of the problems we identified, such as earthquake-related school closures and the complex social context of data collection – and so the extent to which this could have improved data quality is unclear. Fourth, differences between the comparison and intervention groups in this study may compromise comparability. We used a coin toss to select between two programmes in the same school – one in the morning and another in the evening. According to school management, while the sociodemographic characteristics of students were similar, the afternoon (comparison group) has lower entrance examination scores, higher rates of absenteeism, and teachers may have less control over the classroom than in the morning programme. In addition, the earthquake-related school closures resulted in delays to data collection that particularly affected the comparison group. These factors may have exacerbated differences in data quality between the two groups – this or similar differences are likely to affect other studies featuring clustering (e.g. classroom interventions, village interventions etc). Fifth, we do not know the extent to which data collection conditions and challenges ultimately impacted on data quality. Finally, given the described challenges with data collection, the questionnaire data may not be of sufficient quality to merit any analysis at all. For our purposes, however, the flaws are useful because these types of errors and limitations exemplify the many types of weakness of school-based data collection in practice.

Conclusion

Our experience highlights the difficulties of anticipating – or correcting for – the many interacting real-world challenges of school-based sexual health research in practice. Factors that influence data quality include a dynamic study environment, interpersonal interactions during data collection, the evolving and sensitive nature of the concepts being measured, and variation in individual interest or motivation to respond to questions honestly and completely. Data quality will vary across individuals, often based on a range of factors including gender, group and question topic. Collecting complementary sources of data at various timepoints can help mitigate the weaknesses or challenges of each method, and paints a more complete picture of evolving experiences and beliefs. However, this approach also generates challenges with interpretation of data. It is vital that results for all types of studies – quantitative and qualitative – are accompanied

by a reflexive discussion of data collection conditions and challenges that might impact on data quality. This allows for an assessment of data quality and the suitability of the methodological approach to the particular research environment before using potentially meaningless results to inform decision-making.

Chapter 8: Discussion

This thesis contributes to debates about appropriate methods for the evaluation of social and complex interventions by presenting and critiquing a worked example of a real-life evaluation. Using the case of school-based IPV prevention, I explored and critiqued different approaches to evaluation. The focus is not to argue for or against any particular set of methods, but rather to build on and contribute to the growing literature exploring a broadening of acceptable methods that account for complexity and context.

In this chapter, I will first provide a brief summary of the results to orient the reader (8.1) and then elaborate on the implications of the main findings (8.2). I will present a framework comprising the different forms of complexity encountered in the evaluation in Mexico (8.3) and outline research questions that can help account for these different types of complexity in evaluation (8.4). Next, I will consider the role of the theoretical frameworks that influenced this thesis (8.5) and describe the contribution of the thesis (8.6), then outline its strengths and weaknesses (8.7). Finally, I will share concluding thoughts (8.8).

8.1 Summary of Results

I conducted an evaluation in Mexico City of a comprehensive sexuality education programme with a focus on IPV prevention, then used it as a case study to explore methods and approaches suitable for the evaluation of social and complex interventions. Here, I provide a summary of the results from Chapters 4 to 7 to orient the reader before moving to a discussion of the implications in the following section.

This thesis was organised around three main research questions. The first two questions examine pathways through which the intervention influenced participants. The third question explores different approaches to evaluation that help detect processes of change and take complexity into account, using the evaluation in Mexico as a case study. Here, I draw on Chapters 4 to 7 to provide a summary of the key findings of this thesis in relation to the research questions.

- 1. What are the mechanisms through which the intervention can promote gender-equitable and non-violent relationships among young people?**

In Chapter 4, I argue that the school-based comprehensive sexuality education intervention contributes to IPV prevention through various mechanisms. These include promoting critical reflection to influence attitudes and beliefs about gender, violence and relationships; preparing participants to communicate about IPV within their relationships; and encouraging care-seeking behaviour. I illustrate how using a range of methods and sources to gather data at multiple points in time helps build our understanding of the types of change that the intervention appears to generate and the aspects of the course that might contribute to preventing IPV.

2. How does the intervention appear to influence participants' relationship trajectories and interact with their personal context and circumstances?

In Chapter 5, I show that course activities and messages related to jealousy, possessive behaviours and the escalating forms of IPV seemed particularly influential on participants. The results indicate that participants shifted their understandings of IPV, redefined their perceptions about the types of behaviours that were appropriate in a healthy relationship, began to communicate about violence in their relationships, identified their own and their partners' harmful behaviours, and constructed more positive relationship dynamics. I argue that such evidence of shifts along prevention pathways can serve as a proxy for contributions to IPV prevention over the longer term, in particular where measurement is very difficult or when the effects of the intervention may not be detectable until after a study has ended.

3. Which methodological approaches and data collection methods are appropriate and well-suited for the evaluation of social interventions in general, and IPV prevention programming in particular? Specifically, what are the advantages, disadvantages and challenges encountered when using qualitative longitudinal methods, complex adaptive systems, and quantitative survey approaches to account for complexity when evaluating an IPV prevention intervention?

In Chapter 5, I show that repeat interviews implemented during and after an intervention are feasible and acceptable to incorporate into evaluations, and can provide meaningful empirical evidence of participants' processes of change in response to the intervention. I argue that qualitative longitudinal methods can be used to explore pathways to change, identify intervention components that contribute to change among participants, and provide highly contextualised information about participants' lives that help explain their responses to the intervention. This

approach may be particularly useful when evaluating interventions with frequent sessions or when conducting research with young people in relation to their rapidly shifting experiences, such as those involving relationships, sexuality and violence.

In Chapter 6, I argue that a social complex adaptive systems approach to analysing evaluation data helped us look beyond individual processes of change to also examine group-level processes and how these might influence system-level change. The results show that a social complex adaptive systems approach can help explore how the unpredictable and social aspects of an intervention contribute to variation in outcomes. This approach also helped identify system-disruptive messages of the gender-transformative intervention being evaluated.

In Chapter 7, I identified different factors that influenced variability in data quality and inconsistencies in the data: the dynamic study environment, including earthquakes and extreme weather; the complex social setting in which the questionnaire was implemented; the sensitive and evolving nature of sexual health and relationship experiences; data collection timelines; instrument design; and variation in motivation to participate among participants. I argue that it is crucial for all types of studies – quantitative and qualitative – to present research finding alongside a reflexive discussion of the data collection conditions and any other challenges to high quality data. This can ensure that interpretation of data takes these conditions into account, so that decision-making is informed by meaningful data.

8.2 Implications for the evaluation of social and complex intervention

This thesis raises questions about the endeavour of evaluation, specifically about which approaches to measurement make sense in different circumstances in order to build meaningful evidence about social and complex interventions. Using the case study, I identified aspects of complexity that contribute to variability in intervention experiences and outcomes: the volatile study environment – in this case, earthquakes and flooding; the evolving life experiences and circumstances of participants; the sensitive nature of the concepts being measured; and the social and dynamic nature of the school setting. I showed that these aspects of complexity create challenges for data collection about young people's sexuality, relationships and violence and explored complementary evaluation approaches that engage with temporality and change. My central argument is that using evaluation approaches that account for different facets of complexity to build a contextualised understanding of pathways to impact is a fundamental part of

evaluating interventions that aim to influence lived experiences, in particular social change interventions. Depending on the purpose and resources of an evaluation, understanding suggestive pathways to intervention impact may be sufficient, or can be used in combination with complementary approaches, including the quantification of programme effects, to provide a more complete picture of intervention effects.

What comprises 'meaningful' evidence, however, is a political question that varies by audience and context. Beliefs about what types of evidence are 'rigorous' will influence decisions about evaluation design. In considering which methods and approaches can generate meaningful evidence about interventions, this thesis draws on critiques of the hierarchy of evidence as it relates to the evaluation of complex social interventions. The findings of this thesis are in agreement with the importance of aligning study design with methodological aptness rather than the hierarchy of evidence (Petticrew and Roberts, 2003; Liket, Rey-Garcia and Maas, 2014). They are also in line with the literature advocating for a broadening of the scope of evaluation methods for the fields of international development and global health (Adams et al., 2016; AusAID Office of Development Effectiveness (ODE), 2012; Bamberger et al., 2012; Stern et al., 2012). Finally, the results support the assertion that evaluation should prioritise the needs of programme implementors, beneficiaries, and other stakeholders (Eyben, 2013; Adams, Craig and Samen, 2016; Leung *et al.*, 2019) rather than focusing on 'upwards accountability' to donors; this helps ensure that findings are relevant and useful.

In this thesis, I built on and engaged with these literatures in three ways. First, I documented the different facets of complexity that can come into play when implementing complex social interventions, demonstrating the challenges with and limitations of various data collection approaches when collecting data about relationships, sexuality and violence in a volatile and social environment. Second, I explored and critiqued a range of evaluation approaches that engage with the diverse forms of complexity that influence evaluations. Third, I provided an empirical example of an evaluation designed around research questions of relevance to implementing partners and stakeholders, resulting in locally meaningful findings that were translated into practice. I discuss each of these below.

8.2.1 Data collection challenges when examining sexuality, relationship and violence

In this thesis, I demonstrated that variability in intervention experiences and outcomes should be expected because of the complex nature and evolving context of social interventions. First, individual participants have unique intervention trajectories because of their personal history, background, beliefs, or circumstances (Chapters 4, 5 and 6). Second, social interaction, group dynamics, and disruptive behaviour are additional complexities that can impact intervention implementation and contribute to variability in outcomes between groups of participants (Chapters 6, 7). Third, data quality and evaluation processes too evolve and are influenced by the dynamic environments and social settings in which interventions are conducted (Chapter 7). Each of these is described below.

Using IPV prevention as a case study, I showed that individual participants' trajectories – in terms of their relationships, their experiences in the intervention, and their lives more broadly – are constantly evolving in a non-linear and often unpredictable fashion (Chapters 5 and 6). This poses a challenge for evaluating change over time – a measurement approach often used in evaluation – because any data point reflects an arbitrary start or endpoint for a process of change that is evolving, non-linear, and highly dependent on context and initial conditions. Accordingly, an analysis that seeks to quantify the effects of an intervention may be based on the mistaken assumption that a programme will have relatively similar effects on different participants.

It remains an important challenge to understand “what happens when a particular intervention is joined together with an individual, team, organisation or health system” (Howarth *et al.*, 2016). I explored the influence of different types of contextual influences on the study in Mexico to help explain why we should not expect that intervention experiences would be consistent between groups of participants or sites of implementation. The findings showed that different groups of participants will have diverse experiences in an intervention in relation to a range of factors, including who is in the group, who the facilitator is, any external events that influence implementation, and group dynamics (Chapter 6). Complex adaptive systems concepts can help understand these findings. *Context dependency* suggests that individuals and groups will behave differently depending on the unique context within each system (Keshavarz *et al.*, 2010), and *path-dependence* implies that different types of complex systems – including individuals and groups – are ‘sensitive to their initial conditions’ (Lindberg & Schneider, 2013). Similarly, there are challenges in implementing complex social interventions with fidelity (Craig, Dieppe, Macintyre,

Michie, *et al.*, 2008) – raising questions about the extent to which fidelity is a relevant concept for complex interventions, and whether fidelity should reflect aspects of the intervention or only their “function and process” (Hawe, Shiell and Riley, 2004, 2009).

Third, data collection and evaluation processes too evolve and are influenced by the dynamic environments and social settings in which interventions are conducted, creating barriers for measurement (Chapter 7). I illustrated the evolving and dynamic nature of data collection and other evaluation processes, which had to be adapted to unanticipated events such as earthquakes and adverse weather events (Chapters 6 and 7). All research is susceptible to disruption, as has become increasingly evident with data collection and fieldwork in the the novel coronavirus pandemic (Marhefka, Lockhart and Turner, 2020). I described some of the complexities of school-based sexual health research that can hinder data collection; similar challenges have been described elsewhere (Aarons *et al.*, 2001; Young *et al.*, 2016; Naveed *et al.*, 2017). I also showed that data quality often varied between different groups of participants (Chapter 7). This variability appears highly influenced by participant characteristics, including individual interest in or motivation to complete the questionnaire. Variability in the accuracy and completion of data means that grouped analyses may obscure data that are an adequate proxy for ‘reality’ with responses that are meaningless. It is crucial to reflexively describe data collection conditions and challenges, as well as observations about variability in data quality, when reporting on results from all types of data; this information should inform the interpretation of data in context.

I argue that evaluation methods and approaches should be chosen to engage with different facets of complexity and temporality and based on their suitability for local and dynamic implementation contexts. For some evaluations, it will be sufficient and realistic to identify suggestive pathways to change based on empirical evidence to build a highly contextualised understanding of how the intervention works, rather than focusing on quantification of intervention ‘outcomes.’ Doing so necessitates an exploration of a range of complementary approaches to evaluation practice.

8.2.2 Complementary and alternative evaluation approaches

I explored and critiqued a range of evaluation approaches: complex adaptive systems, qualitative longitudinal research to understand participant trajectory, and using different data sources and timepoints to build an understand of intervention effects (‘mixed methods’). I describe reflections on each of these approaches below.

Complex adaptive systems approach

I used a gender-transformative IPV prevention programme as a case study to explore the potential of a complex adaptive systems approach for engaging with complexity in evaluation (Chapter 6). Complex systems approaches are increasingly discussed in the literature as having potential for evaluation of social and complex interventions, but there is little guidance on doing so (Craig & Petticrew, 2013; Datta & Petticrew, 2013; Moore et al., 2019; Smith & Petticrew, 2010). The scarcity of studies adopting this approach is particularly striking with regards to the evaluation of social change interventions, which are particular well-suited to complex systems approaches (Lacayo, Obregón and Singhal, 2008).

This thesis provides a worked example of how a complex systems approach can be integrated into evaluation processes and can help identify system-disruptive elements (Moore *et al.*, 2019) of social change interventions. When an evaluation does not directly address or measure social norms, a complex adaptive systems analysis can help explore the link between observed changes in individual attitudes and shifts in group-level norms – in this case, the formal and informal rules that govern behaviour and interaction in an intervention setting. For example, I used a complex adaptive systems framework to show that core intervention messages appeared to shift perceptions of what was considered acceptable commentary in class (Chapter 6). This analysis helped us indirectly learn about social norms, for example to examine the influence of dominant perceptions of masculinities among the participants and how it may have contributed to resistance among men in the intervention to IPV prevention efforts (Chapter 6). These findings in Mexico are consistent with recent literature that argues for gender-transformative programming to include a wider range of issues – such as the prevention of homophobic bullying – that are underpinned by harmful gender norms promoting hegemonic masculinities (Brush and Miller, 2019; Dworkin and Barker, 2019).

Qualitative longitudinal evaluation

This thesis also provides a worked example of qualitative longitudinal methods as an approach to centre evaluation on different types of temporality and change. I show that repeat interviews, conducted every one to two months during and after an intervention, helped identify shifts in participants' evolving beliefs, intentions, behaviours, and relationship dynamics (Chapter 5). This approach provided insight about how personal context and circumstances influenced participant

experiences and trajectories in the intervention, and helped identify which parts of interventions contribute to shifts among participants.

One benefit of gathering highly contextualised information about intervention processes is that it helps understand which aspects of an intervention influence research participants in a particular setting. For example, I identified critical reflection about gender and violence – and related norms – as key mechanisms supporting IPV prevention among young people in a Mexican school-based setting (Chapter 4 and 5). I also showed that beliefs about jealousy and possessive behaviour in relationships appear tractable, and these behaviours, which are downstream on pathways to violence, can be targeted in IPV prevention efforts. By developing case histories that showed the development of intervention processes, I was able to provide an example of how critical reflection was experienced in practice and over time, and what aspects of the intervention may have triggered it. Such rich data provides insight into the ‘active ingredients’ of interventions, in this case the mechanisms that contribute to IPV prevention. The results of this thesis complement the growing literature from around the world that finds that critical reflection and shifting norms are important components of effective IPV prevention (Jewkes et al., 2019; Kyegombe et al., 2014; Nelson et al., 2010). Evaluation focused on pathways to impact can help understand nuanced and gradual processes of change among intervention participants in different contexts, to help identify programme mechanisms that may have relevance across settings.

Qualitatively driven mixed methods

I have shown that meaningful evaluation findings can be generated by mixing data from different data collection methods, timepoints, participants and perspectives to weave together a credible narrative about how an intervention influences participant trajectories in context (Chapters 4, 5 and 6). The findings suggest that intervention evaluation can be seen as a form of bricolage (Kincheloe, 2005), in which different data collection methods are creatively mixed in a qualitatively-driven way to learn about the lived experiences of participants in an intervention (Mason, 2006).

Using multiple data collection methods can illuminate multidimensional lived experiences and help connect the micro information gathered about individual participants with the macro context in which they are situated (Mason, 2006). However, using multiple data collection methods can also lead to challenges in interpretation. I found inconsistencies between data collected from the same

participant at different points in time and using different data collection methods (Chapter 7). Such inconsistencies can be expected both when using mixed methods of data collection (Slonim-Nevo and Nevo, 2009; Creamer, 2018) and when collecting data at different points in time – particularly when examining sensitive or taboo topics (Loxton *et al.*, 2019; Stern and Heise, 2019). Despite the challenges of interpreting inconsistencies, I found that having data from different sources for each participant, coupled with contextual information to inform the interpretation of these data, provided insight to individual experiences.

8.2.3 Co-production of evaluation

Based on the findings of this thesis, I argue there is a need to engage evaluation stakeholders when defining the types of evidence that are ‘meaningful’ and ‘rigorous’ and identifying and contextualising the nature of ‘complexity’ in their particular context. This is in line with others who have advocated for conceptions of evidence that reflect the ways of knowing, priorities and needs of local organisations and stakeholders rather than those of donors, governments, or researchers based in the Global North (Adams *et al.*, 2016; COFEM, 2018; Eyben, 2013; Leung *et al.*, 2019; Smith, 2012; Zimmerman *et al.*, 2016).

The collaborative theory of change development process in the evaluation in Mexico played a central role in co-production. First, it brought practice-based knowledge to the forefront of theory and hypothesis development. Second, it ensured that research question development was strongly grounded in local context, understandings and needs. Third, it contributed to a research design that was feasible and appropriate for the implementing organisation in the context of local complexities and dynamics.

Co-production brought additional benefits to this study. The discussions we had at the beginning of the study with Mexfam’s teams encouraged us to explore a wide range of approaches to evaluation. For example, these discussions informed our selection of the qualitative longitudinal approach to the study, which they had used in prior evaluations and found to be useful. Local research partners were involved throughout the research process, providing input on interpretation of colloquial Spanish as well as interpreting data in the context of localised understandings of the concepts being examined. In addition, having programmatic staff as part of the research created opportunities to ensure that research outputs were written accessibly and with relevance to programmatic audiences. Finally, because the research questions had been

designed to be relevant to the project partners, this facilitated translation of evidence to practice. The study outputs included policy briefs, informed updated training materials, and findings have been shared throughout the IPPF network in Latin America as well as with ministries of health and education in Mexico and other countries in the region. This scope of dissemination and transfer of evidence is greatly facilitated by the strong relationship with civil society organisations that was built through a co-production approach.

Our experience is in line with findings that co-produced research is likely to increase the utilisation of evaluation results (Jung, Harrow and Pharoah, 2012). Co-production also helped build institutional commitment from all partner organisations. After the earthquakes in Mexico, Mexfam and the school chose to continue with the study despite the challenges, likely reflecting their interest in the study findings – suggesting that the questions at the core of the evaluation remained relevant to the partners and were considered useful. Because Mexfam had an ongoing relationship with the school that served as the site for the study, reporting the research findings back to the school was treated as a priority. When we presented the findings at the school, the staff told us that none of their research partners had ever before returned to share back results – which unfortunately is not uncommon in research partnerships. Co-produced research involves building long-term relationships that help guard against the possibility of the researchers disappearing once data collection is complete.

Involving implementing partners in defining ‘rigour’ in evaluation will necessarily contribute to a shift in the hierarchy of evidence that is used for international development and global health, among other fields. From the perspective of implementing organisations and stakeholders, ‘rigour’ likely means that a method is well-suited to address the diverse facets of complexity that influence social interventions in practice, and meaningful evaluation questions are likely to result in evidence that can inform programmatic improvements and ensure that interventions are suited to their context.

8.3 Conceptualising ‘complexity’ in evaluation practice

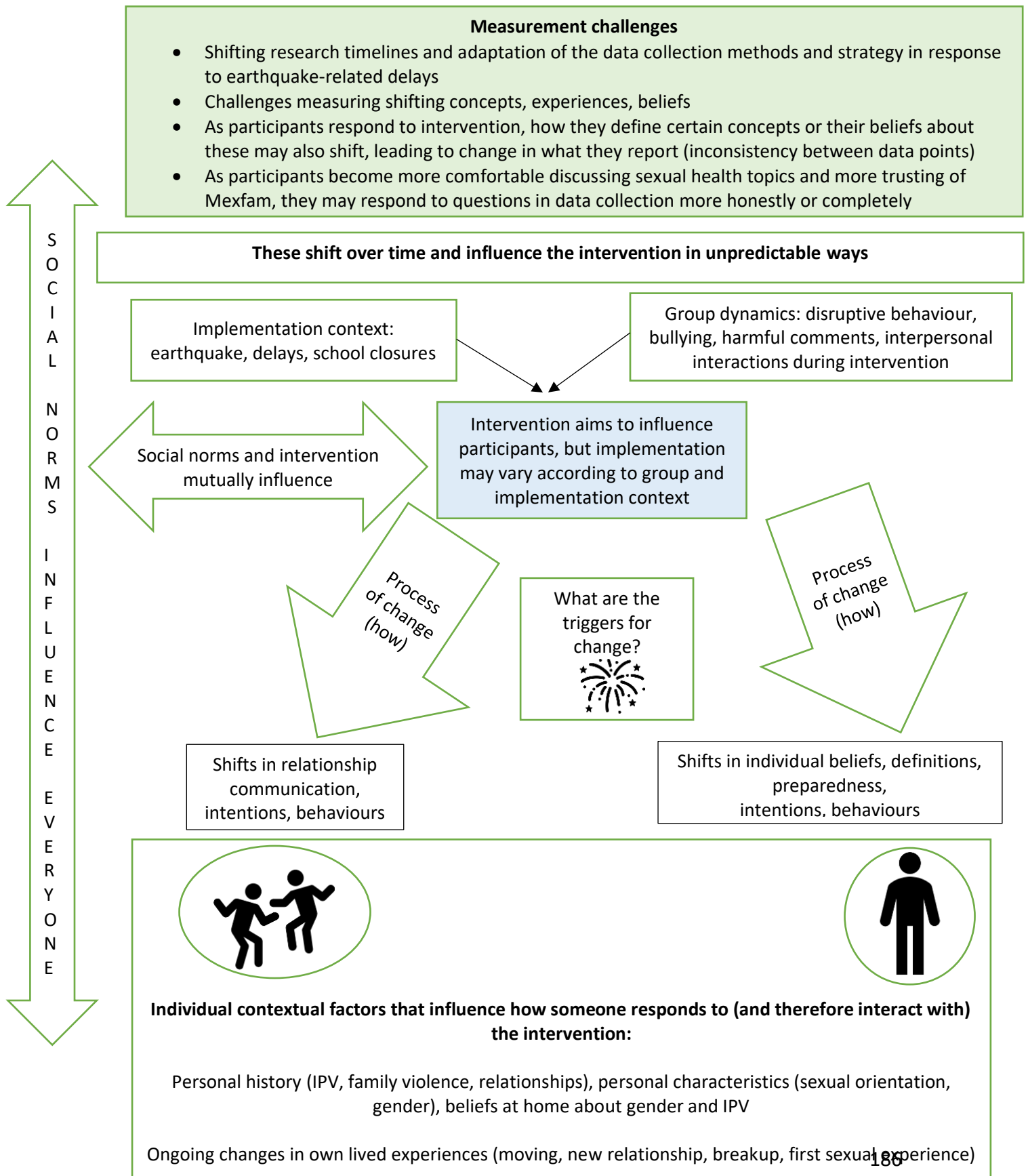
Drawing on the real-world experiences in the evaluation in Mexico, in this section I present a ‘complexity map’ that demonstrates the diverse facets of complexity we encountered (Figure 7). This map could be useful for others who are planning evaluations of social or complex interventions by providing examples of the types of complexities that could be anticipated.

Figure 7 centres on the intervention itself. The gender-transformative intervention evaluated in Mexico both aimed to influence, and is influenced by, social norms. Social norms and the social change interventions that address them are inherently complex. As shown towards the bottom of the diagram, the process of change initiated by the intervention is non-linear and can result in shifts at the individual level – such as in beliefs, definitions, preparedness, intentions and behaviours – and at the relationship level – for example, in communication and interaction between partners. The aspects of the intervention that trigger change may differ between participants, with individuals responding to the messages that seem relevant to them. Each participant’s experience in the intervention will be influenced by their own circumstances, beliefs and context – including experiences of family violence, IPV, sexuality, and their evolving experiences with relationships.

As shown in the diagram, another facet of complexity is the implementation context, which is dynamic and unpredictable. In Mexico, earthquakes and floods led to delays and school closures. In addition, social interactions, such as bullying, aggressive commentary in class, and group dynamics, influenced how the intervention was carried out and its effects on participants.

These different forms and facets of complexity contribute to measurement challenges that influence how evaluations should be conducted and analysed. As shown at the top of Figure 7, a range of complexities influenced data collection and measurement in the evaluation in Mexico City. These included the adaptation of the data collection methods and strategies in response to earthquake-related delays. In addition, it was challenging to measure the shifting concepts, experiences and beliefs related to sexuality, relationships and violence that were the focus of the intervention. As participants responded to the intervention, there were shifts in how they defined certain concepts, and their beliefs or intentions evolved in non-linear ways. This likely influenced what they reported in the study, potentially leading to apparent inconsistencies between data points at different points in time. Similarly, as participants became more comfortable discussing sexual health topics and more trusting of Mexfam, they may have begun to respond to our questions more honestly or completely.

Figure 6: Different facets of complexity encountered in the evaluation in Mexico



Each of the factors in this ‘complexity map’ draws on experiences from the evaluation in Mexico. When examined in aggregate, this diversity of intersecting factors sheds light on the many meanings of ‘complexity’ and highlights diverse aspects that evaluators may need to anticipate. Evaluation methodology should be able to engage with such complexities and take them into account.

8.4 A typology of change for engaging complexity in evaluation

One way to account for complexity is to ensure that evaluations are designed to engage with the different facets of change that comprise complexity in practice. For example, my study comprises four complementary and interrelated analyses, each centred on a different type of change. In Chapter 4, I examined the different types of change that the intervention appears to initiate among participants. In Chapter 5, I considered the use of qualitative longitudinal research as a tool to detect processes of change, such as gradual and subtle shifts over time in individual participants. In Chapter 6, I examined the unpredictable and dynamic nature of complex and social interventions and how they contribute to a variability in outcomes and experiences, both for individual participants and at the group or system level. In Chapter 7, I considered how dynamic intervention contexts and the shifting nature of participants’ experiences and understandings should influence the choice of data collection methods and evaluation approaches. Table 16 presents questions that were central to considerations of temporality and change for each chapter, and highlights what these questions contributed to the analysis.

Table 16: Core questions that engage with complexity in evaluation

Chapter	Questions that can be considered in evaluation to help focus analysis on temporality and change:	This allowed us to...
4	<ul style="list-style-type: none"> • What are the different types of changes observed in participants as they go through an intervention? • Which elements of the intervention appear to contribute to these? 	<ul style="list-style-type: none"> • Refine programme theory • Identify the range of changes we might see across a larger sample
5	<ul style="list-style-type: none"> • How does the intervention influence participant trajectories? • What gradual shifts or processes of change do participants experience as they go through the intervention? 	<ul style="list-style-type: none"> • Observe gradual and subtle shifts on individual pathway to change • Consider what may trigger these

	<ul style="list-style-type: none"> • What aspects of the intervention contributed to change? • How does this differ for different people based on their personal context and situation? 	<ul style="list-style-type: none"> • Identify ways that context affects participant trajectories
6	<ul style="list-style-type: none"> • How and through what processes does the intervention influence group-level or system-level change (disrupt the system)? • How do interpersonal dynamics and unpredictable events influence outcomes? 	<ul style="list-style-type: none"> • Identify system-disruptive aspects of an intervention • Examine variability in implementation and outcomes • Identify barriers to intervention success
7	<ul style="list-style-type: none"> • In what ways do changing data collection timelines and conditions impact on data quality? • What methods are best-suited to assess shifting beliefs and understandings among participants? 	<ul style="list-style-type: none"> • Identify challenges to data quality when using different data collection methods • Develop strategies to overcome barriers to gathering meaningful and high-quality data

By centring questions around these different types of change – which act as a proxy for diverse facets of complexity, evaluation processes can concretely focus on complexity. Building on the core questions that engage with complexity in evaluation (Table 16) and drawing from the complexity map (Figure 7), I present a typology of different changes that together comprise complexity in practice when evaluating social and complex interventions (Table 17).

Table 17: Typology of change to engage with complexity in evaluation

Type of change	Questions to consider in evaluation design	Example from this study
Change in participant understanding of concepts being measured	<ul style="list-style-type: none"> • Did the participant shift their understanding of the concepts being measured? • Was this related to what they learned in the intervention? • How does this influence their responses to questions, if at all? 	Participants shifted their conceptualisation of IPV due to the intervention (Chapters 4, 5) and may begin to understand and answer questions about IPV differently

Individual change supporting IPV prevention	<ul style="list-style-type: none"> • In what ways did the participant change their beliefs, intentions or behaviours? • Is there evidence these changes are linked to intervention? 	Participants began to intervene in IPV around them; felt prepared to seek services or support for IPV; describe intentions not to act jealous in relationships (Chapter 4, 5)
Relationship-level change supporting IPV prevention	<ul style="list-style-type: none"> • In what ways did the intervention influence the participant beliefs about acceptable behaviour in a relationship? • In what ways did the intervention influence relationship dynamics and communication patterns? 	Participants described talking to partners about controlling behaviour, improving relationship dynamics, or leaving relationships (Chapter 4, 5)
Change trajectory (individual or relationship)	<ul style="list-style-type: none"> • What was the process through which individual or relationship-level change happened? • How did the participant's context or circumstance influence their response to the intervention? 	IPV experiences in the family created stronger rejection of IPV for some participants; others perceived certain violent behaviours as 'normal' based on what they observed at home (Chapter 4, 5). These influenced how participants engaged in the intervention.
Change trigger	<ul style="list-style-type: none"> • What aspect(s) of an intervention appear to trigger or initiate individual changes or shifts in relationship dynamics? • How do these relate to the particular context or circumstances of the participant? 	We identified 'sticky' messaging: 'minor' forms of violence can escalate into physical IPV; excessive jealousy and possessive behaviour are IPV triggers or forms of IPV (Chapter 4, 5)
Change in personal circumstance	<ul style="list-style-type: none"> • Is the participant going through any life changes or new experiences that are external to the intervention but appear to influence their experience in the intervention? 	We learned about new relationships or recent sexual experiences and how those lead them to reflect on course messages (Chapter 4, 5); we saw how these changes in life experiences created apparent inconsistencies in the data (Chapter 7)

Change in group norms or dynamics	<ul style="list-style-type: none"> • In what ways did the group dynamics influence intervention experiences among participants? • How did the informal or formal rules of group interaction change in response to the intervention? 	Some participants hesitated to participate because of harmful commentary in class; in other cases, interactions encouraged sharing and learning or shifted the types of comments that were made (Chapter 6)
Change in intervention implementation	<ul style="list-style-type: none"> • Has the intervention implementation changed due to unforeseen circumstances? • How did this influence participants' experiences? 	School closures due to earthquakes interrupted the intervention, but it is difficult to assess how this may have influenced experiences in the course (Chapter 6, 7)
Change in research implementation	<ul style="list-style-type: none"> • In what ways were data collection or other research processes changed due to unforeseen circumstances? • Did this contribute to measurement challenges or compromise data quality? • Were certain data collection methods more influenced than others by these unforeseen circumstances? 	School closures delayed data collection, in particular creating challenges for endline questionnaire implementation and contributed to low quality data; in-depth interviews were more flexible to adapt to these delays (Chapter 7)

The questions reflecting each of these types of change, used together, can help examine different aspects of the complexity that may influence a particular intervention in its context. Incorporating these types of change in evaluation planning and implementation can help understand whether and how interventions work, while also taking into account dynamic contexts and social environments, the interactions between intervention and context, and the non-linear and evolving nature of lived experiences. These types of questions and analytical approaches can be used to evaluate similar (social and complex) interventions, particularly those focused on social change.

8.5 The role of theoretical frameworks in the thesis

The different theoretical frameworks that informed this thesis came together to facilitate an examination of how gender-transformative change can occur. First, social norms theory underpins the design and implementation of gender-transformative work and provided a theoretical foundation. Second, the social ecological model provided a frame through which to consider the different levels at which gender-transformative programmes may influence on participants and their environment – in this case, considering the individual, relationship, family, peer group, classroom and school-wide levels. Third, a complex adaptive systems lens helped identify the system-disruptive elements of gender-transformative interventions. A systems perspective also was useful for considering aspects of the intervention that might sustain the effects of an intervention. A refresher training, for example, might be a way to prevent the dampening of intervention effects. Finally, the concept of a gender system was useful to conceptualise the system in which the intervention was implemented – the classroom groups, and the school itself – as systems in which the intervention aimed to shift the gender order. I found these theoretical frameworks and conceptual models to be flexible, in that they permitted explorations of the evaluation data from different angles without limiting the scope of what could be explored.

When I encountered the concept of system disruption in literature about complex adaptive systems approaches (Moore et al., 2019), I began to explore the potential of this as a lens through which to consider the effects of the intervention examined in this thesis. This approach shaped what types of changes I tried to measure and explore as part of the evaluation process (Chapter 6). I conceptualised classroom groups that experienced the intervention together as a system, and was able to focus on the types of changes observed in that system. This helped focus on individual and group level changes, as well as interpersonal interactions at the dyadic and collective levels. The complex adaptive systems approach helped bring a focus to shifts in group norms in the intervention system, such as what seems to be socially acceptable behaviour or commentary in the intervention groups. These shifts, particularly those addressing gender norms, were potential indicators of changes in the gender system of the group. In these ways, a system-disruption lens to examine incremental or small shifts among groups of participants proved useful for evaluating gender-transformative programming, and in examining how such programmes can influence on the gender system within a classroom. This approach may be useful when evaluating a range of social change interventions.

8.6 Contribution of the thesis

This thesis contributes to the debate about what methods are appropriate to evaluate social and complex interventions, in particular social change interventions. It does so by presenting and critiquing a worked example of a real-life evaluation that examines complexity from different angles and using various methodological approaches.

First, the thesis contributes to the theorising of gender transformative interventions as disruptors of complex adaptive systems. This provides new analytic avenues to identify aspects of intervention that are able to influence and shift beliefs and practices related to gendered social norms, as well as the gender system in which interventions are implemented. Second, the thesis provides a framework for considering 'complexity' in practice and how it influences on evaluation. It does so by offering questions and considerations that help identify how complexity manifests and how can be considered in evaluation. Third, the thesis highlights the importance of documenting and considering not only the intervention context, but also personal context and the context of data collection, when conducting evaluation. The findings show that every participant, and each group of participants, will experience a unique version of the intervention, regardless of fidelity of the intervention, which has implications for evaluation practice and the types of outcomes we try to measure.

This thesis also helps address the gap in guidance on how to account for temporality in evaluation. The UK MRC guidance for evaluating complex interventions identifies temporality as an important aspect of interventions, but it does not provide concrete advice on how to address these in evaluation (Craig, Dieppe, Macintyre, Michie, *et al.*, 2008). Similarly, the MRC guidance for process evaluation recommends that researchers 'consider data collection at multiple time points to capture changes to the intervention over time' but does not provide guidance for doing so, nor does it identify the situations in which qualitative or quantitative longitudinal methods might be appropriate (Moore *et al.*, 2014, 2015). Chapters 5 and 6 provide empirical examples of how evaluation can consider temporality and change through longitudinal qualitative research and a social complex adaptive systems approach to engage different facets of complexity.

The findings have implications for evaluation commissioners. First, the findings add to the literature suggesting that co-production is a strategy to ensure that evaluation responds to locally relevant questions that result in useful – and used – findings. Second, the thesis highlights the

importance of involving beneficiaries and partners in defining the questions that matter and influencing how rigour is defined for an evaluation study, based on the local context and audiences. Third, the results support investments in evaluations that focus on contextualised pathways to change and methods that address complexity.

Beyond the methodological implications, this thesis also has practical implications for policy-makers and practitioners. The findings suggest that gender-transformative comprehensive sexuality education has potential for IPV prevention, and builds our understanding of the mechanisms through which it can do so. Our findings suggest that programmes should address interrelated issues such as homophobic violence, masculinities, interpersonal violence, relationship violence as a holistic approach to prevention. They also show in detail what critical reflection processes can look like and how they can influence understandings and beliefs related to IPV. The findings reinforce the importance of facilitators delivering such interventions, who must be well-trained and supported to promote reflection processes and manage complex group dynamics. The findings highlight jealousy and possessive or controlling behaviour as downstream forms of IPV that may be particularly amenable to change as part of IPV prevention interventions – as others too have found (Stern and Niyibizi, 2018). The findings also show the importance of promoting the right to services for participants and providing referrals to trusted providers and care, which may be utilised after the intervention has ended as participants' needs evolve.

The evaluation in Mexico responds to gaps in the literature about effective IPV prevention programming in low- and middle-income countries (Arango *et al.*, 2014; Ellsberg *et al.*, 2015) – particularly around prevention for adolescent girls (Blanc *et al.*, 2013), who may be at elevated risk of IPV (Abramsky *et al.*, 2011; Stöckl *et al.*, 2014). The findings of the evaluation in Mexico also engage with calls for research examining processes of change and pathways to IPV prevention (Fulu & Kerr-Wilson, 2015; Jewkes *et al.*, 2015; Kyegombe *et al.*, 2014; Michau *et al.*, 2015). Recent studies on IPV prevention – including longitudinal evaluations – have examined prevention with married or cohabitating adults in Africa and Asia (Starmann *et al.*, 2017; McGhee *et al.*, 2019; McLean, Heise and Stern, 2019; Stern and Heise, 2019). The findings in Mexico complement this body of work by focusing on prevention among young unmarried participants in Latin America. The evidence presented in Chapters 4 and 5 from the Mexican context supports of the argument that comprehensive sexuality education is a plausible IPV prevention strategy (Haberland, 2015; Holden, Bell and Schauerhammer, 2015) that merits further exploration in different settings.

8.7 Strengths and limitations of the study

This study has a number of strengths. First, it uses a real-world evaluation as a case study for a methodological exploration, using empirical data to reflect on evaluation theory and critique methods in practice. As such, the findings provide a detailed depiction of how complexity can concretely manifest in evaluation practice and consider multiple approaches to engaging with this complexity. Second, the study provides an empirical example of the application of a complex adaptive systems approach to a social change intervention. To the best of my knowledge, this approach has not been applied to IPV prevention programmes in the past. Third, the study used qualitative longitudinal research during and after an intervention, providing a worked example of this approach to learn about participants experiences in a relatively short-term (ten week) intervention. Other studies using qualitative longitudinal research for evaluation tend to gather data at longer intervals, limiting their ability to learn about and identify gradual and subtle shifts among participants. Fourth, the evaluation in Mexico was co-produced in partnership with local and regional non-government and community-based organisations, with research questions designed to meet the needs of local implementing organisations. The partner organisations have been involved in interpreting the study findings and translating these into practice. Mexfam has already used these findings to inform and improve their programming, and the three partner organisations have been jointly working to disseminate findings through policy briefs, blogs, podcasts, academic publications, and in conferences.

The study also has limitations. First, IPV prevention is only one of many types of social and complex interventions, and the findings of the methodological exploration might not be relevant to other programme topics. Because the programme being evaluated was gender-transformative in its aims, it is likely that the findings are particularly relevant to other interventions addressing social norms related to gender and sexuality. Second, some of the critiques of the methods used in this study are artefacts of the particular data collection instruments, study design, and context. All methods are situated in – and will respond to – the particularities of their context. In this study, we found that the qualitative component was easier to adapt in the context of earthquakes, flooding and school closures. Such disruptions to research are common, and it may be that quantitative methods are more challenging to adapt in more unstable contexts. Third, because of the earthquakes that disrupted the study, the quantitative data collected in the evaluation in Mexico were severely compromised. This limited my ability to analyse quantitative and qualitative

data together to create an integrated 'mixed methods' narrative about intervention effects, as originally intended. However, the effort spent on quantitative data collection was salvaged to some extent through the analysis of data collection challenges and data quality as part of the methodological exploration in the thesis (Chapter 7). Because disruptions to research due to natural hazards, weather events, and pandemics (among other examples) are common, exploring methods that are resilient to unpredictable and dynamic contexts is an important contribution to the evaluation of complex interventions. Fourth, there were many possible methodological explorations that I could have conducted. I selected the particular analyses presented in this thesis based on the particular nature of the evaluation experience in Mexico, which I used as a case study. I did not utilise all of the forms of data that I collected in Mexico for this thesis. For example, I did not examine the focus group data from the comparison group in any of these analyses due to challenges recruiting participants in the comparison group, as well as the challenges described in Chapter 7 with data collection, which also affected focus groups. Finally, this study used individual-level narratives to learn about attitudes, intentions and behaviours, but did not measure shifts in social norms at the group level. The study engaged social norms peripherally, for example by examining narratives about group dynamics and how those were influenced by the intervention. The focus group discussion guides included questions designed to learn about social norms through vignette-based activities, but due to the challenges with recruitment and implementation that affected the focus groups, these data were not analysed. Future studies working at a larger scale to explore how comprehensive sexuality education helps prevent IPV could examine shifts in social norms at the school or community level.

8.8 Conclusion

The findings of this thesis suggest that methods centred around different facets of complexity, and particularly those that engage unpredictability, interaction, temporality and change, can result in meaningful evidence that helps advance understandings of how social interventions work. There are still many unanswered questions about which methods and approaches can be used to account for complexity when evaluating various types of interventions in different contexts. Further work is needed to explore how qualitative longitudinal methods and a complex adaptive systems approach contribute to evaluations in other contexts and for different types of interventions, and to continue to experiment with a range of evaluation approaches that can take the many intersecting forms of complexity into account.

I have shown that for social complex intervention evaluation, what is considered 'meaningful' evidence will often differ from the standard hierarchy of evidence that is currently prioritised in the fields of international development and global health, among others. Consequently, future work could contribute to reshaping the definition of the term 'rigour' to include methods designed to engage with different facets of complexity that often influence the evaluation of social and complex interventions in practice. This can be achieved to some extent by changing the nature of how evaluation is designed – for example, prioritising programmatically relevant questions and engaging stakeholders and beneficiaries as partners in defining what 'meaningful' evidence entails. This may necessitate a culture change among evaluation commissioners and users, to convince them of the value of a range of evaluation approaches that engage explicitly with different facets of complexity and respond to locally relevant research questions and build on practice-based knowledge. Ultimately, a collaborative and inclusive approach to evaluation can help subvert the power differentials that often influence evaluation research and generate more meaningful and useful evaluation findings that are relevant to real-world implementation environments, and which can be used to improve practice.

References

Aarons, S. *et al.* (2001) 'Enhancing recruitment and data quality in a junior high school-based teen pregnancy prevention study', *Evaluation and Program Planning*, 24(3), pp. 277–285. doi: 10.1016/S0149-7189(01)00019-2.

Abramsky, T. *et al.* (2011) 'What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence', *BMC Public Health*. BioMed Central Ltd, 11(109), pp. 1–17. doi: 10.1186/1471-2458-11-109.

Abramsky, T. *et al.* (2014) 'Findings from the SASA! Study: A cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda', *BMC Medicine*, 12(1), p. 122. doi: 10.1186/s12916-014-0122-5.

Adam, T. *et al.* (2012) 'Evaluating health systems strengthening interventions in low-income and middle-income countries: are we asking the right questions?', *Health Policy and Planning*, 27, pp. 9–19. doi: 10.1093/heapol/czs086.

Adams, V., Craig, S. R. and Samen, A. (2016) 'Alternative accounting in maternal and infant global health', *Global Public Health*. Taylor & Francis, 11(3), pp. 276–294. doi: 10.1080/17441692.2015.1021364.

Aguilar Montes de Oca, Y. P. *et al.* (2013) 'Los roles de género de los hombres y las mujeres en el México contemporáneo', *Enseñanza e Investigación en Psicología*, 18(2), pp. 207–224. Available at: <http://www.redalyc.org/html/292/29228336001/>.

Aniekwe, C. C. *et al.* (2012) *Academic-NGO Collaboration in International Development Research: a reflection on the issues*. Available at: http://www.intrac.org/data/files/resources/750/Academic-NGO-Collaboration-in-International-Development_September-2012.pdf (Accessed: 11 November 2014).

Arango, D. J. *et al.* (2014) *Interventions to prevent and reduce violence against women and girls: a systematic review of reviews*. Washington D.C. Available at: <http://documents.worldbank.org/curated/en/2014/01/20426963/interventions-prevent-or-reduce-violence-against-women-girls-systematic-review-reviews>.

Astbury, B. and Leeuw, F. L. (2010) 'Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation', *American Journal of Evaluation*, 31(3), pp. 363–381. doi: 10.1177/1098214010371972.

AusAID Office of Development Effectiveness (ODE) (2012) *Impact Evaluation. A Discussion Paper for AusAID Practitioners*. Canberra, Australia. Available at: <https://www.gov.uk/dfid-research-outputs/impact-evaluation-a-discussion-paper-for-ausaid-practitioners>.

Aventin, Á. *et al.* (2016) 'Recruiting faith- and non-faith-based schools, adolescents and parents to a cluster randomised sexual-health trial: Experiences, challenges and lessons from the mixed-methods Jack Feasibility Trial', *Trials*. Trials, 17(1), pp. 1–13. doi: 10.1186/s13063-016-1506-y.

Bamberger, M. (2007) 'A framework for assessing the quality, conclusion validity and utility of evaluations. Experience from international development and lessons for developed countries.', in *American Evaluation Association*, pp. 1–26.

Bamberger, M., Rugh, J. and Mabry, L. (2012) *RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints*. Second. SAGE Publications.

Barker, G., Aguayo, F. and Correa, P. (2013) *Understanding Men's Violence Against Women: Findings from the IMAGES Survey in Brazil, Chile and Mexico*. Rio de Janeiro. Available at: www.promundo.org.br/en.

Barker, G., Ricardo, C. and Nascimento, M. (2007) *Engaging men and boys in changing gender-based inequity in health: Evidence from programme interventions*. Geneva. Available at: <http://www.gsdrc.org/document-library/engaging-men-and-boys-in-changing-gender-based-inequity-in-health-evidence-from-programme-interventions/>.

Baytiyeh, H. (2018) 'Online learning during post-earthquake school closures', *Disaster Prevention and Management: An International Journal*, 27(2), pp. 215–227. doi: 10.1108/DPM-07-2017-0173.

Becker, D. and Olavarrieta, C. D. (2013) 'Decriminalization of abortion in Mexico City: The effects on women's reproductive rights', *American Journal of Public Health*, 103(4), pp. 590–593. doi: 10.2105/AJPH.2012.301202.

Bernabe-Ortiz, A. *et al.* (2008) 'Handheld computers for self-administered sensitive data collection: A comparative study in Peru', *BMC Medical Informatics and Decision Making*, 8, pp. 1–7. doi: 10.1186/1472-6947-8-11.

Bhugra, D. *et al.* (2007) 'Sexual attitudes and practices in North India: A qualitative study', *Sexual and Relationship Therapy*, 22(1), pp. 83–90. doi: 10.1080/14681990600920770.

Bickman, L. and Reich, S. M. (2009) 'Randomized Controlled Trials: A Gold Standard With Feet of Clay?', in Donaldson, S. I., Christie, C. A., and Mark, M. M. (eds) *What counts as credible evidence in applied research and evaluation practice?* SAGE Publications. doi: <https://dx.doi.org/10.4135/9781412995634.d10>.

Billo, E. and Hiemstra, N. (2013) 'Mediating messiness: Expanding ideas of flexibility, reflexivity, and embodiment in fieldwork', *Gender, Place and Culture*, 20(3), pp. 313–328. doi: 10.1080/0966369X.2012.674929.

Blair, G., Imai, K. and Lyall, J. (2014) 'Comparing and combining list and endorsement experiments: Evidence from Afghanistan', *American Journal of Political Science*, 58(4), pp. 1043–1063. doi: 10.1111/ajps.12086.

Blanc, A. K. *et al.* (2013) *A review of the evidence on multi-sectoral interventions to reduce violence against adolescent girls*. Available at: <http://www.girleffect.org/media?id=3013>.

Blum, R. W. *et al.* (2019) 'Measuring young adolescent perceptions of relationships: A vignette-

based approach to exploring gender equality', *PLoS ONE*, 14(6). doi: 10.1371/journal.pone.0218863.

Blum, R. W., Mmari, K. and Moreau, C. (2017) 'It Begins at 10: How Gender Expectations Shape Early Adolescence Around the World', *Journal of Adolescent Health*. Society for Adolescent Health and Medicine, 61(4), pp. S3–S4. doi: 10.1016/j.jadohealth.2017.07.009.

Bonell, C. *et al.* (2011) 'Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions.', *Journal of Epidemiology and Community Health*, 65(7), pp. 582–7. doi: 10.1136/jech.2008.082602.

Bonell, C. *et al.* (2012) 'Realist randomised controlled trials: A new approach to evaluating complex public health interventions.', *Social Science & Medicine*, 75, pp. 2299–2306. doi: 10.1016/j.socscimed.2012.08.032.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

Braverman, M. T. (2012) 'Negotiating Measurement: Methodological and Interpersonal Considerations in the Choice and Interpretation of Instruments', *American Journal of Evaluation*, 34(1), pp. 99–114. doi: 10.1177/1098214012460565.

Breuer, E. *et al.* (2016) 'Using theory of change to design and evaluate public health interventions: A systematic review', *Implementation Science*. Implementation Science, 11(1). doi: 10.1186/s13012-016-0422-6.

Brush, L. D. and Miller, E. (2019) 'Trouble in paradigm: "Gender transformative" programming in violence prevention', *Violence Against Women*, 25(14), pp. 1635–1656. doi: 10.1177/1077801219872551.

Bryman, A. (2006a) 'Integrating quantitative and qualitative research: How is it done?', *Qualitative Research*, 6(1), pp. 97–113. doi: 10.1177/1468794106058877.

Bryman, A. (2006b) 'Paradigm peace and the implications for quality', *International Journal of Social Research Methodology*, 9(2). doi: <https://doi.org/10.1080/13645570600595280>.

Burke, H. M. C. *et al.* (2019) 'A longitudinal qualitative evaluation of an economic and social empowerment intervention to reduce girls' vulnerability to HIV in rural Mozambique', *Evaluation and Program Planning*. Elsevier, 77(July), p. 101682. doi: 10.1016/j.evalprogplan.2019.101682.

Bush, H. M. *et al.* (2019) 'Do Violence Acceptance and Bystander Actions Explain the Effects of Green Dot on Reducing Violence Perpetration in High Schools?', *Journal of Interpersonal Violence*, pp. 1–22. doi: 10.1177/0886260519888206.

Calman, L., Brunton, L. and Molassiotis, A. (2013) 'Developing longitudinal qualitative designs: Lessons learned and recommendations for health services research.', *BMC medical research methodology*, 13, p. 14. doi: 10.1186/1471-2288-13-14.

Campbell, M. (2000) 'Framework for design and evaluation of complex interventions to improve health', *BMJ*, 321(7262), pp. 694–696. doi: 10.1136/bmj.321.7262.694.

Cartwright, N. and Hardie, J. (2012) *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford University Press.

Castillo, L. G. *et al.* (2010) 'Construction and initial validation of the marianismo beliefs scale', *Counselling Psychology Quarterly*, 23(2), pp. 163–175. doi: 10.1080/09515071003776036.

Catania, J. A. *et al.* (1996) 'Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior', *Public Opinion Quarterly*, 60(3), p. 345. doi: 10.1086/297758.

Chandra-Mouli, V. *et al.* (2017) 'Implications of the Global Early Adolescent Study's formative research findings for action and for research', *Journal of Adolescent Health*. Elsevier Inc., 61(4), pp. S5–S9. doi: 10.1016/j.jadohealth.2017.07.012.

Chandra-Mouli, V. *et al.* (2018) 'Evolution and resistance to sexuality education in Mexico', *Global Health: Science and Practice*, 6(1), pp. 137–149. doi: 10.9745/GHSP-D-17-00284.

Charmaz, K. (2006) *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: SAGE Publications. Available at: <http://www.sagepub.in/textbooks/Book217586>.

Chavez, N. *et al.* (2017) 'Central Mexico earthquake kills more than 200, topples buildings', *CNN*. Available at: <https://edition.cnn.com/2017/09/19/americas/mexico-earthquake/index.html>.

Choi, B. C. K. and Pak, A. W. P. (2006) 'Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness', *Clin Invest Med*, 29(6), pp. 351–64. doi: 10.25011/cim.v31i1.3140.

Christ, T. W. (2014) 'Scientific-Based Research and Randomized Controlled Trials, the "Gold" Standard? Alternative Paradigms and Mixed Methodologies', *Qualitative Inquiry*, 20(1), pp. 72–80. doi: 10.1177/1077800413508523.

Cialdini, R. B., Kallgren, C. A. and Reno, R. R. (1991) 'A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior', *Advances in Experimental Social Psychology*, 24. doi: 10.1016/S0065-2601(08)60330-5.

Cislaghi, B. (2018) 'The story of the "now-women": Changing gender norms in rural West Africa', *Development in Practice*. Taylor & Francis, 28(2), pp. 257–268. doi: 10.1080/09614524.2018.1420139.

Cislaghi, B. and Heise, L. (2018a) 'Theory and practice of social norms interventions: Eight common pitfalls', *Globalization and Health*. Globalization and Health, 14(83), pp. 1–10. doi: 10.1186/s12992-018-0398-x.

Cislaghi, B. and Heise, L. (2018b) 'Using social norms theory for health promotion in low-income

- countries', *Health Promotion International*, pp. 1–8. doi: 10.1093/heapro/day017.
- COFEM (2018) 'TIP SHEET # 5: Feminist approaches to building knowledge and evidence on GBV', in *Feminist Pocketbook*.
- Coker, A. L. *et al.* (2017) 'RCT Testing Bystander Effectiveness to Reduce Violence', *American Journal of Preventive Medicine*. Elsevier Inc., 52(5), pp. 566–578. doi: 10.1016/j.amepre.2017.01.020.
- Collins, P. (1998) 'Negotiating Selves: Reflections on "unstructured" interviewing', *Sociological Research Online*, 3(3). doi: <https://doi.org/10.5153/sro.143>.
- Collins, R. *et al.* (2000) 'Effects of Incentive Size and Timing on Response Rates to a Follow-Up Wave of a Longitudinal Mailed Survey', *Evaluation Review*, 24(4), pp. 347–363. doi: 10.1177/0193841x0002400401.
- Costenbader, E. *et al.* (2019) 'Social norms measurement: Catching up with programs and moving the field forward', *Journal of Adolescent Health*. Society for Adolescent Health and Medicine, 64(4), pp. S4–S6. doi: 10.1016/j.jadohealth.2019.01.001.
- Cousens, S. *et al.* (2011) 'Alternatives to randomisation in the evaluation of public-health interventions: Statistical analysis and causal inference', *Journal of Epidemiology and Community Health*, 65(7), pp. 576–81. doi: 10.1136/jech.2008.082610.
- Cowen, N. and Cartwright, N. (2019) 'Street-level theories of change: Adapting the medical model of evidence-based practice for policing', *SSRN Electronic Journal*, (January). doi: 10.2139/ssrn.3324778.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., *et al.* (2008) *Developing and evaluating complex interventions: New guidance*. Available at: www.mrc.ac.uk/complexinterventionsguidance.
- Craig, P., Dieppe, P., Macintyre, S., Mitchie, S., *et al.* (2008) 'Developing and evaluating complex interventions: The new Medical Research Council guidance', *British Medical Journal*, 337. doi: 10.1136/bmj.a1655.
- Craig, P. and Petticrew, M. (2013) 'Developing and evaluating complex interventions: Reflections on the 2008 MRC guidance', *International journal of nursing studies*. England: Elsevier Ltd, 50(5), pp. 585–7. doi: 10.1016/j.ijnurstu.2012.09.009.
- Creamer, E. G. (2018) 'Paradigms in play: Using case studies to explore the value-added of divergent findings in mixed methods research', *International Journal of Multiple Research Approaches*, 10(1), pp. 30–40. doi: 10.29034/ijmra.v10n1a2.
- Datta, J. and Petticrew, M. (2013) 'Challenges to evaluating complex interventions: A content analysis of published papers', *BMC Public Health*, 13(1). doi: 10.1186/1471-2458-13-568.
- Debnam, K. J. and Temple, J. R. (2020) 'Dating Matters and the Future of Teen Dating Violence

- Prevention', *Prevention Science*. Prevention Science. doi: 10.1007/s11121-020-01169-5.
- Denzin, N. K. (2012) 'Triangulation 2.0*', *Journal of Mixed Methods Research*, 6(2), pp. 80–88. doi: 10.1177/1558689812437186.
- Denzin, N. K. (2019) 'The death of data in neoliberal times', *Qualitative Inquiry*, 25(8), pp. 721–724. doi: 10.1177/1077800419847501.
- Denzin, N. K. and Lincoln, Y. S. (eds) (2005) *The SAGE handbook of qualitative research*. 3rd edn. Sage Publications, Inc.
- Devries, K. M. *et al.* (2015) "I never expected that it would happen, coming to ask me such questions": Ethical aspects of asking children about violence in resource poor settings', *Trials*, 16(516). doi: 10.1186/s13063-015-1004-7.
- Dunkle, K. *et al.* (2019) 'Impact of Indashyikirawa: An innovative programme to reduce partner violence in rural Rwanda', (September), p. 8. Available at: <https://www.whatworks.co.za/documents/publications/352-indash-evidence-brief-aug-2019/file>.
- Durose, C. *et al.* (2012) *Towards Co-Production in Research with Communities*. Available at: https://www.dur.ac.uk/resources/geography/reframing_state/CCDiscussionPaperDurose2etal.pdf.
- Dworkin, S. L. and Barker, G. (2019) 'Gender-transformative approaches to engaging men in reducing gender-based violence: A response to Brush & Miller's "Trouble in Paradigm"', *Violence Against Women*, 25(14), pp. 1657–1671. doi: 10.1177/1077801219872555.
- Dworkin, S. L., Fleming, P. J. and Colvin, C. J. (2015) 'The promises and limitations of gender-transformative health programming with men: Critical reflections from the field.', *Culture, health & sexuality*, 17 Suppl 2(sup2), pp. S128-43. doi: 10.1080/13691058.2015.1035751.
- Ebrahim, A. (2003) 'Accountability in practice: Mechanisms for NGOs', *World Development*, 31(5), pp. 813–829. doi: 10.1016/S0305-750X(03)00014-7.
- Ebrahim, A. and Rangan, V. K. (2014) 'What Impact? A framework for measuring the scale and scope of social performance', *California Management Review*, 56(3), pp. 118–141. Available at: [http://www.hbs.edu/faculty/Publication Files/CMR5603_07_Ebrahim_e3316477-8965-4287-be95-04642982b638.pdf](http://www.hbs.edu/faculty/Publication%20Files/CMR5603_07_Ebrahim_e3316477-8965-4287-be95-04642982b638.pdf).
- Eisenbruch, M. (2018) 'Violence against women in Cambodia: Towards a culturally responsive theory of change', *Culture, Medicine and Psychiatry*. Springer US, 42(2), pp. 350–370. doi: 10.1007/s11013-017-9564-5.
- Ellsberg, M. *et al.* (2015) 'Prevention of violence against women and girls: What does the evidence say?', *The Lancet*. Elsevier Ltd, 385(9977), pp. 1555–1566. doi: 10.1016/S0140-6736(14)61703-7.
- Espinosa-Hernández, G., Vasilenko, S. A. and Bámaca-Colbert, M. Y. (2016) 'Sexual behaviors in Mexico: The role of values and gender across adolescence', *Journal of Research on Adolescence*,

26(3), pp. 603–609. doi: 10.1111/jora.12209.

Eyben, R. (2013) *Uncovering the politics of 'evidence' and 'results'. A framing paper for development practitioners*. Available at: <http://bigpushforward.net/wp-content/uploads/2011/01/Uncovering-the-Politics-of-Evidence-and-Results-by-Rosalind-Eyben.pdf>.

Farrall, S. (1996) *What is qualitative longitudinal research?, Papers in Social Research Methods*. 11. London. Available at: <http://www.lse.ac.uk/methodology/pdf/QualPapers/Stephen-Farrall-Qual-Longitudinal-Res.pdf>.

Fenton, K. A. *et al.* (2001) 'Measuring sexual behaviour: Methodological challenges in survey research', *Sexually Transmitted Infections*, 77(2), pp. 84–92. doi: 10.1136/sti.77.2.84.

Fereday, J. and Muir-Cochrane, E. (2006) 'Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development', *International Journal of Qualitative Methods*, 5(1). doi: 10.1016/j.cger.2010.06.003.

Finlay, L. (2002) "'Outing" the researcher: The provenance, process, and practice of reflexivity', *Qualitative Health Research*, 12(4), pp. 531–545. doi: 10.1177/104973202129120052.

Flecha, A., Puigvert, L. and Redondo, G. (2005) 'Socialización preventiva de la violencia de género', *Feminismo/s*, 6, pp. 107–120.

Fonner, V. A. *et al.* (2014) 'School based sex education and HIV prevention in low- and middle-income countries: A systematic review and meta-analysis', *PLoS ONE*, 9(3), p. e89692. doi: 10.1371/journal.pone.0089692.

Fulu, E. and Kerr-Wilson, A. (2015) *What works to prevent violence against women and girls? Evidence review of interventions to prevent violence against women and girls*. Available at: <http://www.whatworks.co.za/resources/all-resources/publications/item/70-global-evidence-reviews-paper-2-interventions-to-prevent-violence-against-women-and-girls>.

Fund for Shared Insight (no date). Available at: <https://www.fundforsharedinsight.org> (Accessed: 3 May 2020).

Gale, N. K. *et al.* (2013) 'Using the framework method for the analysis of qualitative data in multi-disciplinary health research.', *BMC medical research methodology*. BMC Medical Research Methodology, 13(1), p. 117. doi: 10.1186/1471-2288-13-117.

Garcia-Moreno, C. *et al.* (2005) 'Violence against women', *Science*, 310(5752), pp. 1282–3. doi: 10.1126/science.1121400.

Garcia-Moreno, C. *et al.* (2006) 'Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence', *The Lancet*, 368(9543), pp. 1260–1269. doi: 10.1016/S0140-6736(06)69523-8.

- Garcia-Moreno, C. *et al.* (2014) 'Addressing violence against women: A call to action', *The Lancet*, 385(9978), pp. 1685–1695. doi: 10.1016/S0140-6736(14)61830-4.
- Gear, C., Eppel, E. and Koziol-McLain, J. (2018) 'Advancing complexity theory as a qualitative research methodology', *International Journal of Qualitative Methods*, 17(1), pp. 1–10. doi: 10.1177/1609406918782557.
- Geertz, C. (1973) 'Thick description: Toward an interpretive theory of culture', in *Interpretation of Cultures*. Basic Books, pp. 3–30.
- Ghanem, K. G. *et al.* (2005) 'Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients', *Sexually Transmitted Infections*, 81(5), pp. 421–425. doi: 10.1136/sti.2004.013193.
- Gibbs, A. *et al.* (2019) 'Constructing, reproducing and challenging masculinities in a participatory intervention in urban informal settlements in South Africa', *Culture, Health and Sexuality*. Taylor & Francis, 0(0), pp. 1–16. doi: 10.1080/13691058.2019.1614671.
- Giddings, L. S. and Grant, B. M. (2007) 'A Trojan Horse for positivism?: A critique of mixed methods research', *Advances in Nursing Science*, 30(1), pp. 52–60. doi: 10.1097/00012272-200701000-00006.
- Giordano, P. C. *et al.* (2010) 'The characteristics of romantic relationships associated with teen dating violence', *Social Science Research*. Elsevier Inc., 39(6), pp. 863–874. doi: 10.1016/j.ssresearch.2010.03.009.
- Given, L. M. (2017) 'It's a new year...so let's stop the paradigm wars', *International Journal of Qualitative Methods*, 16(1), pp. 1–2. doi: 10.1177/1609406917692647.
- Gläser, J. and Laudel, G. (2013) 'Qualitative social life with and without coding: Two methods for early stage data analysis in qualitative research aiming at causal explanations', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 14(2), pp. 1–25. doi: 10.1016/j.jenvp.2010.03.006.
- Gomersall, T. (2018) 'Complex adaptive systems: A new approach for understanding health practices', *Health Psychology Review*, 12(4). doi: 10.1080/17437199.2018.1488603.
- Gómez Espinoza, L. M. (2006) 'El desplazamiento de prácticas impresas y la apropiación de prácticas digitales: Un estudio con alumnos del bachillerato tecnológico aprendiendo a usar la computadora en la escuela', *Revista Brasileira de Educação*, 11(31), pp. 58–79. doi: 10.1590/S1413-24782006000100006.
- Green, J. and Thorogood, N. (2018) 'Qualitative Methodology in Health Research', in *Qualitative methods for health research*. 4th edn. London: SAGE Publications.
- Greene, J. C., Caracelli, V. J. and Graham, W. F. (1989) 'Toward a conceptual framework for mixed-method evaluation designs', *Educational Evaluation and Policy Analysis*, 11(3), pp. 255–274. doi:

10.3102/01623737011003255.

Grossoehme, D. and Lipstein, E. (2016) 'Analyzing longitudinal qualitative data: The application of trajectory and recurrent cross-sectional approaches', *BMC Research Notes*. BioMed Central, 9(1), pp. 1–5. doi: 10.1186/s13104-016-1954-1.

Guba, E. G. and Lincoln, Y. S. (1989) *Fourth Generation Evaluation*. Newbury Park, CA: SAGE Publications.

Guest, G. (2013) 'Describing mixed methods research: An alternative to typologies', *Journal of Mixed Methods Research*, 7(2), pp. 141–151. doi: 10.1177/1558689812461179.

Gupta, A. K. and Santhya, K. G. (2019) 'Promoting Gender Egalitarian Norms and Practices Among Boys in Rural India: The Relative Effect of Intervening in Early and Late Adolescence', *Journal of Adolescent Health*. Elsevier Inc., 66(2), pp. 157–165. doi: 10.1016/j.jadohealth.2019.03.007.

Haberland, N. (2015) 'The case for addressing gender and power in sexuality and HIV education: A comprehensive review of evaluation studies', *International Perspectives on Sexual and Reproductive Health*, 41(1), pp. 31–42. doi: 10.1363/4103115.

Haberland, N. and Rogow, D. (2015) 'Sexuality education: Emerging trends in evidence and practice', *Journal of Adolescent Health*. Elsevier Inc., 56(1), pp. S15–S21. doi: 10.1016/j.jadohealth.2014.08.013.

Hall, R. (2013) 'Mixed Methods: In Search of a Paradigm', in Le, T. and Le, Q. (eds) *Conducting Research in a Changing and Challenging World*. Nova Science Publishers, Inc, pp. 71–78.

Hatcher, A. M. *et al.* (2020) 'Process evaluation of a community mobilization intervention for preventing men's partner violence use in peri-urban South Africa', *Evaluation and Program Planning*, 78(September 2019). doi: 10.1016/j.evalprogplan.2019.101727.

Hawe, P. (2015) 'Lessons from complex interventions to improve health', *Annual Review of Public Health*. doi: 10.1146/annurev-publhealth-031912-114421.

Hawe, P., Shiell, A. and Riley, T. (2004) 'Complex interventions: how "out of control" can a randomised controlled trial be?', *Bmj*, 328(7455), pp. 1561–1563. doi: 10.1136/bmj.328.7455.1561.

Hawe, P., Shiell, A. and Riley, T. (2009) 'Theorising interventions as events in systems', *American Journal of Community Psychology*, 43(3–4), pp. 267–276. doi: 10.1007/s10464-009-9229-9.

Heise, L. (1998) 'Violence against women: An integrated, ecological framework', *Violence Against Women*, 4(3), pp. 262–290. doi: 10.1177/1077801298004003002.

Heise, L. (2011) *What works to prevent partner violence? An evidence overview*. London: STRIVE Research Consortium, London School of Hygiene and Tropical Medicine. doi: 978 0 902657 85 2.

Heise, L. *et al.* (2019) 'Gender inequality and restrictive gender norms: Framing the challenges to health', *The Lancet*, 393(10189), pp. 2440–2454. doi: 10.1016/S0140-6736(19)30652-X.

Heise, L., Ellsberg, M. and Gottemoeller, M. (1999) 'Ending violence against women', *Population Reports*. Baltimore, MD, Series L(11), pp. 1–43. doi: 10.1080/15211030802194589.

Henderson, M. *et al.* (2010) 'Retaining young people in a longitudinal sexual health survey: A trial of strategies to maintain participation', *BMC Medical Research Methodology*, 10. doi: 10.1186/1471-2288-10-9.

Henderson, S. *et al.* (2012) 'Storying qualitative longitudinal research: Sequence, voice and motif', *Qualitative Research*, 12(1), pp. 16–34. doi: 10.1177/1468794111426232.

Hesse-Biber, S. (2012) 'Weaving a multimethodology and mixed methods praxis into randomized control trials to enhance credibility', *Qualitative Inquiry*, 18(10), pp. 876–889. doi: 10.1177/1077800412456964.

Hirdman, Y. (1991) 'The Gender System', in Andreassen, T. (ed.) *Moving on: New perspectives on the Women's movement*. Aarhus University Press, pp. 187–207.

Hoffmann, T. C. *et al.* (2014) 'Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide', *BMJ (Online)*, 348(March), pp. 1–12. doi: 10.1136/bmj.g1687.

Holden, J., Bell, E. and Schauerhammer, V. (2015) '*We want to learn about good love*': Findings from a qualitative study assessing the links between comprehensive sexuality education and violence against women and girls. London. Available at: <http://www.awid.org/publications/study-we-want-learn-about-good-love>.

Holland, J., Thomson, R. and Henderson, S. (2006) *Qualitative longitudinal research: A discussion paper*. Available at: https://www.lsbu.ac.uk/__data/assets/pdf_file/0019/9370/qualitative-longitudinal-research-families-working-paper.pdf.

Howarth, E. *et al.* (2016) 'Contextual issues and qualitative research', *Health Services and Delivery Research*, 4(16), pp. 105–120. doi: 10.3310/HSDR04160-105.

Imbusch, P., Misse, M. and Carrión, F. (2011) 'Violence research in Latin America and the Caribbean: A literature review', *International Journal of Conflict and Violence*, 5(1), pp. 87–154.

Instituto Nacional de Estadística y Geografía (2018a) *Anuario estadístico y geográfico de los Estados Unidos Mexicanos 2018*. Aguascalientes, Mexico. Available at: http://internet.contenidos.inegi.org.mx/contenidos/Productos/prod_serv/contenidos/espanol/bvini/inegi/productos/nueva_estruc/AEGEUM_2018/702825107000.pdf.

Instituto Nacional de Estadística y Geografía (2018b) *Mujeres y Hombres en México 2018*. Aguascalientes, Mexico. Available at: http://cedoc.inmujeres.gob.mx/documentos_download/MHM_2018.pdf.

Instituto Nacional de Salud Pública (2015) *Estudio sobre la Prevención del Embarazo en Adolescentes desde las Masculinidades: Informe final*. Cuernavaca. Available at: https://backend.aprende.sep.gob.mx/media/uploads/proedit/resources/estudio_sobre_la_pre_9f57284c.pdf.

Jaspan, H. B. *et al.* (2007) 'Brief report: Methods for collecting sexual behaviour information from South African adolescents - a comparison of paper versus personal digital assistant questionnaires', *Journal of Adolescence*, 30(2), pp. 353–359. doi: 10.1016/j.adolescence.2006.11.002.

Jennings, W. G. *et al.* (2017) 'Dating and intimate partner violence among young persons ages 15–30: Evidence from a systematic review', *Aggression and Violent Behavior*. Elsevier Ltd, 33, pp. 107–125. doi: 10.1016/j.avb.2017.01.007.

Jewkes, R. *et al.* (2008) 'Impact of Stepping Stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: Cluster randomised controlled trial', *BMJ*, 337, p. a506. doi: 10.1136/bmj.a506.

Jewkes, R. (2017) *What Works evidence review: Social norms and violence against women and girls*. doi: 10.1177/1097184X1769617.

Jewkes, R., Dartnall, E. and Sikweyiya, Y. (2012) *Ethical and safety recommendations for research on the perpetration of sexual violence*. Pretoria, South Africa. Available at: <http://www.svri.org/EthicalRecommendations.pdf>.

Jewkes, R., Flood, M. and Lang, J. (2015) 'From work with men and boys to changes of social norms and reduction of inequities in gender relations: A conceptual shift in prevention of violence against women and girls', *The Lancet*, 385(9977), pp. 1580–1589. doi: 10.1016/S0140-6736(14)61683-4.

Jewkes, R., Stern, E. and Ramsoomar, L. (2019) *Preventing violence against women and girls: Community activism approaches to shift harmful gender attitudes, roles and social norms*. Available at: <https://www.whatworks.co.za/resources/evidence-reviews/item/664-community-activism-approaches-to-shift-harmful-gender-attitudes-roles-and-social-norms>.

Jung, T., Harrow, J. and Pharoah, C. (2012) *Co-producing research: Working together or falling apart?* Available at: [http://www.cgap.org.uk/uploads/Briefing Papers/CGAP Briefing Note 8 - Co-producing research.pdf](http://www.cgap.org.uk/uploads/Briefing%20Papers/CGAP%20Briefing%20Note%208%20-%20Co-producing%20research.pdf) (Accessed: 24 November 2014).

Kågesten, A. *et al.* (2016) 'Understanding factors that shape gender attitudes in early adolescence globally: A mixed-methods systematic review', *PLoS ONE*, 11(6), pp. 1–36. doi: 10.1371/journal.pone.0157805.

Keogh, S. C. *et al.* (2018) 'Challenges to implementing national comprehensive sexuality education curricula in low- and middle-income countries: Case studies of Ghana, Kenya, Peru and Guatemala', *PLoS ONE*, 13(7), pp. 1–18. doi: 10.1371/journal.pone.0200513.

Keshavarz, N. *et al.* (2010) 'Schools as social complex adaptive systems: A new way to understand

the challenges of introducing the health promoting schools concept', *Social Science and Medicine*. Elsevier Ltd, 70(10), pp. 1467–1474. doi: 10.1016/j.socscimed.2010.01.034.

Kim, M.-S. and Hunter, J. E. (1993) 'Relationships among attitudes, behavioral intentions, and behavior: A meta-analysis of past research, Part 2', *Communication Research*, 20(3), pp. 331–364. doi: 10.1177/009365093020003001.

Kincheloe, J. L. (2005) 'On to the next level: Continuing the conceptualization of the bricolage', *Qualitative Inquiry*, 11(3), pp. 323–350. doi: 10.1177/1077800405275056.

Kirby, D. B. (2008) 'The impact of abstinence and comprehensive sex and STD/HIV education programs on adolescent sexual behavior', *Sexuality Research and Social Policy*, 5(3), pp. 18–27. doi: 10.1525/srsp.2008.5.3.18.

Kirby, D. B., Laris, B. . and Rolleri, L. (2006) *Sex and HIV education programs for youth: Their impact and important characteristics*. Available at: http://hivhealthclearinghouse.unesco.org/sites/default/files/resources/bie_sex_hiv-education_programs_youth_impact_characteristics_en.pdf.

Koch, P. B., Colaco, C. and Porter, A. W. (2011) 'Sexual Health Practices Self-Efficacy Scale', in Fisher, T. D. et al. (eds) *Handbook of Sexuality-Related Measures*. New York: Routledge Handbooks Online, pp. 345–347. doi: 10.4324/9781315881089.ch115.

De Koker, P. et al. (2014) 'A systematic review of interventions for preventing adolescent intimate partner violence', *Journal of Adolescent Health*. Elsevier Ltd, 54(1), pp. 3–13. doi: 10.1016/j.jadohealth.2013.08.008.

Krug, E. G. et al. (2002) 'The world report on violence and health', *The Lancet*, 360(9339), pp. 1083–1088. doi: 10.1016/S0140-6736(02)11133-0.

Kyegombe, N. et al. (2014) "'SASA! is the medicine that treats violence". Qualitative findings on how a community mobilisation intervention to prevent violence against women created change in Kampala, Uganda', *Global Health Action*, 7, p. 25082. doi: 10.3402/gha.v7.25082.

Kyegombe, N. et al. (2017) 'How did the Good School Toolkit reduce the risk of past week physical violence from teachers to students? Qualitative findings on pathways of change in schools in Luwero, Uganda', *Social Science and Medicine*. Elsevier Ltd, 180, pp. 10–19. doi: 10.1016/j.socscimed.2017.03.008.

De la Rue, L. et al. (2014) 'School-based interventions to reduce dating and sexual violence: A systematic review', *Campbell Systematic Reviews*, 7. doi: 10.4073/csr.2014.7.

De La Rue, L. et al. (2017) 'A Meta-Analysis of School-Based Interventions Aimed to Prevent or Reduce Violence in Teen Dating Relationships', *Review of Educational Research*, 87(1), pp. 7–34. doi: 10.3102/0034654316632061.

De La Rue, L. (2019) 'Intersectionality and resilience: Updating how we address adolescent dating

violence', *American Journal of Public Health*, 109(10), pp. 1324–1325. doi: 10.2105/AJPH.2019.305302.

Lacayo, V., Obregón, R. and Singhal, A. (2008) 'Approaching social change as a complex problem in a world that treats it as a complicated one: The case of Puntos de Encuentro, Nicaragua', *Investigación y Desarrollo*, 16(2), pp. 126–159.

Lakshman, R. *et al.* (2014) 'Using the Medical Research Council framework for the development and evaluation of complex interventions in a theory-based infant feeding intervention to prevent childhood obesity: The baby milk intervention and trial.', *Journal of obesity*. doi: 10.1155/2014/646504.

Lambert, H. (2013) 'Plural forms of evidence in public health: tolerating epistemological and methodological diversity', *Evidence & Policy: A Journal of Research, Debate and Practice*, 9(1), pp. 43–48.

Laybourn, C. (no date) *Evaluations of NGO international development and humanitarian work*.

Learning Collaborative to Advance Normative Change (2019) *Social norms and AYSRH: Building a bridge from theory to program design*. Washington, D.C.

Lee, C. and Wong, J. S. (2020) 'Examining the effects of teen dating violence prevention programs: a systematic review and meta-analysis', *Journal of Experimental Criminology*. doi: 10.1007/s11292-020-09442-x.

Leung, L. *et al.* (2019) 'Making feminism count: Integrating feminist research principles in large-scale quantitative research on violence against women and girls', *Gender and Development*. Taylor & Francis, 27(3), pp. 427–447. doi: 10.1080/13552074.2019.1668142.

Lewis, J. (2007) 'Analysing qualitative longitudinal research in evaluations', *Social Policy and Society*, 6(4), pp. 545–556. doi: 10.1017/s1474746407003880.

Liket, K. C., Rey-Garcia, M. and Maas, K. E. H. (2014) 'Why aren't evaluations working and what to do about it: A framework for negotiating meaningful evaluation in nonprofits', *American Journal of Evaluation*, 35(2), pp. 171–188. doi: 10.1177/1098214013517736.

Lopez, F. G., Morua, W. and Rice, K. G. (2007) 'Factor Structure, Stability, and Predictive Validity of College Students' Relationship Self-Efficacy Beliefs', *Measurement and Evaluation in Counseling and Development*, 40, pp. 80–97. Available at: <http://faculty.coe.uh.edu/flopez/docs/relationship-self-efficacy-2007.pdf>.

Loxton, D. *et al.* (2019) 'Longitudinal inconsistency in responses to survey items that ask women about intimate partner violence', *BMC Medical Research Methodology*. BMC Medical Research Methodology, 19(1), pp. 1–8. doi: 10.1186/s12874-019-0835-4.

Lundgren, R. and Amin, A. (2015a) 'Addressing intimate partner violence and sexual violence among adolescents: Emerging evidence of effectiveness', *Journal of Adolescent Health*. Elsevier

Inc., 56(1), pp. S42–S50. doi: 10.1016/j.jadohealth.2014.08.012.

Lundgren, R. and Amin, A. (2015b) 'Addressing intimate partner violence and sexual violence among adolescents: Emerging evidence of effectiveness', *Journal of Adolescent Health*. Elsevier Inc., 56, pp. S42–S50. doi: 10.1016/j.jadohealth.2014.08.012.

Mackie, G. *et al.* (2015) *What are social norms? How are they measured?* Available at: <http://www.polisci.ucsd.edu/~gmackie/>.

Mackie, G. and Lejeune, J. (2009) *Social dynamics of abandonment of harmful practices: A new look at the theory, Special series on social norms and harmful practices*. Available at: https://www.unicef-irc.org/publications/pdf/iwp_2009_06.pdf.

Makleff, S. *et al.* (2019) 'Exploring stigma and social norms in women's abortion experiences and their expectations of care', *Sexual and Reproductive Health Matters*, 27(3). doi: 10.1080/26410397.2019.1661753.

Mannell, J. *et al.* (2019) 'Why interventions to prevent intimate partner violence and HIV have failed young women in southern Africa', *Journal of the International AIDS Society*, 22(8), p. 6. doi: 10.1002/jia2.25380.

Marhefka, S., Lockhart, E. and Turner, D. (2020) 'Achieve research continuity during social distancing by rapidly implementing individual and group videoconferencing with participants: Key considerations, best practices, and protocols', *AIDS and Behavior*. Springer US, (0123456789). doi: 10.1007/s10461-020-02837-x.

Marston, C. (2004) 'Gendered communication among young people in Mexico: Implications for sexual health interventions', *Social Science & Medicine*, 59(3), pp. 445–456. doi: 10.1016/j.socscimed.2003.11.007.

Marston, C. (2005) 'What is heterosexual coercion? Interpreting narratives from young people in Mexico City.', *Sociology of health & illness*, 27(1), pp. 68–91. doi: 10.1111/j.1467-9566.2005.00432.x.

Marston, C. and King, E. (2006) 'Factors that shape young people's sexual behaviour: A systematic review', *Lancet*, 368, pp. 1581–86. doi: 10.1016/S0140-6736(06)69662-1.

Mason, J. (2006) 'Mixing methods in a qualitatively driven way', *Qualitative Research*, 6(1), pp. 9–25. doi: 10.1177/1468794106058866.

Mathews, C. *et al.* (2016) 'Effects of PREPARE, a Multi-component, School-Based HIV and Intimate Partner Violence (IPV) Prevention Programme on Adolescent Sexual Risk Behaviour and IPV: Cluster Randomised Controlled Trial', *AIDS and Behavior*. Springer US, 20(9), pp. 1821–1840. doi: 10.1007/s10461-016-1410-1.

McCallum, E. B. and Peterson, Z. D. (2012) 'Investigating the impact of inquiry mode on self-reported sexual behavior: Theoretical considerations and review of the literature', *Journal of Sex*

Research, 49(2–3), pp. 212–226. doi: 10.1080/00224499.2012.658923.

McGeeney, E. (2015) 'A focus on pleasure? Desire and disgust in group work with young men', *Culture, Health and Sexuality*. Routledge, 17, pp. S223–S237. doi: 10.1080/13691058.2015.1038586.

McGhee, S. *et al.* (2019) "'Change really does need to start from home": Impact of an intimate partner violence prevention strategy among married couples in Nepal', *Journal of Interpersonal Violence*. doi: 10.1177/0886260519839422.

McLean, L., Heise, L. L. and Stern, E. A. (2019) 'Shifting and transforming gender-inequitable beliefs, behaviours and norms in intimate partnerships: The Indashyikirwa couples programme in Rwanda', *Culture, Health and Sexuality*. Taylor & Francis, 0(0), pp. 1–18. doi: 10.1080/13691058.2019.1679394.

Medical Research Council (2000) *A framework for development and evaluation of RCTs for complex interventions to improve health*. Available at: <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003372>.

Meinck, F. *et al.* (2019) 'Interventions to reduce gender-based violence among young people living with or affected by HIV/AIDS in low-income and middle-income countries', *AIDS*, 33(14), pp. 2219–2236. doi: 10.1097/QAD.0000000000002337.

Mejía, M. L. *et al.* (2010) *Barreras para el acceso de adolescentes y jóvenes a servicios de salud*. Available at: <https://colombia.unfpa.org/sites/default/files/pub-pdf/BarrerasJovenesWeb%281%29.pdf>.

Mensch, B. S., Hewett, P. C. and Erulkar, A. S. (2003) 'The reporting of sensitive behavior by adolescents: A methodological experiment in Kenya', *Demography*, 40(2), pp. 247–268. doi: 10.2307/3180800.

Merchant-Vega, N. (2011) *Practical challenges of rigorous impact evaluation in international governance NGOs: Experiences and lessons from the Asia Foundation, Occasional Paper*. Available at: <http://www.asiafoundation.org/resources/pdfs/OccasionalPaperNo8IEFinal.pdf> (Accessed: 3 January 2015).

Mertens, D. M. and Tarsilla, M. (2015) 'Mixed Methods Evaluation', in Hesse-Biber, S. N. and Johnson, R. B. (eds) *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*. Oxford University Press, pp. 426–446. doi: 10.1093/oxfordhb/9780199933624.013.27.

Michau, L. *et al.* (2015) 'Prevention of violence against women and girls: Lessons from practice', *The Lancet*. Elsevier Ltd, 385(9978), pp. 1672–1684. doi: 10.1016/S0140-6736(14)61797-9.

Miles, M. B. and Huberman, A. M. (1994) *Qualitative data analysis: An expanded sourcebook*. 2nd Ed. Beverly Hills, CA: SAGE Publications.

Miller, D. T. and Prentice, D. (2016) 'Changing norms to change behavior', *Annual Review of*

Psychology, 67, pp. 339–361. doi: 10.1146/annurev-psych-010814-015013.

Miller, E. *et al.* (2020) 'An Athletic Coach-Delivered Middle School Gender Violence Prevention Program: A Cluster Randomized Clinical Trial', *JAMA Pediatrics*, 174(3), pp. 241–249. doi: 10.1001/jamapediatrics.2019.5217.

Milnes, L. and Kendal, S. (2012) 'Ethical challenges of conducting health research in UK school setting', *British Journal of Nursing*, 21(5), pp. 294–298. doi: 10.12968/bjon.2012.21.5.294.

Mmari, K. *et al.* (2017) 'Exploration of gender norms and socialization among early adolescents: The use of qualitative methods for the Global Early Adolescent Study', *Journal of Adolescent Health*. Elsevier Inc., 61(4), pp. S12–S18. doi: 10.1016/j.jadohealth.2017.07.006.

Moffatt, S. *et al.* (2006) 'Using quantitative and qualitative data in health services research - What happens when mixed method findings conflict?', *BMC Health Services Research*, 6, pp. 1–10. doi: 10.1186/1472-6963-6-28.

Mohanty, C. (1988) 'Under Western Eyes: Feminist Scholarship and Colonial Discourses', *Feminist Review*, 30(30), pp. 61–88. doi: 10.4324/9780203825235.

Montgomery, P. and Knerr, W. (2018) *Review of the evidence on sexuality education: Report to inform the update of the UNESCO International Technical Guidance on Sexuality Education*. Paris. Available at: <http://unesdoc.unesco.org/images/0026/002646/264649e.pdf>.

Moore, G. F. *et al.* (2014) *Process evaluation of complex interventions: UK Medical Research (MRC) guidance*. Available at: <https://mrc.ukri.org/documents/pdf/mrc-phsrn-process-evaluation-guidance-final/>.

Moore, G. F. *et al.* (2015) 'Process evaluation of complex interventions: Medical Research Council guidance', *BMJ*, 350(h1258), pp. 1–7. doi: 10.1136/bmj.h1258.

Moore, G. F. *et al.* (2019) 'From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation', *Evaluation*, 25(1), pp. 23–45. doi: 10.1177/1356389018803219.

Moore, G. F. and Evans, R. E. (2017) 'What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions', *SSM - Population Health*. Elsevier, 3, pp. 132–135. doi: 10.1016/j.ssmph.2016.12.005.

Moosa, Z. *et al.* (2012) *A theory of change for tackling violence against women and girls*. Available at: https://www.actionaid.org.uk/sites/default/files/doc_lib/theory_of_change_on_vawg.pdf.

Morrison, A., Ellsberg, M. and Bott, S. (2004) 'Addressing gender-based violence in the Latin American and Caribbean region: A critical review of interventions', *Policy Research Working Paper Series*. (Policy Research Working Paper Series), pp. i–70. doi: 10.1093/wbro/lkm003.

Morrison, A., Ellsberg, M. and Bott, S. (2007) 'Addressing gender-based violence: A critical review of interventions', *The World Bank Research Observer*, 22(1), pp. 25–51. doi: 10.1093/wbro/lkm003.

Moseson, H. *et al.* (2017) 'The List Experiment for Measuring Abortion: What We Know and What We Need', *Studies in Family Planning*, 48(4), pp. 397–405. doi: 10.1111/sifp.12042.

Mūkoma, W. *et al.* (2009) 'Development and test-retest reliability of a research instrument designed to evaluate school-based HIV/AIDS interventions in South Africa and Tanzania', *Scandinavian Journal of Public Health*, 37(SUPPL. 2), pp. 7–15. doi: 10.1177/1403494809103995.

Muñoz-Fernández, N. *et al.* (2019) 'The efficacy of the "Dat-e Adolescence" prevention program in the reduction of dating violence and bullying', *International Journal of Environmental Research and Public Health*, 16(3). doi: 10.3390/ijerph16030408.

Naveed, A. *et al.* (2017) 'Understanding, embracing and reflecting upon the messiness of doctoral fieldwork', *Compare*. Routledge, 47(5), pp. 773–792. doi: 10.1080/03057925.2017.1344031.

Nelson, A. *et al.* (2010) 'Eliciting behavior change in a US sexual violence and intimate partner violence prevention program through utilization of Freire and discussion facilitation', *Health Promotion International*, 25(3), pp. 299–308. doi: 10.1093/heapro/daq024.

Nightingale, A. (2009) 'Triangulation', in *International Encyclopedia of Human Geography*, pp. 489–492.

Oakley, A. *et al.* (2003) 'Using random allocation to evaluate social interventions: Three recent UK examples', *The ANNALS of the American Academy of Political and Social Science*, 589(170). doi: 10.1177/0002716203254765.

Oakley, A. (2016a) 'A small sociology of maternal memory', *Sociological Review*, 64(3), pp. 533–549. doi: 10.1111/1467-954X.12367.

Oakley, A. (2016b) 'Interviewing women again: Power, time and the gift', *Sociology*, 50(1), pp. 195–213. doi: 10.1177/0038038515580253.

Oakley, A. N. N. (1999) 'Paradigm wars: Some thoughts on a personal and public trajectory', *International journal of social research methodology*, 2(3), pp. 247–254.

Parker, R. G. (1996) 'Behaviour in Latin American men: Implications for HIV/AIDS interventions', *International Journal of STD & AIDS*, 7(Suppl. 2), pp. 62–65. doi: 10.1258/0956462961917663.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. SAGE Publications. Available at: <http://www.uk.sagepub.com/books/Book205276>.

Peacock, D. and Barker, G. (2014) 'Working with men and boys to prevent gender-based violence: Principles, lessons learned, and ways forward', *Men and Masculinities*, 17(5), pp. 578–599. doi: 10.1177/1097184X14558240.

- Perrin, N. *et al.* (2019) 'Social norms and beliefs about gender based violence scale: A measure for use with gender based violence prevention programs in low-resource and humanitarian settings', *Conflict and Health*. *Conflict and Health*, 13(1), pp. 1–12. doi: 10.1186/s13031-019-0189-x.
- Peters, D. H. (2014) 'The application of systems thinking in health: Why use systems thinking?', *Health Research Policy and Systems*, 12(51). doi: 10.1186/1478-4505-12-51.
- Petticrew, M. *et al.* (2012) 'Damned if you do, damned if you don't: Subgroup analysis and equity.', *Journal of Epidemiology and Community Health*, 66(1), pp. 95–8. doi: 10.1136/jech.2010.121095.
- Petticrew, M. *et al.* (2013) 'Synthesizing evidence on complex interventions: How meta-analytical, qualitative, and mixed-method approaches can contribute.', *Journal of clinical epidemiology*, 66(11), pp. 1230–43. doi: 10.1016/j.jclinepi.2013.06.005.
- Petticrew, M., Chalabi, Z. and Jones, D. R. (2012) 'To RCT or not to RCT: Deciding when "more evidence is needed" for public health policy and practice', *Journal of Epidemiology and Community Health*, 66(5), pp. 391–396. doi: 10.1136/jech.2010.116483.
- Petticrew, M. and Roberts, H. (2003) 'Evidence, hierarchies, and typologies: Horses for courses', *Journal of Epidemiology and Community Health*, (57), pp. 527–529. doi: 10.1136/jech.57.7.527.
- Picciotto, R. (2012) 'Experimentalism and development evaluation: Will the bubble burst?', *Evaluation*, 18(2), pp. 213–229. doi: 10.1177/1356389012440915.
- Picciotto, R. (2014) 'Have development evaluators been fighting the last war... And if so, what is to be done?', *IDS Bulletin*, 45(6), pp. 6–16. doi: 10.1111/1759-5436.12109.
- Pierotti, R. S., Lake, M. and Lewis, C. (2018) 'Equality on his terms: Doing and undoing gender through men's discussion groups', *Gender and Society*, 32(4), pp. 540–562. doi: 10.1177/0891243218779779.
- Plsek, P. E. (1997) *Working paper: Some emerging principles for managers of complex adaptive systems (CAS)*.
- Plsek, P. E. and Greenhalgh, T. (2001) 'The challenge of complexity in health care', *British Medical Journal*, 323(7313), pp. 625–628. doi: 10.1136/bmj.323.7313.625.
- Pluye, P. *et al.* (2009) 'Understanding divergence of quantitative and qualitative data (or results) in mixed methods studies', *International Journal of Multiple Research Approaches*, 3(1), pp. 58–72. doi: 10.5172/mra.455.3.1.58.
- Pluye, P. and Hong, Q. N. (2014) 'Combining the Power of Stories and the Power of Numbers: Mixed Methods Research and Mixed Studies Reviews', *Annual Review of Public Health*, 35(1), pp. 29–45. doi: 10.1146/annurev-publhealth-032013-182440.
- Popay, J. (1998) 'Qualitative research and evidence-based healthcare', *Journal of the Royal Society of Medicine, Supplement*, 91(35), pp. 32–37. doi: 10.1177/014107689809135s08.

Price, M. *et al.* (2016) 'Young love: Romantic concerns and associated mental health issues among adolescent help-seekers', *Behavioral Sciences*, 6(2), pp. 7–10. doi: 10.3390/bs6020009.

Pronyk, P. M. *et al.* (2006) 'Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: A cluster randomised trial', *Lancet*, 368(9551), pp. 1973–1983. doi: 10.1016/S0140-6736(06)69744-4.

Pulerwitz, J. *et al.* (2006) *Promoting More gender-equitable norms and behaviors among young men as an HIV/AIDS prevention strategy, Horizons Final Report*. Available at: http://pdf.usaid.gov/pdf_docs/PNADF883.PDF#sthash.o1AUvB2Q.dpuf.

Pulerwitz, J. *et al.* (2010) *Promoting gender equity for HIV and violence prevention: Results from the PEPFAR Male Norms Initiative evaluation in Ethiopia*. Washington D.C.: PATH. Available at: [http://www.hiwot.org.et/Resource/Hiwot Ethiopia MNI Poster.pdf](http://www.hiwot.org.et/Resource/Hiwot%20Ethiopia%20MNI%20Poster.pdf).

Pulerwitz, J. and Barker, G. (2008) 'Measuring attitudes toward gender norms among young men in Brazil: Development and psychometric evaluation of the GEM Scale', *Men and Masculinities*, 10(3), pp. 322–338. doi: 10.1177/1097184X06298778.

Pulerwitz, J., Gortmaker, S. L. and DeJong, W. (2000) 'Measuring Sexual Relationship Power in HIV/STD Research', *Sex Roles*. Kluwer Academic Publishers-Plenum Publishers, 42(7), pp. 637–660. doi: 10.1023/A:1007051506972.

Rao, V. and Woolcock, M. (2003) 'Integrating qualitative and quantitative approaches in program evaluation', in *The impact of economic policies on poverty and income distribution: Evaluation techniques and tools*, pp. 165–190. doi: 10.1007/BF02360471.

Ravallion, M. (2018) *Should the randomistas (continue to) rule?* CDG Working Paper 492. Washington D.C. Available at: www.cgdev.org.

Report of the International Conference on Population and Development, Cairo, 5-13 September 1994 (1995). Available at: <http://www.refworld.org/docid/4a54bc080.html> (Accessed: 14 September 2018).

Ricardo, C., Eads, M. and Barker, G. (2011) *Engaging boys and young men in the prevention of sexual violence*. Pretoria, South Africa. Available at: <http://www.svri.org/menandboys.pdf>.

Ridgeway, C. L. and Correll, S. J. (2004) 'Unpacking the gender system: A theoretical perspective on gender beliefs and social relations', *Gender and Society*, 18(4), pp. 510–531. doi: 10.1177/0891243204265269.

Ridgeway, C. L. and Smith-Lovin, L. (1999) 'The gender system and interaction', *Annual Review of Sociology*, 25(1), pp. 191–216. doi: 10.1146/annurev.soc.25.1.191.

Rijsdijk, L. E. *et al.* (2011) 'The World Starts With Me: A multilevel evaluation of a comprehensive sex education programme targeting adolescents in Uganda.', *BMC public health*, 11(1), p. 334. doi: 10.1186/1471-2458-11-334.

Ritchie, J. and Spencer, L. (1994) 'Qualitative data analysis for applied policy research', in Bryman, A. and Burgess, R. G. (eds) *Analysing Qualitative Data*. London: Routledge, pp. 173–194. doi: <https://www.taylorfrancis.com/books/e/9780203413081>.

Roetman, E. (2011) *A can of worms? Implications of rigorous impact evaluations for development agencies*, 3ie. Available at: <https://www.3ieimpact.org/evidence-hub/publications/working-papers/can-worms-implications-rigorous-impact-evaluations> (Accessed: 3 January 2015).

Rogers, M. (2012) 'Contextualizing theories and practices of bricolage research', *The qualitative report*, 17(48), pp. 1–17.

Rottach, E., Schuler, S. R. and Hardee, K. (2009) *Gender perspectives improve reproductive health outcomes: New evidence*. doi: 10.1002/anie.197906331.

Ruane-McAteer, E. *et al.* (2019) 'Interventions addressing men, masculinities and gender equality in sexual and reproductive health and rights: An evidence and gap map and systematic review of reviews', *BMJ Global Health*, 4(5). doi: 10.1136/bmjgh-2019-001634.

Ruiz, M. Á. B. (2015) 'Implicaciones del uso de las redes sociales en el aumento de la violencia de género en adolescentes', *Comunicación y Medios*, 30, pp. 124–141. doi: 10.5354/0719-1529.2014.32375.

Sakata, N. *et al.* (2019) 'Addressing the messiness of data analysis: Praxis, readiness and tips from doctoral research', *Compare: A Journal of Comparative and International Education*. Routledge, 49(2), pp. 318–336. doi: 10.1080/03057925.2018.1562676.

Saldaña, J. (2003) *Longitudinal qualitative research: Analyzing change through time*. AltaMira Press.

Saldaña, J. (2009) *The coding manual for qualitative researchers*. SAGE Publications. doi: 10.1017/CBO9781107415324.004.

Santhya, K. G. and Jejeebhoy, S. J. (2015) 'Sexual and reproductive health and rights of adolescent girls: Evidence from low- and middle-income countries', *Global Public Health*. Taylor & Francis, 10(2), pp. 189–221. doi: 10.1080/17441692.2014.986169.

Sarnquist, C. *et al.* (2014) 'Rape prevention through empowerment of adolescent girls', *Pediatrics*, 133(5), pp. e1226–e1232. doi: 10.1542/peds.2013-3414.

Saul, J. *et al.* (2018) 'The DREAMS core package of interventions: A comprehensive approach to preventing HIV among adolescent girls and young women', *PLoS ONE*, 13(12), pp. 1–18. doi: 10.1371/journal.pone.0208167.

Savedoff, W., Levine, R. and Birdsall, N. (2006) *When will we ever learn? Improving lives through impact evaluation, Report of the Evaluation Gap Working Group*. Available at: https://www.cgdev.org/sites/default/files/7973_file_WillWeEverLearn.pdf (Accessed: 3 January 2015).

- Schoonenboom, J., Johnson, R. B. and Froehlich, D. E. (2018) 'Combining multiple purposes of mixing within a mixed methods research design', *International Journal of Multiple Research Approaches*, 10(1), pp. 271–282. doi: 10.29034/ijmra.v10n1a17.
- Schutt, R. K. and Chambliss, D. F. (2013) 'Qualitative data analysis', in *Making sense of the social world: Methods of investigation*. 5th edn. SAGE Publications, pp. 320–357. doi: 10.1136/ebnurs.2011.100352.
- Shannon-Baker, P. (2016) 'Making Paradigms Meaningful in Mixed Methods Research', *Journal of Mixed Methods Research*, 10(4), pp. 319–334. doi: 10.1177/1558689815575861.
- Shiell, A., Hawe, P. and Gold, L. (2008) 'Complex interventions or complex systems? Implications for health economic evaluation', *BMJ*, 336, pp. 1281–1283. doi: 10.1103/PhysRevB.61.3085.
- Silva, D. *et al.* (2014) 'Theory of Change: A theory-driven approach to enhance the Medical Research Council's framework for complex interventions', *Trials*, 15. doi: 10.1186/1745-6215-15-267.
- Sinkovics, R. R. and Alfoldi, E. a. (2012) 'Progressive focusing and trustworthiness in qualitative research: The enabling role of computer-assisted qualitative data analysis software (CAQDAS)', *Management International Review*, 52(6), pp. 817–845. doi: 10.1007/s11575-012-0140-5.
- Slonim-Nevo, V. and Nevo, I. (2009) 'Conflicting findings in mixed methods research: An illustration from an Israeli study on immigration', *Journal of Mixed Methods Research*, 3(2), pp. 109–128. doi: 10.1177/1558689808330621.
- Smith, L. T. (2012) *Decolonizing methodologies: Research and indigenous peoples*. 2nd edn. London: Zed Books, Ltd.
- Smith, P. H., White, J. W. and Holland, L. J. (2003) 'A longitudinal perspective on dating violence among adolescent and college-age women', *American Journal of Public Health*, 93(7), pp. 1104–1109. doi: 10.2105/AJPH.93.7.1104.
- Smith, R. D. and Petticrew, M. (2010) 'Public health evaluation in the twenty-first century: Time to see the wood as well as the trees', *Journal of Public Health*, 32(1), pp. 2–7. doi: 10.1093/pubmed/fdp122.
- Smith, W. *et al.* (2016) 'Social Norms and Stigma Regarding Unintended Pregnancy and Pregnancy Decisions: A Qualitative Study of Young Women in Alabama', *Perspectives on Sexual and Reproductive Health*, 48(2), pp. 73–81. doi: 10.1363/48e9016.
- Sosa-Rubi, S. G. *et al.* (2016) 'True Love: Effectiveness of a school-based program to reduce dating violence Among adolescents in Mexico City', *Prevention Science*. doi: 10.1007/s11121-016-0718-4.
- Stake, R. E. (1981) 'The art of progressive focusing', in *Paper presented at the Annual Meeting of the American Educational Research Association*. Los Angeles. Available at: <http://eric.ed.gov/?id=ED204358> (Accessed: 20 December 2015).

- Starmann, E. *et al.* (2017) 'Exploring couples' processes of change in the context of SASA!, a violence against women and HIV prevention intervention in Uganda', *Prevention Science*. *Prevention Science*, 18(2), pp. 233–244. doi: 10.1007/s11121-016-0716-6.
- Stern, E. *et al.* (2012) *Broadening the range of designs and methods for impact evaluations*. Working Paper 38. Department for International Development. Available at: <http://r4d.dfid.gov.uk/Output/189575/Default.aspx> (Accessed: 15 November 2014).
- Stern, E. and Heise, L. (2019) 'Sexual coercion, consent and negotiation: Processes of change amongst couples participating in the Indashyikirwa programme in Rwanda', *Culture, Health and Sexuality*. Taylor & Francis, 21(8), pp. 867–882. doi: 10.1080/13691058.2018.1521991.
- Stern, E. and Niyibizi, L. L. (2018) 'Shifting perceptions of consequences of IPV among beneficiaries of Indashyikirwa: An IPV prevention program in Rwanda', *Journal of Interpersonal Violence*, 33(11), pp. 1778–1804. doi: 10.1177/0886260517752156.
- Stöckl, H. *et al.* (2014) 'Intimate partner violence among adolescents and young women: Prevalence and associated factors in nine countries: A cross-sectional study', *BMC Public Health*, 14(1). doi: 10.1186/1471-2458-14-751.
- Szasz, I. (1998) 'Masculine identity and the meanings of sexuality: A review of research in Mexico', *Reproductive Health Matters*, 6(12), pp. 97–104. doi: 10.1016/S0968-8080(98)90012-X.
- Temmerman, M. (2014) 'Research priorities to address violence against women and girls', *The Lancet*, 385(9978), pp. e38–e40. doi: 10.1016/S0140-6736(14)61840-7.
- Testa, A. C. and Coleman, L. M. (2006) 'Accessing research participants in schools: A case study of a UK adolescent sexual health survey', *Health Education Research*, 21(4), pp. 518–526. doi: 10.1093/her/cyh078.
- The World Bank (2013) *Women, business and the law 2014: Removing restrictions to enhance gender equality*. London: Bloomsbury Publishing.
- Thompson, D. S. *et al.* (2016) 'Scoping review of complexity theory in health services research', *BMC Health Services Research*. *BMC Health Services Research*, 16(1). doi: 10.1186/s12913-016-1343-4.
- Thomson, R. (2007) 'The qualitative longitudinal case history: Practical, methodological and ethical reflections', *Social Policy and Society*, 6(4), pp. 571–582. doi: 10.1017/s1474746407003909.
- Thomson, R. and McLeod, J. (2015) 'New frontiers in qualitative longitudinal research: An agenda for research', *International Journal of Social Research Methodology*, 18(3), pp. 243–250. doi: 10.1080/13645579.2015.1017900.
- Toomey, R. B., McGuire, J. K. and Russell, S. T. (2012) 'Heteronormativity, school climates, and perceived safety for gender nonconforming peers', *Journal of Adolescence*. Elsevier Ltd, 35(1), pp. 187–196. doi: 10.1016/j.adolescence.2011.03.001.

- Tronco Rosas, M. A. (2012) *No sólo ciencia y tecnología... Ahora, el IPN a la vanguardia en perspectiva de género*. Mexico City. Available at: <http://www.genero.ipn.mx/Conocenos/Documents/MemoriaPIGPG.pdf>.
- Tronco Rosas, M. A. and Ocaña López, S. (2012) *Género y Amor: Principales aliados de la violencia en las relaciones de pareja que establecen estudiantes del IPN*. Mexico City. Available at: http://www.genero.ipn.mx/Materiales_Didacticos/Documents/ARTICULO3BCD.pdf.
- UNAIDS (2004) *Seen but not heard ... Very young adolescents aged 10-14 years*. Available at: https://www.unicef.org/Seen_but_not_heard-Draft.pdf.
- UNFPA (2015) *The evaluation of comprehensive sexuality education programmes: A focus on gender and empowerment outcomes*. New York. Available at: <https://www.unfpa.org/publications/evaluation-comprehensive-sexuality-education-programmes>.
- United Nations Children's Fund (2015) *UNICEF Data: Monitoring the situation and of children and women*. Available at: https://www.unicef.org/infobycountry/mexico_statistics.html.
- United Nations Educational Scientific and Cultural Organization (2018) *International technical guidance on sexuality education: An evidence-informed approach*. Paris. Available at: <http://unesdoc.unesco.org/images/0026/002607/260770e.pdf>.
- United Nations Educational Scientific and Cultural Organization (2019) *Behind the numbers: Ending school violence and bullying*. Paris. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000366483> (Accessed: 3 November 2019).
- UNODC (2019) *Global Study on Homicide*. Available at: <https://www.unodc.org/documents/data-and-analysis/gsh/Booklet1.pdf>.
- Unzelman, A. (2011) 'Latin America update: The development of same-sex marriage and adoption laws in Mexico and Latin America', *Law and Business Review of the Americas*, 17(1), p. 135.
- Verma, R. K. et al. (2008) *Promoting gender equity as a strategy to reduce HIV risk and gender-based violence among young men in India, Horizons Final Report*. Washington, D.C. Available at: <http://menengage.org/wp-content/uploads/2014/06/Promoting-Gender-Equity-as-a-Strategy.pdf>.
- Vincent, K. A. (2013) 'The advantages of repeat interviews in a study with pregnant schoolgirls and schoolgirl mothers: Piecing together the jigsaw', *International Journal of Research and Method in Education*, 36(4), pp. 341–354. doi: 10.1080/1743727X.2012.705276.
- Wagman, J. A. et al. (2015) 'Effectiveness of an integrated intimate partner violence and HIV prevention intervention in Rakai, Uganda: Analysis of an intervention in an existing cluster randomised cohort', *The Lancet Global Health*, 3(1), pp. e23-33. doi: 10.1016/S2214-109X(14)70344-4.
- Walker, S., Read, S. and Priest, H. (2013) 'Use of reflexivity in a mixed-methods study', *Nurse Researcher*, 20(3), pp. 38–43. doi: 10.7748/nr2013.01.20.3.38.c9496.

Walters, V. (2004) 'The social context of women's health', *BMC Women's Health*, 4(Suppl 1), p. S2. doi: 10.1186/1472-6874-4-s1-s2.

Walton, M. (2014) 'Applying complexity theory: A review to inform evaluation design', *Evaluation and Program Planning*. Elsevier Ltd, 45, pp. 119–126. doi: 10.1016/j.evalprogplan.2014.04.002.

Weber, A. M. *et al.* (2019) 'Gender norms and health: Insights from global survey data', *The Lancet*, 393(10189), pp. 2455–2468. doi: 10.1016/S0140-6736(19)30765-2.

Wells, M. *et al.* (2012) 'Intervention description is not enough: Evidence from an in-depth multiple case study on the untold role and impact of context in randomised controlled trials of seven complex interventions', *Trials*, 13, pp. 1–17. doi: 10.1186/1745-6215-13-95.

Westhorp, G. (2012) 'Using complexity-consistent theory for evaluating complex systems', *Evaluation*, 18(4), pp. 405–420. doi: 10.1177/1356389012460963.

White, H. (2014) 'Current Challenges in Impact Evaluation', *European Journal of Development Research*. Nature Publishing Group, 26, pp. 18–30. doi: 10.1057/ejdr.2013.45.

Wong, K. K. *et al.* (2014) 'Why is school closed today? Unplanned K-12 school closures in the United States, 2011-2013', *PLoS ONE*, 9(12). doi: 10.1371/journal.pone.0113755.

World Health Organization/London School of Hygiene and Tropical Medicine (2010) *Preventing intimate partner and sexual violence against women: Taking action and generating evidence*, World Health Organization. Geneva: World Health Organization. Available at: http://www.who.int/violence_injury_prevention/publications/violence/9789241564007_eng.pdf.

World Health Organization (2013) *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. Geneva. Available at: <https://www.who.int/reproductivehealth/publications/violence/9789241564625/en/>.

Wu, Y. *et al.* (2020) 'Intimate partner violence: A bibliometric review of literature', *International Journal of Environmental Research and Public Health*, 17(15), pp. 1–17. doi: 10.3390/ijerph17155607.

Youdell, D. (2005) 'Sex-gender-sexuality: How sex, gender and sexuality constellations are constituted in secondary schools', *Gender and Education*, 17(3), pp. 249–270. doi: 10.1080/09540250500145148.

Young, H. *et al.* (2016) 'Sexual Health questions included in the Health Behaviour in School-aged Children (HBSC) Study: An international methodological pilot investigation', *BMC Medical Research Methodology*, 16(1), pp. 1–12. doi: 10.1186/s12874-016-0270-8.

Zimmerman, C. *et al.* (2016) 'Rigged or rigorous? Partnerships for research and evaluation of complex social problems: Lessons from the field of violence against women and girls', *Journal of Public Health Policy*, 37(1), pp. S95–S109. doi: 10.1057/s41271-016-0006-3.

Appendices

APPENDIX 1: RESEARCH QUESTIONS TO EXPLORE THE PATHWAYS SET OUT IN THE THEORY OF CHANGE FOR MEXFAM'S COMPREHENSIVE SEXUALITY EDUCATION (CSE) PROGRAMME	222
APPENDIX 2: FINAL PROGRAMME THEORY OF CHANGE, INCORPORATING STUDY FINDINGS (SPANISH)	223
APPENDIX 3: FOCUS GROUP DISCUSSION GUIDE FOR STUDENTS.....	225
APPENDIX 4: BASELINE AND ENDLINE QUESTIONNAIRE FOR STUDENTS.....	228
APPENDIX 5: ILLUSTRATIVE EXCERPT FROM LONGITUDINAL QUALITATIVE ANALYSIS MATRIX	237
APPENDIX 6: ILLUSTRATIVE EXCERPT FROM CONSOLIDATED CHANGE MATRIX FOR LONGITUDINAL QUALITATIVE ANALYSIS	240
APPENDIX 7: ETHICS APPROVAL DOCUMENTS	243
APPENDIX 8: DELAYS TO DATA COLLECTION AND IMPLICATIONS ON THE RESEARCH	251

Appendix 1: Research questions to explore the pathways set out in the theory of change for Mexfam's comprehensive sexuality education (CSE) programme

Research questions related to the curriculum

- Do the CSE participants increase their level of knowledge of key course messages (the different types of IPV; the implications of IPV on health, relationships, families and communities; gender equality; sexual diversity; the characteristics of equitable and unequal relationships) during the CSE course?

Research questions related to participation and experiences in the CSE sessions

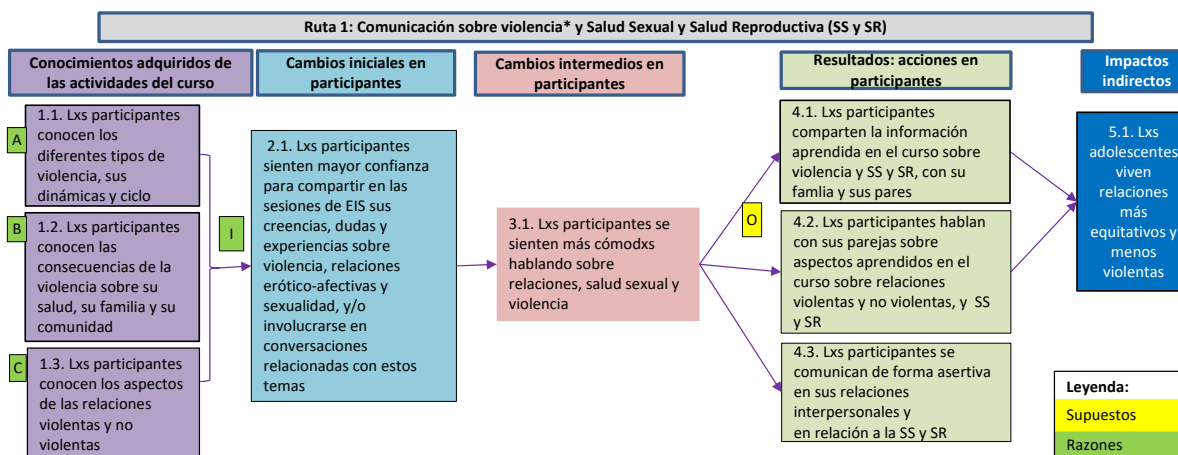
- How much of the CSE curriculum are participants receiving?
- Do participants appear to pay attention to the content that is being presented in the sessions?
- Do the participants appear to understand the information in the curriculum?
- In what ways do the participants appear to reflect on key course themes (the types of behaviours that comprise IPV; equitable and inequitable behaviours and gender norms and roles; what it means to express your sexuality; the importance of using assertive communication in relationships; the importance of building equitable relationships) in class? Do participants appear to identify aspects of these themes in their own life, in their family, in their relationships, in their peers and in their context?
- Do the participants appear to develop assertive communication skills?
- Do participants appear to learn new ways to build more positive and equitable relationships with their partners and/or peers?

Research questions related to intermediate changes in participants related to the CSE course

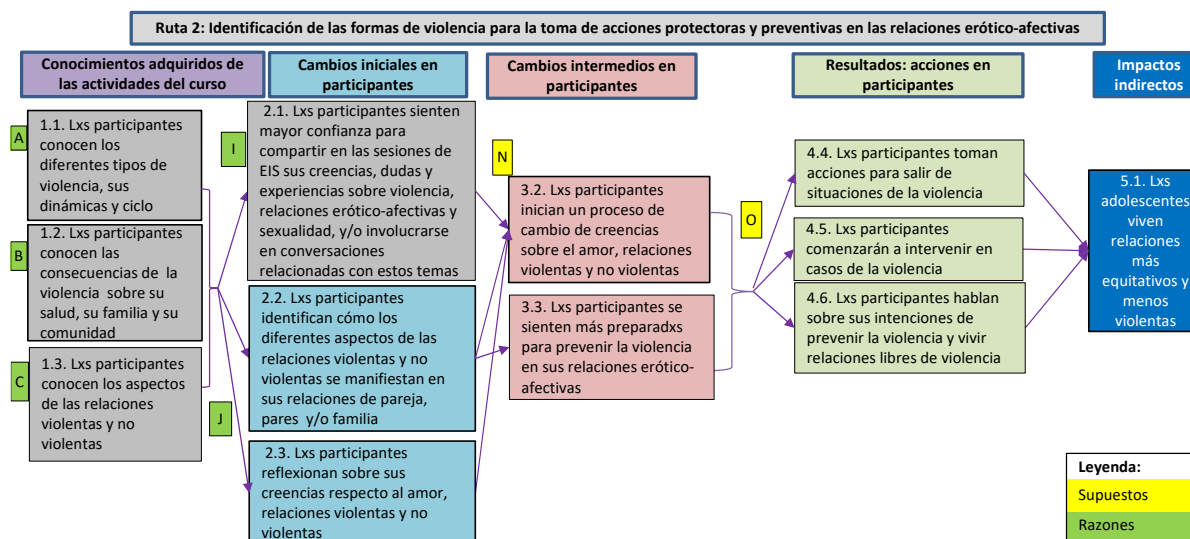
- Do the participants appear to begin to engage a process of change in their beliefs and attitudes in relation to the key course topics (the acceptability of IPV; gender roles and gender equality; expression of sexuality; equitable relationships) that is related their participation in the intervention? How does this process of change occur?
- Do the teachers perceive their students as more critical regarding IPV and gender inequality due to their participation in the intervention?
- Do the participants appear to shift their self-acceptance in terms of their own sexuality and the sexuality of others due to their participation in the intervention?
- Do participants appear more prepared to form more equitable and less violent partnerships due to their participation in the intervention?

Appendix 2: Final programme theory of change, incorporating study findings (Spanish)

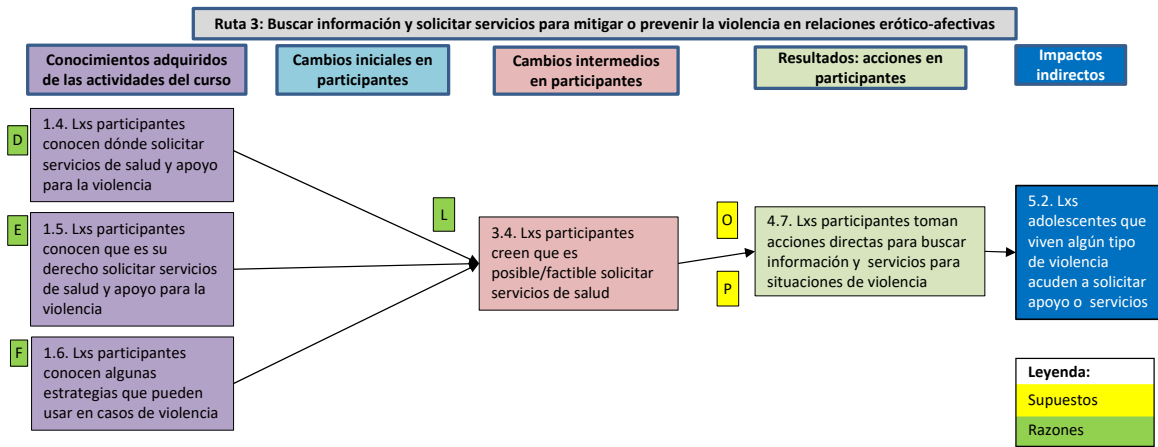
Theory of change pathway 1: Communicating about relationships, sexuality and violence



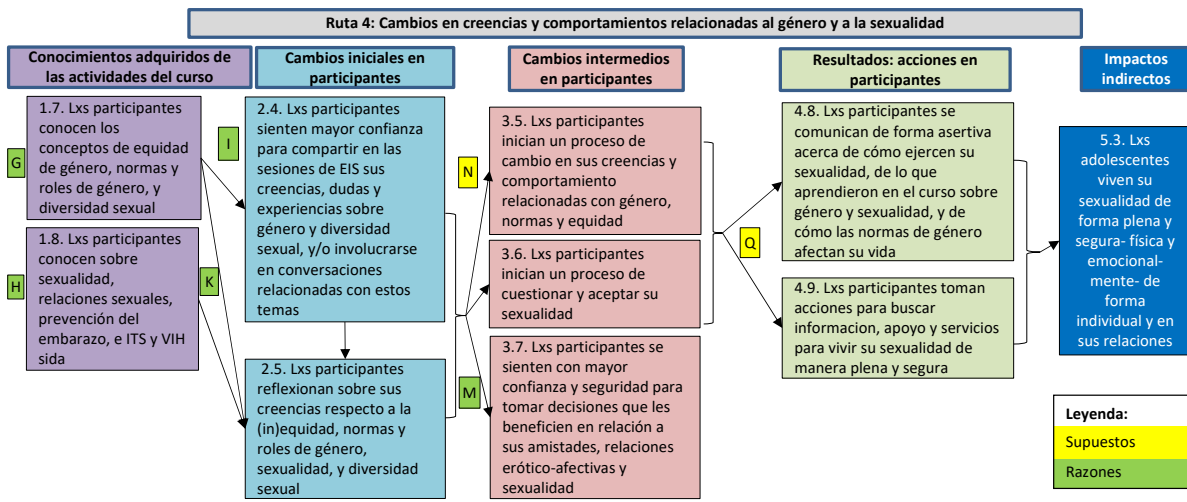
Theory of change pathway 2: Taking protective and preventative actions to promote equitable and less violence relationships



Theory of change pathway 3: Accessing violence-related and sexual and reproductive health services



Theory of change pathway 4: Shifting attitudes and behaviours related to gender, sexuality and violence



Appendix 3: Focus group discussion guide for students

Focus Group Guide: Students

[In cases when the information sheet was reviewed in another moment before the focus group, the entire introduction should be read out loud. In cases when the information sheet was just reviewed, skip the first paragraph and read only the last two paragraphs.]

Introduction

Thanks to all of you for being here. My name is _____ (moderator name) _____ and s/he (note taker) is _____. We are trying to learn about young people’s experiences with their friendships and dating, and with sexual and reproductive health topics. In today’s conversation, we invite you to openly discuss your opinions. What we hear from you today will be used to improve school programs that support young people to take care of their health and well-being. It is very important for us to hear your thoughts and opinions, since all of you are the experts in these topics and in the experiences that young people have.

During the conversation, [name of note taker] will take notes and will remind me if I forget to ask something. But s/he can’t write down everything, which is why we would like to record the conversation, so that we don’t lose the details of what you say. Only members of the research team will listen to it. Do you give me permission to record this interview?

Please put your cell phones on silent mode during our conversation, try to let everyone participate, and please respect one another’s point of view and ideas. Lastly, we ask that everyone keep what we talk about today confidential. Does everyone agree to this pact of confidentiality? We’d like to remind you that we are in a group, so please only share your experiences if you feel comfortable doing so. Finally, we ask that if you share any experiences of other people, including people who aren’t here, please don’t use their names to respect their privacy.

The session will last approximately 60-90 minutes. Are there any questions before we begin?

I. Introductions

Before we begin recording, we can start by introducing ourselves – please say your name and your favourite fruit. *[Go around the table and each person speaks].*

Now we will start recording.

II. Opening questions (approximately 10 minutes)

1. Young people sometimes use different terms like “boyfriend/girlfriend”, “friends with benefits,” etc. to refer to two people who go out together and have a relationship beyond friendship.
 - Which of these types of relationships are most common among young people your age?
 - [Of the most common that are mentioned:] What does each one mean?

[For the most common types, one by one:] From what you have seen at the school, at what age do people start having these types of relationships?

- What do you think about this? What age is too early and why? What age is too late and why?

Moving forward, we will talk about different situations that can happen in a relationship, and we ask that when you talk about relationships, please think of and discuss all the different types we just discussed.

2. What do you all think is expected of women and of men in the following contexts?
 - School: What is expected from men and women, is it similar or different, when is it different and when not, can you share examples?
 - Home: What is expected from men and women, is it similar or different, when is it different and when not, can you share examples?
 - Relationships: What is expected from men and women, is it similar or different, when is it different and when not, can you share examples?
3. What do you think about the different rules that sometimes exist for daughters compared to for sons? In your opinion, should there be different rules for daughters than for sons? When yes and when no, and why?

Do you have other settings you'd like to discuss?

III. Dating violence (40 minutes approximately)

Now we are going to talk about the different types of violence that can exist in dating relationships.

4. What types of violence can exist between a couple / people dating? [give examples from the things they mention, without using names to protect confidentiality].
5. In the following examples, what opinions do you think that the other students at the school, or other people you are close to, have about a situation where one person is violent with their partner? Why do you think that they believe that? Have you heard something like that? Has anyone said something to you?
 - If the violent person is under a lot of stress
 - If they are drunk
 - If they violent person repents after hurting someone
 - If the partner admits that they had sex with someone else
 - If the partner ends or tries to end the relationship.
6. Imagine this situation: Another student at the school is in a violent relationship
 - In general, do you all think that when young people of your age are in this kind of situation, they talk to anyone about it? Why or why not? In what cases would they talk about it, and in what cases not? Why do you think sometimes they don't talk to anyone about it?
 - In your opinion, what should the young person do if they are in a violence relationship? If they were in a formal relationship, versus being friends with benefits, would it be different?

- If the young person talked to you about their situation, what would you tell them? What would you do? Why?

[Probe:] Is there any place you would recommend they go to ask for help and counselling? [Probe]: clinic, Mexfam, other programs.] Why?

7. What do you think of a situation when one person in the relationship wants to have sex and the other does not? What can happen in this situation?

[Probe:] Would it be different in a steady relationship compared to a casual one? Why? If the person paid for dinner, in that case what would happen, and why? Or what if they plan to get married? What if it's their first date?

8. Before this semester, did you ever receive information or education about sexuality, contraceptives, relationships, etc?

- What topics did you discuss?
- Who were the educators?
- How many sessions were they?

IV. Recommendations [This section only for school intervention] (5 minutes approximately)

Now we would like to ask you what you think about the CSE course you received from Gente Joven this semester.

9. In your opinion, what did you like most from all the CSE content? Why? Can you give examples?

- What did you like less? Why? Can you give examples?
- What additional things would you want us to include in the CSE course?

10. Do you have any other recommendations about how we can improve the CSE course?

V. Closing [This section for all groups] (5 minutes approximately)

We are almost done with the conversation.

11. Out of everything we spoke about today, which topics are top priorities for you?

- What else?

12. In your opinion, what can be done to prevent dating violence among young people?

13. Is there anything else that you would like to share or ask before we finish the session?

Finally, we'd like to ask again that you keep everything that was said in this discussion confidential.

Thank you for sharing your experiences and ideas with us

[Distribute thank you gift, thank them for their time, and invite them to come to Mexfam for services if they need help.]

Appendix 4: Baseline and endline questionnaire for students

Individual PRE-POST Survey

SEMESTER you are studying: _____

GROUP: _____

SEX: _____

AGE: _____

TECHNICAL COURSE YOU WOULD LIKE TO COMPLETE: _____ [pre only]

WHO WAS THE GENTE JOVEN EDUCATOR WHO LED YOUR GROUP? _____ [post only]

IMPORTANT – READ BEFORE ANSWERING
Remember that if you don’t want to participate, you can leave the survey blank and no one will know.

Every time you see a gray box, read the instructions carefully. Please remember that we will not share your responses with anyone.

Please respond alone to the questionnaire, without looking at the answers of the other students in your class.

In the following section, **make an (X)** inside the parenthesis of the option that corresponds to your answer. **Mark only one correct response for each question.**

	1. Does this method have a high level of protection against pregnancy?			2. How sure are you that you know how to use this method?
a. IUD (intrauterine device)	Yes	No	I don’t know	I don’t know how to use it / more or less sure/ very sure
b. Contraceptive injections	Yes	No	I don’t know	I don’t know how to use it / more or less sure/ very sure
c. Implant (chip)	Yes	No	I don’t know	I don’t know how to use it / more or less sure/ very sure
d. Patches	Yes	No	I don’t know	I don’t know how to use it / more or less sure/ very sure
e. Male condom	Yes	No	I don’t know	I don’t know how to use it / more or less sure / very sure

In the following section, **make an (X)** inside the parenthesis of the option that corresponds to your answer. **Mark only one correct response for each question.**

3	How often should you take a contraceptive pill?	<input type="checkbox"/> Once a week <input type="checkbox"/> Once daily <input type="checkbox"/> I don’t know
----------	---	--

4	When is the emergency contraceptive pill most effective?	<input type="checkbox"/> Immediately after having unprotected sex <input type="checkbox"/> Between 1 and 3 days after unprotected sex <input type="checkbox"/> After the 3 rd day following unprotected sex <input type="checkbox"/> I don't know
5	On what part of the body is an implant placed?	<input type="checkbox"/> In the leg <input type="checkbox"/> In the arm <input type="checkbox"/> I don't know
6	To protect yourself during anal sex, should you use a condom with men as well as with women?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> I don't know
7	Can you use a female condom at the same time as you use a male condom?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> I don't know

In the following section, **make an X in the parenthesis** next to the option that corresponds to your answer.

8. Which reactions in your body or your genitals can appear if you have a sexually transmitted infection? You can mark all that apply.	<input type="checkbox"/> Secretions <input type="checkbox"/> Ulcers <input type="checkbox"/> Burning <input type="checkbox"/> Bumps on the skin <input type="checkbox"/> Fever <input type="checkbox"/> Itching <input type="checkbox"/> Swelling and pain <input type="checkbox"/> Bad odour <input type="checkbox"/> Sometimes there are no effects <input type="checkbox"/> I don't know
9. What is the name of the virus that, if left untreated over time, can cause AIDS?	Write your answer here:
10. How can you transmit the virus that causes AIDS? Please mark all that apply.	<input type="checkbox"/> Through sexual contact <input type="checkbox"/> In the air <input type="checkbox"/> Using public toilets <input type="checkbox"/> Blood transfusions <input type="checkbox"/> Using needles that are not sterilized <input type="checkbox"/> From an insect bite <input type="checkbox"/> During a pregnancy if the woman has the virus that causes AIDS <input type="checkbox"/> Through daily contact with someone who has AIDS <input type="checkbox"/> I don't know
11. What can one do to reduce the risk of getting a sexually transmitted infection? Please mark all that apply.	<input type="checkbox"/> Use a condom <input type="checkbox"/> Use an implant <input type="checkbox"/> Ask your partner to be loyal

	<input type="checkbox"/> Not have sexual relations <input type="checkbox"/> There is nothing you can do <input type="checkbox"/> I don't know
--	---

In the following section, **make an X in the parenthesis** next to the option that corresponds to your answer.

12. Have you ever gone out with someone, as in a dating relationship? It could be with a boyfriend, girlfriend, friends with benefits, etc.

- Yes
- No
- I would rather not say

13. Are you currently going out with someone, as in a dating relationship? It could be with a boyfriend, girlfriend, friends with benefits, etc.

- Yes
- No
- I would rather not say

14. Read the following things that young people like you might do on the weekend.

Sleep late, Go to the movie theatre, Go to a party, Wash your clothes

Of these 4 things, HOW MANY have you done? Please only mark the total number, but do NOT tell us which you have done.

- 0
- 1
- 2
- 3
- 4
- Prefer not to say

15_a. Read the following things that can happen in different kinds of dating relationships.

We kissed, We experienced violence, We were angry at each other, We got married

Of these 4 things, HOW MANY have you experienced in your own relationships (past or present)? Please only mark the total number, but do NOT tell us which you have done.

- 0
- 1
- 2
- 3
- 4
- Prefer not to say

15_b. Read the following things that can happen in different kinds of dating relationships.

We kissed, We were angry at each other, We got married

Of these 3 things, HOW MANY have you experienced in your own relationships (past or present)? Please only mark the total number, but do NOT tell us which you have done.

- 0
- 1
- 2
- 3
- Prefer not to say

16_a. Read the following things that can happen in different kinds of dating relationships.

We held hands, We have lived together, Sometimes they didn't respond to my text messages
Of these 3 things, HOW MANY have you experienced in your own relationships (past or present)? Please only mark the total number, but do NOT tell us which you have done.

- 0
- 1
- 2
- 3
- Prefer not to say

16_b. Read the following things that can happen in different kinds of dating relationships.

We held hands, We experienced violence, We have lived together, Sometimes they didn't respond to my text messages
Of these 4 things, HOW MANY have you experienced in your own relationships (past or present)? Please only mark the total number, but do NOT tell us which you have done.

- 0
- 1
- 2
- 3
- 4
- Prefer not to say

17. With who would you say that you prefer to have a dating relationship? It could be serious, casual, friends with benefits, etc.

- With someone of the same sex as me
- With someone of the opposite sex as me
- With people of both sexes
- I don't know
- I would rather not say

18. Do you think the majority of your **female** friends have already had sex?

- Yes
- No
- I would rather not say

19. Do you think the majority of your **male** friends have already had sex?

- Yes
 No
 I would rather not say

20. Have you ever had sexual contact? For example, touching or being touched on the genitals, anal sex, vaginal sex or oral sex?

- Yes
 No
 I would rather not say

21. Have you ever felt agreed to have sexual contact with someone even though you didn't want to?

- Yes
 No
 I would rather not say

22. In the following section, read each statement and **make an X** in the box that best reflects your opinion.

	Totally Agree	Somewhat Agree	Disagree
a. Due to their nature, men need to have more sex than women.			
b. Pregnant adolescents and young women who are already mothers, should keep studying.			
c. Women should be understanding and sensitive.			
d. Men, as well as women, should be responsible for housework, for example: changing diapers every day, making breakfast, washing the dishes, etc.			
e. Men are violent by their nature.			
f. I would have a homosexual friend.			
g. Parents should give more freedom to their sons than to their daughters.			
h. Men will only go as far as women want them to.			

In the following section, **make an X** next to the option that corresponds to your answer.

23. As far as you know, has there been any type of dating violence at your school? It could be among boyfriend/girlfriend, friends with benefits, etc).

- Yes
- No
- I would rather not say

24. Have you ever experienced dating violence in your own relationship (past or present)?

- Yes
- No
- I would rather not say

25. Have you ever heard of, seen or experienced sexual harassment in your school? For example, whistles, sexual advances, inappropriate looks, unwanted touching, kissing, pressure to have sex through threats, etc.

- Yes
- No
- I would rather not say

26. In the following section, read each statement and **make an X** in the box that best reflects your opinion.

Dating violence can be justified if	Totally Agree	Somewhat Agree	Disagree
a. the partner admits to having sex with someone else.			
b. the partner ends or tries to end the relationship.			
c. the violent person apologizes after being violent.			
d. the violent person was under a lot of stress.			
e. the violent person was drunk.			

27. In the following section, read each statement and **make an X** in the box that best reflects your opinion.

	Totally Agree	Somewhat Agree	Disagree
a. Rapes are a result of the inability of men to control their sexual desires.			
b. Domestic violence (in intimate relationships) is a private situation that must be resolved by the family.			
c. Most women could leave a violent relationship if they really wanted to.			
d. A woman can not be raped by someone with whom she has an intimate (dating) relationship			

28. Below, **mark with an X** all of the actions in the list that represent types of violence that can take place in a dating relationship (any type of dating relationship).

- ___ Ignore someone / give the cold shoulder
- ___ Trust someone
- ___ Forbid someone from talking to someone
- ___ Ridicule / Offend someone
- ___ Read all of someone's texts and email
- ___ Call someone to say hello
- ___ Play hit someone
- ___ Push/Pull someone
- ___ Hurtful jokes about someone
- ___ Aggressive touching of someone
- ___ Negotiate with someone
- ___ Blackmail someone
- ___ Lie / Cheat someone
- ___ Pinch / Scratch someone
- ___ Watch over / Spy on someone
- ___ Sexual abuse of someone
- ___ Jealousy of someone
- ___ Blame someone
- ___ Humiliate someone in public
- ___ Intimidate / Threaten someone
- ___ Kick someone
- ___ Admire someone
- ___ Destroy someone's personal things
- ___ Grope someone
- ___ Demand from someone
- ___ Slap someone

29. In your opinion, is partner or dating violence common in Mexico?

- Yes
- No
- I don't know

30. Do you know where you can go for help in case you experience violence, including rape and dating violence?

- No
- Yes (specify: _____)

In the following section, read each statement and **make an X** in the box that best reflects your opinion.

31. Which of these do you consider yourself?

- Not at all religious
- A little religious
- Very religious

32_pre. Have you ever received education or sessions about sexual and reproductive health? For example, about relationships, pregnancy, sexual transmitted infections.

- Yes
- No
- I don't remember/I don't know

If yes: Who gave you the information and on what topics?

32_post. How many sessions of the Gente Joven CSE program did you attend?

- All (10 or more sessions)
- Almost all (7-9 sessions)
- Some (4-6 sessions)
- Few (3 or less)

33_post. On a scale of 1 to 5, how satisfied are you with each aspect of the course?
Mark with an X the option that best reflects your response.

	1 Very satisfied	2 Satisfied	4 Not that satisfied	5 Unsatisfied
--	----------------------------	-----------------------	--------------------------------	-------------------------

The course in general				
The parts about violence				
The parts about gender				
The parts about contraceptives				
The parts about sexually transmitted infections				
The parts about sexual and reproductive rights				
The parts about pregnancy				
The parts about life plans				
The way that the course was facilitated				

34_post. What did you like most about the course? Why?

35_post. What did you like least about the course? Why?

Thank you for your time

Appendix 5: Illustrative excerpt from longitudinal qualitative analysis matrix

Lizbeth	Interview 1: Sept 15 2017	Interview 2: Oct 13 2017	Interview 3: Dec 8 2017	Interview 4: Feb 16 2018	Change
Comfort discussing	49:25 with best friend comfortable talking about everything, but not as much with others. 53:26 doesn't talk about SRH with family; doesn't trust mother not to tease her- her mother wouldn't let her have HPV vaccine because its for sexually active girls.	3:20 feels more comfortable talking to best friend than mother, because has more trust.	41:48 her mom tells her not to get pregnant and ruin her life, but doesn't teach her how, and she won't talk to her mom about having sex, doesn't think she will support it.	talks to boyfriend about methods	Consistently says more comfortable talking to best friends, not with mom or others who have less trust. This aspect doesn't seem to be much changed or influenced by course over time. Mom disapproves of the course.
Information and care-seeking	43:05 sought info online about contraception - on FB, google. 45:54 - not in this school, but in old school asked teacher about sex. 51:49 said was walking with friends and one went to a pharmacy to ask about morning after pill, just for sake of gathering info.	32:16 found info about condoms online. 44:17 also looked into info about the implant.		44:14 shared info about methods with her bf, openly discussed.	doesn't talk much about seeking services or info or changes in how she did that related to course. She was already looking up info only and sexually active before course.

Gender norms	5:39 thinks her mom defends and gives more liberty to her brother. Lizbeth is the one who stays home to take care of grandma. 22:05 her mom assumes she will behave same way as her brother, restricts her.	27:50 in CSE talks about double standard where boys can date two people, but girls are sluts if they do. 28:12 she thinks machismo will never end.	2:19 brother is the favourite. 6:43 brother gets to go out, she doesn't - he has more freedom and preferential treatment.	8:07 after describing fights with brother and mother, how she does all the chores and he doesn't, is asked if this has to do with gender roles.	She doesn't seem to have shifted beliefs about gender norms in the course. She spent most of the time in the interview talking about preferential treatment at home, and how it angered her.
Sexual diversity	n/a	n/a	n/a	n/a	n/a
Relationships and how to behave	16:20 she would like her boyfriend to be faithful, she likes to be faithful to him.	24:41 doesn't like too much sweet talk in relationships, but wants them to be loving.	25:12 back with bf, but still doesn't trust him. 30:30 said boys more likely to trick girls- if a boy cheats, their friends accept that, but if a girl is cheating, they get called a slut by the boy's friends.	30:38 she is getting at how with her boyfriend she is taking him less seriously. 31:24 said CSE helped her see relationship differently, respect herself more.	She talked about the course teaching her to respect herself and her decisions, which helped her think differently about her relationship- if he cheated and didn't respect her, was it love?
Violence	29:04 defines violence as a lack of physical respect, hitting if someone tells you not to. She thinks violence starts with insults, mistreatment.	5:43 said it bothered her when boy she was going out with used <i>groserias</i> , thinks people should talk more respectfully.		17:20 when asked, said thinks the way brother treats her is violent [verbally abusive, calls her puta, etc.]	Didn't talk a lot about this or changing about it

Jealousy	33:31 she has logged into her boyfriend's messages before. 34:14 her bf let her see his phone, she saw messages, she said people lie to you and mislead you.	18:05 said again that she has looked at her boyfriend's messages, but also, he reviewed her phone and FB messages.		39:33 asked directly, said thinks jealousy is a form of violence. Said she is jealous too, it starts violence, and that she is violent in that way.	Didn't talk a lot about this, but mentioned only after course that jealousy is a form of violence, and then said jealousy and insecurity go hand in hand.
Group dynamics		36:14 thinks male classmates take CSE as a joke, but women take it more seriously.			Didn't talk about shift in dynamics much, maybe a few less machista comments.
Impact			35:06 describes shift in machista behaviour of a male classmate who assigned all responsibility for pregnancy to women. This aligns with stories other interviewees told us].	29:37 said again that male classmate(s) became less machista in how they talk- some more than others- particularly example of one [J.]. 42:32 said without the course wouldn't have learned to respect herself and her own decisions.	Same as above, two main shifts are classmate saying less machista things, and her learned to respect herself and her own decisions.

Appendix 6: Illustrative excerpt from consolidated change matrix for longitudinal qualitative analysis

	Lizbeth	Gerardo
Comfort discussing	consistent, doesn't seem to be much changed or influenced by course.	He starts off saying he is comfortable talking about SRH and relationships with people he really trusts, but over the course he says he becomes more comfortable with it and now sees it like any other topic- and he also starts to disseminate the information he learns in CSE to others, even to consider replicating the course with a church group. He also tells his parents about the topics. He also observes classmates becoming more mature in their discussions of SRH, joking less.
Information and care-seeking	doesn't talk much about seeking services or info or changes in how she did that related to course. She was already looking up info only and sexually active before course.	He starts off - either before or early in the class- seeking information online, for example about condoms, and asking sisters for advice about relationships and sexuality, but as the course progressed, he has more in-depth discussions with friends about using condoms, contraception, avoiding sexually transmitted infections. He also seeks support / counselling at Mexfam but doesn't go. He knows of one girl who went to Mexfam for support for a controlling relationship. And he generally feels that the course provided tools to help young people seek care.
Gender norms	She doesn't seem to have shifted beliefs about gender norms in the course. She spent most of the time in the interview talking about preferential treatment at home, and how it angered her. she used term ' <i>machista</i> ' even before the course, and it doesn't much seem like she changed beliefs, though the course seems to be somewhat related to the topic, and may have helped her think a bit about it. The change on this topic that she did note multiple times was in a male classmate who made <i>machista</i> comments and then was called on it and stopped over time, she said he changed. She felt that this was good, and surprising- and she talked about it in multiple interviews. A Gente Joven educator also discussed this and some other interviewees, it might be a good case study to highlight in the article?	When he talks about gender, he often refers to sexual diversity/his sexuality, also to gendered norms. He took that part of the course seriously, and said it reiterated that having a penis doesn't mean you need to want to date women. It seems the course helped him reflect on his sexual preferences and come more to terms with them.
Sexual diversity	Nothing mentioned	Same as above- plus said that he used to see his sexuality as strange or a sin and now he accepts it more. Also observed that a classmate became more accepting of sexual diversity, thinks classmates took the topic well.

<p>How to behave in relationship</p>	<p>she talked about the course teaching her to respect herself and her decisions, which helped her think differently about her relationship- if he cheated and didn't respect her, was it love? And open herself up to maybe being with someone else. Less feelings of needing to stay with him. She also says when asked that jealousy is a form of violence, and in that way, she is a violent person, but she contextualised it by saying that it is linked with insecurity, and her boyfriend cheated on her and made her insecure which stoked the jealousy. so, while she was on and off with him in the interviews, it seems that she gained perspective and began to consider her own needs more through the influence of the course.</p>	<p>Says the course helped him learn more clearly how one should treat a partner, what is good and bad behaviour, such as jealousy being bad, controlling behaviour too, fighting isn't cool. He talked in Interview 1 about being proud of having had healthy relationships in the past - but as the other interviews took place, he disclosed some past relationships that were not healthy and where he didn't live up to his own standards. He feels clearer on how to behave in relationships now.</p>
<p>Violence</p>	<p>Didn't talk a lot about this or changing about it</p>	<p>Both of his parents have clearly educated him to reject partner violence- especially towards women. Also, he has had some form of presentation at school about violence last year- with <i>Violentómetro</i>- so this isn't new information. In the first interview he talks about violent <i>machista</i> men who drink, so he has clearly been exposed to that idea. His stories seem to favour the man as perpetrator of violence in Interview 1. He is affected by violence sessions, considers it very pertinent to their age group, has recall of the course contents months after the course ends. He thinks the course may have shifted some participants' minds about jealousy, how to behave in relationships, and did shift things for him- he now thinks jealousy and fighting are bad.</p>
<p>Jealousy</p>	<p>Didn't talk a lot about this, but mentioned only after course that jealousy is a form of violence, and then said jealousy and insecurity go hand in hand, said she herself is violent in this way but that it was built through her boyfriend cheating on her and not respecting her.</p>	<p>While he begins in Interview 1 saying that he rejects jealousy and tries to avoid it, over the course of the interviews he talks about how his perception of jealousy has shifted. He seems, in recollection, to see more of a shift in rejecting jealousy, though even in Interview 1 he already strongly rejected it. May have been what he wanted to communicate in the interviews- but then his deeper feelings shifted further? Either way he clearly has reflected on jealousy a lot through the course and seems to have shifted his views on it.</p>
<p>Group dynamics</p>	<p>Didn't talk about shift in dynamics much, maybe a few less <i>machista</i> comments.</p>	<p>Some classmates took it seriously, many didn't. Lots of disruptive comments as noted by others. Doesn't talk about concrete shifts in the dynamic in the group.</p>

<p>Impact on others</p>	<p>Same as above, two main shifts are classmate saying less <i>machista</i> things, and her learned to respect herself and her own decisions.</p>	<p>He talks about his own shifts- more confident, sure of his sexuality, reflections about jealousy and how to act in relationships. Also, about observations of classmates- more comfortable talking about SRH, taking it seriously, recognising controlling behaviour as violence and seeking support for it, raising consciousness generally and about sexual diversity.</p>
--------------------------------	---	---

Appendix 7: Ethics approval documents

The following documents are included in the following pages:

1. Approval from the LSHTM Research Ethics Committee, granted 6 January 2017.
2. Approval from *Comité de Bioética y Ciencia Para la Investigación, Centro de Investigación Clínica Acelerada (CICA)* in Mexico, granted 23 December 2016.
3. Renewal of approval from *Comité de Bioética y Ciencia Para la Investigación, Centro de Investigación Clínica Acelerada (CICA)* in Mexico, dated 27 January 2018.

London School of Hygiene & Tropical Medicine

Keppel Street, London WC1E 7HT
 United Kingdom
 Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

**Observational / Interventions Research Ethics Committee**

Ms Shelly Makleff
 LSHTM

6 January 2017

Dear Shelly,

Study Title: A process evaluation examining the link between comprehensive sexuality education and intimate partner violence prevention in Mexico City

LSHTM Ethics Ref: 11864

Thank you for responding to the Observational Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Investigator CV	Marston short CV Sept 2016 IPPF	30/09/2016	1
Investigator CV	Makleff Shelly CV 2016	19/10/2016	1
Information Sheet	4 En Assent form IDIs and case studies students Oct 23	23/10/2016	1
Information Sheet	5 En Consent form for GJ promoters and coordinators 18 years and older Oct 23	23/10/2016	1
Information Sheet	7 En Consent and assent form for GJ promoters under 18 October 23	23/10/2016	1
Protocol / Proposal	1 En In-depth interview guide teachers	23/10/2016	1
Protocol / Proposal	2 En Focus group Guide Educators and coordinators	23/10/2016	1
Protocol / Proposal	3 En Focus Group Guide Students	23/10/2016	1
Information Sheet	2 En Assent form pre post survey Oct 23	24/10/2016	1
Protocol / Proposal	5 En Case study follow up interview guide	24/10/2016	1
Advertisements	11 En Recruitment contact sheet Oct 25	25/10/2016	1
Protocol / Proposal	4 En Case study initial interview guide	28/10/2016	1
Protocol / Proposal	6 En In-depth interview guide students	28/10/2016	1
Protocol / Proposal	7 Pre post test EN	29/10/2016	1
Information Sheet	3 En Assent form FG students Oct 30	30/10/2016	1

Information Sheet	6 En Information letter for parents of GJ promoters under 18 Oct 30	30/10/2016	1
Information Sheet	10 En Info sheet for FGD promoters and educators Oct 30	30/10/2016	1
Information Sheet	12 En Info sheet for students pre post survey CSE Oct 30	30/10/2016	1
Information Sheet	13 En Info sheet for students pre post survey control Oct 30	30/10/2016	1
Information Sheet	14 En Info sheet for student FG Oct 30	30/10/2016	1
Information Sheet	15 En Info sheet for student IDIs Oct 30	30/10/2016	1
Information Sheet	16 En Info sheet for case studies Oct 30	30/10/2016	1
Information Sheet	1 En Information letter for parents Oct 30	30/10/2016	1
Information Sheet	8 En Consent form IDI teachers Oct 30	30/10/2016	1
Information Sheet	9 En Info sheet for IDI teachers Oct 30	30/10/2016	1
Protocol / Proposal	En Protocol for LSHTM ethics Oct 30	30/10/2016	1
Protocol / Proposal	7 En Pre post test V2	09/12/2016	2
Information Sheet	1 En Information letter for parents V2	09/12/2016	2
Information Sheet	2 En Assent form pre post survey V2	09/12/2016	2
Information Sheet	3 En Assent form FG students V2	09/12/2016	2
Information Sheet	4 En Assent form IDIs and case studies students V2	09/12/2016	2
Information Sheet	5 En Consent form for GJ promoters and coordinators 18 years and older V2	09/12/2016	2
Information Sheet	6 En Information letter for parents of GJ promoters V2	09/12/2016	2
Information Sheet	7 En Consent and assent form for GJ promoters under 18 V2	09/12/2016	2
Information Sheet	8 En Consent form IDI teachers V2	09/12/2016	2
Information Sheet	9 En Info sheet for IDI teachers V2	09/12/2016	2
Information Sheet	10 En Info sheet for FGD promoters and educators V2	09/12/2016	2
Information Sheet	12 En Info sheet for students pre post survey CSE V2	09/12/2016	2
Information Sheet	13 En Info sheet for students pre post survey control V2	09/12/2016	2
Information Sheet	14 En Info sheet for student FG V2	09/12/2016	2
Information Sheet	15 En Info sheet for student IDIs V2	09/12/2016	2
Information Sheet	16 En Info sheet for case studies V2	09/12/2016	2
Covering Letter	Response to LSHTM ethics committee	12/12/2016	1

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: <http://leo.lshtm.ac.uk>

Additional information is available at: www.lshtm.ac.uk/ethics

Yours sincerely,



Professor John DH Porter
Chair

ethics@lshtm.ac.uk
<http://www.lshtm.ac.uk/ethics/>

Improving health worldwide



COMITÉ DE ÉTICA EN INVESTIGACIÓN

Ciudad de México a 23 de Diciembre de 2016.

M.S.P. Shelly Makleff
Investigador Principal
Fundación Mexicana para la Planificación Familiar, A.C.
Federación Internacional de Planificación Familiar/Región del Hemisferio Occidental
(IPPF/RHO)
Universidad de Londres (London School of Hygiene and Tropical Medicine)
Presente

Una evaluación de proceso que analiza la relación entre la Educación Integral en Sexualidad y la prevención de la Violencia en el noviazgo en la Ciudad de México

Estimada M.S.P. Makleff

Por medio de la presente le informo, que el Comité de Ética en Investigación (Comité de Bioética y Ciencia para la Investigación) del Centro de Investigación Clínica Acelerada, SC ha revisado los documentos del protocolo de referencia y ha determinado dar su APROBACIÓN Y AUTORIZACIÓN a los documentos que a continuación se enlistan:

Documentos de asentimiento y consentimiento

1. Carta informativa para padres
2. Asentimiento para encuesta (Cédula individual) con estudiantes
3. Asentimiento para grupo focal con estudiantes
4. Asentimiento para entrevistas (A profundidad y estudios de caso) con estudiantes
5. Consentimiento para grupo focal con coordinadoras y promotoras (es) Gente Joven (GJ) mayores de edad
6. Carta informativa para padres de promotoras (es) GJ menores de edad
7. Consentimiento y asentimiento para promotoras (es) GJ menores de edad
8. Consentimiento para entrevistas a profundidad con profesores
9. Hoja de información para entrevistas a profundidad con profesores
10. Hoja de información para grupo focal con coordinadoras y promotoras (es) GJ
11. Documento de información de contacto para reclutamiento
12. Hoja de información para encuesta (cédula individual) escuela EIS
13. Hoja de información para encuesta (cédula individual) escuela control
14. Hoja de información para grupo focal con estudiantes
15. Hoja de información para entrevistas a profundidad con estudiantes
16. Hoja de información para estudios de caso con estudiantes

Protocolo e instrumentos

1. Protocolo
2. Guía de entrevistas a profundidad para profesores
3. Guía de grupo focal para coordinadoras y promotoras (es) GJ
4. Guía de grupo focal para estudiantes
5. Guía inicial de estudios de caso para estudiantes



6. Guía subsecuente de estudios de caso para estudiantes
7. Guía de entrevistas a profundidad para estudiantes
8. Cédula individual para estudiantes

El estudio se ha autorizado para que sea conducido en las instalaciones de la Fundación Mexicana para la Planificación Familiar, A.C., ubicada en Juárez 208, colonia Tlalpan, Delegación Tlalpan, CDMX, C.P. 14000.

Al mismo tiempo, se le informa que deberá entregar informes semestrales en el que se reporte los avances del estudio y renovar su aprobación a más tardar el 18-Noviembre-2017.

Sin más por el momento, le envío un cordial saludo.

Atentamente,

Sra. Guadalupe Guzmán Espinoza
Vocal Secretario
Comité de Ética en Investigación
Comité de Bioética y Ciencia para la Investigación



CENTRO DE INVESTIGACIÓN CLÍNICA ACELERADA, SC

COMITÉ DE ÉTICA EN INVESTIGACIÓN



Ciudad de México a 27 de Enero de 2018.

M.S.P. Shelly Makleff
Investigador Principal
Fundación Mexicana para la Planificación Familiar, A.C.
Federación Internacional de Planificación Familiar/Región del Hemisferio Occidental
(IPPF/RHO)
Universidad de Londres (London School of Hygiene and Tropical Medicine)
Presente

Una evaluación de proceso que analiza la relación entre la Educación Integral en Sexualidad y la prevención de la Violencia en el noviazgo en la Ciudad de México

Estimada M.S.P. Makleff

Por medio de la presente le informo, que el Comité de Ética en Investigación, del Centro de Investigación Clínica Acelerada, SC certifica que en la reunión ordinaria del 27 de Enero de 2018 se ha re-evaluado la documentación del estudio de referencia y ha determinado **Renovar la Aprobación** del mismo, para que continúe siendo conducido en las instalaciones de la Fundación Mexicana para la Planificación Familiar, A.C., ubicada en Juárez 208, colonia Tlalpan, Delegación Tlalpan, CDMX, C.P. 14000.

Documentación contemplada:

V.7 14/08/2017. Protocolo

• **Instrumentos**

V.2 2017 06 20. Guía grupo focal para profesores

V.2 2017-06-20. Guía grupo focal coordinadores y promotores

V.2 2017-06-20. Guía grupo focal estudiantes

V.2 2017-06-20. Guía inicial estudio de caso

V.2 2017-06-20. Guía subsecuente estudio de caso



V.3 2017-08-14. Guía entrevista a profundidad estudiantes

V.3 2017-08-14. Cédula

- **Consentimientos, asentimientos y hojas de información**

V.3 14 agos 2017. Carta de información para padres

V.2 2017-06-20. Asentimiento para encuesta (Cédula individual) con estudiantes

V.2 2017-06-20. Asentimiento para grupo focal con estudiantes

V.2 2017-06-20. Asentimiento para entrevistas (A profundidad y estudios de caso) con estudiantes

V.2 2017-06-20. Consentimiento para grupo focal con coordinadoras y promotoras (es) Gente Joven (GJ) mayores de edad

V.1 2016-10-28. Carta informativa para padres de promotoras (es) GJ menores de edad

V.2 2017-06-20. Consentimiento y asentimiento para promotoras (es) GJ menores de edad

V.2 20 jun 2017. Formulario de consentimiento para grupo focal con profesores

V.2 20 jun 2017. Hoja de información para profesores

V.2 8 dic 2016. Hoja de información grupo focal coordinadores

V.1 2016-10-25. Documento de información de contacto para reclutamiento

V.2 2017-06-20. Hoja de información para encuesta (cédula individual) escuela EIS

V.2 2017-06-20. Hoja de información para encuesta (cédula individual) escuela control

V.3 25 ene 2017. Hoja de información para estudiantes grupo focal

V.3 25 ene 2017. Hoja de información para entrevistas con estudiantes

V.3 25 ene 2017. Hoja de información estudios de caso



CENTRO DE INVESTIGACIÓN CLÍNICA ACELERADA, SC

COMITÉ DE ÉTICA EN INVESTIGACIÓN



Al mismo tiempo, se le exhorta a que continúe entregando informes semestrales en el que se reporte los avances del estudio.

La presente renovación tiene una vigencia del 27 de Enero de 2018 al 27 de Enero de 2019.

Sin más por el momento, le envío un cordial saludo.

Atentamente,



Lic. María Verónica García Flores
Presidente
Comité de Ética en Investigación

Appendix 8: Delays to data collection and implications on the research

	Originally planned event	Interruptions or delays	Implications
Baseline questionnaire implementation			
Week of August 28, 2017	Parent meetings (intervention group)	One of the two parent meetings interrupted by heavy rains and flooding, causing delays and preventing some parents from arriving	One meeting as scheduled. Second meeting cancelled. Paperwork sent home to parents who missed session; school staff followed up with parents to ensure receipt.
	Baseline questionnaire (three intervention groups)		As scheduled.
Week of September 4, 2017	CSE course started for all three intervention groups		As scheduled.
	Parents meetings (comparison group)	One of the three parent meeting delayed by torrential rains, then interrupted by evacuation of building due to earthquake drill.	Two meetings as scheduled. One was cancelled and rescheduled to Friday.
	Baseline questionnaire (three comparison groups)	Earthquake on Thursday September 7.	Rescheduled to following week.
	Rescheduled parent meeting (from early cancellation due to rains and earthquake)	School closed on Friday for building inspection after earthquake.	Rescheduled meeting cancelled again. Paperwork sent home to parents who missed session due to school closure; school staff followed up with parents to ensure receipt.

Week of September 11, 2017	Rescheduled baseline questionnaire (two comparison groups) due to earthquake		Data collected one week later than scheduled.
Week of September 18, 2017	Baseline questionnaire (third comparison group)	Earthquake on Tuesday September 19. School closures for rest of the week pending building inspection.	Delayed for undetermined time until school reopens.
	Ongoing CSE implementation through end of semester	School closure pending building inspection.	CSE delayed for undetermined time until school reopens.
Week of September 25, 2017	Ongoing CSE implementation through end of semester	School closure all week.	CSE delayed for undetermined time until school reopens.
Week of October 2, 2017	Ongoing CSE implementation through end of semester	School closure all week. Received building inspection, school scheduled to reopen on October 9.	CSE delayed for undetermined time until school reopens.
Week of October 9, 2017	Ongoing CSE implementation through end of semester		Course resumed after 3 weeks of school closures.
	Baseline questionnaire implemented in third comparison class.		Data collected 5 weeks later than scheduled.
Endline questionnaire implementation			

<p>Week of December 11, 2017</p>	<p>Ongoing CSE implementation through end of semester</p>	<p>Once school reopened after earthquake, the administration decided to extend the semester by two weeks. Mexfam planned to recuperate hours and deliver full CSE curriculum. Three weeks before end of semester, the administration reversed the decision and decided to extend by only one week.</p>	<p>Endline questionnaire implemented in all three intervention groups (Monday- Wednesday) and all three comparison groups (Wednesday-Friday). Implemented concurrent with student examinations immediately before the Christmas holidays. Students were focused on examinations and there were high levels of absenteeism (common at the end of the semester). This meant less time for the intervention, as we had to end the course early to implement the endline questionnaire.</p>
<p>Week of December 18, 2017</p>	<p>Endline questionnaire for intervention and comparison groups</p>	<p>Last minute change in semester dates (see above); school closed.</p>	<p>We had planned endline questionnaire implementation this week with both intervention and comparison groups. Instead, we had to implement these one week early.</p>