Contact Interventions in Intractable Conflicts:

Long-Term Attitude and Behavior Intention Changes in Israel

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften (Dr. rer. nat.)

dem

Fachbereich Psychologie der Philipps-Universität Marburg

(Hochschulkennziffer 1180) vorgelegt

von

Dipl.-Psych. Kerstin Guffler (geb. Hammann) aus Mainz

Marburg/Lahn im Mai 2016

Erstgutachter: Prof. Dr. Ulrich Wagner (Philipps-Universität Marburg)

Zweitgutachter: Prof. em. Dr. Peter Schmidt (Justus-Liebig-Universität Gießen)

Einreichungstermin: 02.05.2016; Prüfungstermin: 21.07.2016

Diese Arbeit wurde gefördert durch ein Promotionsstipendium der Philipps-Universität Marburg im Zusammenhang mit dem DFG-Graduiertenkolleg „Gruppenbezogene Menschenfeindlichkeit“ (GRK 884) an den Universitäten Marburg und Bielefeld, eine halbjährige Unterstützung durch die Arbeitseinheit Sozialpsychologie der Philipps-Universität Marburg, ein Reisestipendium der Marburg University Research Academy und ein Auslaufstipendium zur Auslauffinanzierung des Promotionsvorhabens am Fachbereichs Psychologie der Philipps-Universität Marburg.

Contact Interventions in Intractable Conflicts:

Long-Term Attitude and Behavior Intention Changes in Israel

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften (Dr. rer. nat.)

dem

Fachbereich Psychologie der Philipps-Universität Marburg

(Hochschulkennziffer 1180) vorgelegt

von

Dipl.-Psych. Kerstin Guffler (geb. Hammann) aus Mainz

Marburg/Lahn im Mai 2016

Erstgutachter: Prof. Dr. Ulrich Wagner (Philipps-Universität Marburg)

Zweitgutachter: Prof. em. Dr. Peter Schmidt (Justus-Liebig-Universität Gießen)

Einreichungstermin: 02.05.2016; Prüfungstermin: 21.07.2016

Diese Arbeit wurde gefördert durch ein Promotionsstipendium der Philipps-Universität Marburg im Zusammenhang mit dem DFG-Graduiertenkolleg „Gruppenbezogene Menschenfeindlichkeit“ (GRK 884) an den Universitäten Marburg und Bielefeld, eine halbjährige Unterstützung durch die Arbeitseinheit Sozialpsychologie der Philipps-Universität Marburg, ein Reisestipendium der Marburg University Research Academy und ein Auslaufstipendium zur Auslauffinanzierung des Promotionsvorhabens am Fachbereichs Psychologie der Philipps-Universität Marburg.

Für meine Familie

TABLE OF CONTENT

I. INTRODUCTION 1

1. Intractable Conflicts 3

1.1. Definitions of Intractable Conflicts 3

1.2. Contact in the Context of Intractable Conflicts 6

2. Theoretical and Empirical Background 8

2.1. Contact Theory 8

2.1.1. Positive Contact 8

2.1.2. Negative Contact 9

2.1.3. Interrelation of Positive and Negative Contact 10

2.2. Intergroup Contact Interventions in Intractable Conflicts 11

2.2.1. Face-to-Face Intergroup Contact Interventions in Intractable Conflicts 11

2.2.2. Short-Term Intergroup Contact Intervention Effects in Intractable Conflicts 14

2.2.3. Long-Term Intergroup Contact Intervention Effects in Intractable Conflicts 14

2.2.4. Fading Intergroup Contact Intervention Effects in Intractable Conflicts 20

2.2.5. Repeated Intergroup Contact Interventions in Intractable Conflicts 20

2.2.6. Minority-Majority Differences in Intergroup Contact Interventions 22

2.2.7. Relevance of Positive Contact Experience in Intergroup Interventions 23

2.2.8. Excursus: Intergroup Contact Intervention Effects in Young Age Groups 24

2.3. Status Differences Between Jewish Israeli and Arab Israeli in Israel 26

3. Methodological Issues in Longitudinal (Quasi-)Experiments 28

3.1. Research Studies and Research Evaluations 28

3.2. Structural Equation Modeling With (Quasi-)Experimental Data 29

3.3. Longitudinal and Cross-Group Measurement Invariance Analysis 30

3.4. Hierarchical Data Structure Analysis 35

3.5. Missing Data Analysis 36

3.6. Causality Assumptions in (Quasi-)Experimental Analyses 38

4. Research Questions 39

II. CURRENT RESEARCH MANUSCRIPTS 44

1. Manuscript #1 44

1.1. Introduction to Manuscript #1 44

1.2. Submitted Manuscript #1 44

2. Manuscript #2 81

2.1. Introduction to Manuscript #2 81

2.2. Submitted Manuscript #2 81

III. ANCILLARY ANALYSIS 106

1. Assessment of the Israel-Gaza Conflict 2012 106

2. Analysis of the Relevance of Students' Intergroup Experience 110

2.1. Analysis of the Relevance of Students' Intergroup Experience of the Remaining Samples 110

2.2. Overview of all Results Regarding the Relevance of Students' Intergroup Experience 112

3. Measurement Invariance and Noninvariance Analysis 114
3.1. Measurement Invariance Analysis 115
3.2. Measurement Noninvariance Analysis 116
3.3. Overview of all Measurement Invariance and Noninvariance Analysis 120
4. Multiple Imputation as Alternative Missing Data Technique 122
5. Analysis of Gender Differences 125

IV. CONCLUSION AND DISCUSSION 126
1. Conclusion 126
2. Discussion 129
3. Limitations 133
4. Implications 135
5. Future Research 136

V. REFERENCES 139

ZUSAMMENFASSUNG 155

DANKSAGUNG 159

ANGABEN ZUR PERSON 161

ERKLÄRUNG DER AUTORIN 164

LIST OF TABLES

Table 1: Overview of Studies Analyzing Contact Intervention Effects in Conflicts With at Least Three Points of Measurement 16
Table 2: Characteristics of Israeli Population 28
Table 3: Characteristics of Maximum Likelihood Estimation (ML) and Multiple Imputation Analysis (MIM) 38
Table 4: Overview of all six Samples Analyzed Within the Thesis 42
Table 5: Overview of the Research Questions 43
Table 6: Israeli Public Opinion Before and After the Israel-Gaza Conflict 2012 109
Table 7: Overview of the Analyses of Students' Intergroup Experiences 113
Table 8: Overview of Measurement Invariance and Noninvariance Analyses 121
Table 9: Comparison of the Study Results Regarding two Missing Data Techniques 123

LIST OF FIGURES

Figure 1: Assessment of Tensions Between Jews and Arabs in 2012 107
Figure 2: Assessment of Tensions Between Jews and Arabs in 2013 108

I. INTRODUCTION

"Conflicts based in ethnic, religious, and racial differences continue to erupt around the world, despite decades of interventions and scholarly research."
(Tropp, 2012, p. 3)

As apparent in recent news, several conflicts between groups of different backgrounds exist. To name just a few: Hindus and Muslims in Kashmir ("Kashmir Struggle", 2015; "Kashmir Tensions", 2013), Nuer and Dinka in South Sudan ("Massacres S. Sudan", 2015), Jews and Muslims in Israel and Palestine (Benhaida, 2015), Kurds and Turks in Turkey and Iraq (Pope, 2014), and Greek and Turkish Cypriots in Cyprus (Sak, 2013). According to the Heidelberg Institute for International Conflict Research (HIIK; 2015a), there were 21 wars and 25 limited wars in the year 2014 adding up to a total of 46 severe conflicts.[1] The consequences of these conflicts for the populations are manifold and threaten physical integrity, well-being, safety, security, self-esteem and personal fulfillment, and sometimes even the possibility to satisfy basic human needs (e.g., food, water, shelter; Maiese, 2003, Segment 5). One way to improve intergroup relations was introduced by Allport (1954) in *The Nature of Prejudice.* In this book, he wrote that intergroup contact under certain conditions can reduced prejudice between conflicting groups. Several scholars extended this understanding empirically and discovered that contact in general can reduce prejudice respectively improve intergroup relations among conflicting groups. Especially for contact programs (i.e., intergroup interventions), positive effects on reduced prejudice and improved intergroup relations were discovered (e.g., Lemmer & Wagner, 2015; Pettigrew & Tropp, 2006). Thus, their results pointed out that intergroup relations can be influenced by intergroup contact starting on the individual level (i.e., through contact between individuals of different groups). Nevertheless, the quotation from Tropp (2012) at the beginning of this paragraph illustrates that existing interventions and scholarly research has not yet been able to stop the eruption of further conflicts.

Considering the high number of ongoing conflicts, their associated negative consequences for the involved populations, and existing knowledge about the impact of intergroup interventions, the question arises how interventions can be realized meaningfully in order to improve intergroup relations among conflicting groups. While there is strong

[1] Limited wars and wars are classified as violent conflicts with a high intensity (HIIK, 2015a). Not all of these conflicts are considered to be based on ethnic, racial, and/or religious differences.

evidence about positive short-term outcomes of contact interventions (Lemmer & Wagner, 2015), knowledge on how to implement interventions in a manner that enables sustainable, long-term effects is limited (Salomon, 2009b). Although, profound knowledge about short-term effects is crucial, longitudinal analyses are needed to improve knowledge about sustainable changes. The consciousness about the relevance of sustainable effects is recently increasing not only in contact research but also in almost every other area inside and outside scientific research. Two prominent examples are the agenda of the United Nations that recently started to focus on Sustainable Development Goals (GA Res 70/1) and the report of the Intergovernmental Panel on Climate Change (2014), which illustrates the impact of climate change on sustainable development.

This doctoral thesis follows Allport's approach and strives to increase theoretical and practical knowledge about intergroup contact interventions between conflicting groups. We focus on those interventions that are implemented in severe conflicts.[2] We build on and extend current contact sustainability research by investigating long-term contact intervention effects. Accordingly, we evaluate, in the context of the current intractable Israel-Palestine conflict, an intergroup contact intervention involving Jewish Israeli and Arab Israeli students between the age of 10 and 13.

The chapters within this thesis are structured as follows: Chapter I introduces background information about intractable conflict concepts and specific conflict aspects. The chapter also presents the theoretical background as well as up-to-date findings from contact research, points to methodical aspects that are relevant in longitudinal (quasi-) experimental studies, and outlines the research questions. Chapter II presents the majority of the thesis results, which can be found in our submitted manuscripts. Chapter III provides ancillary analyses adding to a deeper understanding of the findings and complementing the outcome of our manuscripts. Chapter IV illustrates the conclusion and discussion. Finally, Chapter V provides an overview of all references used in this thesis except for those references that have already been listed within the two submitted manuscripts.

[2] The term *we* refers either to the author of this thesis or to the authors of Manuscript #1 and Manuscript #2 (essentially related to the content of Chapter 2).

1. Intractable Conflicts

1.1. Definitions of Intractable Conflicts

The analyses of this thesis are based on the implementation of intergroup contact interventions that strive to improve attitudes and behavior intentions between conflicting groups in the context of severe conflicts. To obtain a deeper understanding about the concept of severe conflicts, we present several definitions of severe conflicts and related constructs and thereby show differences and similarities among the underlying concepts and the overall understanding. There is a wealth of publications that deal with definitions of severe conflicts (e.g., Azar, 1990; Bar-Tal, 2007; Burton, 1987; Coleman, 2003, 2006; Crocker, Hampson, & Aall, 2005; Kriesberg, 1998). Upon a closer examination of these explications, there is no congruent understanding among researchers; although, definitions include several similar components.

Intractable conflicts. In 1998, Kriesberg described intractable conflicts as long lasting (i.e., even longer than one generation), irreconcilable (i.e., most people who are involved perceive the conflict to be not solvable), violent (e.g., physically damaging), and entangled (i.e., some conflict members have interests in the continuation of the conflict). These aspects were complemented by Bar-Tal (1998) who identified three further components: Intractable conflicts are also total (i.e., existential for the affected parties), perceived as zero-sum in nature (i.e., any loss sustained by the other side is perceived as own gain and vice versa), and central (i.e., the conflict holds a central role in the lives of affected individuals and the respective societies). In 2005, Kriesberg clarified his former statement, highlighting that conflicts could be regarded as more or less intractable and not entirely intractable underlining that conflicts that are declared as being intractable are not determined as such over time.

In 2003, Coleman also used the term intractable conflicts and his understanding is based on similar phenomena. However, his descriptions are slightly less explicit about its components and focus more on its consequences. In his characterization, intractable conflicts are destructive, "persist for long periods of time and resist every attempt to resolve them constructively" (p. 533), involve many parties, and threaten both basic human needs and values. Typically, they result in negative outcomes including mutual alienation, disrespect, and violent atrocities such as homicides and genocides.

Protracted social conflict. In 1990, Azar declared that "the focus of these conflicts is religious, cultural, or ethnic communal identity, which in turn is dependent upon the

satisfaction of basic needs (…), most contemporary conflicts are about developmental needs expressed in terms of cultural values, human rights, and security" (p. 2; see also Cohen & Azar, 1981). Again, these conflicts include "hostile interactions which extend over long periods of time with sporadic outbreaks of open warfare fluctuating in frequency and intensity (…) [,] involve whole societies and act as agents for defining the scope of national identity and social solidarity" (Azar, Jureidini, & McLaurin, 1978, p. 50). Within Azar's explication of protracted social conflicts, similar aspects are listed as in the above mentioned description of intractable conflicts.

Deep-rooted conflicts. Burton (1987) emphasized that deep-rooted conflicts are an integral part of social relationships. In his definition, deep-rooted conflicts are not restricted to the ethnic, racial, or cultural level, but can also be found in the family and at work. Burton (1987) further adds that these conflicts

> cannot be settled by an order from some outside authority, such as a court, an arbitrator or a more powerful nation. These are conflicts which may seem endless, erupting into emotional displays and even violence from time to time, contained only by imprisonment or social, political and sometimes military pressures. (p. 3)

This concept of deep rooted conflicts differs from the above mentioned ones, because it can also be applied to smaller entities such as the family and work units, and because it is not necessarily related to ethnicity, race and/or religion.

Enduring rivalries. Goertz and Diehl (1993) stated that enduring rivalries are repeated conflict situations among specific states or, as Thies (2001) framed it, "the notion of an *enduring rivalry* requires us to examine interstate behavior as a time-dependent *process*, rather than as a series of discrete events" (p. 693). While enduring rivalries are by definition also long lasting, they explicitly only involve states as actors, which makes them stand out in comparison to the other mentioned definitions (intractable conflicts, protracted social conflicts, and deep-rooted conflicts). Beyond that, these rivalries are not described more explicitly.

Political conflict. According to the HIIK (2015b) political conflicts are *"positional difference between at least two assertive and directly involved actors regarding values relevant to a society (...) which (...) threaten core state functions, the international order, or hold the prospect of doing so (Paragraph 4)."* The HIIK refers to five levels of intensity in conflicts: disputes, non-violent crises, violent crises, limited wars, and wars. As in enduring rivalries, these levels focus on the macro level.

The illustration of these above mentioned definitions is not exhaustive; however, it includes some of the main concepts used in research that are related to intergroup relations literature. According to most of these definitions, the respective conflicts have a long duration, include forms of violence, and stress negative consequences for society; some additional characteristics are irreconcilability, the involved actors, and specific conflict components. Overall most definitions seem straightforward. However, categorizing a specific conflict to one of these definitions is inherently difficult, because explicit criteria are missing. For example, what do authors intend when defining severe conflicts as long lasting? Kriesberg (1998) wrote that the conflict has to persist over more than one generation. Goertz and Diehl (1993), however, talked about repetition of conflicts, and Azar (1990) stated that conflicts "continue to be pursued in the long term" (p. 2). Thus, no exact time span is offered within the definitions. The same holds true for all other aspects, such as violence, the way of suffering within the societies, and the concept of irreconcilability. However, a definition with more specific criteria might also not portray the phenomenon adequately because severe conflicts underlie unique dynamics, for example, they are more irreconcilable or more violent at one time than at another time. Therefore, the concept of a severe conflict and related concepts is more based on a common understanding than on a set of criteria that has to be fulfilled.

Within this thesis, we focus on intergroup conflicts that are based on different ethnic, religious, and/or racial backgrounds. These conflicts are located either on an interstate or on an intrastate level. In the following, we use the term intractable conflict, because this term is commonly used in intergroup contact literature and focuses on the fact that the conflict is not easy to solve at the moment (Crocker et al., 2005). We regard intractable conflicts as persistent, sporadically violent, existential for societal members, and currently not solvable. We focus specifically on the Jewish-Arab Conflict in Israel since our studies were located in this area. The Jewish-Arab Conflict fulfills many of these criteria on a regular basis and is therefore a good example for an intractable conflict.[3] The conflict is persistent; It can be traced back to the foundation of the State of Israel in the year 1945 and even further back in time and it is also sporadically violent, meaning there are repeated violent outbreaks between the groups, for example, the Six Day War or the Yom Kippur War (Peace Research Institute in the Middle East, 2003). The Jewish-Arab Conflict is existential for societal members as can be seen in the fact that in public polls Jewish-Arab Tensions are

[3] According to the HIIK (2015a) the Israel-Palestine Conflict was classified as a war.

considered to be the strongest tensions within the Israeli society (Hermann, Heller, Atmor, & Lebel, 2013a). Additionally, the conflict is currently not solvable, for example, the State of Israel does not acknowledge the State of Palestine until now.

1.2. Contact in the Context of Intractable Conflicts

Again, this thesis focusses on intergroup contact interventions that are located in intractable conflict areas. Thus, the above mentioned aspects have an influence on participants before, during, and after their participation in the intergroup intervention. This shows that intergroup interventions struggle against the difficult circumstances, which are inherent in a conflict area to achieve the goal of improved intergroup relations (e.g., Salomon, 2009b). Especially sustainable improvements of intergroup relations that should remain after the end of contact interventions seem to be hardly realizable against the background of this specific context situation (i.e., they also have to resist the impact of the contextual influence, which occurs after their participation in the intervention; e.g., Salomon, 2011). The underlying assumption that the context in which intergroup contact takes place has a substantial impact on individuals' outgroup prejudice was already confirmed by Christ et al. (2014) who demonstrated the impact of the context on prejudice in lower level conflict societies.

As stated above, intractable conflicts constitute specific challenging circumstances for successful contact interventions to reduce intergroup prejudice (e.g., Bar-Tal, 1990; Bar-Tal, Rosen, & Nets-Zehngut, 2010; Salomon, 2004, 2006). In order to enable a deeper understanding, we introduce these challenges in more detail. The challenging situation is usually based on both a long history of the conflict (e.g., Deutsche Gesellschaft für die Vereinten Nationen, 2011; Moltmann, 2014) and competing narratives about the conflict history from each group involved (e.g., Hammack, 2006; Peace Research Institute in the Middle East, 2006; Ron & Maoz, 2013; Roundtable, 2002). Jewish and Palestinian school textbooks, for example, describe narratives (historical understanding of the conflict) according to their respective perspective, depicting the outgroup in a negative and the ingroup in a positive way (Council of Religious Institutions of the Holy Land, 2013). Bar-Tal (2013) called this the *collective memory of intractable conflicts* and explained, "a common past provides a commonality and a continuation of experiences across time, which are crucial ingredients for group formation, survival, and identity construction" (p. 137). Additionally, Bar-Tal (2000) stated that intractable conflicts are challenging because they

comprise an *ethos* of conflict, which is a set of societal beliefs that provide an orientation for society and its understanding of the conflict at present and future. Thus, the past-directed collective memory is complemented by this ethos of a common understanding of the present and future. In his view, there are eight such societal beliefs (Bar-Tal, 1998, see also Bar-Tal, Raviv, Raviv, & Dgani-Hirsh, 2009): Societal beliefs about …the justness of one's own goals, … the delegitimization of the opponent, … victimization, … a positive collective self-image, … security, … patriotism, … unity, and … peace. Elcheroth, Doise, and Reicher, (2011) described a similar phenomenon called *social representations*. The authors combined four aspects: shared knowledge and collective elaboration of the world, a meta-knowledge (i.e., the assumption of individuals about the thinking of relevant others), enacted communication and the respective discourse about it, as well as world-making assumptions (i.e., realities are brought into existence due to these representations). Ethos of conflict and social representations both include a current understanding of the world, whereby Bar-Tal (1998, 2013) concentrates on specific content aspects and Elcheroth et al. (2011) focus on a general perception of world affairs. A further component of the challenging situation for intergroup intervention in intractable conflicts is the *collective emotional orientation* (Bar-Tal, 2013), meaning that societies cultivate emotions, such as hatred, fear, or anger which are essential for the stimulation, motivation, control, and interpretation of the dynamics of intractable conflicts. Some other authors argue in the same direction but refer to this phenomenon as *collective emotions* (e.g., Bar-Tal, Halperin, & Rivera, 2007; Halperin, Bar-Tal, Nets-Zehngut, & Drori, 2008). Niedenthal and Brauer (2012) stated that "for the group, emotion processes seem necessary for the creation and maintenance of group viability and for long-term commitment to action as that achieve the goals of the group" (p. 269; see also Frijda & Mequita, 1994). Literature shows, there are even further aspects in intractable conflicts that create a specifically challenging context for the success of intergroup relations by intergroup interventions. *Routinization*, for example, describes that the daily life of Israeli people is affected by the conflict. The term routinization refers to "expressions and symbols of the conflict" (Vered & Bar-Tal, 2014, p.44) Israeli people are exposed to on a daily basis such as the celebration of national holidays that are related to the conflict (e.g., celebration of the *Independence Day* in Israel and commemoration of the *Nakba Day* in Palestine), the practice of using bomb shelters, the perception of monuments that are related to the conflict, obligatory military duty, the exposure to information about the conflict through radio, television, newspaper, as well as

internet, or school policies which instruct students to use the bus in order to prevent walking unaccompanied to and from the school (e.g., Kilpatrick & Leitch, 2004; Vered & Bar-Tal, 2014). These daily life components provoke a permanent consciousness about the conflict and prohibit ordinary contact between conflicting groups. Further aggravating aspects are violent outbreaks such as air raids or suicide bombings, which occur from time to time and provoke considerably negative intergroup relations (Bar-Tal & Labin, 2001). Opportunities that enable positive interactions between the conflicting groups are also limited due to segregated residential areas and some types of informal segregations (e.g., spatial patterns of social interactions; Dixon & Durrheim, 2003). As shown, several conflict specific components influence people living in an intractable conflict area.

As stated earlier, the goal of the thesis is to offer new insights about long-term contact intervention effects in intractable conflicts. Therefore, the investigated interventions within this thesis were implemented in Israel so that the focus of this thesis is on this specific conflict area.

2. Theoretical and Empirical Background

2.1. Contact Theory

This thesis is based on the contact theory (originally referred to as contact hypothesis; Allport, 1954) which describes the phenomenon that intergroup contact can reduce prejudice between members of two opposing groups. While Allport (1954) did not differentiate between positive and negative intergroup contact situations but implied positive contact, current research revealed contrasting outcomes between contact valences. Given that negative contact is likely to occur in intractable conflicts (given all its specific challenges), we divide this chapter into three subchapters. Two subchapters present aspects and empirical findings of respectively positive and negative contact. A third subchapter illustrates the interrelation between the two.

2.1.1. Positive Contact

Allport's (1954) *The Nature of Prejudice* includes the following paragraph, which describes optimal conditions for intergroup contact to reduce prejudice between groups.

> Prejudice (…) may be reduced by equal status contact between majority and minority groups in the pursuit of common goals. The effect is greatly enhanced if this contact is sanctioned by institutional supports (i.e., by law, custom, or local atmosphere), and provided it is of a sort that leads to the perception of common interests and common humanity between members of the two groups. (p. 281)

Scholars mainly refer to these optimal conditions as the following four concepts: *equal status* (i.e., individuals meet on an equal footing within the contact situation), pursuit of *common goals* (i.e., individuals interact to achieve the same outcome), *cooperation* instead of competition between group members (i.e., joint efforts without competition), as well as *institutional support* (i.e., social sanctions, norms, support of authorities; e.g., Pettigrew, 1971, 1998). In 1998, Pettigrew supplemented these four conditions with a fifth condition, namely the *opportunity to become friends*. In this article Pettigrew (1998) pointed out that the opportunity to become friends is an important aspect within the intergroup contact situation because it enables "self-disclosure and further friendship-developing mechanisms" (p. 76). The above mentioned five conditions, if given, imply a friendly and favorable atmosphere and thus a positive contact situation between opposing groups. Moreover, researchers found that (positive) contact led not only to reduced prejudice toward the specific outgroup member that was involved in the contact situation but also generalized to other members of the same outgroup (i.e., primary transfer effect) and even to members of other outgroups (i.e., secondary transfer effect); indicating the high impact of positive contact (Pettigrew & Tropp, 2011). In 2006, Pettigrew and Tropp empirically supported the contact theory in their meta-analysis.

2.1.2. Negative Contact

While most research in the past has focused on the effects of positive intergroup contact (which is why until recently positive and negative contact has not necessarily been explicitly differentiated in theory), researchers currently also investigate intergroup contact in case of negative contact situations. Since negative contact is by now not explicitly integrated within the contact theory, insights into negative contact and its effects are explorative and not inferable from theory.[4] Nevertheless, the research strain is growing rapidly (e.g., Bekhuis, Ruiter, & Coenders, 2013; Graf, Paolini, & Rubin, 2014; Paolini, Harwood, & Rubin, 2010) and the latest state of the art should be introduced.

First indicators of negative contact were found in the meta-analysis of Pettigrew and Tropp (2011) who demonstrated that in 4% of the examined studies (21 out of 515 studies) the contact situation led to increased prejudice. However, whether these effects were caused by negative contact could not be investigated ex post facto. Researchers who examine

[4] Negative contact is already embedded in the integrated threat theory (Stephan, Stephan, & Gudykunst, 1999). However, this theory focuses more on threat than on increased empathy and enhanced knowledge that are related to prejudice reduction.

negative contact situations today mainly understand negative contact as a negative interaction between an outgroup member and an ingroup member. Stephan, Stephan, and Gudykunst (1999), for example, described negative contact as "disagreements, fights, losing team efforts, unpleasant intergroup activities" (p. 621). Thus, while positive contact leads to less prejudice, negative contact is related to increased prejudice. In the context of a conflict situation, for example, contact that takes place under unfavorable conditions is related to increased prejudice, which can intensify the conflict (Ben-Ari & Amir, 1986). Additionally, it was revealed that negative attitudes also generalize to the whole outgroup, equally to the primary transfer effect, which was found for positive contact (Stark, Flache, & Veenstra, 2013).

2.1.3. Interrelation of Positive and Negative Contact

Within previous studies it could be shown that negative and positive contact have independent effects on prejudice (e.g., Barlow et al., 2012; Christ, Ulrich, & Wagner, 2008; Graf et al., 2014). This illustrated that, although the wording suggests that positive and negative contact are two poles of one (measured) contact variable, they are not mutually exclusive, so that it is possible for contact experiences to consists of much/little positive and much/little negative contact (e.g., Techakesari et al., 2015). It is also not reasonable to assume that a positive contact situation (e.g., making friends) is the exact opposite of a negative one (e.g., fighting). Thus, in case positive and negative contact is manipulated within an experiment it cannot be concluded whether the positive situation is as substantive as the negative or vice versa. However, there is a heavy debate in research literature whether negative contact has a stronger effect than positive contact. Baumeister, Bratslavsky, Finkenauer, and Vohs (2001) illustrated that negative (bad) is stronger than positive (good) within several psychological phenomena. They wrote that "events that are negatively valenced (e.g., losing money, being abandoned by friends, and receiving criticism) will have a greater impact on the individual than positively valenced events of the same type (e.g., winning money, gaining friends, and receiving praise)" (p. 323). Within the intergroup contact context, some researchers assume that positive contact is a stronger predictor for prejudice than negative contact (Christ et al., 2008; Pettigrew & Tropp, 2011, Chapter 12) while others claim the opposite (e.g., Barlow et al., 2012; Graf et al., 2014). Since both positive and negative contact is given concurrently in real life, it seems worth to examine both contact situations (Pettigrew & Tropp, 2011).

In line with the idea that negative and positive contact is co-occurring, some researchers have already investigated the outcome in case both contact situations were given. As shown in some studies, the simultaneous occurrence of positive and negative contact situations resulted in a reduction of prejudice (e.g., Christ et al., 2008; Fell, 2014; Paolini et al., 2014; Pettigrew & Tropp, 2011). The authors suggested that positive contact might buffer negative contact effects, which would be given if negative contact occurred alone. Paolini et al. (2014), for example, speculated that previous positive or diverse contact experience might buffer against adverse effects of new, negative contact experiences. This finding is especially interesting concerning contact interventions in intractable conflict areas in which a hostile atmosphere between opposing groups is permanently given (Bar-Tal, 2000) and where unfavorable preconditions for intergroup contact exist and negative contact situations might occur more often than in no-conflict area.

In conclusion, by now scholars have shown that positive intergroup contact can reduce prejudice and buffer the effects of negative contact interactions, especially whenever the optimal contact conditions (Pettigrew, 1971, 1998) are given. Nevertheless, negative contact can increase prejudice and has an independent impact on prejudice, and thus worsens intergroup relations. Until now, it is neither known which conditions within negative contact situations especially generate prejudice nor whether extremely negative interactions can outperform buffer effects. However, Pettigrew and Tropp (2011) stated, "negative contact relates to raised prejudice largely in the absence of positive contact. This is an important but often ignored point: Positive contact can counter much of the harmful effect of negative contact" (p. 190). This quote indicates that in contexts in which negative contact exists, such as in intractable conflict areas, the realization of positive intergroup contact is essential to counteract increased prejudice and worsened intergroup relations among conflicting groups.

2.2. Intergroup Contact Interventions in Intractable Conflicts

2.2.1. Face-to-Face Intergroup Contact Interventions in Intractable Conflicts

Within this thesis, we only considered *face-to-face* interventions, meaning participants meeting each other physically. All other types of contact such as imagined (e.g., Turner, Crisp, & Lambert, 2007), extended (e.g., Wright, Aron, McLaughlin-Volpe, & Ropp, 1997), or virtual contact (e.g., Tavakoli, Hatami, & Thorngate, 2010) have not

been considered. Lemmer and Wagner (2015) discovered that direct (face-to-face) interventions have a high impact on the reduction of ethnic prejudice; however, the authors did not find differences between the effects of direct ($\hat{\mu}$ = .29; k = 63 number of comparisons) and indirect interventions[5] ($\hat{\mu}$ = .23; k = 16 number of comparisons).[6] Despite the fact that there is no explicit evidence to suggest a stronger efficacy of face-to-face contact, there are indicators that lead in this direction. As Fazio and Zanna (1981) declared, "attitudes formed through direct experience are stronger than those formed through indirect experience. There is evidence to suggest that direct experience attitudes are more clearly defined, held with greater certainty, more stable over time, and more resistant to counterinfluence" (p. 185). Thus, interventions in which members of conflicting groups meet face-to-face seem promising to achieve positive contact intervention outcomes. Therefore, we explicitly investigated face-to-face contact in this thesis.

In addition, we focused on *intergroup contact interventions* in intractable conflict areas. These interventions can be considered as a specific type of intergroup contact in which interventions are (actively) structured between members of opposing groups in specific geographical areas. Most often these interventions include (explicitly or implicitly) at least some of Allport's and Pettigrew's optimal conditions: Participants strive for common goals (e.g., get to know each other's narratives), the contact interventions enable equal status for the participants (e.g., guides of both groups are assigning the group tasks), participants work cooperatively (e.g., working on a task with an outgroup member), and the contact experience is supported by institutions (e.g., schools; Allport, 1954). Additionally, participants have the chance to meet with outgroup members and become friends (Pettigrew, 1998). Thus, existing intergroup contact interventions usually provide almost ideal circumstances for contact to reduce prejudice and improve intergroup relations between conflicting groups (Stephan, Hertz-Lazarowitz, Zelniker, & Stephan, 2004). In their meta-analysis, Pettigrew and Tropp (2006) also demonstrated that structured contact programs led to stronger prejudice reduction effects within racial and ethnic samples (r = -.26; k = 40)[7] than contact outside of any structured program (r = -.21; k = 322); however, this difference was not significant. When investigating nonracial and nonethnic

[5] Lemmer and Wagner (2015) combined extended and virtual contact as indirect contact.

[6] These results are related to studies including a control group design. Findings in studies without a control group design also showed no differences between direct ($\hat{\mu}$ = .41; k = 37 number of comparisons) and indirect intervention effects ($\hat{\mu}$ = .33; k = 5 number of comparisons).

[7] Pettigrew and Tropp (2006) refer to k as "number of samples associated with the mean effect size" (p. 756); Lemmer and Wagner (2015) indicate k as "number of comparisons" (p. 158).

samples, Pettigrew and Tropp (2006) revealed differences between outcomes in regard to structured programs (r = -.30; k = 94) and no programs (r = -.19; k = 240; $Q_B[1]$ = 19.67, p < .001).

The interventions analyzed in this thesis are located *in the field* (i.e., an intractable conflict area), meaning we did not conduct laboratory experiments. Although many researchers refer to the high standard of laboratory experiments (e.g., Falk & Heckman, 2009), Paluck and Green (2009) concluded in their overview of intervention studies that field interventions have the advantage to examine "whether an intervention's effects emerge and endure among the cacophony of real-world influences, including larger political and economic changes and proximal special pressures and distractions" (p. 184). Also Bar-Tal (2004) emphasized that the investigation of social behavior is incomplete when real world contexts are not included in the analysis, indicating the relevance of ecological validity.

In general, there is a large amount of literature that explores face-to-face intergroup contact interventions in conflict areas. These publications focus on various aspects: general overview of intervention studies (e.g., Maoz, 2010; Paluck, 2012; Paluck & Green, 2009), classification of interventions (e.g., Maoz, 2002, 2011; Ross, 2016), general outcomes of interventions such as prejudice reduction (e.g., Cuhadar, Genc, & Kotelis, 2015; Eshel, 1999; Eshel & Dicker, 1995; Halabi & Zak, 2006; Lazarus, 2011; Lazovsky, 2007; Maoz, 2000; Maoz, Bar-On, Bekerman, & Jaber-Massarwa, 2004; Sagy, 2002), status differences between the participants (e.g., Maoz, Steinberg, Bar-On, & Fakherelden, 2002), facilitators of contact interventions (e.g., Maoz, Bekerman, & Sheftel, 2007), the role of identity and narratives (e.g., Hammack, 2006, 2010; Pilecki & Hammack, 2014; Ron, Maoz, & Bekerman, 2010), and effects of bilingual or -cultural schools (e.g., Bekerman & Horenczyk, 2004; Hayes, McAllister, & Dowds, 2007).[8] Although there are many more publications on this topic and most consider not only one but several of the above mentioned aspects and therefore this classification is neither exhaustive nor distinct, the enumeration conveys an idea about the various facets of existing literature. In sum, only parts of these studies explicitly address the reduction of prejudice respectively the improvement of intergroup relations. Besides, most of these studies use qualitative or mixed-method analysis and only few are evaluated based on quantitative data.

[8] Bilingual or -cultural schools cannot be considered as contact intervention in the narrower sense, because their main goal is to teach students and not to make them like one another more. Nevertheless, we list them due to their broader relation to contact literature in intractable conflict areas.

2.2.2. *Short-Term Intergroup Contact Intervention Effects in Intractable Conflicts*

Research findings based on qualitative and quantitative data analysis have found that contact intervention led to positive intergroup relations in intractable conflict areas. For example, Nevo and Brem (2002) reviewed 79 studies qualitatively and concluded that about 80-90% of these studies showed a positive or partially positive effect. In their meta-analysis regarding existing quantitative studies, Lemmer and Wagner (2015) found five documents that focus on short-term contact intervention effects in intractable conflict areas. These documents revealed that contact interventions are related to an improvement of inter-group relations – comparing intervention group (IG) and control group (CG) participants ($\hat{\mu}$ = .31, k = 11 number of comparisons).[9] Thus, structured programs for intergroup contact improve intergroup relations even in these challenging contexts.

2.2.3. *Long-Term Intergroup Contact Intervention Effects in Intractable Conflicts*

Regarding long-term contact intervention effects, Salomon (2009a) stated that

> not many evaluations of peace education programs measure their impact beyond the 'morning after' effect. When measured immediately after the conclusion of a program, its effects are found to be positive. The picture is often different when long-term effects are measured. (p. 114)

There are some studies, which investigated long-term contact intervention effects in intractable conflict areas using quantitative methods by including one posttest in the data analysis (Connolly, 1992; Cuhadar, Genc, & Kotelis, 2015; Kamfer & Venter, 1994; Luiz & Krige, 1985; Malhotra & Liyanage, 2005).[10] In these studies posttests were conducted between one to twelve months after the end of the intervention; no direct posttest at the exact end of the intervention was implemented. This non-direct posttest is referred to as follow-up test, posttest, or posttest 2 indicating that there is no consistent terminology across researchers. However, without findings of a direct posttest, follow-up effects[11] could also be related to spontaneous changes or sleeper effects (i.e., effects that develop some time after the end of the intervention; Kumkale & Albarracín, 2004). Additionally, the development of the contact intervention effects from post- to follow-up test are not

---

[9] Lemmer and Wagner (2015) also revealed contact intervention effects within conflict areas in studies that did not use a control group design ($\hat{\mu}$ = .47, k = 20 number of comparisons; n = 9 studies).

[10] From the listed studies the conflict areas and the time spans between the end of the intervention and the occasions of measurement of the follow-up tests (posttests) are as follows: Connolly (1992): six weeks (South Africa); Cuhadar, Genc, and Kotelis (2015): less than a year (Greece and Turkey); Kamfer and Venter (1994): one month (South Africa); Luiz and Krige (1985): one year (South Africa); Malhotra and Liyanage (2005): one year (Sri Lanka).

[11] In this thesis, the word posttest refers to the occasion of measurement directly after the intervention whereas follow-up test means a second occasion of measurement conducted after the intervention.

examined in these studies. Because we were interested in these developments within this thesis (i.e., fading effects: decrease of positive posttest effects to the follow-up test), we especially considered studies including at least two occasions of measurement after the end of the intervention. An overview of existing contact intervention studies using a longitudinal design (including a pre-, post-, and follow-up test) is illustrated in Table 1. Studies were found using a PSYCINFO literature research.[12] We looked for studies that investigated contact interventions between different ethnicities, races, or cultures in an intractable conflict area using quantitative methods. Relevant dependent variables were related to prejudice and intergroup relations.[13] In addition, we screened studies mentioned in the reference lists of relevant articles and included studies recommended by researchers familiar with our topic.[14]

[12] The database PSYCINFO was browsed using terms related to six different components at the same time: 1. Intervention (intervent*, treat*, train*, workshop*, program*, camp*, encount*, contact*, meeting*, cooperat*, dialogu*) 2. Ethnicity, race, or culture (ethnic*, racial*, cultur*, Black*, White*, Color*) 3. Conflict (conflict*, war*, clash*, struggl*, fight*, intractab*) 4. Prejudice (prejudic*, attitud*, stereotyp*, behavior*, discrim*, coexist*) 5. Longitudinal data (long*, endur*, sustain*, constant*, perman*, last*, dura*) 6. Number of measurements (three*, four*, five*, follow*, post*, repeat*, multi*). All components were connected with *and* within each components *or* connectors were used. The literature research discovered 1008 documents and included all papers that were available at January 24th, 2016.

[13] Within this thesis we focused on attitude and behavior intentions specifically.

[14] We did not include (educational) approaches (i.e., integrated, multicultural schools) that enable contact but are not designed as an intervention to specifically improve intergroup relations.

Table 1: Overview of Studies Analyzing Contact Intervention Effects in Conflicts With at Least Three Points of Measurement

Reference	Contact Intervention				Participants			Methods		Reported Outcomes[d]
Authors (year)	Conflict	No. of Meetings	Duration per Meeting	Approx. Time Span t2-t3 in Month	Status	*N*	Approx. Age in Years	DV	CG Design	
Arnon (2010)[a,b]	Israel/ Palestine	1	2 days	2	Majority Minority	259	16-17	Attitudes toward the outgroup	NCG	▪ Combined outcomes for status groups ▪ t1-t2: Positive short-term effects ▪ t2-t3: Fading effects occurred
Bar-Natan, Rosen, and Salomon (2010)[b]	Israel/ Palestine	1	2 days	6	Majority Minority	110 100	-	Readiness for social contact	NCG	▪ Combined outcomes for status groups ▪ t1-t2: Positive short-term effects ▪ t2-t3: Fading effects occurred
Berger, Benatov, Abu-Raiya, and Tadmor (2016)[a]	Israel/ Palestine	12	4 hours	15	Majority Minority	159 163	9-11	▪ Readiness for social contact ▪ Neg. feelings about the other ▪ Neg. stereo-typing of the other	RCG	▪ Combined outcomes for status groups ▪ t1-t2-t3: Positive effects (between condition) regarding all dependent variables
Jayusi (2009)[a,b]	Israel/ Palestine	1	2 days	2	Majority Minority	120 120	15-18	Attitudes toward the outgroup	NRCG	▪ t1-t2: Positive short-term effects ▪ t2-t3: Fading effects occurred

Table 1: Overview of Studies Analyzing Contact Intervention Effects in Conflicts With at Least Three Points of Measurement (cont.)

Reference	Contact Intervention				Participants			Methods		Reported Outcomes[d]
Authors (year)	Conflict	No. of Meetings	Duration per Meeting	Approx. Time Span t2-t3 in Month	Status	*N*	Approx. Age in Years	DV	CG Design	
Kropiunigg and Pabst (2007)[a]	Israel/ Palestine	1	10 days	5	Majority	21	15-17	Belief in superiority of own ethnicity	NCG	▪ Combined outcomes for all investigated groups ▪ t1-t2-t3: Positive effects ▪ t1-t2: Positive short-term developments (Δ*M*) ▪ t2-t3: Positive effects (Δ*M*)
					Minority	38				▪ Combined outcomes for all investigated groups ▪ t1-t2-t3: Positive effects ▪ t1-t2: No short-term changes (Δ*M*) ▪ t2-t3: Positive effects (Δ*M*)
Schleien (2007)[c]	Israel/ Palestine	1	3.5 weeks	10	Majority	31	14-18	Closeness to outgroup	NCG	▪ t1-t2: Positive short-term effects (combined outcomes for all investigated groups ▪ t1-t3: Negative effects (minority)
					Minority	16				
Schroeder and Risen (2016) Intervention 2011	Israel/ Palestine	1	3 weeks	9-12	Majority	27	14-16	Positivity to outgroup (different items than in 2012)	NCG	▪ Combined outcomes for status groups ▪ t1-t2: Positive short-term effects ▪ t2-t3: Fading effects occurred ▪ t1-t3: Marginal long-term effects
					Minority	14				
Schroeder and Risen (2016) Intervention 2012	Israel/ Palestine	1	3 weeks	9-12	Majority	31	14-16	Positivity to outgroup (different items than in 2011)	NCG	▪ Combined outcomes for status groups ▪ t1-t2: Positive short-term effects ▪ t2-t3: Fading effects occurred ▪ t1-t3: No long-term effects
					Minority	4				

Table 1: Overview of Studies Analyzing Contact Intervention Effects in Conflicts With at Least Three Points of Measurement (cont.)

Reference	Contact Intervention				Participants			Methods		Reported Outcomes[d]
Authors (year)	Conflict	No. of Meetings	Duration per Meeting	Approx. Time Span t2-t3 in Month	Status	*N*	Approx. Age in Years	DV	CG Design	
Shani (2015)[b]	Israel/ Palestine	1	2 days	12	Majority	22	15-18	Overall coexistence orientation (different items in both status groups)	NRCG	▪ t1-t2: Positive short-term effects ▪ t2-t3: Fading effects occurred ▪ t1-t3: No long-term effects
					Minority	20				▪ t1-t2: No short-term effects ▪ t2-t3: Fading effects occurred ▪ t1-t3: Negative long-term effects

Note. t2-t3 = time span from the direct posttest (t2) to the follow-up test (t3); DV = dependent variable; CG = control group; NCG = no control group; NRCG = no randomized control group; RCG = randomized control group. [a] These studies deviate in regard to the selection criteria: Arnon (2010) and Jayusi (2009) are doctoral dissertations, which are written in Hebrew. Thus, the information is based on the English summaries of these doctoral dissertations and not explicitly on their quantitative findings; Berger et al. (2016) investigated a contact intervention including several separate informational sessions, so that results might not indicate contact intervention effects only; Kropiunigg and Pabst (2007) analyzed their data including groups from non-conflict areas, so that their findings might not be related to conflict areas specifically (therefore we additionally provide information about mean differences of the conflict groups). [b] These studies (probably) refer to dialogue encounters/peace education workshops, which were organized by the Jewish-Arab Center Givat Haviva; although the summary by Jayusi (2009) does not refer to Givat Haviva explicitly. [c] Schleien (2007) examined their data including three groups (Israeli, Palestinian, Non-Palestinian Arabs). [d] Due to dropouts, short-term effects are in some studies represented by a larger sample than long-term effects.

As shown in Table 1, only few studies have investigated long-term contact intervention effects (including a post- and a follow-up test) in intractable conflict areas using quantitative methods. The presented studies share some similarities, but are also different in many aspects. To elaborate on the similarities, most of the results are based on one-time contact interventions (i.e., participants meet one time for several days) and participants are between the age of 14-18 years (i.e., middle and late adolescents). In addition, many studies include either no control group or refer to a non-randomized control group design. Additionally, all of these studies investigate contact intervention effects in the context of the Israel/Palestine conflict. There are also studies investigating long-term contact intervention effects with one only posttest (follow-up test), as mentioned in the beginning of this subchapter, and within these studies further conflict areas such as South Africa, Sri Lanka, or Greece and Turkey are examined. Main differences are in the used methods of the data analysis: some studies investigate overall effects across time and groups (experimental and control group or minority and majority), some refer to effects between pre- and follow-up tests, and some investigate differences between the post- and follow-up test. Therefore, results are hardly comparable. Further differences are distinguishable in the time span between post- and follow-up tests, which encompasses a time interval of 2-15 months from after the end of the intervention until the follow-up test. Though not outlined in Table 1, but also an interesting aspect, is that either no information about missing data was given or the method predominantly used to deal with missing data was listwise deletion.

The existing number of studies evaluating short- and long-term as well as fading effects of face-to-face intergroup contact interventions in intractable conflict areas is limited and the illustration of results in the field is heterogeneous. Thus, the thesis seeks to shed more light into this research gap and accordingly we investigated two intergroup contact interventions regarding long-term contact intervention effects in Israel.[15] Thereby, we focused on a detailed illustration of all effects: short- and long-term as well as fading effects.

[15] The Ein Dor Museum of Archaeology in Ein Dor (Israel) organized the two investigated contact interventions between Jewish and Arab Israeli. The program name is “Learning from the Past – Building Bridges Today”.

2.2.4. Fading Intergroup Contact Intervention Effects in Intractable Conflicts

Several authors have stated that contact intervention effects in conflict areas decrease (i.e., fade) after a while (Kupermintz & Salomon, 2005; Rosen & Salomon, 2011; Salomon, 2006, 2009a). Salomon (2009a) wrote "the observation that the effects of peace education can quite easily be reversed implies that the challenge facing peace education is not just how to effect change, but how to *sustain* it" (p. 115). Arnon (2010) showed that improved outcomes of a contact intervention in Israel at the post-test returned to the value of the original level only two months after the end of the intervention. In contrast to long-term outcomes, which mostly focus on the development of contact intervention effects between the pre- and the follow-up tests, fading effects examine changes between post- and follow-up test.[16] As apparent from Table 1, some researchers already found fading effects within contact intervention studies in conflict areas (e.g., Arnon, 2010; Bar-Natan et al., 2010; Jayusi 2009; Schroeder & Risen, 2016; Shani, 2015). In order to get a more thorough understanding of enduring contact intervention effects in intractable conflict areas, we believe it is necessary to investigate all time intervals (i.e., pre-posttest changes, pre-follow-up test changes, post-follow-up test changes).

2.2.5. Repeated Intergroup Contact Interventions in Intractable Conflicts

Given that contact intervention effects mostly decrease in intractable conflict areas, the question arises how sustainable contact intervention outcomes can be enabled in intractable conflict areas. Some researchers state that to facilitate these long-term contact intervention effects the repetition or continuation of an intervention might lead to prolonged positive findings (e.g., Kilpatrick & Leitch, 2004; Kupermintz & Salomon, 2005; Rosen, 2006 as cited in Salomon, 2009b; Rosen & Salomon, 2011; Salomon, 2009a, 2011). Repetition can be either an additional distinct intervention or an additional element of the same intervention. Some researchers have already investigated whether or not extra interventions induce a more sustainable outcome. Rosen (2006, as cited in Salomon, 2009b), for example, used a forced compliance paradigm, meaning participants role-played the outgroup's perspective in front of other participants, to consolidate the previous outcome. Although Rosen (2006, as cited in Salomon, 2009b) investigated an intragroup intervention, meaning not a face-to-face intervention, results indicated that the implementation of a further intervention (implemented some weeks after the first

[16] These terms are not used consistently throughout all studies.

intervention) that included a forced compliance element, led to a prolongation of the positive attitudes toward outgroup members. Additionally, Berger et al. (2016), who did not explicitly focus on investigating repeated intergroup contact intervention effects, also revealed long-term contact intervention effects in their study after students met twelve times. Even so, it is important to mention that the intervention in their study was not limited to intergroup contact meetings but included extra information sessions. Participants were additionally encouraged to maintain contact after the end of the intervention, so the research findings also might be related to this contact (Berger et al., 2016). Thus, it is not possible to draw clear conclusions about repeated intergroup contact effects. A study conducted by Jayusi (2009) discovered that a peer tutoring placed four to five weeks after the end of the intergroup contact intervention led to an increased long-term effect for those students who participated in the extra intervention compared to those who did not. The peer tutoring included a preparation workshop as well as planning and instructing an activity that enabled peers to understand the other side. Thereby, Jayusi (2009) demonstrated that an additional intervention did not only restore the previous positive outcome but also increased intergroup intervention outcomes in the long run. These studies indicate that extra interventions might have a positive impact on long-term contact intervention effects in conflict areas. Accordingly, Arnon (2010) suggested that a "peace education program should include at least three parts: preliminary preparation, the program itself, and follow-up/reinforcement activity after the completion of the program" (p. 197).[17] However, many open questions remain such as: Are these effects detectable for other interventions? Is the implementation of additional interventions necessary or do repeated meetings of the same intervention have the equivalent impact on long-term effects? Should repeated interventions have a short or long duration?

To shed more light on this field of research, we investigated two contact interventions, which included several contact meetings in Israel (two and four face-to-face meetings respectively). We investigated repeated meetings within the same intervention and not an additional separate intervention. We examined short- and long-term as well as fading effects of attitudes and behavior intentions toward the outgroup.

[17] In her study, Arnon (2010) also implemented an extra intervention, which took place about two months after the first encounter. About half of the participants took part in the extra intervention (intervention group), the other half did not (control group). All participants were questioned before the first encounter, after the first encounter and at the time of the end of the extra intervention. Thus, long-term effects were only measured for the control group as the measurement at the end of the extra intervention can be considered as a direct posttest for the intervention group.

2.2.6. Minority-Majority Differences in Intergroup Contact Interventions

For a long time, contact studies focused on majority group outcomes (Pettigrew & Tropp, 2011). In recent years, contact research has also concentrated on minority groups and minority-majority status differences. A meta-analysis by Tropp and Pettigrew (2005) revealed that contact-prejudice effects are effective for both groups; however, they are less strong for the minority than for the majority group (minority: $r = -.18$; $k = 125$ number of samples; majority: $r = -.24$; $k = 205$ number of samples). Within their meta-analysis Lemmer and Wagner (2015) confirmed positive outcomes for both groups specifically regarding intergroup contact intervention effects. Differences between the two groups were significant within samples that included a control group design (minority: $\hat{\mu} = .20$; $k = 20$ number of comparisons; majority: $\hat{\mu} = .38$; $k = 39$ number of comparisons; $Q_{\text{model}} = 4.39$, $p < .05$).[18]

So far status differences can be explained by the fact that both groups perceive the same contact situation from different perspectives (e.g., Salomon, 2011). Thus, minority group members assume that the contact situation confronts them with their devaluated societal status within the intergroup intervention, whereas majority members are concerned about being perceived as having prejudices (Pettigrew & Tropp, 2011; Shnabel & Nadler, 2008). Therefore, the minority perspective might inhibit contact effects (Tropp & Pettigrew, 2005), because minority group members are more aware of their status within the contact situation than majority members (Bastian, Lusher, & Ata, 2012). Status is also related to different preferences in the contact situation: Minority group members are more interested in talking about power to improve the position of their ingroup (Saguy, Dovidio, & Pratto, 2008), whereas majority group members strive more to talk about commonalities. A further discovery is that the argumentation patterns between both groups are deviating – showing that Palestinians (minority) created more developing points (elaborations), whereas Jewish Israeli (majority) used more prompters (i.e., objections) and delimiters (i.e., frames that delimit the points) during communication (Ellis & Maoz, 2002). Given the fact that contact effects are less strong for the minority than for the majority, these differences might possibly explain some of the deviating effects; although, no precise understanding of the underlying processes is yet established. Another explanation for status differences focuses on the deviating baseline prejudice score, meaning minority group members often

[18] There was no difference between the findings of the minority ($\hat{\mu} = .37$; $k = 15$ number of comparisons) and the majority ($\hat{\mu} = .46$; $k = 12$ number of comparisons) within studies that did not include a control group (Lemmer & Wagner, 2015).

show less prejudice in their baseline scores than majority members, which might lead to a smaller range of improvement possibilities. Whether these deviating baseline scores are associated with more previous intergroup interactions is not clear yet. However, Pettigrew and Tropp (2011) could not find any differences in the reported amount of contact in American survey data. The authors could not rule out a difference in the amount of contact even so because the absence of group differences could also be related to a deviating assessment of intergroup contact frequencies between both groups.

Some authors discuss the implementation of different interventions for each group (e.g., Abu-Nimer, 2004), corresponding to the idea that minority and majority group members have different perspectives, preferences, and communication patterns within a contact situation. Referring to Salomon (2011):

> rather than striving to attain a common goal - such as mutual acknowledgement, empathy, or reduced prejudice - peace education would need to accept the possibility that programs serve very different needs and goals for the parties involved, allowing one side to 'have a voice,' strengthen its adherence to its own collective narrative, or become empowered; and the other side to acknowledge its role in the conflict and give legitimacy to the other's collective narrative. (p. 53, see also Salomon, 2002)

Finally, Bastian et al. (2012) showed that intergroup contact had an impact on different dependent variables for the minority and the majority.

Considering the studies outlined in Table 1, only nine interventions were investigated regarding long-term contact effects in an intractable conflict area. Many of these studies did not distinguish between minority or majority samples and data were analyzed concerning overall findings. Table 1 illustrates that despite the increased focus on status differences within the last years there is still little research respecting status differences and long-term contact intervention effects in conflict areas.

In this thesis, we investigated two intergroup contact interventions. In each intervention, both status groups participated. Hence, interventions in both status groups were equivalent. We also asked participants from all groups to answer items related to the same dependent variables, because we were interested in comparing the results. In order to discover status differences, we analyzed both groups separately in a first step and examined whether the effects deviated significantly in a second.

2.2.7. Relevance of Positive Contact Experience in Intergroup Interventions

As mentioned above, Pettigrew (1998) described the opportunity to become friends as a fifth optimal condition for intergroup contact to reduce prejudice. He mentioned that

friendship implies the potential for extensive and repeated contact in various contexts for individuals from different groups. Thus, individuals who become friends with outgroup members might have more contact leading to reduced prejudice in the long run. Indeed, in their meta-analysis Pettigrew and Tropp (2006) showed that studies which tested the effects of friendship on prejudice reduction illustrated a stronger effect (r = -.25; k = 154 tests) than studies that did not include friendship as a measure of contact (r = -.21; k = 1211 tests; $Q_B[1] = 4.42$, $p < .05$). Within another meta-analysis, the authors showed that two aspects of friendships particularly had a strong impact on attitudes, namely time spent with outgroup members and self-disclosure (Davies, Tropp, Aron, Pettigrew, & Wright, 2011). With reference to different conflict areas, Hughes, Lolliot, Hewstone, Schmid, and Carlisle (2012) revealed that a higher proportion of friendship also led to improved intergroup attitudes and positive action tendencies in Northern Ireland (see also Sharing Education Programme, 2012).

Seeing that there is empirical evidence that (aspects of) cross-group friendships have a positive impact on prejudice reduction, we examined the valence of students' overall intergroup experience. Based on previous findings, we expected positive experiences to be related to stronger positive outcomes. To investigate the relation of the valence of students' experience to attitudes and behavior intentions toward the outgroup, we categorized students' written comments on their questionnaires as positive, neutral, or negative. Cross-group friendship was thereby included as one aspect of positive intergroup contact experiences, so that students who reported about intergroup friendship were categorized as having a positive intergroup experience. Other positive experiences were for example positive feelings about the contact activities or reported positive interactions with other participants. Given that positive contact was found to lead to positive contact intervention effects and, as shown in Chapter I.2.1.2., negative contact is related to negative effects (e.g., Ben-Ari & Amir, 1986), we assumed that the valence of students' experience (positive, negative, neutral) was associated to short- and long-term contact intervention effects.

2.2.8. Excursus: Intergroup Contact Intervention Effects in Young Age Groups

In this thesis, we investigated 4th and 6th grade students in an intractable conflict area (i.e., students between the age of 10 and 13 years).[19] To reflect the contact-prejudice

[19] Some studies refer to age cohorts in their investigations whereas others refer to school classes. The descriptions within the excurses relate to the original studies and therefore might switch between these expressions.

relation and the contact-attitude and -behavior intentions relation respectively, we briefly illustrate research findings related to these specific age cohorts.

Age-prejudice relation. Labeling of ethnic categories is already detectable in early childhood (e.g., Bergen, 2001; Brown, 2010; Farhan, 2008). Bar-Tal (1996) illustrated that Jewish Israeli acquire the social category "Arab" as one of the first social groups between the young age of 2.5 and 3.5 years. He further presented that a negative perception of Arabs is found in majority members as early as at the age of 6 years. Additionally, Bar-Tal and Teichman (2005) revealed that majority members' positive attitudes toward Arabs decrease between the age of 2 and 6 years. Perceptions of minority group children are less clear. In regard to ingroup favoritism, some children prefer children of the ingroup, some of the outgroup, and others are unbiased (Aboud & Brown, 2013). Thus, we can draw no clear conclusion concerning minority members. Outgroup attitudes probably develop through the influence of the child's home, peers, media, and social living environment (e.g., Bar-Tal & Teichman, 2005; Bergen, 2001; Brown, 2010). Additionally, Raabe and Beelmann (2011) did not discover any systematic development of prejudices concerning the relative differences between age groups above the age of 10, meaning they found no relative changes of prejudice within the age cohorts relevant to this thesis.[20] Furthermore, Bar-Tal and Teichman (2005) assumed that living in an intractable conflict area increases children's ethnic awareness as well as their ingroup favoritism and outgroup negativity.

Contact-prejudice relation for different age groups. Referring to the development of prejudice in children, Bergen (2001) stated that the "exposure to different value systems stimulates racial critical thinking" (p. 160). He also reported that it might be easier to neutralize prejudices at an early age, preventing their development rather than changing them afterwards (see also Aboud & Brown, 2013). In a meta-analysis, Beelmann and Heinemann (2014) revealed that educational programs that aim to reduce prejudice are effective for children ($d = 0.30$, $k = 122$ comparisons). Contrary to Bergen's statement, age group did not have any effect on the effect size variability in their findings.[21] Within a further meta-analysis, Lemmer and Wagner (2015) demonstrated that contact interventions

[20] Investigated age groups: 2-4 years (early childhood), 5-7 years (middle childhood), 8-10 years (late childhood), 11-13 years (early adolescents), 14-16 years (middle adolescents), and 17-19 years (late adolescents). Raabe and Beelmann (2011) found relative changes in the effect sizes in regard to the development of prejudice among 2-4 versus 5-7 ($d = 0.29^{**}$) and 5-7 versus 8-10 years old ($d = -0.11^{*}$).

[21] Investigated age groups: 3:6-7:11 years; 8-9:11 years; 10-13:11 years; 14-18 years. Beelmann and Heinemann (2014) did not solely focus on contact interventions within their analysis, but also included information/knowledge acquisition programs, as well as programs that strive to develop individual social-cognitive competencies.

improve intergroup relations; again, no moderating effect of age in the contact intervention-prejudice relation could be found.[22]

Age-context-prejudice relation. Considering the fact that within the time of the data gathering in Israel the Israel-Gaza Conflict 2012 occurred (i.e., outbreak of violence between Jews and Palestinians), we examine whether contextual influences (e.g., political events, war) have a particular impact on different age groups. According to Muldoon and Trew (2000), children are exposed to specific conflict experiences in situations of violent conflict. More precisely, Bar-Tal and Labin (2001) investigated the effect of a terrorist attack on stereotyping in 5th and 8th graders and found that negative feelings increased toward Palestinians especially in the 5th grade cohort. Furthermore, Leonardi, Magos, and Oikonomou (2014) discovered that increased mortality salience was related to negative trait statements and stereotypes toward the outgroup for 6th grade students. The authors mentioned that their findings are similar to the ones found in samples including adults. They argue that this is due to the fact that 11-year-old children have already reached a cognitive developmental stage in which they have a concept of death. In sum, studies indicate that context has an influence on prejudice; however, whether this is especially high for specific age cohorts remains unclear.

Further findings. In contrast to the findings above, one aspect which seems to be related to age is negative behaviors such as name-calling, exclusion, and avoidance, which have a peak in middle childhood.

> Verbal discrimination, in the form of nasty name-calling, increases in primary school, with 30 percent of children in grades 3 and 4, and more than 50 percent of children in grades 5 and 6 saying that they were bullied in the past month.
> (Aboud & Brown, 2013, p. 179)

Even though no differences in the contact-prejudice relation for different age groups were found, this quotation indicates that negative contact situations might occur more often in certain age groups (5th and 6th grade) compared to other age groups.

2.3. Status Differences Between Jewish Israeli and Arab Israeli in Israel

To provide insight into the current minority-majority status situation in Israel, we briefly outline some information regarding this specific conflict context. As mentioned before, the territorial partition between the State of Israel and the State of Palestine is not solved until now. Within this thesis we regard the State of Palestine as the West Bank and

[22] Investigated age groups: 5-9 years, 10-13 years, 14-18 years, and > 18 years.

the Gaza Strip territory. The thesis focuses on the comparison of Jewish Israeli and the Arab Israeli students (i.e., both groups are living in the State of Israel) and not of Jewish Israeli and Palestinian students. Arab Israeli students may hold some benefits compared to Palestinians in regard to their Israeli citizenship. However, because "[they] share an ethnic and ideological bond with the Palestinians in the Palestinian Authority, many of them … [prefer] to be identified as Israeli Palestinians" (Teichman, Bar-Tal, & Abdolrazeq, 2007, p. 425). Because of their identification as Palestinians and the perception as Palestinians by many Jewish Israeli, we consider this thesis to be related to the Israeli-Palestinian Conflict (Hammack, 2010).

In 2013, the Israeli population consisted of 6.451.300 "Jews & Others" (79.3%) and 1.683.200 "Arabs" (20.7%) showing that Arab Israeli citizens are a minority in reference to their number (The Jerusalem Institute for Israel Studies, 2015c; Table 2). Additional, Arab Israeli and Jewish Israeli belong to different ethnicities and have a deviating religious background. Thus, status, ethnicity, as well as religion are confounded and cannot be considered separately within this thesis. Therefore, we refer to these cumulated differences as status differences. As illustrated in Table 2 the "Arab" population in Israel is less represented in higher education than "Jews & Others" and less employed (The Jerusalem Institute for Israel Studies, 2015b, 2015d). Although in the recent past Arabs have attained more management positions (Khattab, Miaari, Manor, Nabwani, & Kagya, 2013), referring to management roles they are still extremely underrepresented. Furthermore, a greater part of the Arab population lives in poverty (The Jerusalem Institute for Israel Studies, 2015a). According to Teichman et al. (2007) it is almost impossible to cross the impermeable border between the Arab Israeli minority and Jewish Israeli majority.

Table 2: Characteristics of Israeli Population

	Jews & Others	Arabs
Population[a]	79.3%	20.7%
Education[a]		
First Degree	84.9%	15.1%
Second Degree	91.5%	8.5%
Third Degree	95.1%	4.9%
Employement	94.0%	91.0%
Unemployment[a]	6.0%	9.0%
Poverty[a,b]	14.0%	52.0%

Note. [a] Information is based on The Jerusalem Institute for Israel Studies (2015a, 2015b, 2015c, 2015d). [b] Percentage of the respective population.

Besides these differences, Shani (2015) stated

> while Jews and Palestinian are politically polarized and the current political arrangements perpetuate existing gaps and power asymmetries, the two groups enjoy pragmatic daily cooperation in a largely calm and peaceful atmosphere. At the sociopsychological level, the majority on both sides is interested in improving relations and promoting coexistence, but the two societies still largely exhibit negative beliefs and attitudes that are incompatible with coexistence. (p. 31)

This description illustrates that there are still major differences between the minority and majority group living in Israel and that the above mentioned separate investigation of both groups is necessary to enable meaningful insights into the impact of contact interventions regarding both status groups.

3. Methodological Issues in Longitudinal (Quasi-)Experiments

In the following, we discuss six important aspects of longitudinal (quasi-)experimental studies: research studies and research evaluations, structural equation modeling, longitudinal and cross-group measurement invariance, hierarchical data structure, missing data, and causality assumption.

3.1. Research Studies and Research Evaluations

While research studies focus predominantly on the development and advancement of an underlying theory (e.g., contact theory), this thesis aims at improving knowledge in regard to contact theory (Allport, 1954) and identifying practical implications for future

interventions. Due to the fact that the analysis could also be considered as a research evaluation[23] (i.e., analyzing an evaluation object and striving to optimize the object and future action; e.g., Gollwitzer & Jäger, 2014) utilizing the contact theory as a specific theoretical background, we briefly integrate our investigation into the evaluation terminology.

Our analysis could be regarded as having three evaluation goals: First, the investigation of the effectivity of the school program (i.e., *summative output evaluation*: Did students who participated in the school program have more positive attitudes and behavioral intentions toward the outgroup than students who did not take part in the program?). Second, the analysis of the long-term effects of the school program (i.e., *sustainability*: Could the impact of the school program be sustained?). Third, the reasonableness of the school program's structure (i.e., *quality of the concept*: Did the repetition of the intergroup meetings have an impact on the sustainability of the school program?; Gollwitzer & Jäger, 2014). The present investigation could be classified as a *field evaluation* (i.e., it was implemented in the field and not in a laboratory) as well as an *external evaluation* (i.e., it was conducted - mainly - by a neutral person)[24], and was based on *self-reports* (Gollwitzer & Jäger, 2014).

3.2. Structural Equation Modeling With (Quasi-)Experimental Data

Until now, longitudinal (quasi-)experimental research studies have been predominantly analyzed using repeated measures ANOVA (using manifest variables[25]). Although recent literature highlights the superiority of latent variable modeling (i.a., variables taking into account measurement errors) in comparison to traditional methods using manifest variables, only very few studies within the field of longitudinal (quasi-)experimental research apply this modeling (e.g., Bagozzi & Yi, 1989; Kano, 2001; McArdle & Prindle, 2008; Pugesek, Tomer, & von Eye, 2009; Russell, Kahn, Spoth, & Altmeier, 1998). In this thesis, we analyzed quasi-experimental data on the basis of latent variables using structural equation modeling (SEM). The fact that measurement errors are taken into account within SEM is a crucial advantage for research analysis, because it enables to consider this bias within the calculations (Brown, 2006). However, there are several more

[23] The expression is not used consistently in scientific literature.

[24] The data was gathered by individuals that were involved in the school program, whereas the analyses were conducted by the author of this thesis.

[25] We use the term manifest variable when the value of the variable is calculated as the mean of the values of several indicators.

advantages of SEM: SEM allows for the evaluation of model fit information (i.e., whether data match a theoretical model) and provides modification indices that indicate which changes in the model could be implemented to improve the model fit; it supports testing of alternative models; it enables the simultaneous examination of the measurement and the structural model; it facilitates complex analyses; and it enables cross-group and longitudinal measurement invariance (MI) calculations (e.g., Brown, 2006; Kline, 2011; MacCallum & Austin, 2000; Thompson & Green, 2013; Ullman, 2006; Werner & Schermelleh-Engel, 2009).[26]

In order to calculate the internal consistency of latent variables, the reliability coefficient omega can be used (ω; McDonald, 1999). Omega has advantages in comparison to the commonly used Cronbach's Alpha, since it is based on a congeneric and not an essential tau-equivalent model, meaning that true score means and item error variances are allowed to vary (e.g., Dunn, Baguley, & Brunsden, 2014). The size of omega can be interpreted equally to the size of the Chronbach's Alpha coefficient.

For analyzing quasi-experimental data, we utilized a specific type of structural equation model, namely the latent change model (LCM) also known as latent difference score model, latent true score model (Geiser, 2010; Gollwitzer, Christ, & Lemmer, 2014; Steyer, Eid, & Schwenkmezger, 1997), or change regression model[27] (McArdle, 2009). In our models, latent difference scores were created to measure the change between two occasions of measurement, for example, the change in the dependent variable between pre- and posttest. The basic version of this model was extended by including the (quasi-)experimental condition (IG and CG) as a predictor variable of the latent difference score (e.g., McArdle & Prindle, 2008). By including this extension, the LCM is similar to the often used repeated measures ANOVA, but, again, based on latent variables (McArdle, 2009). A description including a figure of the latent change models which were used within this thesis can be found in Manuscript #1.

3.3. Longitudinal and Cross-Group Measurement Invariance Analysis

As mentioned above, measurement invariance (MI) calculations across time and between groups can be regarded as an advantage of SEM over traditional techniques.

[26] If not stated otherwise, we used latent variables in the analysis.

[27] McArdle (2009) used this expression explicitly when a regression - instead of a correlation - is drawn from the dependent variable at time 1 on the difference score (time 2-time 1).

Therefore, we introduce this aspect in the following in more detail. It is important to investigate MI to either ascertain the comparability of measured constructs and the respective means (e.g., Brown, 2006; Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014; Horn & McArdle, 1992; Little, Preacher, Selig, & Card, 2007; Maede, Lautenschlaeger, & Hecht, 2005; Vandenberg & Lance, 2000) or to discover meaningful differences in regard to the understanding of the underlying constructs. Horn and McArdle (1992) described the importance of MI as follows:

> The general question of invariance of measurement is one of whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute. If there is no evidence indicating presence or absence of measurement invariance - the usual case - or there is evidence that such invariance does not obtain, then the basis for drawing scientific inference is severely lacking: findings of differences between individuals and groups cannot be unambiguously interpreted. (p. 117)

Within this thesis, we tested whether MI was given over time (pre-, post-, and follow-up measurement) and between groups (IG and CG; Arab Israeli and Jewish Israeli; 4th and 6th grade).[28] We focused predominantly on longitudinal MI, as this thesis is concerned with long-term changes of attitudes and behavioral intentions; however, particularly within quasi-experimental studies in which participants are not randomly assigned, ascertaining cross-group MI is also essential to verify an equal understanding of the measured constructs across IG and CG.

In order to evaluate whether the measured scales are equivalent, certain steps have to be conducted. Configural, metric, and scalar MI are required in case latent means are to be compared across time (Davidov, 2008). Configural invariance (pattern invariance)[29] demonstrates that an equal pattern of the factor structure is given; metric invariance (loading invariance, weak factorial invariance) shows that equal factor loadings exist between groups or across time; scalar invariance (intercept invariance, strong factorial invariance) indicates that intercepts are identical across groups or throughout time (e.g., Brown, 2006; Chen, 2007; Geiser, 2010; Little, 2013; Little et al., 2007; Vandenberg & Lance, 2000; van de Vijver, 2011).[30] To examine whether these preconditions were met in the measured constructs in this thesis, configural MI had to be ensured first. Thus, it had to

[28] Cross-group measurement invariance analysis between different age cohorts (4th and 6th grade students) of the same ethnic background are displayed within the ancillary analysis (Chapter III.3.).

[29] The terms used for the single MI steps vary across articles (e.g., Little, 2013; Vandenberg & Lance, 2000). Therefore, expressions presented in the parentheses illustrate synonyms found in the literature. Sometimes researchers also report residual MI results (error variance invariance, strict factorial invariance). We skip this step because it is not necessary for latent mean comparisons.

[30] A detailed description about MI practices can be found in Vandenberg and Lance (2000).

be ascertained that the model fit of the configural model was acceptable. To test whether the constructs showed metric MI, the model fit of the configural model with unconstrained indicators was then compared to the model fit of a model in which factor loadings were constrained to be equal across groups or throughout time (using the χ^2-difference test; Issmer, 2012). Since we used MLR estimators for the analyses, which are robust to non-normality (Muthén & Muthén, 1998-2012), the correction of the χ^2-difference test statistic as shown by Satorra and Bentler (2001) was applied (Kline, 2011).[31] In order to consider data as being metric MI, the two model fits are not allowed to differ significantly. In a last step, a model fit comparison had to be realized between the metric MI model and a model fit of a model in which additionally the intercepts were constrained to be equivalent across groups or over time. Significant model fit differences would lead to a rejection of the assumption of scalar invariance. Bontempo, Grouzet and Hofer (2012) described this procedure as follows: "If these equivalence constraints can be placed on the necessary parameters without unacceptable decreases in model fit, then measurement equivalence is demonstrated and construct trajectories (or group differences) can be interpreted without the shadow of internal validity threat" (p. 109).

It is also possible that within the MI computations a few factor structures and/or intercepts are not equal across groups or time. Several scholars refer to this phenomenon as partial MI. To check whether data are partially measurement invariant or not invariant at all, single factor loadings and/or intercepts are released and the χ^2-difference test is conducted again (e.g., Byrne, Shavelson, & Muthén, 1989), meaning that model fit comparisons including the new partially released model are conducted. Little (2013) stated

> when only a few of the loadings and intercepts of a construct show a lack of invariance (while the majority of the indicators are consistent with invariance), one can generally proceed with making comparisons of the constructs' key parameters (construct means, variances, and covariances/correlations). (p. 178)

Steenkamp and Baumgartner (1998) defined partial MI more precisely and stated that the measured construct has to consist of a minimum of two invariant factor loadings and intercepts in case group differences are compared. Some other authors disagree and propose partial MI to be unacceptable (e.g., De Beuckelaer & Swinnen, 2011). Up to now, there is no clear consensus whether partial MI is acceptable and if yes, under which conditions. As Davidov et al. (2014) explain, "further simulations are needed to provide a more solid

[31] It is also possible to use the CFI model fit indices for model comparisons (Chen, 2007).

recommendation for how applied researchers should handle partial MI when full equivalence is not given" (p. 66). Within this thesis, we applied latent means when partial MI, as defined by Steenkamp and Baumgartner (1998), was given.

It has to be mentioned that recently, a new approach to MI has been presented: *Approximate MI*. According to van de Schoot et al. (2013), approximate MI allows "for some wiggle room for the intercept differences between groups" (p. 2). Cieciuch, Davidov, Schmidt, Algesheimer, and Schwartz (2014) described the difference between partial MI and approximate MI as follows:

> approximate measurement invariance differs from the partial measurement invariance approach, because in the latter some parameters are constrained to be exactly equal and others are released entirely, while in the former all parameters are constrained; however, the restrictions are more liberal and refer to the concept of approximate equality. (p. 4)

Applying approximate MI, Zercher, Schmidt, Cieciuch, and Davidov (2015) realized longitudinal and cross-group MI calculations simultaneously.[32] Given this method to be a recent innovation, MI analysis were calculated separately within this thesis. However, approximate MI seems to pave a way for future MI calculations.

In case neither partial nor approximate MI is achieved, measurement noninvariance exists. There are different reasons why measurement noninvariance appears in the data. Van de Vijver (1998, see also van de Vijver, 2011) described three biases that might lead to cross-group measurement noninvariance and which are not mutually exclusive: *method bias* (i.e., measurement noninvariance is based on the used methods which are different for groups), for example, incomparability of samples, stimulus differences within the conditions, and communication problems; *item bias* (i.e., measurement noninvariance occurred due to anomalies on the item level), such as translation issues[33], deviating familiarity with the content of the item among different groups; and lastly *construct bias* (i.e., the understanding of the constructs is deviating across groups). However, it is also conceivable that the understanding of a measured construct is not only differing across groups but also changing throughout time (i.e., response shifts). A response shift indicates "a change in the meaning of one's self-evaluation of a target construct" (Sprangers & Schwartz, 1999,

[32] In case a large number of groups are represented in the data, the *alignment method* facilitates finding MI among the groups. As Asparouhov and Muthén (2014) stated, the alignment method provides "a comparison of factor means and factor variances across groups while allowing for approximate measurement invariance" (p. 499). Due to the fact that within this thesis only a small number of groups (i.e., two groups) is given, this method was not relevant.

[33] Davidov and De Beuckelaer (2010) also reported that it is more difficult to achieve measurement invariance across questionnaires using different languages than across questionnaires that are based on the same language.

p. 1508). Response shifts, which are especially important within longitudinal studies, might appear due to *maturation* (Bontempo et al., 2012), *external incidents*, or as a result of a *manipulation/intervention*. As shown in several studies, response shifts have already been observed in the context of health interventions (Fokkema, Smits, Kelderman, & Cuijpers, 2013; Oort, 2005), organizational interventions (Millsap & Hartog, 1988), and violence prevention programs (Rosen, 2015) indicating that participants who took part in a manipulation/intervention changed their understanding of the measured constructs in the context of their participation. When referring to response shifts, most researchers use the terms alpha, beta, and gamma change to indicate the specific change that appeared over time (Golembiewski, Billingsley, & Yeager, 1976).[34] Alpha change can be understood as true change in the values of the measured construct, meaning MI is given and latent means of the constructs can be compared over time (e.g., Maede et al., 2005). Beta change is usually declared as change in the measured construct calibration (Bontempo et al., 2012), and gamma change is a major change of the understanding of the measured constructs, meaning a redefinition of an underlying construct across time. Millsap and Hartog (1988) described gamma change as follows:

> gamma change is expected to occur only in the treatment group, as a result of the intervention. But if we define gamma change as change in factorial structure, this change is possible even in the absence of an intervention. For example, factorial change could result from maturation or from environmental influences operating between pretest and posttest. (p. 575)

Aside from the reference to alpha, beta, and gamma change, Schwartz and Sprangers (1999; see also Sprangers & Schwartz, 1999) addressed the terms recalibration, reprioritization, and reconceptualization, which are directly related to the scalar, metric, and configural level of MI respectively.

In other research areas recalibration, reprioritization, and reconceptualization in regard to possible response shifts have already been analyzed (Barclay & Tate, 2014; Rosen, 2015), while these investigations are not common in the contact literature yet. Within this thesis, longitudinal and cross-group MI analyses were conducted to either ascertain that latent mean comparisons were allowed or to investigate which type of measurement noninvariance was given.

[34] For a list of studies that investigated alpha, beta, and gamma change see Vandenberg and Lance (2000).

3.4. Hierarchical Data Structure Analysis

In many (quasi-)experimental field studies, participants are nested in subgroups (e.g., school classes, working teams, departments). In these cases, the assumption of independence of observations is violated because participants within a subgroup are supposed to respond more similarly to a questionnaire than participants between subgroups (Hox, 2010). In order to investigate whether these violations are meaningful the design effect (deff) can be calculated. In contrast to the more utilized intraclass coefficient (ICC), the deff also considers the average cluster size: Deff = 1 + (Average Cluster Size – 1) x ICC. A deff can be regarded as meaningful in case it is greater than 2 (Maas & Hox, 2005), which indicates that measures against an inflation of the Type-I error should be implemented (Bovaird, 2007). In general, the hierarchical data structure can be taken into account by either a random or a fixed model approach (Snijders & Bosker, 2012). In case higher-level units are considered as a sample from a population and conclusions regarding the population shall be drawn, the random approach is appropriate. In contrast, if higher-level units are regarded as unique categories and specific conclusions regarding each of these categories shall be examined, the fixed approach shall be utilized. In addition to these conceptual differences, there are also practical requirements that have an impact on the choice of the used approach. According to a simulation study by Maas and Hox (2005), a random approach requires at least 50 level-2-units. Correspondingly, Snijders and Bosker (2012) stated that in case less than 10 level-2-units are given, a fixed approach is appropriate. Therefore, in case a small number of level-2-untis exists, a fixed approach should be implemented. Although, school classes (level-2-units within this thesis) are considered to be a sample from a population, we applied a fixed approach based on the small number of subgroups in this thesis.[35] In order to implement this fixed approach and control for[36] the level-2-units, we included several level-2 control variables into the LCM.

[35] The statistic software M*plus* provides tools to take nested (clustered) data into account (e.g., Type = Complex; Muthén & Muthén, 1998-2012). However, these tools can only be applied in case 30-50 clusters (level-2-units) are given; in case this number is not reached, Muthén (2012) stated that dummy variables have to be included into the analysis.

[36] Based on Miller and Chapman (2001), it is not possible to *control for* a covariate in quasi-experimental studies (Lord's Paradox), because based on a correlation between the control and the independent variable, the independent variable might change qualitatively. According to the fact, that no alternative proceeding is known, we calculated the data including so called "control variables".

A description about the specific coding of these level-2 control variables that were inserted into the latent change models can be found in Manuscript #1 and Manuscript #2.[37]

3.5. Missing Data Analysis

In longitudinal studies, researchers usually have to deal with unplanned missing values, "in fact it is difficult to imagine a longitudinal study without at least some unplanned missing data" (Collins, 2006, p. 521). According to Schafer and Graham (2002), there are two main reasons for missing data: individuals that leave the questioning for one or more waves of questioning (unit nonresponse), which is a specific missing data problem within longitudinal data; and items that were not answered (item nonresponse).

Types of missing data. To handle missing data, it is necessary to discover whether missingness is related to the data or not. In 1976, Rubin created a typology that described three different missing data classifications. Schafer and Graham (2002) described these three types as follows: *missing completely at random* (MCAR) "means that the probability that Y is missing for a participant does not depend on his or her own values of X and Y" (p. 151); *missing at random* (MAR) indicates that missing data may be related to X but not to Y; *missing not at random* (MNAR) shows that missing data are related to Y. Most up-to-date missing data techniques (e.g., maximum likelihood estimation, multiple imputation) require MCAR or MAR to analyze data including missing values (e.g., Collins, Schafer, & Kam, 2001; Feng, Cong, & Silverstein, 2012; Newman, 2003; Schafer & Graham, 2002). In order to test whether data are MCAR, Little's MCAR test can be used (Little, 1988) and a non-significant outcome indicates that data are MCAR. There is no further option to test whether data are MAR or MNAR. In case MCAR is not reached, Graham (2009) stated

> the best way to think of all missing data is as a continuum between MAR and MNAR. Because all missingness is MNAR (i.e., not purely MAR), then whether it is MNAR or not should never be the issue (…), we should answer the question of whether the violation is big enough to matter to any practical extent. (p. 567)

Thus, it has to be ruled out that data are overall MNAR, meaning that missingness is related to Y (Schafer & Graham, 2002). Collins et al. (2001) also declared that when "the correlation between the cause of missingness and the variable subject to missingness was .4,

[37] A group of researchers, namely Ulrich Wagner, Mario Gollwitzer, Stefan Thörner, Lisa Gutenbrunner, and the author of this thesis, was involved in identifying the solution (i.e., coding of the level-2 control variables) for this fixed effects approach including only a small number of level-2-units.

omitting the cause of missingness had a negligible effect" (p. 333). Under these circumstances, even a combined assumption of MAR and MNAR can hold as prerequisite for the respective missing data analysis.

Missing data techniques. After exploring the type of missing data, a technique to handle missing data has to be chosen. In the past, missing values were handled using ad hoc methods which tried to fix the data ahead of the calculations (Peugh & Enders, 2004). These ad hoc measures usually applied listwise as well as pairwise deletion or the missing values were substituted with the variable mean. However, as Wilkinson and the Task Force on Statistical Inference of the American Psychological Association stated already in 1999 "the two popular methods for dealing with missing data that are found in basic statistics packages–listwise and pairwise deletion of missing values–are among the worst methods available for practical applications" (p. 598). Accordingly, Schafer and Graham (2002) recommended using all available data to recover missing information because especially in the case of longitudinal data, repeated measures of individuals are probably related. Currently two other missing-data techniques are regarded as state of the art: maximum likelihood estimation (ML; sometimes also referred to as full-information-maximum-likelihood [FIML]) and multiple imputation (MIM[38]; Peugh & Enders, 2004; Schafer & Graham, 2002). Although the two missing data techniques ML and MIM are recommended by several authors (e.g., Baraldi & Enders, 2010; Brown, 2006; Feng et al., 2012; Peugh & Enders, 2004; Schafer & Graham, 2002), they are barely used in case of longitudinal experimental data in contact research. Differences between the two recommended missing data techniques are considered as small (Peugh & Enders, 2004) and according to Enders (2013), the choice for either one of the missing data techniques depends on personal preferences and the utilized software. Nevertheless, there are some differences between the two missing data techniques, which are outlined in Table 3 (Graham, 2003; Peugh & Enders, 2004; Schafer & Graham, 2002).

[38] Usually the abbreviation of multiple imputation is MI; however, due to the fact that measurement invariance is also shortened to MI in research literature and measurement invariance was already introduced in this text, we abbreviate multiple imputation with the letters MIM.

Table 3: Characteristics of Maximum Likelihood Estimation (ML) and Multiple Imputation Analysis (MIM)

Issues	Maximum Likelihood Estimation	Multiple Imputation
Estimator	Maximum likelihood	Bayesian
Data Basis	Data is estimated directly in the model, no complete data sets exist	Data is used to compute several complete data sets
Number of Steps	One step: analysis	Two steps: imputation and analysis
Auxiliary variables	Relatively difficult to implement in analysis	Relatively easy to implement in imputation process
Standard errors	In general slightly smaller	In general slightly larger

Within this thesis, we decided to calculate missing data using ML when calculating latent variables via the computer software M*plus* (Version 7.2; Muthén & Muthén, 1998-2014). In M*plus,* "standard errors for the parameter estimates are computed using the observed information matrix" (Muthén & Muthén, 1998-2012, p. 387; see also Kenward & Molenberghs, 1998). We used MIM when data was analyzed with manifest variables via IBM SPSS Statistics (Version 21.0; IBM Corp., 2012). In general, in case MIM is conducted, predominantly between $m = 5$ to $m = 10$ new imputations are generated (Collins et al., 2001). Within this thesis, we generated $m = 10$ data sets.[39]

3.6. Causality Assumptions in (Quasi-)Experimental Analyses

Causality is given in case three conditions are fulfilled: the cause presumes the effect, cause and effect are interrelated, and there are no plausible alternative explanations (e.g., Antonakis, Bendahan, Jacquart, & Lalive, 2010). Within our quasi-experiments, the causes (interventions) presumed the effects (improved attitude and behavior intentions toward the outgroup) and were also related to them. However, given that we used a quasi-experimental design, our groups were non-randomized, so that alternative explications could not be excluded without limitations. As Smith (2013) stated, "randomization is the principal factor dividing experimental from *quasi-experimental* research designs, the latter

[39] Van Buuren (2012) complemented that under certain circumstances (e.g., high amount of missing information) a higher number of imputations $m = 20\text{-}100$ is also reasonable.

of which may have strong comparative components, but whose *validity* can be jeopardized by the presence of 'extraneous variables'" (p. 46, see also Campbell & Stanley, 1963). Due to the fact that it is not always possible to randomize groups, as was the case within this thesis, one has to deal with internal validity threats. According to Campbell and Stanley (1963) there are the following internal validity threats: *history* (e.g., events that occur concurrently with the treatment), *maturation* (e.g., natural change over time), *testing* (e.g., exposure to the test), *instrumentation* (e.g., nature of a measure), *regression* (e.g., high initial scores that are less extreme after the treatment), *selection* (e.g., systematic differences across the groups), *mortality* (attrition; i.e., loss of participants), and *additive interaction effects*. In case there is no reason to assume that intervention group (IG) and control group (CG) differ systematically (e.g., history, maturation, testing, and instrumentation effects), IG and CG are exposed to the same conditions, which indicates that inferences about the treatment can be drawn from the data. However, in case specific differences between the groups cannot be ruled out (e.g., selection, mortality, and regression effects), additional investigations should be carried out to enable a more substantial understanding of the data. In order to approach whether selection effects are given, pre-test differences between groups in regard to the dependent variable provide a deepened understanding of the data; in regard to mortality effects, the analysis of pretest-values (dependent variable) between those individuals who left the experiment and those who remained are helpful; regarding regression effects, the examination of means across time provide more insights into the data. The last threat, namely the interaction between different components, is a vulnerable threat to internal validity, especially - as in this thesis - when many interactions are conceivable (e.g., Treatment x History, Treatment x Mortality, or Treatment x Selection interactions). These internal validity threats cannot be ruled out. Therefore, it has to be analyzed and discussed to what extent internal validity threats are prohibiting causal assumptions. Within this thesis, we discussed possible validity threats for all six samples in the submitted publications (see Manuscript #1, Manuscript #2).

4. Research Questions

As shown above, few studies have analyzed long-term face-to-face contact intervention effects within intractable conflict areas using quantitative methods. This thesis contributes to a more systematic investigation of contact intervention effects over time in

these specific areas (i.e., short- and long-term, as well as fading effects of intergroup contact interventions). We investigated two face-to-face contact interventions between Jewish Israeli and Arab Israeli students in Israel (4th and 6th grade). The intergroup contact intervention for 4th grade students consisted of four, the intervention for 6th grade students of two intergroup meetings. Within this thesis, we studied the impact of these intergroup contact interventions on attitudes and behavior intentions toward the outgroup across time (pre-, post-, follow-up test). We also examined characteristics of interventions and their participants that influence sustainable intergroup contact intervention effects (e.g., group status, repetition of intergroup meetings, and valence of reported personal experience within the contact interaction) and thereby extended previous research studies. In the following, we outline our specific research questions.

In order to analyze short- and long-term changes and thus confirm findings from previous studies, our primary research question was:

- *Do intergroup contact interventions, which include repeated intergroup meetings, cause short- and long-term attitude and behavior intention changes in intractable conflict areas?*

As mentioned, previous studies discovered that improved intergroup relations after intergroup interventions in intractable conflict areas decreased again after several weeks (e.g., Arnon, 2010). Hence, we wanted to replicate these findings within our sample and therefore proposed an additional research question:

- *Do improved attitude and behavior intentions at the end of an intergroup contact intervention worsen to their original level after several weeks in intractable conflict areas?*

To discover whether specific characteristics of intervention programs and its participants have an impact on short- and long-term contact intervention effects, we examined three additional aspects: the role of contact repetition, the role of students' intergroup experience within the intervention, and the role of status differences. By our knowledge, only one study has investigated whether the implementation of an additional intergroup contact intervention (repetition of contact experience) extended long-term intergroup contact effects within intractable conflict areas (Jayusi, 2009). It was discovered that contact effects can be extended when a further contact scenario was implemented about two to three months after the end of the intervention. Therefore, we tested whether short- and long-term effects differed if students participated in one or in four intergroup contact meetings of the same intergroup contact intervention. Our research question was:

- *Does the repetition of intergroup contact meetings has an impact on short- and long-term attitude and behavior intention changes within intractable conflict areas?*

We also evaluated if student's personal experience – positive, neutral, or negative –, which they reported at the end of the contact intervention, influenced short- and long-term contact intervention effects. Due to the fact that positive contact (e.g., friendship) showed positive influences on intergroup relations (Schroeder & Risen, 2016), and negative contact can be related to negative intergroup relations, we assumed the valence of the contact experience to be related to short- and long-term contact intervention effects. Our further research question was:

- *Does students' personal experience have an impact on short- and long-term attitude and behavior intention changes within intractable conflict areas?*

The third aspect we examined was that of status differences. Larger contact intervention effects are usually discovered for majority members (Lemmer & Wagner, 2015), in accordance with this fact, a further research question focused on this aspect:

- *Does students' status (minority or majority group membership) have an impact on short- and long-term attitude and behavior intention changes?*

In regard to methodical issues, measurement equivalence throughout time and between groups (intervention and control group, age groups, status groups) is a necessary precondition for longitudinal quasi-experimental analysis. Therefore, we investigated this underlying assumption explicitly.

- *Does students' understanding of the attitude and behavior intention variable differ across time and across groups, such as intervention and control group, age groups, status groups?*

In order to give an overview of the studies conducted in this thesis, Table 4 represents a summary of all six samples that were analyzed regarding the above mentioned research questions.

Table 4: Overview of all six Samples Analyzed Within the Thesis

Sample	Contact Intervention	Status Group	Grade	POM	No. of School classes	Design
1	1	Majority	4	3	2 IG; 2 CG	Pre-Post-Follow-Up Test with non-randomized CG
2	1	Minority	4	3	2 IG; 2 CG	Pre-Post-Follow-Up Test with non-randomized CG
3	1	Majority	4	5	1 IG	Pre-Post-Follow-Up Test
4	1	Minority	4	5	1 IG	Pre-Post-Follow-Up Test
5	2	Majority	6	3	3 IG; 2 CG	Pre-Post-Follow-Up Test with non-randomized CG
6	2	Minority	6	3	3 IG; 2 CG	Pre-Post-Follow-Up Test with non-randomized CG

Note. POM = points of measurement; IG = intervention group, CG = control group.

Table 5 outlines which research question was examined within which manuscript and for which sample.

Table 5: Overview of the Research Questions

Research Question	Sample	Manuscripts
Do intergroup contact interventions, which include repeated intergroup meetings, cause short- and long-term attitude and behavior intention changes in intractable conflict areas?	▪ 1 ▪ 2 ▪ 3 ▪ 4 ▪ 5 ▪ 6	▪ #1 ▪ #1 ▪ #1 ▪ #1 ▪ #2 ▪ #2
Do improved attitude and behavior intentions at the end of an intergroup contact intervention worsen to their original level after several weeks in intractable conflict areas?	▪ 1 ▪ 2 ▪ 3 ▪ 4 ▪ 5 ▪ 6	▪ #1 ▪ #1 ▪ #1 ▪ #1 ▪ #2 ▪ #2
Does the repetition of intergroup contact meetings has an impact on short- and long-term attitude and behavior intention changes within intractable conflict areas?	▪ 3 ▪ 4	▪ #1 ▪ #1
Does students' personal experience has an impact on short- and long-term attitude and behavior intention changes within intractable conflict areas?	▪ 1 ▪ 2 ▪ 3 ▪ 4 ▪ 5 ▪ 6	▪ #1 ▪ Synopsis ▪ #1 ▪ Synopsis ▪ #2 ▪ #2
Does students' status (minority or majority group membership) has an impact on short- and long-term attitude and behavior intention changes?	▪ 1 ▪ 2 ▪ 5 ▪ 6	▪ #1 ▪ #1 ▪ #2 ▪ #2
Does students' understanding of the attitude and behavior intention variable differ across time and across groups, such as intervention and control group, age groups, status groups?	▪ 1 ▪ 2 ▪ 5 ▪ 6	▪ #1, Synopsis ▪ #1, Synopsis ▪ #2, Synopsis ▪ #2, Synopsis

II. CURRENT RESEARCH MANUSCRIPTS

1. Manuscript #1

1.1. Introduction to Manuscript #1

The first manuscript investigated whether an implemented contact intervention achieved long-term positive attitude and positive behavior intention changes. To approach this analysis we examined an intergroup intervention consisting of four intergroup meetings between Jewish Israeli and Arab Israeli 4th grade students in Israel. The intervention included multiple meetings due to previous research findings, which discovered that additional contact interventions are necessary to achieve long-term effects over more than two months in Israel (e.g., Jayusi, 2009). The article attempts to extend these findings by showing that not only an additional intervention but also a repetition of the same intervention (repeated meetings) enable long-term contact intervention effects in Israel. To examine long-term contact intervention effects systematically, we also analyzed short-term as well as fading effects. Moreover, we investigated whether or not the valence of students' personal experience and the repetition of the contact meetings had an impact on the contact intervention outcome. We evaluated four samples: two samples were questioned three times (Sample 1: 113 Jewish Israeli 4th grade students [n = 60 IG students; n = 53 CG students]; Sample 2: 108 Arab Israeli students [n = 60 IG students; n = 48 CG students]) and two samples were questioned five times (Sample 3: 33 Jewish Israeli IG students; Sample 4: 30 Arab Israeli IG students). Hereafter, we investigated the following research questions of this thesis: Do intergroup contact interventions in intractable conflict areas that include repeated intergroup meetings influence intergroup relations in regard to short- and long-term as well as fading effects in intractable conflict areas? Do students' personal intergroup experiences (positive, neutral, or negative) influence intergroup relations in intractable conflict areas (short- and long-term effects)? Does the repetition of intergroup contact meetings influence intergroup relations in intractable conflict areas (short- and long-term effects)? Does the understanding of the measured constructs (attitudes and behavior intentions) differ between IG and CG at each point in time and across all points of measurement? Data of Sample 1 and 2 were analyzed using latent change models and measurement invariance was ensured ahead of our calculations.

1.2. Submitted Manuscript #1

Long-Term Contact Intervention Effects in a Conflict Area: The Role of Personal Experiences and the Repetition of Intergroup Contact Meetings

Kerstin Guffler and Ulrich Wagner
Philipps-University Marburg

Author Note

Kerstin Guffler, Department of Psychology, Philipps-University Marburg; Ulrich Wagner, Department of Psychology, Philipps-University Marburg.

Corresponding Author Contact Details

Correspondence concerning this article should be addressed to Kerstin Guffler, Philipps-University Marburg, Psychology Department, Gutenbergstr. 18, 35032 Marburg, Germany; tel.: 0049-(0)6421-23636; mail: guffler@uni-marburg.de

(Submission date: March 28th, 2016; under review in *Group Processes & Intergroup Relations*)

Abstract

Previous research gives reason to assume that positive effects of single intergroup contact interventions in intractable conflict areas cannot be sustained for a very long time (Arnon, 2010). This has led to the idea to improve long-term effectiveness by adding booster sessions. Study 1 illustrates that multiple short intergroup meetings between Jewish Israeli and Arab Israeli students in Israel led to sustainable changes at least for majority members. We also discovered that the valence of personal intergroup experience had an impact on short- and in case of the majority long-term effects. Taking the intercultural context of the data sampling into account, the data were analyzed with latent difference score modeling. In Study 2, we showed that long-term intervention effects were in fact related to the repetition of contact meetings.

Keywords: long-term contact intervention effect, repeated meetings, latent difference score modeling, intractable conflict

Long-Term Contact Intervention Effects in a Conflict Area: The Role of Personal Experiences and the Repetition of Intergroup Contact Meetings

When thinking about severe intergroup conflicts, examples such as the Middle East conflict between Jews and Muslims in Israel and Palestine, “The Troubles” between Catholics and Protestants in Northern Ireland, or the post-Apartheid conflict between Whites, Coloureds, Blacks, and Asians in South Africa cross one’s mind. The overlapping characteristics of these severe conflicts are violence and longevity (Bar-Tal, 1998; Kriesberg, 1998), which implies that the conflicts threaten the physical integrity and the psychological health of individuals who are affected by them over a long time (Schmid & Muldoon, 2013). One promising way to improve intergroup relations, even in conflict areas, are intergroup contact interventions (Lemmer & Wagner, 2015). However, positive intervention effects in such areas have been shown to fade within a few months after the intervention (e.g., Shani, 2015). One suggestion to extend the sustainability of contact interventions is to repeat intergroup meetings (e.g., Salomon, 2011). Until now, we only know about one study that investigated long-term contact intervention effects in an intractable conflict area before and after implementing an extra contact situation sometime after the end of the initial contact (Jayusi, 2009). In order to broaden knowledge about repeated contact effects in these adversarial contexts, we investigated an intergroup program consisting of four short meetings between Jewish Israeli and Arab Israeli students. More precisely, we examined whether intervention group (IG) students who participated in the program increased their short- and long-term attitudes and behavior intentions toward the outgroup in comparison to control group (CG) students and whether the above mentioned fading effects appeared. We further investigated whether the valence of student’s contact experience and the repetition were related to the long-term contact intervention outcome.

Theoretical Background

In 1954, Allport postulated that prejudice can be reduced by intergroup contact, especially when the contact situation is characterized by the following conditions: same status of the participants, institutional support, cooperation instead of competition, and a common goal (Pettigrew, 1971). In 1998, Pettigrew complemented these conditions with a further important aspect, namely the opportunity to become friends within the contact situation. The meta-analysis by Pettigrew and Tropp (2006) exhibited that the contact theory was empirically supported by a great number of studies and that studies which

measured intergroup friendship as an independent variable illustrated higher prejudice reduction effects than studies that used other independent variables. In regard to contact interventions, Lemmer and Wagner (2015) confirmed these findings in their meta-analysis showing that intergroup interventions lead to improved intergroup relations as well.

Short- and Long-Term Contact Intervention Effects in Conflict Areas

Lemmer and Wagner (2015) revealed that contact interventions improve intergroup relations even in conflict areas comparing IG to CG participants ($\hat{\mu} = .31$, $k = 11$ number of comparisons, $n = 5$ studies). However, meta-analytic results were measured up to one month after the contact intervention took place, illustrating *short-term* outcomes. The meta-analysis also investigated *long-term* contact intervention effects; however, due to the small number of studies not only intergroup interventions in conflict but also in non-conflict areas were included in the analysis.[1] Thus, no clear insight in regard to long-term contact intervention effects in conflict regions could be drawn. Looking at studies, which examined long-term contact intervention effects (Arnon, 2010; Bar-Natan, Rosen, & Salomon, 2010; Berger, Benatov, Abu-Raiya, & Tadmor, 2016; Connolly, 1992; Guffler & Wagner, 2016; Jayusi, 2009; Kropiunigg & Pabst, 2007; Luiz & Krige, 1985; Schleien, 2007; Schroeder & Risen, 2016; Shani, 2015)[2] most of them discovered positive short-term but only few revealed positive long-term outcomes (e.g., Berger et al., 2016).

Fading Contact Intervention Effects in Conflict Areas

Upon a closer examination of these studies, it seems that, in most cases, positive changes decreased (faded) with the passing of time (e.g., Shani, 2015). The contextual conditions of the conflicts are possible reasons for these fading effects (Salomon, 2011). For example, Israeli are constantly exposed to conflict symbols so that they remain aware

[1] None of these studies included a control group design.

[2] Studies that are listed and which are not included in the meta-analysis by Lemmer and Wagner (2015) did not meet the inclusion criteria of the meta-analysis: Arnon (2010) and Jayusi (2009) are doctoral dissertations written in Hebrew; therefore, our information is based on the English summary and the methodical procedures of these articles cannot be questioned. Moreover, Berger et al. (2016), Guffler and Wagner (2016), Schroeder and Risen (2016) as well as Shani (2015) were only published recently. Due to the fact that we are interested in long-term effects (i.e., effects, which remain from the post- until the follow-up measurement), only studies including at least two posttests are presented.

of the conflict in their every-day life (Vered & Bar-Tal, 2014). Thus, improved intergroup relations, which are present shortly after the end of such interventions, have little chance to persist where contextual influences emphasize and preserve group differences. There is only a vague idea about how much time passes before positive intervention effects start to fade, because the time intervals between the end of an intervention and the follow-up measurement varies across studies. In regard to interventions between Jewish and Arab Israeli students in Israel, researchers identified that contact intervention effects scored approximately at the pre-intervention level already about two months after the intergroup meeting took place (e.g., Arnon, 2010).

Extended Contact Intervention Effects in Conflict Areas

Given that sustainable effects are intended in most contact interventions, researchers looked for means to accomplish this goal. In order to establish long-term effects and restrict fading processes, some researchers suggested to repeat contact meetings (e.g., Kupermintz & Salomon, 2005).[3] To our knowledge, only Jayusi (2009) implemented an additional peer-tutoring two months after a two-day intergroup contact intervention in Israel and showed that long-term effects were prolonged to a two-month timespan. Additionally, researchers revealed that students who made intergroup friends during their participation showed the strongest long-term effects. Therefore, enabling intergroup friendship might also lead to prolonged extended effects (Schroeder & Risen, 2016).

The Present Studies

In Study 1, we analyzed short-, long-term, and fading effects for Jewish Israeli and Arab Israeli students who participated in an intergroup intervention consisting of four short consecutive intergroup meetings in Israel. Thus, we were able to examine the repetition of intergroup contact and attempted to replicate previous findings. However, in contrast to the study of Jayusi (2009) we did not implement an additional contact intervention, but investigated the effect of repeated meetings within one intervention. Additionally, we examined whether the valence of students' intergroup experiences (e.g., making friends)

[3] Another opportunity to assure long-term contact intervention effects might be the establishment of contact interventions with a long durability. Accordingly, Schroeder and Risen (2016) reported long-term contact effects (around 9-12 months after the end of the intergroup meetings), at least in one out of two investigated interventions for students who met for the duration of three weeks.

had an impact on short- and long-term intergroup relations. To analyze our data, we used state-of-the-art methods including structural equation modeling and measurement invariance calculations.

In Study 2, we surveyed students who participated in the same intervention five times, meaning two extra times compared to Study 1. The additional questionings were placed in between the four meetings and enabled us to examine the role of repeated intergroup meetings more precisely.

Based on the fact that there are status differences between Jewish Israeli (majority) and Arab Israeli (minority) in Israel and thus deviating experiences in regard to the contact experiences are expected (Salomon, 2011), we examined both groups separately.

Study 1

We investigated whether an intergroup school program that included four consecutive intergroup meetings between Jewish and Arab Israeli students led to short- and long-term changes of intergroup attitudes and behavior intentions and whether fading effects appeared a few weeks after the end of the intervention. Moreover, we examined the role experiences that students made during the intergroup meetings. We obtained the permission to survey students in this particular school year from the Israeli Ministry of Education in loco parentis. Correspondingly, we informed students at the beginning of the questionings that they could stop answering the questionnaire at any time and that their information was treated anonymously.

Hypotheses

Hypothesis 1: Intervention effects. H1a: Participation in multiple intergroup meetings leads to improved short-term attitudes and behavior intentions toward the outgroup for the IG compared to CG students. H1b: Participation in multiple intergroup meetings leads to improved long-term attitudes and behavior intentions toward the outgroup for the IG compared to the CG. H1c: Improved attitudes and behavior intentions toward the outgroup fade about three months after the end of the intervention.

Hypothesis 2: Relevance of students' experiences. H2a: Students who report positive intergroup experience at the posttest, show more positive attitudes and behavior intentions than students who report negative or neutral experiences (short-term effect). H2b: Expressed positive experiences at the posttest predict more sustainable intervention effects (long-term effect) than negative or neutral experiences.

Measures. Besides other variables unrelated to the present study, we surveyed students with respect to their attitudes and behavior intentions toward the outgroup. Additionally, we asked IG students to report about their intergroup experiences. Questionnaires were translated from English into Hebrew by a translation agency. In order to attain accurate translations, wording and phrasing were discussed with translators and staff of the Israeli institution and back-translations were conducted (Harkness, 2003).

Attitudes toward Jewish (Arab) Israeli students. We measured this variable with three items on a five-point scale each (1 = *not very much* to 5 = *very much*) adapted from Lanphen (2011). Students were requested to respond to the following statements: "Jewish (Arab) Israeli students are… … nice; ... friendly; … kind." Thereby, higher scores indicated more positive attitudes.

Behavior intentions toward Jewish (Arab) Israeli students. This three-item scale was based on an adoption of items used by Jayusi (2009): "I would like to … … speak to students from Jewish (Arab) Israeli schools; … meet students from Jewish (Arab) Israeli schools; … do common activities with students from Jewish (Arab) Israeli schools." Students answered each item on a five-point scale (1 = *not very much* to 5 = *very much*). Again, higher scores indicated more positive behavior intentions.

Reported Personal Experience. To learn about students' intergroup experiences, we asked IG students to respond to the open ended question: "Did you tell your parents about the meetings? If yes, what did you tell them?" The phrasing was originated in cooperation with local staff and based on their experience. Although this question was not validated before, we expected students to answer more openly than when responding to a direct question in the presence of the program's staff. A similar procedure was used by Graf, Paolini, and Rubin (2014). Students wrote their answers in Hebrew; comments were translated into English for our analysis.

Research design. We used a quasi-experimental pre-post-follow-up-test design. Students replied to the questionnaire three times: first, at their school one week before the first meeting (pretest; t1); second, exactly at the end of the last meeting at the place of the meetings (posttest; t2); third, about eleven weeks after the end of the program at school (follow-up-test; t3). The follow-up timespan of about two to three months (t2-t3) was chosen according to empirical findings in previous studies, which showed fading effects within two months after the contact intervention (e.g., Arnon, 2010). CG students were questioned around the same dates and filled out questionnaires at their school.

Intervention. The intervention took place between Jewish and Arab Israeli students with the goal of improving relations among Israeli citizens irrespective of their ethnic background. Throughout the intervention students were divided into small groups (about half Jewish and half Arab Israeli students). Each group was guided by one Hebrew and one Arab speaking instructor to ensure exchange between the students. In the course of the program students did sport or creative activities collaboratively and discussed differences and similarities between their cultures. The program took place at a location outside students' schools and included four consecutive intergroup meetings between October and December 2012. Each of the meetings lasted about four hours. Only IG students participated in the school program. Within the course of our study, the Israel-Gaza Conflict 2012 occurred (Haaretz, 2012a), lasting for eight days in November 2012. The violent outbreak included aerial assaults from both sides (Haaretz, 2012b). However, based on the fact that we conducted a quasi-experimental study, intergroup intervention effects could be examined independent from this incident.

Jewish Israeli Students

Method

Participants. 116 Jewish Israeli 4th grade students coming from one intervention and one control school in Northern Israel participated in the questionings. The intervention school was chosen according to its participation in the intergroup program; the control school was selected based on its agreement to participate in the questionings, its location in Northern Israel (same geographical area), and because of its inclusion of students from the same academic year. Students with a Christian or Muslim background were excluded from the analyses (n = 3). Thus, 113 Jewish students participated: 60 IG (45% females); 53 CG students (45% females). The number of students who participated at each questioning was N_{t1} = 106, N_{t2} = 96, N_{t3} = 97.

Preliminary analysis. We tested whether students who quit participating in the questionings at t2 or t3 differed in their pretest values from those who remained (Shadish, Cook, & Campbell, 2002). No differences were discovered regarding attitudes (t2: t[104] = -1.32, p = .377; t3: t[98] = 1.564, p = .129)[4] and behavior intentions (t2: t[104] = .80, p = .426; t3: t[101] = 1.483, p = .141). Data were missing due to wave and item

[4] Levene's test for equality of variances was violated; therefore, t-values are shown when equal variances are not assumed (Field, 2009).

nonresponse (Graham, Cumsille, & Shevock, 2013) and the exclusion of five questionnaires that could not be matched based on their codes. A non-significant Little's MCAR test revealed data were missing completely at random (MCAR; $\chi^2 = 54.18$, $df = 53$, $p = .429$; Little, 1988). Therefore, we calculated missing data via maximum-likelihood estimation (Schafer & Graham, 2002). No pretest differences between IG and CG were discovered for attitudes ($M_{IG} = 3.07$; $M_{CG} = 3.15$; $b = -0.08$, $SE = 0.11$, $p = .491$) but higher values were revealed for behavior intentions within the IG ($b = 0.37$, $SE = 0.12$, $p = .003$; $M_{IG} = 3.19$; $M_{CG} = 2.83$; Shadish et al., 2002). According to our nested data (IG and CG participants were from two school classes each), we controlled for level-2-units (Snijders & Bosker, 2012).[5] The reliability coefficient omega (ω; McDonald, 1999) was satisfactory for attitudes ($\omega_{t1} = .82$, $\omega_{t2} = .75$, $\omega_{t3} = .88$) and for behavior intentions ($\omega_{t1} = .77$, $\omega_{t2} = .81$, $\omega_{t3} = .79$).[6] Model fit of the CFA consisting of two dependent variables (attitudes and behavior intentions) was good at t1 ($\chi^2_{corrected}$[7] $= 9.99$; $df = 8$; $p = .266$; CFI = .99; RMSEA = .05; SRMR = .03),[8] t2, and t3. Longitudinal measurement invariance (MI) was ascertained to enable latent mean comparison across time. To compare latent means, longitudinal scalar MI is required (Davidov, 2008).[9] Longitudinal partial metric partial strict scalar MI was reached for attitudes ($\chi^2_{corrected} = 31.61$; $df = 27$; $p = .247$; CFI = .98; RMSEA = .04; SRMR

[5] To control for hierarchical effects, we generated two variables (variable 1: ig_1 = -1, ig_2 = 1, cg_1 = 0, cg_2 = 0; variable 2: ig_1 = 0, ig_2 = 0, cg_1 = -1, cg_2 = 1; see Figure 1).

[6] Omega was calculated using M*plus* input instructions developed by Yang and Green (2011).

[7] Within all model estimations the χ^2 value was corrected using the *maximum likelihood estimation with robust standard errors* (MLR), which is robust to non-normality (Muthén & Muthén, 1998-2012, p. 603).

[8] To calculate the model fit we applied the χ^2 test, the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). A satisfactory model fit was indicated when χ^2 p-value $\geq .01$, CFI $\geq .95$, RMSEA $\leq .08$, and SRMR $\leq .10$ (Schermelleh-Engel, Moosbrugger, & Mueller, 2003, p. 52).

[9] We followed Steenkamp and Baumgartner (1998) and ascertained at least partial MI (i.e., at least two invariant factor loadings and two invariant intercepts per construct).

= .08) and longitudinal strict scalar MI for behavior intentions ($\chi^2_{corrected}$ = 37.13; *df* = 29; *p* = .143; CFI = .97; RMSEA = .05; SRMR = .06).[10]

Results

For our analysis we used latent difference score models, also known as latent change models (LCM). LCM use latent variables, meaning variables which are adjusted for measurement errors (Brown, 2006). In general, LCM indicate within-person changes in the dependent variables (Jajodia, 2012). In our study, we applied two different LCM (Steyer, Partchev, & Shanahan, 2000): First, we calculated a baseline-change-version of the LCM (Figure 1). Thereby, change was defined in relation to the first point of measurement: difference score 1 (Diff1) was the difference between latent means at t1 and t2 (t2-t1; short-term effect); difference score 2 (Diff2) represented the difference between latent means at t1 and t3 (t3-t1; long-term effect). We tested whether Diff1 and Diff 2 varied between IG and CG (McArdle, 2007): more positive difference scores for the intervention group compared to the control group would indicate short- and long-term intervention effects.

[insert Figure 1.]

Second, we analyzed a neighbor-change-version of the LCM to examine the above mentioned fading effects (Figure 1). Within this model, change was analyzed in reference to neighboring points of measurement: Again, Diff1 was examined, while we also calculated latent difference score 3 (Diff3), which represented the difference between latent means at t2 and t3. A less positive score of Diff3 in the IG compared to the CG would indicate fading effects. Data were analyzed using M*plus* (Version 7.2; Muthén & Muthén, 1998-2014).

[10] To investigate whether MI was given between IG and CG, we also tested cross-group MI for each point of questioning. At least partial metric partial scalar MI between the two groups for the positive attitude (pretest: [$\chi^2_{corrected}$ = 5.64; *df* = 5; *p* = .343; CFI = .99; RMSEA = .05; SRMR = .03]; posttest: [$\chi^2_{corrected}$ = 6.14; *df* = 7; *p* = .524; CFI = 1.00; RMSEA = .00; SRMR = .09]; follow-up: [$\chi^2_{corrected}$ = 2.55; *df* = 4; *p* = .635; CFI = 1.00; RMSEA = .00; SRMR = .04]) and the positive behavior intention variable was reached (pretest: [$\chi^2_{corrected}$ = 2.97; *df* = 3; *p* = .396; CFI = 1.00; RMSEA = .00; SRMR = .05]; posttest: [$\chi^2_{corrected}$ = 1.39; *df* = 5; *p* = .924; CFI = 1.00; RMSEA = .00; SRMR = .08]; follow-up: [$\chi^2_{corrected}$ = 1.62; *df* = 4; *p* = .806; CFI = 1.00; RMSEA = .00; SRMR = .06]).

Attitudes toward Arab Israeli students. The model fit of the two LCM was satisfactory ($\chi^2_{corrected}$ = 62.57; *df* = 47; *p* = .064; CFI = .96; RMSEA = .05; SRMR = .08). Diff1 was 0.41 (σ^2 = .25; *p* = .007). Results revealed that Diff1 differed between the two conditions (β = .59, *p* < .001) showing that IG students' attitudes toward Arab Israeli students developed more positively from t1 to t2 than those of CG students (short-term effect). Diff2 was 0.30 (σ^2 = .71, *p* < .001), also showing that the change in attitudes between t1 and t3 was more positive for IG students compared to CG students (β = .23, *p* = .017; long-term effect). Diff3 was -0.11 (σ^2 = .38; *p* = .020) and varied between the two conditions (β = -.34, *p* = .002), pointing to a reduction of positive attitudes toward Arab Israeli students from t2 to t3 in the IG in comparison to the CG. Latent means are depicted in Table 1.

In sum, in line with H1a+b, the participation in four short intergroup meetings increased students' attitudes toward Arab Israeli from t1 to t2 and from t1 to t3, indicating short- and long-term (eleven weeks) intervention effects. However, as hypothesized (H1c) this effect decreased from t2 to t3.[11] In a next step, we examined whether behavior intentions changed accordingly.

[insert Table 1.]

Behavior intentions toward Arab Israeli students. The model fit was satisfactory ($\chi^2_{corrected}$ = 62.94; *df* = 49; *p* = .087; CFI = .96; RMSEA = .05; SRMR = .06). Diff1was -0.13 (σ^2 = .66; *p* < .001). Thus, our findings revealed that Diff1 did not vary between the two conditions (β = .11, *p* = .396). Diff2 was -0.35 (σ^2 = 1.22, *p* < .001), again, no differences were found between groups (β = .02, *p* = .885). Diff3 was -0.22 (σ^2 = .77; *p* = .007) and not different between IG and CG (β = -.09, *p* = .508).

Therefore, H1a-c were not supported. Overall, H1a-c were partly confirmed: Interestingly, our hypotheses were confirmed referring to the attitude but not the behavior intention variable.

[11] When adjusting the *p*-values according to the Bonferroni correction (i.e., divide α by two; because we investigated two dependent variables separately within our latent change analyses) we also discovered significant short- and long-term as well as fading effects for the three β-values within the positive attitude variable (Bender, Lange, & Ziegler, 2007).

Reported Students' Experience. We explored whether the valence of students' intergroup experience reported at the posttest had an effect on short- and long-term contact intervention outcomes. Students' open answers were classified into categories: comments reporting positive or neutral personal experiences, using the qualitative data software MAXQDA (Kuckartz, 1995-2014). Positive comments included aspects such as *positive emotions related to the activities* (e.g., "I told them [parents] that I really enjoyed doing the [activity]"), aspects about *pleasant interactions* with Arab Israeli students (e.g., "I met with the Arab children and realized that Arabs are not so bad as I thought"), and aspects about *friendship* (e.g., "I had a lot of fun and made new friends"). Neutral comments referred to a description of the program (e.g., "We meet with children from other schools"). There were no negative comments at t2. 85% of the IG students responded to the open ended question.[12] To compute interrater reliability, 30% of all comments were randomly selected and coded by a second coder. Codings indicated a good consensus between the two raters (Krippendorff's $\alpha = .89$; Hayes & Krippendorff, 2007; Krippendorff, 2004). Therefore, the first rater's categorization was maintained and transformed into quantitative data (neutral comment = 0; positive comment = 1).We computed several regression analyses[13]: First, we tested whether t2 comments had an effect on quantitative outcomes at t2 (short-term effect); second, we analyzed whether t2 comments also predicted quantitative outcomes at t3 (long-term effect). Both times we controlled for quantitative t1 values. As can be seen in Table 2, expressed positive experiences had an effect on short- and long-term attitudes (t2: $\beta = .25$, $p = .055$; t3: $\beta = .33$, $p = .016$) as well as on short- and long-term behavior intentions (t2: $\beta = .28$, $p = .018$; t3: $\beta = .48$, $p < .001$). Thus, data supported H2a+b.

[insert Table 2.]

Arab Israeli Students

Method

Participants. 108 Arab Israeli 4th grade students took part in the questionings: 60 IG students (52% females); 48 CG students (46% females). The number of students participating at each occasion of measurement was $N_{t1} = 103$, $N_{t2} = 87$, $N_{t3} = 99$. IG students attended the exact same four intergroup meetings as Jewish Israeli IG students. T1 and t2

[12] The number shows the percentage of IG students that reported their opinion in relation to all IG students that attended this specific time of questioning.

[13] We used listwise deletion because qualitative answers could not be imputed.

questionings were at similar points in time as for the Jewish sample. Due to feasibility, the follow-up measurement was slightly earlier (i.e., about ten weeks later). The same questionnaire was used but translated into Arabic to enable students answering the questions in their first language. As a prerequisite to our analysis, we tested again whether the two constructs (attitudes and positive behavior intentions) were equivalent over time. Due to the fact that longitudinal partial scalar MI was not found within the attitude variable, results are only presented for behavior intentions.

Preliminary Analysis. Data showed no selective attrition (t2: $t[94] = 0.20$, $p = .841$; t3: $t[94] = -.380$, $p = .705$; Shadish et al., 2002). Missing data occurred due to wave and item non-response (Graham et al., 2013) and problems in matching questionnaires ($n = 4$ questionnaires). Missing data were MCAR ($\chi^2 = 82.54$, $df = 66$, $p = .082$; Little, 1988) and analyzed via maximum-likelihood estimation (Schafer & Graham, 2002). Pretest differences were found ($M_{IG} = 2.47$, $M_{CG} = 2.70$; $b = -0.23$, $SE = 0.11$, $p = .045$). Again, we controlled for the hierarchical data structure. Reliability for behavior intentions was satisfactory ($\omega_{t1} = .86$, $\omega_{t2} = .83$, $\omega_{t3} = .74$). Longitudinal partial metric partial strict scalar MI was achieved for behavior intentions ($\chi^2_{corrected} = 29.13$; $df = 28$; $p = .406$; CFI = .99; RMSEA = .02; SRMR = .06).[14] Additionally, longitudinal MI between Jewish Israeli and Arab Israeli students was achieved so that latent means from the majority and the minority were comparable: $\chi^2_{corrected} = 19.54$; $df = 29$; $p = .907$; CFI = 1.00; RMSEA = .00; SRMR = .04.[15]

[14] At least cross-group scalar MI (between IG and CG) was given at t1 and t3, no cross-group MI was found at t2: pretest: ($\chi^2_{corrected} = 2.18$; $df = 4$; $p = .642$; CFI = 1.00; RMSEA = .00; SRMR = .04); posttest: (metric MI: $\chi^2_{corrected} = 7.84$; $df = 3$; $p = .049$; CFI = .91; RMSEA = .20; SRMR = .09); follow-up: ($\chi^2_{corrected} = 4.52$; $df = 4$; $p = .340$; CFI = .99; RMSEA = .05; SRMR = .10).

[15] Also at least cross-group partial metric partial scalar measurement equivalence between Jewish Israeli and Arab Israeli students at each time of measurement was achieved: pretest: ($\chi^2_{corrected} = 5.98$; $df = 4$; $p = .201$; CFI = .99; RMSEA = .07; SRMR = .07); posttest: ($\chi^2_{corrected} = 6.13$; $df = 4$; $p = .190$; CFI = .98; RMSEA = .08; SRMR = .10); follow-up: ($\chi^2_{corrected} = 2.97$; $df = 5$; $p = .705$; CFI = 1.00; RMSEA = .00; SRMR = .07).

Results

As described before, three latent difference scores were computed to analyze short-term (Diff 1: pre- to posttest), long-term (Diff 2: pre- to follow-up-test), and fading effects (Diff3: post- to follow-up-test; Figure 1). Results illustrated a satisfactory model fit ($\chi^2_{corrected} = 47.98$; $df = 46$; $p = .393$; CFI = .99; RMSEA = .02; SRMR = .06). Latent means and model indices are outlined in Table 1. Findings showed that Diff1 ($\sigma^2 = 1.01$, $p < .001$) varied between IG and CG ($\beta = .19$, $p = .008$).[16] Diff2 ($\sigma^2 = .99$, $p < .001$; $\beta = .11$, $p = .120$) and Diff3 illustrated no differences between IG and CG ($\sigma^2 = 1.53$, $p < .001$; $\beta = -.13$, $p = .311$). Thus, H1a was confirmed, whereas H1b+c were not confirmed, meaning that the participation in multiple intergroup meetings led to increased short- but not to long-term behavior intentions. Also no fading effects appeared. Findings showed that behavior intentions within the IG developed in accordance to our assumptions, however, also a slightly positive development within the CG occurred over time.

Reported Students' Experience. Again, students' answers were coded as positive or neutral. Interrater agreement was good (Krippendorff's $\alpha = .81$; Krippendorff, 2004). 73% of the students who attended the inquiry wrote a comment about his or her experience. Qualitative and quantitative answers were again interrelated. Results indicated a marginal short- ($\beta = .31$, $p = .085$) but no long-term effect ($\beta = .07$, $p = .752$).[17] Thus, H2a was whereas H2b was not confirmed.

Discussion

Jewish Israeli Students. After ensuring longitudinal (partial) scalar MI, our data demonstrates that participation in repeated intergroup meetings in a conflict context led to short- and long-term improvements in intergroup attitudes for the IG in comparison to the CG. Additionally, we showed that contact intervention effects also faded after repeated contact interventions in conflict areas; although, contact effects still remained significant from t1 to t3. Thus, these results extended previous research findings (Jayusi, 2009) by showing that an intergroup program which includes four short intergroup meetings also achieves long-term effects of about two to three months. However, data exhibited effects

[16] Even when the Bonferroni correction was applied and the *p*-value adjusted, significant results for the β-values regarding to Diff1 were found (Bender et al., 2007).

[17] Again, in both calculations listwise deletion was used.

only with regard to attitudes and not to behavior intentions. Following this, our hypotheses were only partially confirmed.

We are uncertain why participants' attitudes were affected by the intervention while their behavior intentions were not. Probably, there are variables that are influenced more easily through intergroup contact than others (Rosen & Salomon, 2011).

We additionally discovered that students who expressed positive intergroup experiences showed more positive short- and long-term attitudes and behavior intentions than those reporting neutral experiences at t2. Thus, results emphasized the important role of the valence of intergroup contact experiences in a conflict area. It also illustrated that positive experiences are manifold and do not necessarily have to be related to outgroup friendships only. While the measurement of personal experiences has a high ecological validity by permitting all conceivable answers, the used questions were phrased based on local staff experiences and not validated ahead of our analyses. Although, other researchers have used similar instruments (Graf et al., 2014) the instrument's missing validation presents a limitation to our study, because there is no guarantee that students' answers represent their actual contact experience. Therefore, results should be interpreted cautiously.

We also observed pretest differences within the behavior intention variable (i.e., IG showed more positive behavior intentions than CG). These differences may be explained by school differences (e.g., school climate) and probably occurred due to our nonequivalent comparison group design. Finally, we revealed that IG and CG students had an equivalent understanding of the measured constructs and thus strengthened the assumption of a possible intergroup comparison.

Arab Israeli Students. Within the Arab Israeli sample, only behavior intentions were investigated, because the attitude scale did not show measurement invariance over time. Results illustrated that participating in an intergroup intervention that included repeated meetings led to improved short- but not long-term behavior intentions. Results might be based on the changes within the CG who also demonstrated slightly more positive behavior intentions throughout time. Also, no fading effects were found. In sum, there were no long-term effects of the intergroup contact program for the Arab minority members.

Also, pretest differences were found between the IG and the CG (i.e., IG students had smaller pretest values than CG). These differences might again occur due to our non-randomized CG (Paluck & Green, 2009). However, positive behavior intentions were

smaller for the IG than for the CG and not the other way around. Thus, the effect was changed after participation in the intervention.

No cross-group MI was attained at t2. This indicates that IG and CG had a different understanding of the behavior intention variable at the posttest. Changes within the understanding of a construct over time might be depending on a response shift within the IG due to the participation in the intergroup intervention (e.g., Millsap & Hartog, 1988). Looking strictly at these results, latent mean values of the IG and CG should not be compared at t2, though longitudinal MI was achieved.

An important strength of our analysis is that we revealed Jewish Israeli and Arab Israeli students to have the same understanding of the behavior intention variable, meaning that latent means of both samples are comparable. Against the background that questionnaires existed in two languages it is especially interesting that an equal understanding of this variable was found (Davidov & De Beuckelaer, 2010).

Study 2

In Study 2, we investigated whether the repetition of intergroup meetings was crucial for the intervention's effectiveness. By surveying students five times (i.e. two additional times compared to Study 1) we were able to contrast the effect of the first meeting alone with the cumulative effect of all meetings.

Hypotheses

Hypothesis 1: H1a: Participation in multiple intergroup meetings in Israel leads to stronger positive intervention effects exactly at the end of the intervention (short-term effect after four intergroup meetings; t4-t1) than participation in solely one intergroup meeting (short-term effect after the first intergroup meeting; t2-t1). H1b: Participation in multiple intergroup meetings in Israel leads to stronger positive intervention effects at the follow-up measurement (long-term effect after four intergroup meetings, t5-t1) than participation in solely one intergroup meeting (long-term effects after the first intergroup meeting; t3-t1). H1c: Positive contact effects fade (decrease) less strongly after participation in multiple intergroup meetings (fading effect after four intergroup meetings; t5-t4) than after participation in solely one intergroup meeting (fading effect after the first intergroup meetings; t3-t2).

Measures and research design. Students were given the same questionnaires as students in Study 1. The points of measurement were as follows: one week before the intergroup program started (t1; same as t1 in Study 1); one week after the *first* intergroup

meeting (t2); about four weeks after, before the *second* intergroup meeting (t3); at the end of all four intergroup meetings (t4; same as t2 in Study 1); and about eleven weeks after the *fourth* intergroup meeting (t5; same as t3 in Study 1).

Jewish Israeli Students

Method

Participants. We collected data from a Jewish Israeli 4th grade class that consisted of 33 students (48% female). Participating students attended the same school and academic year as Jewish IG students in Study 1. The number of students attending the program at each point of measurement was as follows: $N_{t1} = 30$, $N_{t2} = 31$, $N_{t3} = 30$, $N_{t4} = 27$, $N_{t5} = 29$. No CG was surveyed.[18]

Preliminary Analyses. According to the small sample size data were analyzed using manifest variables. Results indicated selective attrition regarding attitudes at t5 (t4: $t[24] = -.84$, $p = .934$; t5: $t[24] = -.2.37$[19], $p = .026$) but not for behavior intentions (t4: $t[24] = .55$[20], $p = .607$; t5: $t[24] = -1.61$, $p = .119$). Data were missing due to wave and item nonresponse (Graham et al., 2013, p. 110) and failed matching of two questionnaires. Data were MCAR ($\chi^2 = 82.81$, $df = 68$, $p = .107$; Little, 1988), thus, we dealt with missing data using multiple imputation (i.e., outcomes of ten imputed data sets were pooled by using the arithmetic mean of each data set; Baraldi & Enders, 2010). Cronbach's α showed uneven values for attitudes (t1: α = .59; t2: α = .59; t3: α = .48; t4: α = .89; t5: α = .90) and

[18] We did not question the other two IG classes from Study 1 five times, because we were concerned about test effects according to small time intervals in between the questionings (Shadish et al., 2002).

[19] Findings are based on four participants that were present at t1 and missed t5 questionings.

[20] Levene's test for equality of variances was violated within the behavior intention variable; therefore, we reported *t*-values when equal variances are not assumed (Field, 2009).

behavior intentions (t1: α = .82; t2: α = .66; t3: α = .90; t4: α = .82; t5: α = .95). Longitudinal MI testing was not tested due to the small sample size.[21]

Results

We conducted repeated measures ANOVA and examined two short- and two long-term effects. The first short-term effect referred to the measurement directly after students participated in the *first* intergroup meeting; the second short-term effect was related to the measurement directly at the end of all *four* intergroup meetings. Accordingly, the first long-term effect was investigated after the participation in one intergroup meeting and before the second meetings (about four weeks later) whereas the second was investigated after the participation in all four intergroup meetings at the follow-up test (about eleven weeks later). Analogously, two fading effects were examined. All means, standard deviations and time spans are delineated in Figure 2.

[insert Figure 2.]

Attitudes toward Arab Israeli students. Outcomes revealed that time had an impact on attitudes ($F[2.90, 92.74] = 8.92$, $p < .001$, $\eta^2 = .22$).[22]

Short- and long-term effects. We used simple contrasts with t1 as reference point (i.e., results of each time of measurement were contrasted to t1) to demonstrate short- and

[21] However, to investigate whether results of Jewish Israeli students (Study 1 and Study 2) were comparable, longitudinal MI calculations were conducted ($N = 146$ students) and partial metric partial strict scalar MI was reached for attitudes ($\chi^2_{corrected} = 24.48$; $df = 27$; $p = .604$; CFI = 1.00; RMSEA = .00; SRMR = .06) and strict scalar MI for behavior intentions ($\chi^2_{corrected} = 44.26$, $df = 29$; $p = .035$; CFI = .97; RMSEA = .06; SRMR = 06). According to the fact that all Jewish Israeli IG students attended the same school there was no reason to assume that cross-group MI between the school classes was not given. Thus, both variables were assumed to be understood equally throughout three points of measurement by all Jewish Israeli students and are thus comparable. The MI testing did not include the two additional times of measurement of Study 2, because no data existed for these two times of questionings in Study 1.

[22] Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2[9] = 22.37$, $p = .014$). Therefore, we corrected degrees of freedom using Greenhouse-Geisesr estimates of sphericity (ε = .72; Field, 2009).

long-term developments (Field, 2009). Simple contrasts enable similar analyses as the baseline-change-version of the latent change model in Study 1.

Effects after one intergroup meeting. Results showed that attitudes at t2 (exactly at the end of the first intergroup meeting) were more positive than at t1 (before the first intergroup meeting; $F[1,32] = 14.38$, $p = .001$, $\eta^2 = .30$). T1 values were not different from t3 values (before the second intergroup meeting; $F[1,32] = 0.30$, $p = .590$, $\eta^2 = .01$). This indicates that the positive effect of the first intergroup meeting could not be sustained for four weeks.

Effects after four intergroup meetings. Attitude values improved from t1 until t4 (exactly at the end of all four intergroup meetings; $F[1,32] = 37.07$, $p < .001$, $\eta^2 = .54$). This increase was higher than the increase from t1 to t2 and implies a stronger effect after participation in four intergroup meetings compared to one meeting. Attitude scores also improved from t1 to t5 (about eleven weeks after all four intergroup meetings; $F[1,32] = 3.55$, $p = .069$, $\eta^2 = .10$). Findings indicate a long-term effect after participation in four intergroup meetings, which was not discovered after participation in one meeting.

Fading effects. To illustrate fading effects we used repeated contrasts (i.e. results of each point of measurement were compared to the previous point of measurement; Field, 2009). This proceeding resembles the neighbor-change-version of the latent change model in Study 1, only with manifest calculations.

Effects after one intergroup meetings. T2 (exactly at the end of the first intergroup meeting) was different to t3 (before the second intergroup meeting four weeks later; $F[1,32] = 12.06$, $p = .002$, $\eta^2 = .27$), indicating decreases attitude scores.

Effects after four intergroup meetings. Results also indicated fading effects from t4 (exactly at the end of all four intergroup meetings) until t5 (about eleven weeks after the

forth intergroup meetings; $F[1,32] = 9.16$, $p = .006$, $\eta^2 = .22$). [23] As expected, the strength of the fading effect was slightly smaller after the participation in all intergroup meetings.[24]

In conclusion, we found short-term effects after one and after four meetings; however, contact intervention effects after four meetings were stronger than after one meeting. We discovered long-term effects only after four intergroup meetings. Fading effects were found after one and four intergroup meetings, but were slightly less strong after students had taken part in four meetings. Therefore, H1a-c were confirmed. To complete our examination, we tested whether behavior intentions changed similarly.

Behavior intentions toward Arab Israeli students. Results indicated that the time of questioning influenced students' behavior intentions ($F[3.47, 110.96] = 6.42$, $p < .001$, $\eta^2 = .17$).[25] Results are illustrated in Figure 2.

Short- and long-term effects. As before, two short- and two long-term effects were calculated.

Effects after one intergroup meeting. Behavior intention values at t2 were higher than at t1 ($F[1,32] = 17.50$, $p < .001$, $\eta^2 = .35$). There were no differences between t1 and t3 ($F[1,32] = 0.07$; $p = .793$, $\eta^2 = .00$).

---

[23] According to the relatively high number of missing values, we also analyzed the data using listwise deletion to cross-check our findings. Results showed no diverging data structure and almost all values that were significant before were also significant when we applied listwise deletion. In fact, results using listwise deletion were more positive between t1 and t5 ($p = .014$) than using multiple imputations ($p = .069$), indicating an even stronger long-term effect. Only fading effects between t4 and t5 were non-significant when applying listwise deletion ($p = .108$). This was, too, more in favor of our hypotheses.

[24] When applying the stricter Bonferroni correction, the data structure of short-term and fading effects were the same as the structure corrected by the originally used method; however, the long-term effect that had been marginally significant when calculating with multiple imputations ($p = .069$) changed into a non-significant effect ($p = .138$; Bender et al., 2007). When looking at the findings using listwise deletion ($p = .014$), the results remained significant ($p = .028$) even when the Bonferroni correction was applied.

[25] As before, the Mauchly's test indicated a violation of the assumption of sphericity ($\chi^2[9] = 22.11$, $p = .024$). We adjusted degrees of freedom with Huynh-Feldt estimates of sphericity ($\varepsilon = .77$; Field, 2009).

Effects after four intergroup meeting. Behavior intention values increased again and were more positive at t4 than at t1 ($F[1,32] = 8.71$, $p = .006$, $\eta^2 = .21$). However, effects of the intervention were not stronger after students participated in four intergroup meetings. There were also no long-term effects ($F[1,32] = 0.04$, $p = .843$, $\eta^2 = .00$).

Fading effects.

Effects after one intergroup meeting. Findings demonstrated that behavior intention values at t3 were lower than at t2 ($F[1,32] = 15.93$, $p < .001$, $\eta^2 = .33$), indicating fading effects after one contact meeting.

Effects after four intergroup meeting. There was also a reduction in positive behavior intentions between t4 and t5 ($F[1,32] = 13.43$, $p = .001$, $\eta^2 = .30$)[26] also showing that fading effects were slightly smaller than those discovered after one intergroup meeting.

Thus, H1a+b were not whereas H1c was confirmed in regard to behavior intentions.[27]

In sum, we found short-term effects to be stronger after participation in four meetings than after participating in one meeting for the attitude variable. Long-term effects were found only after participation in four intergroup meetings for the attitude variable, showing that only participation in repeated intergroup meetings led to sustainable changes. Fading effects occurred after one and after four intergroup meetings and were less strong after repeated participation in both dependent variables. These results support results from the Jewish Israeli Sample in Study 1 demonstrating the importance of repetition of the intergroup meetings for the majority. Thus, H1a+b were partly confirmed (i.e., within the attitude but not the behavior intention variable). H1c was confirmed in both variables. To investigate whether repeated intergroup meetings had the same impact on the minority, we also surveyed one additional Arab Israeli school class.

[26] As before, we additionally analyzed the data using listwise deletion. Data showed no different structure and all values that were significant before were also significant when applying listwise deletion.

[27] All results that showed significant effects remained significant even after doubling the *p*-value of the β-coefficient (Bonferroni correction; Bender et al., 2007).

Arab Israeli Students

Method

Participants. One further Arab Israeli 4th grade IG school class, which attended the same school as the Arab Israeli IG in Study 1, was investigated (N = 30 students; 53% females). The number of students at the different points of measurement was: N_{t1} = 28, N_{t2} = 27, N_{t3} = 24, N_{t4} = 25, N_{t5} = 24. The school class was questioned five times at the same time as the additional Jewish Israeli school class, except for t5, which was ten (not eleven) weeks after the fourth intergroup meeting (related to Arab Israeli students in Study 1).

Preliminary Analyses. No results were displayed in regard to attitudes due to missing longitudinal MI in Study 1. No selective attrition were found regarding behavior intentions related to t4 (t[25] = .66, p = .514) but to t5 (t[25] = -2.23, p = .039).[28] Missing data were missing according to wave and item nonresponse (Graham et al., 2013, p. 110) and were MCAR (Little's MCAR test: χ^2 = 22.26, df = 26, p = .674; Little, 1988). Thus, we computed missing data via multiple imputations (Schafer & Graham, 2002). Cronbach's α showed very low and uneven values for behavior intentions (t1: α = .50; t2: α = .72; t3: α = .35; t4: α = .82; t5: α = .50).[29]

Results

Analogously to the Jewish Israeli sample, two short- and long-term as well as fading effects were analyzed. Means, standard deviations, and time intervals are shown in Figure 2.

Short-, Long-Term, and Fading Effects. Results showed no changes within behavior intentions over time (F[4, 116] = 1.85, p = .124, η^2 = .06). Therefore, H1a-c were not confirmed. However, mean developments were in the expected direction (Figure 2).

---

[28] T-test values are reported when equal variances are not assumed due to the violation of Levene's test for equality of variances (Field, 2009).

[29] As in Sample 1 (Study 2) no longitudinal MI was investigated due to the small sample size. Nevertheless, it was examined if the combined sample of Arab Israeli 4th grade students that were questioned three and five times indicated longitudinal MI at the common three occasions of measurement. Longitudinal partial metric partial strict scalar MI was reached for positive behavior intentions ($\chi^2_{corrected}$ = 29.91, df = 27; p = .318; CFI = .99; RMSEA = .03; SRMR = .05). Again, no cross-group MI was tested according to the fact that all Arab Israeli IG students attended the same school. In consequence, latent means of all Arab Israeli students are comparable.

Discussion

Jewish Israeli Students. Research findings demonstrated that intergroup interventions led to positive short-term effects after one and after four intergroup meetings within attitudes and behavior intentions. We also discovered that effects after one intergroup meeting could not be sustained when measured four weeks later whereas participation in four intergroup meetings led to a long-term effect, which was still detectable eleven weeks later, at least for the attitude variable. Thus, results indicate that long-term effects might develop in consequence of repeated participation in intergroup meetings. However, this was not shown within all investigated variables. Findings confirmed that fading effects for both variables were slightly less strong after students had taken part in four intergroup meetings compared to participation in one intergroup meeting. Thus, our research confirmed former studies by showing that intervention effects in an intractable conflict area are not sustainable per se (e.g., Arnon, 2010).

Data showed a very low Cronbach's α coefficient within the attitude variable at some occasions of measurement. Due to the fact that we discovered results in direction of our hypotheses despite low reliability, we assume the few low reliability coefficients did not interfere with our research findings. However, this is a clear limitation and results have to be interpreted cautiously. Additionally, selective attrition within the attitude variable was found between the pretest and the follow-up test. Because no selective attrition effects were found in regard to the pretest and the direct posttest, we assume these results not to interfere with our findings.

Time spans in our design were distributed unevenly. However, since we were interested in specific time spans that we derived from previous studies instead of a general trend analysis, we regard missing equidistance not as a problem.

Arab Israeli Students. Even though mean values increased after students participated in four intergroup meetings, no significant short- and long-term or fading effects were discovered. Results might have occurred due to ceiling effects, because the average behavior intentions values were already above 4 (5 being the most positive value) at almost all occasions of measurement. Additionally, preliminary analysis showed some non-ideal outcomes, such as selective attrition effects as well as uneven and very low reliability coefficients. Therefore, findings are questionable.

General Discussion

Our data confirmed previous research findings with another intergroup contact intervention showing that contact effects within intractable conflict areas already fade some weeks after the contact intervention ends (e.g., Shani, 2015). Our results supported the idea that contact intervention effects can be extended within conflict areas when intergroup meetings are repeated (e.g., Salomon, 2011) at least for majority group members. As before, long-term contact intervention effects were only found for the majority group (e.g., Jayusi, 2009). This may be explained by status differences of both groups and accordingly with a deviating perception of the contact situation (e.g., Salomon, 2011).

Moreover, we revealed that the valence of students' reported intergroup experience (positive emotions related to the activities, pleasant interactions, and friendship) had a positive impact on the contact intervention outcome in regard to short- and long-term effects. Thus, we broadened previous findings from Schroeder and Risen (2016) by exhibiting that positive intergroup experience can include multiple aspects, which are not necessarily related to friendship. In contrast to repetition effects, a relation between positive personal experience and the posttest findings was shown for both majority and minority group members. These results indicate that the valence of the contact experience is essential and a negative experience should be prevented as already shown by Guffler and Wagner (2016).

Although our research provides several new insights into sustainable contact intervention effects in an intractable conflict area, we also acknowledge some limitations. First, we used a quasi-experimental design in Study 1, which threatens the internal validity (Shadish et al., 2002). For example, there is a threat to internal validity due to attrition (mortality). To face this problem, we analyzed selective attrition effects within each cohort and discussed the findings respectively. Finally, we ascertained that IG and CG students understood our variables equally at each occasion of measurement and between groups. Within Study 2, we did not investigate a CG due to feasibility and our apprehension of testing effects when asking students five times. Nevertheless, we assume students from Study 2 to be similar to those from Study 1. Second, participation in intergroup programs is probably highly selective because only schools that are open to intergroup exchange take part in intergroup programs or questionings (Paluck & Green, 2009). This implies that findings might be limited to this specific group of questioned individuals. In addition, our research was located in two schools in Northern Israel. Research findings from this specific

area and schools might not be generalizable to other conflict areas because until now there is no investigation whether contact interventions in different conflict contexts are comparable. Finally, our assumption that the contact intervention effects faded due to the Israeli-specific conflict context was only derived theoretically and not proven by empirical research findings.

Regarding practical implications, our results support the idea that future interventions should include multiple intergroup meetings that do not necessarily have to be of long continuance, at least for the majority group. Possibly virtual contact (e.g., via social media) could function as an intergroup contact repetition as well and could provide an interesting opportunity for repetition in case financial resources are limited (e.g., Amichai-Hamburger & McKenna, 2006; Tavakoli, Hatami, & Thorngate, 2010) or the opportunity to meet face-to-face is not given.

In conclusion, our study is one of very few studies that analyzed short-, long-term, and fading effects of intergroup contact interventions within a conflict area systematically. This study revealed the relevance of positive intergroup experiences for the majority and the minority group as a strong predictor for long-term contact intervention effects as well as the important role of intergroup contact repetition for the majority group. We confirmed that contact intervention effects within a conflict area are not sustainable per se, meaning that they fade already after a few weeks. Furthermore, we followed a strict methodical procedure, ensuring latent variable modeling and measurement invariance analysis; thereby, presenting state-of-the-art data analysis.

References

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Amichai-Hamburger, Y., & McKenna, K. Y. A. (2006). The contact hypothesis reconsidered: Interacting via the internet. *Journal of Computer-Mediated Communication*, *11*(3), 825–843. doi:10.1111/j.1083-6101.2006.00037.x

Arnon, M. (2010). *Reflection on the activities of a peace education program as a tool for reinforcing/rehabilitating its long term effects* (Unpublished doctoral dissertation). University of Haifa, Haifa, Israel. [Hebrew]

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5–37. doi:10.1016/j.jsp.2009.10.001

Bar-Natan, I., Rosen, Y., & Salomon, G. (2010). *Does friendship between adversaries generalize?* Unpublished manuscript, University of Haifa, Haifa, Israel.

Bar-Tal, D. (1998). Societal beliefs in times of intractable conflict: The Israeli case. *The International Journal of Conflict Management*, *9*(1), 22–50. doi:10.1108/eb022803

Bender, R., Lange, S., & Ziegler, A. (2007). Multiples Testen. Artikel Nr. 12 der Statistik-Serie in der DMW [Multiple testing. Article 12 of the statistic series in the DMW]. *Deutsche Medizinische Wochenschrift*, *132*(S 01), e26–e29. doi:10.1055/s-2007-959035 [German]

Berger, R., Benatov, J, Abu-Raiya, H., & Tadmor, T. C. (2016). *Reducing prejudice and promoting positive intergroup attitudes among elementary-school children in the context of the Israeli-Palestinian conflict*. Manuscript submitted for publication.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

Connolly, E. (1992). *The Edu-Train project and the reduction of prejudice* (Unpublished master thesis). Rand Afrikaans University, Johannesburg, South Africa.

Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. *Survey Research Methods, 2*(1), 33–46. doi:10.18148/srm/2008.v2i1.365

Davidov, E., & de Beuckelaer, A. (2010). How harmful are survey translations? A test with Schwartz's human values instrument. *International Journal of Public Opinion Research*, *22*(4), 485–510. doi:10.1093/ijpor/edq030

Field, A. (2009). *Discovering statistics using SPSS*. Los Angeles, CA: Sage Publications Ltd.

Graf, S., Paolini, S., & Rubin, M. (2014). Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five Central European countries. *European Journal of Social Psychology*, *44*(6), 536–547. doi:10.1002/ejsp.2052

Graham, J. W., Cumsille, P. E., & Shevock, A. E. (2013). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology. Research methods in psychology. Volume 2* (pp. 109–141). Hoboken, NJ: John Wiley & Sons.

Guffler, K. & Wagner, U. (2016). *Backfire of good intentions: Unexpected long-term contact intervention effects in an intractable conflict area.* Manuscript in preparation.

Haaretz (2012a, November 14). Live blog: Day 1 of Israel-Gaza conflict 2012. *Haaretz*. Retrieved from http://www.haaretz.com/news/diplomacy-defense/live-blog-day-1-of-israel-gaza-conflict-2012-1.477850

Haaretz (2012b, December 15). Live blog: Day 2 of Israel-Gaza conflict 2012. *Haaretz*. Retrieved from http://www.haaretz.com/news/diplomacy-defense/live-blog-day-2-of-israel-gaza-conflict-2012-1.477960

Harkness, J. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: John Wiley & Sons.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. doi:10.1080/19312450709336664

Jajodia, A. (2012). Dynamic structural equation models of change. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis. A practical guide for researchers in aging, health, and social sciences* (pp. 291–328). New York, NY: Routledge.

Jayusi, W. (2009). *Restoring the attitudes of peace education participants through peer-tutoring* (Unpublished doctoral dissertation). University of Haifa, Haifa, Israel. [Hebrew]

Kriesberg, L. (1998). Intractable conflicts. In E. Weiner (Ed.), *The handbook of* interethnic coexistence (pp. 332–342). New York, NY: Continuum Books.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. Human Communication Research, 30(3), 411–433. doi:10.1111/j.1468-2958.2004.tb00738.x

Kropiunigg, U., & Pabst, B. (2007). Selbstwertsteigerung und Vorurteilsreduktion bei Jugendlichen eines multiethnischen Peace Camps [Improvement of self-esteem and reduction of prejudice by multiethnic peace camps for juveniles]. *Psychotherapie Forum, 15*, 63–72. doi:10.1007/s00729-007-0194-1 [German]

Kuckartz, U. (1995-2014). MAXQDA (Version 11) [Computer software]. Berlin, Germany: VERBI GmbH.

Kupermintz, H., & Salomon, G. (2005). Lessons to be learned from research on peace education in the context of intractable conflict. *Theory into Practice*, *44*(4), 293–302. doi:10.1207/s15430421tip4404_3

Lanphen, J. (2011). *Kooperatives Lernen und Integrationsförderung: Eine theoriegeleitete Intervention in ethnisch heterogenen Schulklassen [Cooperative learning and promoting integration: A theory-based intervention in ethnically heterogeneous classes]*. Muenster, Germany: Waxmann. [German]

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European* Journal of Social Psychology, 45(2), 152–168. doi:10.1002/ejsp.2079

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. doi:10.2307/2290157

Luiz, D., & Krige, P. (1985). The effect of social contact between South African White and Colored adolescent girls: A follow-up study. *Journal of School Psychology*, *125*(3), 407–408. doi:10.1080/00224545.1985.9922909

McArdle, J. J. (2007). Dynamic structural equation modeling in longitudinal experimental studies. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 159–188). Mahwah, NJ: Lawrence Erlbaum Associates.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, *73*(3), 574–584. doi:10.1037/0021-9010.73.3.574

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998–2014). M*plus* (Version 7.2) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*(1), 339–367. doi:10.1146/annurev.psych.60.110707.163607

Pettigrew, T. F. (1971). *Racially separate or together?*. New York, NY: McGraw-Hill.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85. doi:10.1146/annurev.psych.49.1.65

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. doi:10.1037/0022-3514.90.5.751

Rosen, Y., & Salomon, G. (2011). Durability of peace education effects in the shadow of conflict. *Social Psychology of Education*, *14*(1), 135–147. doi:10.1007/s11218-010-9134-y

Salomon, G. (2011). Four major challenges facing peace education in regions of intractable conflict. *Peace and Conflict: Journal of Peace Psychology*, *17*(1), 46–59. doi:10.1080/10781919.2010.495001

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. doi:10.1037//1082-989X.7.2.147

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.

Schleien, S. M. (2007). Confronting the intractable: An evaluation of the Seeds of Peace experience (Unpublished doctoral dissertation). University of Waterloo, Canada, Waterloo.

Schmid, K., & Muldoon, O.T. (2013). Perceived threat, social identification, and psychological well-being: The effects of political conflict exposure. *Political Psychology, 36*(1), 75–92. doi:10.1111/pops.12073

Schroeder, J., & Risen, J. L. (2016). Befriending the enemy: Outgroup friendship longitudinally predicts intergroup attitudes in a coexistence program for Israelis and Palestinians. *Group Processes and Intergroup Relations, 19*(1), 72–93. doi:10.1177/1368430214542257

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Bosten, MA: Houghton Mifflin.

Shani, M. (2015). *A theory and practice of coexistence: Improving coexistence orientation through mixed-model encounters between Jews and Palestinians in Israel* (Unpublished doctoral dissertation). University of Bremen, Bremen, Germany.

Sharing Education Programme (2012). *A report on the sharing education learning forum Conference 2012*. Retrieved from

http://www.schoolsworkingtogether.com/documents/SELF%20Conference%20Report%20December%202012.pdf

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling. Second Edition*. London, UK: Sage Publications Ltd.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90. doi:10.1086/209528

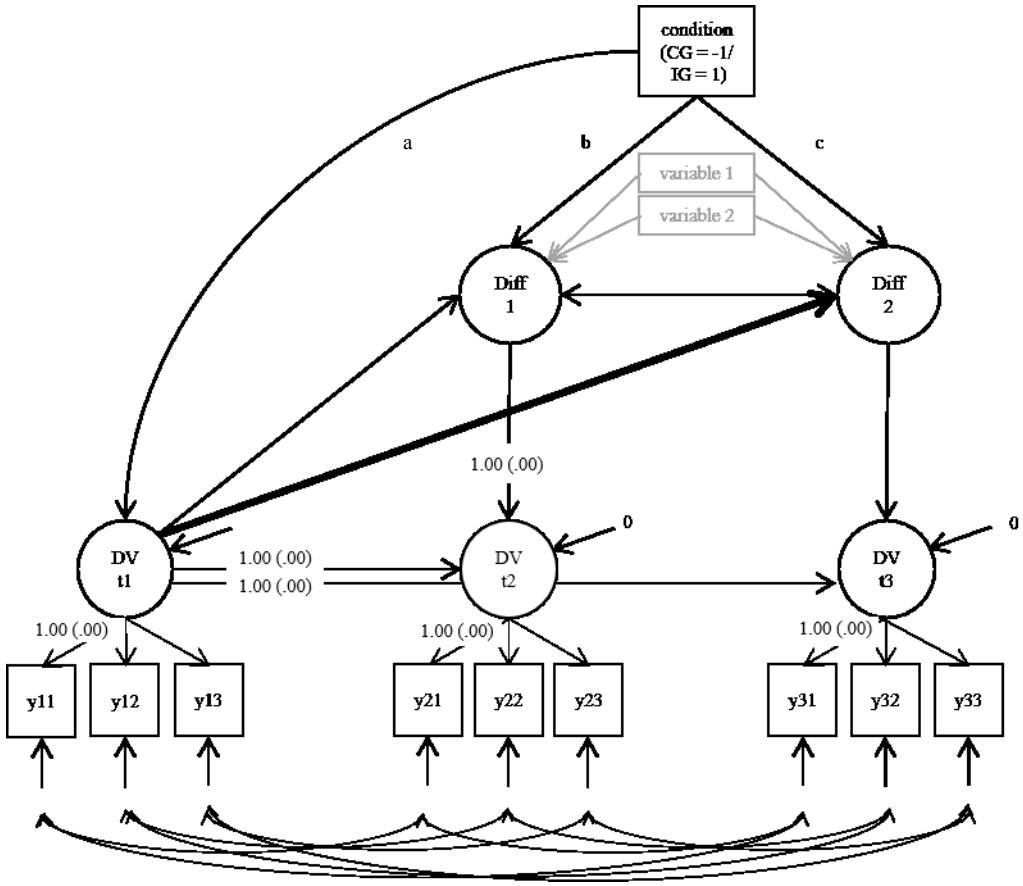
Steyer, R., Partchev, I., & Shanahan, M. J. (2000). Modeling true intraindividual change in structural equation models: The case of poverty and children's psychosocial adjustment. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data. Practical issues, applied approaches and specific examples* (pp. 109–126). Malden, MA: Lawrence Erlbaum Associates.

Tavakoli, M., Hatami, J., & Thorngate, W. (2010). Changing stereotypes in Iran and Canada using computer mediated communication. *Journal of Intercultural Communication*, *23*, 5.

Vered, S., & Bar-Tal, D. (2014). Routinization of the Israeli-Arab conflict: The perspective of outsiders. *Israel Studies Review*, *29*(1), 41–61. doi:10.3167/isr.2014.290104

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, *29*(4), 377–392. doi:10.1177/0734282911406668

(1)


condition
(CG = -1/
IG = 1)
a
b
c
variable 1
variable 2
Diff
1
Diff
2
1.00 (.00)
0
0
DV
t1
1.00 (.00)
1.00 (.00)
DV
t2
DV
t3
1.00 (.00)
1.00 (.00)
1.00 (.00)
y11
y12
y13
y21
y22
y23
y31
y32
y33

(2)

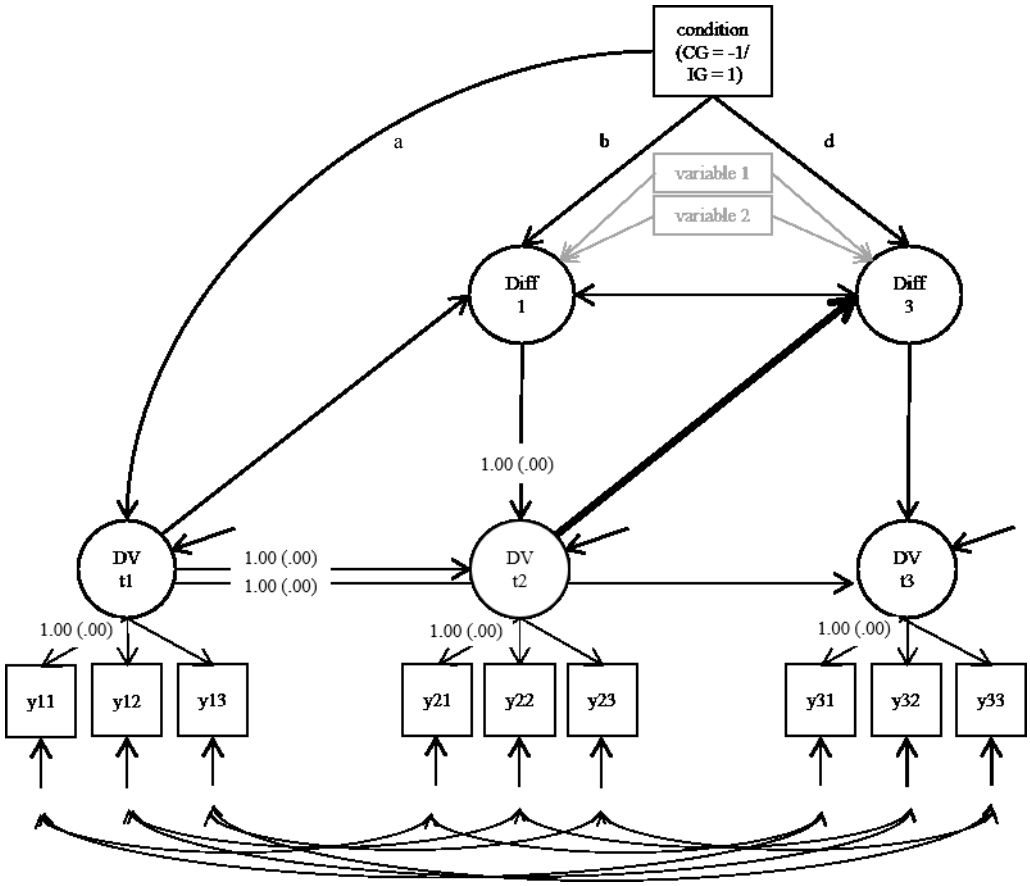


Figure 1. Latent change models: (1) Baseline-change-version of the latent change model (2) Neighbor-change-version of the latent change model; DV = dependent variable (either positive attitudes or positive behavior intentions toward the outgroup); t1 = pretest (before the first intergroup meeting); t2 = posttest (exactly at the end of all four intergroup meetings); t3 = follow-up (about ten weeks [Arab Israeli] or eleven weeks [Jewish Israeli] after the end of all four intergroup meetings); Diff1 = latent difference score between t1 and t2; Diff2 = latent difference score between t1 and t3; Diff3 = latent difference score between t2 and t3; condition = control group (CG) and intervention group (IG); variable 1 + 2 = variables included in the LCM to account for hierarchical data effects; a = pretest differences; b = effect of the condition on Diff1; c = effect of the condition on Diff2; d = effect of the condition on Diff3.

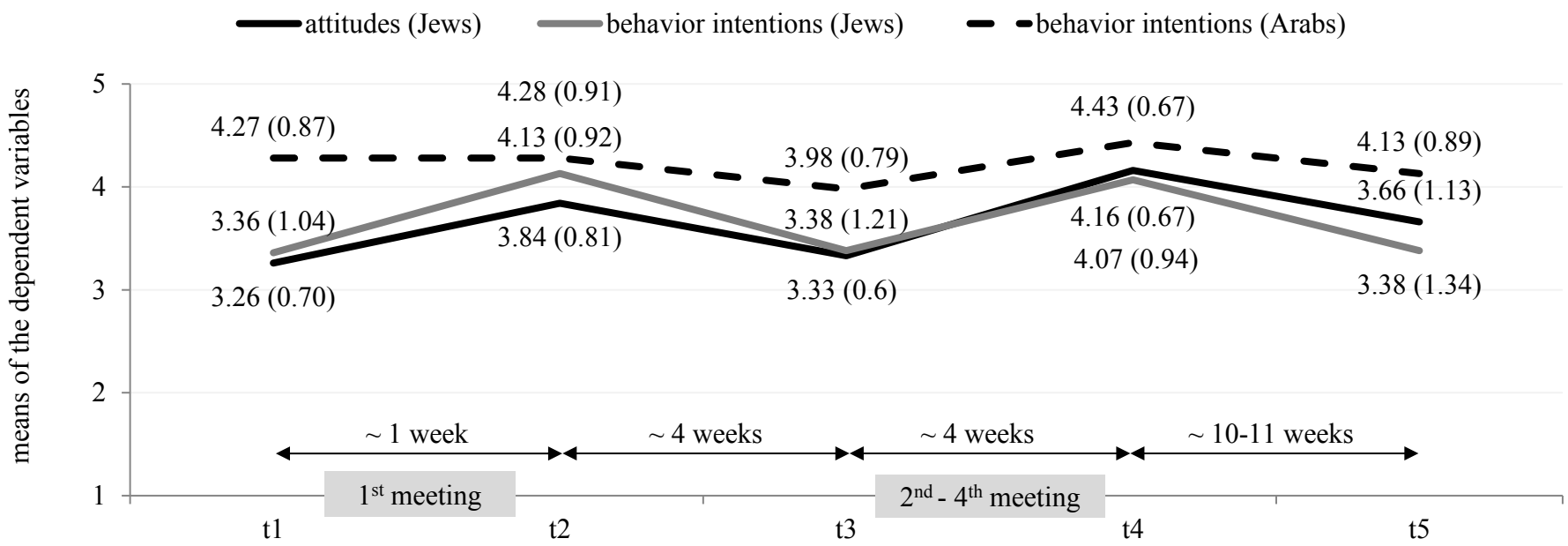


Figure 3. Means, standard deviations, and time spans of Study 2; *N* = 33 Jewish Israeli students; *N* = 30 Arab Israeli students; means = values between 1 and 5 with 1 signifying lowest and 5 highest agreement; t1 = pretest (one week before the first intergroup meeting); t2 = first time between intergroup meetings (exactly at the end of the first intergroup meeting); t3 = second time between the intergroup meetings (before the second intergroup meeting); t4 = posttest (exactly at the end of all four intergroup meetings); t5 = follow-up (about ten weeks [Arab Israeli] or eleven weeks [Jewish Israeli] after the end of all four intergroup meetings); 1st meeting = the first intergroup meeting; 2nd - 4th meetings = the second, third, and fourth intergroup meeting.

Table 1

Latent Means and Latent Change Model Indices

				Latent Means			Latent Change Model Indices					
Sample	DV	Group	*n*	t1	t2	t3	Diff1 - Condition (b), β	Diff1 -Condition (b), *b* (*SE*)	Diff2 - Condition (c), β	Diff2 - Condition (c), *b* (*SE*)	Diff3 - Condition (d), β	Diff3 - Condition (d), *b* (*SE*)
Jewish Israeli	Attitudes	IG	60	3.07	4.01	3.66	.59***	.49 (.08)***	.23*	.25 (.10)*	-.34**	-.24 (.10)*
		CG	53	3.15	3.03	3.15						
	Behavior Intentions	IG	60	3.19	2.99	2.69	.11	.10 (.13)	.02	.02 (.14)	-.09	-.08 (.13)
		CG	52	2.83	2.78	2.65						
Arab Israeli	Behavior Intentions	IG	60	2.47	3.62	3.37	.19**	.36** (.14)	.11	.20 (.12)	-.13	-.17 (.17)
		CG	48	2.70	2.89	2.98						

Note. IG = intervention group; CG = control group; latent means = values between 1 and 5 with 1 signifying lowest and 5 highest agreement; t1 = pretest (one week before the first intergroup meeting); t2 = posttest (exactly at the end of all four intergroup meetings); t3 = follow-up (about ten [Arab Israeli] or eleven [Jewish Israeli] weeks after the end of all four intergroup meetings); Diff1 = latent difference score between t1 and t2; Diff2 = latent difference score between t1 and t3; Diff3 = latent difference score between t2 and t3; b = path b in Figure 1; c = path c in Figure 1; d = path d in Figure 1; condition = IG and CG.
*$p < .05$. ** $p < .01$. ***$p < .001$.

Table 2

Relation of Reported Intergroup Experiences at t2 in Regard to Short- and Long-Term Attitudes and Behavior Intentions

Sample	DV t2		*B*	*SE B*	β	DV t3		*B*	SE *B*	β
Jewish Israeli	Attitudes	Constant	2.58	.31		Attitudes	Constant	2.29	.42	
		Attitudes t1	.40	.10	.50***		Attitudes t1	.37	.13	.37**
		Experiences t2	.50	.25	.25†		Experiences t2	.95	.38	.33*
	Behavior intentions	Constant	1.28	.39		Behavior intentions	Constant	1.81	.53	
		Behavior t1	.55	.11	.54***		Behavior t1	.27	.15	.20
		Experiences t2	.74	.30	.28*		Experiences t2	1.67	.44	.48***
Arab Israeli	Behavior intentions	Constant	4.32	.41		Behavior intentions	Constant	3.75	.55	
		Behavior t1	-.22	.10	-.37*		Behavior t1	.02	.14	.03
		Experiences t2	.65	.36	.31†		Experiences t2	.15	.48	.07

Note. t1 = pretest (one week before the first intergroup meeting); t2 = posttest (exactly at the end of all four intergroup meetings); t3 = follow-up (about ten [Arab Israeli] or eleven [Jewish Israeli] weeks after the end of all four meetings).
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Acknowledgements

We thank Lisa Gutenbrunner and Mathias Kauff for their comments on an earlier version of the paper, as well as Lisa Gutenbrunner and Stefan Thörner for their support with the statistics and Jan Weber for the English editing of the article.

Funding

Deutsche Forschungsgemeinschaft [GRK 884]

Conflict of Interest Statement

No potential or perceived conflicts of interest.

2. Manuscript #2

2.1. Introduction to Manuscript #2

The second manuscript investigated whether our second intergroup contact intervention between Jewish and Arab Israeli 6th grade students led to short- and long-term contact intervention effects in an intractable conflict area in Israel; more precisely, positive attitudes and positive behavior intentions. As within Manuscript #1, emphasis was laid on a strict methodical approach including latent variable modeling and measurement invariance calculations. Two samples were analyzed: Sample 1 consisted of 137 Jewish Israeli (n = 83 IG students; n = 54 CG students); Sample 2 counted 139 Arab Israeli students (n = 83 IG students; n = 56 CG students). In this manuscript we evaluated the following research questions: Do intergroup contact interventions in intractable conflict areas that include repeated intergroup meetings influence intergroup relations (short- and long-term)? Do student's personal experience (positive, neutral, or negative) influence intergroup relations sustainably within intractable conflict areas? Are there differences between minority and majority group members in regard to contact interventions effects?

Manuscript #2 enlarges Manuscript #1 by illustrating results of a second contact intervention (6th grade). Due to the fact that the intervention was experienced negatively by the majority group, short- and long-term as well as fading effects in reference to negative contact were analyzed. Interestingly, the minority group experienced the exact same intervention as positive.

2.2. Submitted Manuscript #2

Backfire of Good Intentions: Unexpected Long-term Contact Intervention Effects in an Intractable Conflict Area

Kerstin Guffler and Ulrich Wagner
Philipps-University Marburg

Authors Note

Kerstin Guffler, Department of Psychology, Philipps-University Marburg; Ulrich Wagner, Department of Psychology and Center for Conflict Studies, Philipps-University Marburg.

This study was financially supported by the DFG research training group *group-focused enmity* of the University of Bielefeld and the Philipps-University Marburg (GRK 884) and the Department of Psychology of the Philipps-University Marburg. We thank Thomas F. Pettigrew, Judith Lanphen, and Frank Asbrock for their helpful feedback, Lisa Gutenbrunner and Stefan Thörner for their support with the statistics, and Jan Weber for the English editing of the article.

Correspondence concerning this article should be addressed to Kerstin Guffler, Philipps-University Marburg, Psychology Department, Gutenbergstr. 18, 35032 Marburg, Germany. E-mail: guffler@uni-marburg.de

(Submission date: April 30th, 2016; revised version under review in *Peace and Conflict: Journal of Peace Psychology*)

Abstract

We investigated whether an intergroup contact intervention that included two intergroup meetings between Jewish and Arab Israeli students in Israel led to improved long-term intergroup relations. Besides our interest in long-term intergroup contact effects in an intractable conflict, we analyzed minority-majority differences in this context. Before we calculated our results, measurement equivalence across time was ascertained. Data were computed using latent change models. Results for minority members (Arab Israeli) showed positive short- but no long-term contact intervention effects, that is, attitudes and behavior intentions improved shortly after the intervention but went back to the before-intervention level after nine weeks. Contrary to our expectation for majority members (Jewish Israeli), the intervention led to a reduction of positive intergroup relations both in the short- and long-run. By merging quantitative data with students' comments about their intergroup experience, we identified that this effect was probably based on negative contact experiences. Results are discussed.

Keywords: contact intervention, negative contact, longitudinal design, intractable conflict

Backfire of Good Intentions: Unexpected Long-term Contact Intervention Effects in an Intractable Conflict Area

Intergroup contact is widely known to reduce prejudice between conflicting groups (e.g., Pettigrew & Tropp, 2006). However, "very little research has been conducted under 'hot' conditions, such as intergroup contact in the context, or against the background, of major political upheaval, war, civil war, ethnic cleansing or genocide" (Wagner & Hewstone, 2012, p. 198). Meanwhile, there are vigorous efforts to increase knowledge about intergroup relations in intractable conflicts (e.g., Bar-Tal, 2013). Intractable conflicts are conflicts that are persisting for a long time, often for generations, perceived as irreconcilable, violent, and zero-sum in nature (Bar-Tal, 2013; Kriesberg, 1998). We aim to contribute to a deeper understanding regarding long-term improvement of intergroup relations after the participation in an intergroup contact intervention in an intractable conflict area. Therefore, we investigated the effects of an intervention consisting of two short meetings on attitudes and behavior intentions for the Jewish Israeli majority and the Arab Israeli minority in Israel.

Effects of Intergroup Contact on Intergroup Relations

Allport (1954) postulated that contact can reduce intergroup prejudice. Over the years Allport's assumption was empirically supported by many research findings (e.g., Pettigrew & Tropp, 2006). However, scholars recently criticized that the traditional research on intergroup contact effects had concentrated on positive intergroup contact situations only. Thus, neglecting the effects of negative intergroup contact, which might lead to greater prejudice (for an overview, see Pettigrew & Tropp, 2011, Chapter 12). Negative contact is considered as a negative interaction that can range from an unpleasant intergroup activity and disagreement to a fight (Stephan, Stephan, & Gudykunst, 1999). In one of the few existing studies, Paolini, Harwood, and Rubin (2010), for example, manipulated negative contact as reserved nonverbal behavior in a face-to-face interaction. Due to the negative atmosphere between the parties in conflict areas, in our opinion, the existence of positive and negative intergroup contact should be considered in studies related to contact in intractable conflicts.

Long-term Contact Intervention Effects in Conflict Areas

In conflict areas where members of conflicting parties usually have little contact (e.g., due to segregated residential areas, separate schools) the implementation of intergroup contact interventions provides an opportunity to improve intergroup relations on the

grassroots level. In a meta-analysis, Lemmer and Wagner (2015) revealed that interventions lead to positive *short-term* effects (until one month after the intervention) in conflict areas ($\hat{\mu}$ = .47 for studies without a control group design, k = 20 comparisons). They also investigated *long-term* contact intervention effects including studies related to either a conflict or a non-conflict area ($\hat{\mu}$ = .35 for studies without a control group design, k = 17 comparisons). Out of these 17 comparisons 13 were related to conflict areas. None of these studies included a control group. The small number of comparisons within the meta-analysis and the missing control groups underline the difficulty to conduct controlled experimental research under circumstances that are characterized by an ongoing conflict.

Looking at the Israel-Palestine conflict specifically, there are very few studies which investigated *long-term* intergroup contact intervention effects quantitatively (e.g., Arnon, 2010; Bar-Natan, Rosen, & Salomon, 2010; Berger, Benatov, Abu-Raiya, & Tadmor, 2016; Jayusi, 2009; Kropiunigg & Pabst, 2007; Schleien, 2007; Schroeder & Risen, 2016; Shani, 2015).[1] Almost all of these studies discovered positive short-term effects; however, most of them also revealed that the short-term effects had faded (decreased) at the follow-up measurements.[2] These fading effects might be related to the negative atmosphere of the conflicts; as revealed by Christ et al. (2014), context effects can have an impact on intergroup relations. In order to impede fading effects, Jayusi (2009) analyzed the role of repeated contact. She discovered that short-term effects following a two-day intervention faded within eight weeks in Israel, but when participants took part in an additional peer-tutoring after eight weeks, the effects sustained for the following two months. Therefore, it seems reasonable to implement repeated contact meetings to enable long-term contact intervention effects in Israel.

[1] The studies conducted by Bar-Natan et al. (2010), Kropiunigg and Papst (2007), and Schleien, (2007) were included in the meta-analysis. The remaining five studies did not meet Lemmer and Wagner's (2015) inclusion criteria: the studies implemented by Arnon (2010) and Jayusi (2009) were only available in Hebrew; thus, we relied on English summaries and were not able to question their methodical procedures; the studies conducted by Berger et al. (2016), Schroeder and Risen (2016) as well as Shani (2015) were published only recently.

[2] Timespans between the intervention and the follow-up measurement varied for the respective evaluations.

Investigation of the Minority and Majority

In general, contact interventions were found to reduce prejudice for minority and majority members directly after the intervention (short-term effects) with interventions being more beneficial for the majority (Lemmer & Wagner, 2015).[3] However, status differences in regard to long-term effects in conflict areas are less straightforward as only few studies examined long-term effects, and even less analyzed them for both status groups separately. In the few existing studies different findings were reported: some studies found positive long-term effects for the majority but not for the minority group (e.g., Jayusi, 2009) whereas others discovered no long-term effects for the majority but negative effects for the minority (e.g., Schleien, 2007), and again others revealed the same outcomes for both status groups (e.g., Shani, 2015). However, deviating outcomes for minority and majority members are reasonable as both groups perceive the contact situation from a different perspective. Thus, differences might be related to power relations (e.g., Maoz, Bar-On, Bekerman, & Jaber-Massarwa, 2004) so that minority and majority members have deviating goals and needs within intergroup interventions (Salomon, 2011). Salomon (2011) speculated that in the programs, Jews try to carry out the ideas of peace education whereas Palestinians attempt to maintain their position and become empowered. Moreover, Maoz et al. (2004) stated that the Jewish dominance in intergroup meetings reflects the society (macro level) and might be reduced only when Palestinians are empowered to challenge the status quo. In order to get a deepened understanding about status differences, we analyzed short- and long-term effects for both groups separately and tested whether status moderated our outcomes.

The Present Study

This study illustrates the evaluation of a contact intervention targeting Arab and Jewish youth in Israel. We attempt to expand previous findings by showing that repeated intergroup meetings consisting of two short repeated encounters can contribute to long-term contact effects (attitudes and behavior intentions) even nine weeks after the end of the main intervention and by showing that the outcomes are different for both status groups. For methodological reasons, we compared the effects observed in the intervention group with a non-intervention control group. Our research hypotheses were as follows:

[3] This was found for studies using a control group but not in studies without a control group design.

H1 Contact interventions that include repeated short intergroup meetings lead to long-term contact effects (i.e., more positive attitudes and behavior intentions) for the Jewish Israeli majority in Israel.

H2 Contact interventions that include repeated short intergroup meetings lead to long-term contact effects for the Arab Israeli minority in Israel.

H3 Short- and long-term intergroup contact effects are deviating for both status groups.

We present data from a longitudinal study including Jewish (Sample 1) and Arab students (Sample 2). Within both samples, condition (intervention group [IG], control group [CG]) was our independent variable whereas attitudes and behavior intentions were dependent variables.

Sample 1

Method

Sampling procedure and participants. We questioned 139 Jewish Israeli 6th grade students (approximately 13 years old). Two students were excluded because they did not identify themselves as Jewish Israeli. About two-thirds of the students attended a participating school (IG: n = 83, females = 51%). The other third were students of a school that did not participate in the program (CG: n = 54; females = 50%). The intervention school was selected due to their participation in the program; the control school was chosen because it was located in the same geographic area, agreed to participate in the questionings, and taught students of the same age. Neither IG nor CG students got monetary rewards. We received the consent from the Israeli Ministry of Education, which acted for the parents. Students were informed that they could quit the questionings at any time and that their data were treated confidentially.

Measures. IG and CG students answered the same questionnaire. To enable students reading and answering in their first language, questionnaires were translated from English into Hebrew. Translations were realized by a translation agency and in consultation with local managers; back-translations were devised (Harkness, 2003). Among other items, which are not subject of this study, we used the following measurements.

Attitudes toward Arab Israeli students. We used three items adapted from scales by Lanphen (2011). Students were asked to answer the statements "Arab Israeli students are… … nice; ... friendly; … kind" on five-point scales, ranging from 1 = *not very much* to 5 = *very much*. Higher scores indicated more positive attitudes toward Arab Israeli students.

Behavior intentions toward Arab Israeli students. Three items were used adapted from Jayusi (2009): "I would like to … … speak to students from Arab Israeli schools; … meet students from Arab Israeli schools; … do common activities with students from Arab Israeli schools." Again, students responded on a five-point scale (1 = *not very much* to 5 = *very much*). Higher scores indicated more positive behavior intentions.

Reported Personal Experience. To get insights into IG students' opinion, we asked them to answer the question: "Did you tell your parents about the meetings? If yes, what did you tell them?" Students wrote their responses in Hebrew on three blank lines. Answers were translated into English and analyzed. Our assumption was that students would report in this way about the most salient aspects of their intergroup experiences.

Research design. We used a quasi-experimental pre-post-follow-up-test design. IG students were questioned three times: at school about one week before the first meeting (t1, October 2012); at the place of the intervention exactly after the second (last) meeting (t2; December 2012); at school about nine weeks after t2 (t3; February 2013). CG students responded at their school nearly at the same times.

Intervention. The goal of the intervention was to increase tolerance and understanding toward the other; thus, the intervention could be regarded as a coexistence program (Maoz, 2011). Students met each other at a place outside their communities. During the program students were divided into small groups, each consisting of 20-30 students (half Jewish, half Arab). The composition of the small groups was probably similar but not identical throughout both meetings. Within the small groups, students had discussions (e.g., about Israeli citizenship, conflict related topics, and mutual understanding), interacted cooperatively (e.g., constructed houses made of cardboard, played ball games), and ate breakfast together. Groups were instructed by one Hebrew and one Arab speaking guide who were independent from students' teachers.[4] IG students had participated in the program already in their 4th and 5th grade; thus, the intervention was part of a three-year intergroup program. The general program included exchange about the common history and cultures, visits of the local areas of the two intervention schools, and interactive games. All three years of the program were based on the coexistence idea,

[4] We have no information about the educational background of the guides other than that there were male and female guides and most of them were involved in the program for years. Questionnaires were discussed with them ahead of the questionings.

meaning striving to improve intergroup relations among citizens irrespective of the ethnic background.

The societal context. From 14th-21st November (between the meetings) an eight-day violent outbreak between Jews and Muslims occurred (Israel-Gaza Conflict 2012; Pillar of Defense). The ceasefire gave about ”1.7 million Gazans respite from days of ferocious air strikes and halt rocket salvoes from militants that unnerved a million people in southern Israel and reached Tel Aviv and Jerusalem for the first time” (Reuters, 2012). The violent outbreak took place between pre- and posttest. Due to our CG design, intervention effects could be investigated independent from the violent outbreak.

Preliminary analyses. No selective attrition was found for attitudes (t2: $t[123] = -1.07$, $p = .285$; t3: $t[123] = -.98$, $p = .331$) and behavior intentions (t2: $t[125] = -.86$, $p = .390$; t3: $t[125] = -.04$, $p = .967$; Shadish, Cook, & Campbell, 2002).[5] Missing data were given due to wave and item nonresponse (Graham, Cumsille, & Shevock, 2013) and unsuccessful matching of questionnaires ($n = 5$). Data were missing completely at random (Little’s MCAR test: $\chi^2 = 40.73$, $df = 44$, $p = .612$; Little, 1988); thus, missing values were estimated via maximum-likelihood estimation (Schafer & Graham, 2002). Pretest differences (IG, CG) existed for attitudes ($b = 0.62$, $SE = 0.16$, $p < .001$; $M_{IG} = 3.51$, $M_{CG} = 2.89$) but not for behavior intentions ($b = 0.19$, $SE = 0.19$, $p = .328$; Shadish et al., 2002). Given the nested data structure (five school classes), we considered hierarchical effects (Snijders & Bosker, 2012).[6] The CFA (attitudes and behavior intentions being two separate variables) was satisfactory at t1 ($\chi^2_{corrected}$[7] $= 6.32$; $df = 8$; $p = .611$; CFI = 1.00; RMSEA =

[5] The number of participants at each occasion of measurement was $N_{t1} = 130$, $N_{t2} = 135$, and $N_{t3} = 121$.

[6] To control for hierarchical effects we created three variables for the five school classes (SC). Variable_1: IG SC 1 = 1, IG SC 2 = -1/2, IG SC 3 = -1/2, CG SC 1 = 0, CG SC 2 = 0; variable_2: IG SC 1 = 0, IG SC 2 = 1, IG SC 3 = -1, CG SC 1 = 0, CG SC 2 = 0; variable_3: IG SC 1 = 0, IG SC 2 = 0, IG SC 3 = 0, CG SC 1 = 1, CG SC 2 = -1.

[7] *Maximum likelihood estimation with robust standard errors* was used within CFA and latent change model calculations (Muthén & Muthén, 1998-2012). According to using MLR the χ^2 value was corrected (Muthén & Muthén, 1998-2012).

.00; SRMR = .03),[8] t2, and t3. Omega (ω; McDonald, 1999) illustrated an acceptable reliability for attitudes (ω_{t1} = .79, ω_{t2} = .86, ω_{t3} = .84) and behavior intentions (ω_{t1} = .86, ω_{t2} = .88, ω_{t3} = .90).[9] To enable latent mean comparisons across time, we ensured measurements of the used constructs to be equivalent over time (Little, 2013).[10] Partial strict factorial measurement invariance (MI) was reached for attitudes ($\chi^2_{corrected}$ = 27.69; *df* = .28; *p* = .481; CFI = 1.00; RMSEA = .00; SRMR = .05) and strict factorial MI for behavior intentions ($\chi^2_{corrected}$ = 37.95, *df* = 29; *p* = .124; CFI = .98; RMSEA = .05; SRMR = .05).[11]

Results

As recommended by Gollwitzer, Christ, and Lemmer (2014) we analyzed our data using latent change models (LCMs). According to the three occasions of measurement, we computed two latent difference scores (Diff1, Diff2) both illustrating change in reference to t1. Diff1 was the latent mean difference from t1 to t2, Diff2 the latent mean difference from t1 to t3 (Figure 1). Within our analysis, we were interested whether Diff1 and Diff2 varied between IG and CG (McArdle, 2007), assuming the IG to exhibit more changes than the CG. LCMs were calculated using M*plus* (Muthén & Muthén, 1998-2014).

<< Figure 1 about here >>

Attitudes. To analyze whether Jewish IG students changed their attitudes more strongly in a positive direction compared to CG students from t1 to t2, and from t1 to t3, we calculated the distinction of latent difference scores in reference to the condition (IG; CG). The model fit was acceptable ($\chi^2_{corrected}$ = 82.29; *df* = 55; *p* = .010; CFI = .95; RMSEA = .06; SRMR = .09). Diff1 (ΔM = -.056; σ^2 = .47, $p < .001$) varied between the conditions

[8] According to Schermelleh-Engel, Moosbrugger, and Mueller (2003) a satisfactory model fit is indicated when χ^2 *p*-value ≥ .01, CFI ≥ .95, RMSEA ≤ .08, and SRMR ≤ .10 (p. 52).

[9] To calculate omega we used M*plus* templates from Yang and Green (2011).

[10] At least (partial) strong factorial MI is required (Little, 2013). Steenkamp and Baumgartner (1998) specified partial MI to consist of at least two invariant factor loadings and intercepts per construct.

[11] We also tested MI across IG and CG at each occasion of measurement. At least (partial) strong factorial MI was given for both dependent variables at all times except for behavior intentions at t3.

(β[12] = -.69, p = .001). However, results indicated differences in the opposite direction of our expectations: IG students demonstrated a less positive attitude development in comparison to CG students from t1 to t2 and also from t1 to t3 (Diff2; ΔM = -0.50; σ^2 = .46, p < .001; β = -.63, p = .002). Thus, we discovered *negative* short- and long-term effects (see Table 1).[13]

Behavior intentions. Model fit was acceptable ($\chi^2_{corrected}$ = 79.64; df = 56; p = .021; CFI = .97; RMSEA = .06; SRMR = .07). Again, Diff1 (ΔM = -0.51; σ^2 = .36, p < .001) varied negatively between the two conditions (β[14] = -1.00, p < .001). Diff2 (ΔM = -0.55; σ^2 = .42, p < .001) also showed negative differences (β = -1.05, p < .001).[15] Thus, results exhibited negative short- and long-term behavior intentions for the IG in comparison to the CG. Latent means of the CG hardly changed (Table 1). In sum, H1 was not supported. Although we found significant results, findings were contrary our expectations.

<< Table 1 about here >>

Reported Personal Experiences. We analyzed IG students' comments about what they told their parents about the meeting similarly to the procedure used by Graf, Paolini, and Rubin (2014) utilizing the qualitative data analysis software MAXQDA (Kuckartz, 1995-2014). Students' answers were assigned to three categories: positive (n = 20), negative (n = 21), or neutral (n = 35). Five comments included both a positive and negative aspect and were excluded from further analysis. Interrater reliability showed a good consistency (Krippendorff's α = .86; Hayes & Krippendorff, 2007) based on 30% randomly

[12] The independent variable was dummy coded; thus, we standardized only the dependent variable. Therefore, β is the change in the dependent variable, when the independent variable switches from 0 to 1 (Muthén & Muthén, 1998-2012).

[13] The discovered pretest differences for attitudes (i.e., IG students showing more positive attitudes toward the outgroup) might have been related to positive experiences of the IG within the two preceding years of the program. We assume this to be possible due to the fact that student's comments at t1 included positive statements about experiences within previous years.

[14] β can be > 1 because Diffs and dependent variable can correlate (Deegan, 1978).

[15] We exhibited significant results for all β-values, even if we adjust p-values according to the Bonferroni correction (i.e., dividing α by two; Bender, Lange, & Ziegler, 2007).

selected English comments which were coded by two independent raters. Thus, we maintained categorization of the first rater.[16]

According to the high number of negative comments, we decided to analyze our data in regard to negative contact. Pettigrew and Tropp (2011) stated "such [negative] contact typically involves not only both individual and collective threat but is also often involuntary" (p. 205). Students' negative responses included in fact aspects referring to anxiety and unpleasant interactions or feelings (e.g., beating, cursing, the ingroup being laughed at, being afraid, being scared, feeling uncomfortable, presenting Arabs as being unfriendly, indicating that meetings were no fun). To strengthen our assumptions, we created a dichotomous variable "experiences t2" by turning negative (= -1) and positive comments (= 1) into numeric values. In a next step, we used this new independent variable to examine its impact on attitudes at t2 and t3, respectively, while controlling for attitudes at t1. Data were in fact related to both attitudes at t2 ($\beta = .38$, $p = .011$; $M_{neg} = 2.33$, $M_{pos} = 3.60$) and at t3 ($\beta = .41$, $p = .062$; $M_{neg} = 2.76$, $M_{pos} = 3.78$).[17] Thus, we revealed that students who reported negative experiences at t2 had more negative attitudes at t2 and t3; no relations were found for behavior intentions (Table 2).[18]

<< Table 2 about here >>

Discussion

Altogether, the high number of negative reports ($N = 21$) and the results of our analysis based on students' experiences give reason to assume that the negative attitude effects observed for the Jewish participants might be rooted in their negative contact experiences during the program. An extended explanation for the negative program effects for the Jewish participants could be the Israel-Gaza Conflict 2012. Although, only one student might have referred to this event ("I like them very much even though I'm a little scared of Arabs at the moment") and no changes within CG scores over time were found, the influence of the Israel-Gaza Conflict 2012 can never be ruled out.

---

[16] About 90% of the students reported about their experiences ($N = 79$ IG students were attending t2 questionings and $N = 71$ IG students wrote a comment).

[17] We used listwise deletion; pairwise deletion exhibited a slightly different data structure for attitudes (t2: $\beta = .23$, $p = .152$; t3: $\beta = .46$, $p = .039$).

[18] We asked guides to tell us about any special incidents throughout the meetings but did not receive any specific feedback.

Sample 2

Method

We surveyed 139 Arab Israeli 6th grade students (IG: n = 83, females = 54%; CG: n = 56, females = 43%). IG students attended the exact same intergroup intervention and responded to an Arabic version of the same questionnaire. They were questioned about their attitudes and behavior intentions toward Jewish Israeli students. Questionings were approximately at the same time as in Sample 1.

Preliminary analyses. No selective attrition was found for attitudes (t2: $t[122] = .56$, $p = .579$; t3: $t[122] = -.73$, $p = .466$) and behavior intentions (t2: $t[126] = .87$, $p = .384$; t3: $t[126] = -1.71$, $p = .090$).[19] Three questionnaires were excluded due to problems in matching the questionnaires. Missing data were not MCAR ($\chi^2 = 101.03$, $df = 68$, $p = .006$). We assumed our data to be missing at random and calculated missing data via maximum-likelihood estimation (Schafer & Graham, 2002). No pretest differences were found regarding attitudes ($b = 0.10$, $SE = 0.29$, $p = .670$) and behavior intentions ($b = -0.04$, $SE = 0.24$, $p = .879$). Again, we controlled for hierarchical effects. Model fit was satisfactory at t1 ($\chi^2_{corrected} = 13.49$; $df = 8$; $p = .096$; CFI = .97; RMSEA = .07; SRMR = .04), t2, and t3. Reliability was acceptable for attitudes ($\omega_{t1} = .78$, $\omega_{t2} = .75$, $\omega_{t3} = .79$) and behavior intentions ($\omega_{t1} = .77$, $\omega_{t2} = .81$, $\omega_{t3} = .79$). Strict factorial invariance across all occasions of measurement was given for attitudes ($\chi^2_{corrected} = 41.44$; $df = 29$; $p = .063$; CFI = .96; RMSEA = .06; SRMR = .08) and behavior intentions ($\chi^2_{corrected} = 21.39$, $df = 29$; $p = .845$; CFI = 1.00; RMSEA = .00; SRMR = .05).[20]

Results

As in Sample 1, we analyzed LCMs.

Attitudes. Model fit was satisfactory ($\chi^2_{corrected} = 74.43$; $df = 59$; $p = .085$; CFI = .95; RMSEA = .04; SRMR = .07). Data showed that Diff1 ($\Delta M = -.20$; $\sigma^2 = .63$, $p = .002$) deviated between conditions ($\beta = .40$, $p = .056$). Thus, IG students liked Jewish students more than CG students from t1 to t2. The effect was generated by less positive attitudes within the CG, and stable IG values across time. Long-term results (Diff2; $\Delta M = -.24$; $\sigma^2 =$

[19] The number of participants at each occasion of measurement was: $N_{t1} = 131$, $N_{t2} = 126$, $N_{t3} = 128$.

[20] Cross-group (partial) scalar MI was given for both variables at all times except for attitudes at t1.

.67, $p < .001$) revealed no significant deviation ($\beta = .31$, $p = .127$; Table 1). Overall, participation in the intergroup program led to short-, but not to long-term differences between IG and CG students regarding attitudes.

Behavior intentions. Model fit was satisfactory ($\chi^2_{corrected} = 48.05$; $df = 59$; $p = .845$; CFI = 1.00; RMSEA = .00; SRMR = .06). Diff1 ($\Delta M = .17$; $\sigma^2 = .72$, $p < .001$) differed between conditions ($\beta = .60$, $p = .002$) showing latent means of the CG to become more negative at t2, whereas latent means of the IG developed in a positive direction. Diff2 ($\Delta M = -.10$; $\sigma^2 = .78$, $p < .001$) illustrated no long-term changes ($\beta = .11$, $p = .597$).[21] In sum, results did not confirm H2. IG students did not have more long-term attitudes and behavior intentions than CG students even after participation in repeated meetings.

Personal Experiences. Students' comments were classified into three categories (positive: $n = 44$; negative: $n = 2$; neutral: $n = 18$).[22] One comment was assigned to more than one category and excluded from further analysis. Interrater reliability was acceptable (Krippendorff's $\alpha = .79$; Hayes & Krippendorff, 2007). Qualitative findings portrayed a rather positive experience as almost no indication for a negative contact was given based on students' comments. We again analyzed the relation between comments and quantitative measures. Results revealed no relations for attitudes but for behavior intentions (t2: $\beta = .25$, $p = .065$; t3. $\beta = .24$, $p = .075$).[23] More positive comments at t2 were associated with more positive behavior intentions at t2 and t3.

Analysis Between Minority and Majority Members. In addition, we investigated whether data from Arab and Jewish students were invariant throughout time to enable comparisons between the samples. We created a data set including Jewish and Arab data and attained partial strict factorial invariance for attitudes ($n = 276$; $\chi^2_{corrected} = 25.88$; $df = .28$; $p = .580$; CFI = 1.00; RMSEA = .00; SRMR = .04) and strict factorial invariance for behavior intentions ($\chi^2_{corrected} = 26.89$, $df = 29$; $p = .578$; CFI = 1.00; RMSEA = .00; SRMR

[21] We obtained the same data structure for all β-coefficients when applying the Bonferroni correction (Bender et al., 2007). Only the short-term effects for attitudes were not significant.

[22] At least 86% of all IG students wrote comments at t2 ($N = 73$ IG students were attending t2 questionings; $N = 63$ IG students reported about their experiences).

[23] Results using pairwise deletion were slightly different (t2: $\beta = .27$, $p = .048$; t3: $\beta = .30$, $p = .025$).

= .03).[24] In a next step, we investigated whether Diff1 and Diff2 differed between Jewish and Arab Israeli students. In order to investigate whether a moderation was given, we directed two further variables (Ethnicity; Group x Ethnicity) on Diff1 and Diff2. The model fit for the attitude ($\chi^2_{corrected}$ = 95.73, *df* = 69; *p* = .018; CFI = .97; RMSEA = .04; SRMR = .06) and the behavior intention variable ($\chi^2_{corrected}$ = 86.78, *df* = 70; *p* = .085; CFI = .99; RMSEA = .03; SRMR = .05) was good. Results revealed that the interaction term was significant regarding the attitude variable for Diff1 (β = .23, *p* = .001) and Diff2 (β = .18, *p* = .007) as well as for the behavior intention variable for (Diff β = .33, *p* < .001) and Diff2 (β = .24, *p* < .001). Thus, H3 was confirmed and both status groups showed deviating short- and long-term effects.

Discussion

We found positive short- but no long-term effects for Arab IG students that participated in repeated short intergroup meetings in comparison to CG students. Thus, our hypothesis concerning long-term intervention effects was not confirmed. Results were discovered because the CG developed more negative attitudes and behavior intentions from t1 to t2. In contrast, IG students remained at about the same level within attitudes and became more positive only for behavior intentions at t2. Searching for an explanation of the attitude effect, it might be that the more negative attitudes and behavior intentions within the CG were a result of the Israel-Gaza Conflict 2012. Thus, it is possible that attitudes and behavior intentions of students who did not experience positive intergroup contact diminished due to the external incident whereas attitudes and behavior intentions in the IG were bolstered by the contact intervention. This would go along with Paolini et al. (2014) who speculated the "positive and diverse contact experiences in the past can buffer [individuals] against the harmful effects of new, discrete experiences of negative contact in the present" (p. 560). Due to the fact that the scale ranged from 1 to 5 and the mean values for the IG were far from extreme values (*M*: t1 = 3.59; t2 = 3.53; t3 = 3.46), ceiling effects are excluded empirically.

General Discussion

Aim of the present study was to analyze short- and long-term effects of contact interventions on majority and minority participants in an intractable conflict region. We

[24] We also achieved at least partial strong MI across groups at each occasion of measurement regarding both variables.

used a non-equivalent control group design with three measurement points. And, by using LCMs we did everything possible to ensure high reliability of the results. Nevertheless, our study revealed that the repeated intervention between Jewish and Arab 6th grade students in Israel did lead to long-term negative effects for the (Jewish) majority and to positive short- but not long-term effects for the (Arab) minority. This data pattern shows that contact interventions might be helpful to improve intergroup attitudes and behavior intentions – as shown in numerous studies before (e.g., Lemmer & Wagner, 2015) – but that the effect of contact interventions depends on at least three aspects, all connected to the societal and ideological atmosphere in intractable conflict regions. First, the negative contact effect in the Jewish sample indicates that contact is not per se equivalent to positive experiences. Even though there is good reason to assume that contact in "calm" regions of the world, like Western Europe or North America is mostly positive (e.g., Pettigrew & Tropp, 2011), this does not hold true for intractable conflict regions. When looking back to a long history of mutual rejection and violence, a contact situation can easily acquire a negative character from the perspective of the participants (see also Paolini et al., 2014), which then leads to more negative attitudes and behavior intentions. Second, looking at the Arab participants' data, only short-term positive contact effects were demonstrated. This might also be connected to the specific societal circumstances in which the evaluated intervention was embedded. Having the opportunity to really get to know members of the outgroup might reduce prejudice and improve positive behavior intentions, however, returning to a segregated social environment reestablishes the former outgroup images and automatic behavior intentions. Moreover, it is important to consider status within the analysis of intergroup contact interventions in conflicting areas, in particular as both groups have strongly deviating perceptions and needs in regard to the intergroup contact situation.

As a conclusion, we would not say that contact interventions in intractable conflicts are without merit. Instead, on the basis of our results, we proclaim, first, that such contact interventions have to be prepared carefully to inhibit negative contact experience. Second, also measures against the relapse of intervention effects must be taken especially for regions that are characterized by continued and hardened conflict, for example, the integration of the entire network of families and communities into the program could develop a more positive view about the outgroup (Rosen & Perkins, 2013).

A limitation of the present real-life evaluation was the non-randomized control group design. However, our study is one of very few providing CG information at

all – a strong improvement in internal validity compared to the usual evaluations without control groups (Shadish et al., 2002). Nevertheless, our design is associated with internal validity threats. For example, intervention group students might have been generally more open to intercultural encounters compared to control group students. This, however, would not explain the negative development of the Jewish students' attitudes and behavior intentions. Further research should try to realize truly random assignment to intervention and control group – if this is possible in a complicated research field.

References

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Arnon, M. (2010). *Reflection on the activities of a peace education program as a tool for reinforcing/rehabilitating its long term effects* (Unpublished doctoral dissertation). University of Haifa, Haifa, Israel. [Hebrew]

Bar-Natan, I., Rosen, Y., & Salomon, G. (2010). *Does friendship between adversaries generalize?* Retrieved from http://peach.haifa.ac.il/images/7/7b/Irit_Bar-Natan.pdf

Bar-Tal, D. (2013). *Intractable Conflicts. Socio-psychological foundations and dynamics*. Cambridge: Cambridge University Press.

Bender, R., Lange, S., & Ziegler, A. (2007). Multiples Testen. Artikel Nr. 12 der Statistik-Serie in der DMW [Multiple testing. Article 12 of the statistic series in the DMW]. *Deutsche Medizinische Wochenschrift*, *132*(S 01), e26–e29. [German]

Berger, R., Benatov, J, Abu-Raiya, H., & Tadmor, T. C. (2016). *Reducing prejudice and promoting positive intergroup attitudes among elementary-school children in the context of the Israeli-Palestinian conflict.* Manuscript submitted for publication.

Christ, O., Schmid, K., Lolliot, S., Swart, H., Stolle, D., Tausch, N., … Hewstone, M. (2014). Contextual effect of positive intergroup contact on outgroup prejudice. *Proceedings of the National Academy of Sciences*, *111*(11), 3996–4000. doi:10.1073/pnas.1320901111

Deegan, J., Jr. (1978). On the occurrence of standardized regression coefficients greater than one. *Educational and Psychological Measurement*, *38*(4), 873–888.

Gollwitzer, M., Christ, O., & Lemmer, G. (2014). Individual differences make a difference: On the use and the psychometric properties of difference scores in social psychology. *European Journal of Social Psychology*, *44*(7), 673–682.

Graf, S., Paolini, S., & Rubin, M. (2014). Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and

contact prevalence in five Central European countries. *European Journal of Social Psychology*, *44*(6), 536–547.

Graham, J. W., Cumsille, P. E., & Shevock, A. E. (2013). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology. Research methods in psychology. Volume 2* (pp. 109–141). Hoboken, NJ: John Wiley & Sons.

Harkness, J. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: John Wiley & Sons.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89.

Jayusi, W. (2009). *Restoring the attitudes of peace education participants through peer-tutoring* (Unpublished doctoral dissertation). University of Haifa, Haifa, Israel. [Hebrew]

Kriesberg, L. (1998). Intractable conflicts. In E. Weiner (Ed.), *The handbook of interethnic coexistence* (pp. 332–342). New York, NY: Continuum Books.

Kropiunigg, U., & Pabst, B. (2007). Selbstwertsteigerung und Vorurteilsreduktion bei Jugendlichen eines multiethnischen Peace Camps [Increase in self-worth and reduction of prejudice in youth participants of a multi-ethnic peace camp]. *Psychotherapie Forum*, *15*, 63–72. [German]

Kuckartz, U. (1995-2014). MAXQDA (Version 11) [Computer software]. Berlin, Germany: VERBI GmbH.

Lanphen, J. (2011). *Kooperatives Lernen und Integrationsförderung: Eine theoriegeleitete Intervention in ethnisch heterogenen Schulklassen [Cooperative learning and promoting integration: A theory-based intervention in ethnically heterogeneous classes]*. Münster, Germany: Waxmann. [German]

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology, 45*(2), 152–168.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: The Guilford Press.

Maoz, I. (2011). Does contact work in protracted asymmetrical conflict? Appraising 20 years of reconciliation-aimed encounters between Israeli Jews and Palestinians. *Journal of Peace Research, 8*(1), 115–125.

Maoz, I., Bar-On, D., Bekerman, Z., & Jaber-Massarwa, S. (2004). Learning about 'good enough' through 'bad enough': A story of a planned dialogue between Israeli Jews and Palestinians. *Human Relations, 57*(9), 1075–1201.

McArdle, J. J. (2007). Dynamic structural equation modeling in longitudinal experimental studies. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 159–188). Mahwah, NJ: Lawrence Erlbaum Associates.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Miller, G. A., & Chapman, J. P., (2001). Misunderstanding Analysis of Covariance. *Journal of Abnormal Psychology, 110*(1), 40–48. doi:10.1037//0021-843X.110.1.40

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998–2014). M*plus* (Version 7.2) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Paolini, S., Harwood, J., & Rubin, M. (2010). Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin*, *36*(12), 1723-1738.

Paolini, S., Harwood, J., Rubin, M., Husnu, S., Joyce, N., & Hewstone, M. (2014). Positive and extensive intergroup contact in the past buffers against the disproportionate impact of negative contact in the present. *European Journal of Social Psychology*, *44*(6), 548-562.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783.

Pettigrew, T. F., & Tropp, L. R. (2011). *When groups meet. The dynamics of intergroup contact.* New York, NY: Psychology Press.

Rosen, Y., & Perkins, D. (2013). Shallow roots require constant watering: The challenge of sustained impact in educational programs. *International Journal of Higher Education, 2*(4), 91–100.

Rosen, Y., & Salomon, G. (2011). Durability of peace education effects in the shadow of conflict. *Social Psychology of Education*, *14*(1), 135–147.

Saleh, Y., & al-Mughrabi, N. (2012, November 21). Hamas-Israel ceasefire takes hold but mistrust runs deep. *Reuters*. Retrieved from http://www.reuters.com/article/2012/11/21/us-palestinians-israel-hamas-idUSBRE8AD0WP20121121

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

Schermelleh-Engel, K., Moosbrugger, H., & Mueller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schleien, S. M. (2007). *Confronting the intractable: an evaluation of the Seeds of Peace experience* (Unpublished doctoral dissertation). University of Waterloo, Waterloo, Canada.

Schroeder, J., & Risen, J. L. (2016). Befriending the enemy: Outgroup friendship longitudinally predicts intergroup attitudes in a coexistence program for Israelis and Palestinians. *Group Processes and Intergroup Relations, 19*(1), 72-93. doi:10.1177/1368430214542257

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shani, M. (2015). *A theory and practice of coexistence: Improving coexistence orientation through mixed-model encounters between Jews and Palestinians in Israel* (Unpublished doctoral dissertation). University of Bremen, Bremen, Germany.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling. Second Edition*. London: Sage Publications Ltd.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90.

Stephan, W. G., Stephan, C. W., & Gudykunst, W. B. (1999). Anxiety in intergroup relations: a comparison of anxiety/uncertainty management theory and integrated threat theory. *International Journal of Intercultural Relations*, *23*(4), 613–628.

Wagner, U., & Hewstone, M. (2012). Intergroup Contact. In L. R. Tropp (Ed.), *The Oxford Handbook of Intergroup Conflict* (pp. 193–209). New York, NY: Oxford University Press.

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, *29*(4), 377–392.

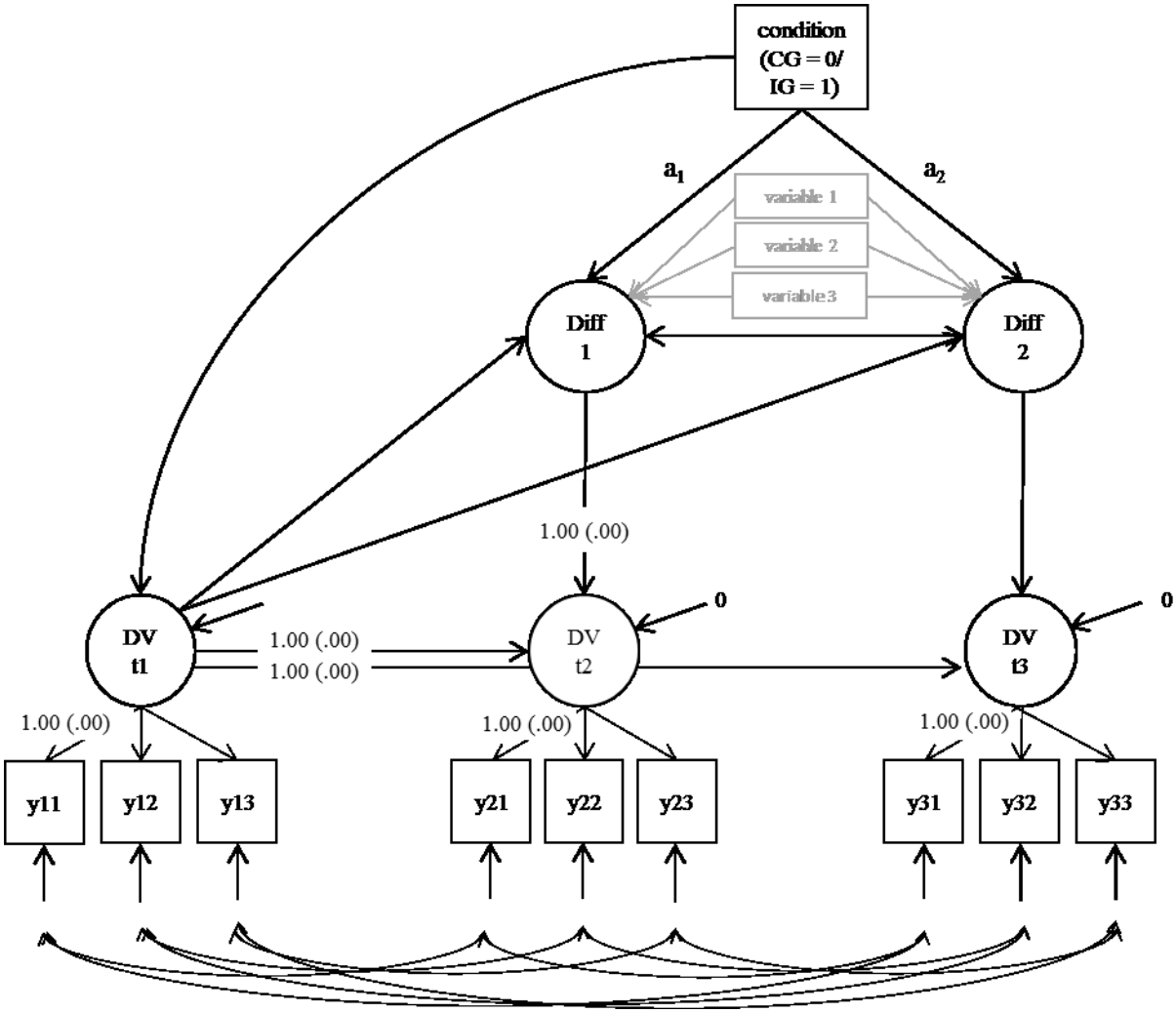


Figure 1. Depiction of the used LCM; condition = independent variable; IG = intervention group; CG = control group; DV = latent dependent variable (either attitudes or behavior intentions toward the outgroup); t1 = pretest (before the first intergroup meeting); t2 = posttest (exactly at the end of the last intergroup meeting); t3 = follow-up (about nine weeks after the last intergroup meeting); Diff1 = latent difference score between t1 and t2; Diff2 = latent difference score between t1 and t3; variable 1-3 = control for hierarchical effects; a_1 = effect of the condition on Diff1; a_2 = effect of the condition on Diff2.

Table 1

Latent Means and Latent Change Model Coefficients for Minority and Majority Group Students

				Latent Means			Latent Change Model Coefficients			
Sample	DV	Group	*n*	t1	t2	t3	Diff1 - condition (a_1) β	Diff1 - condition (a_1) *b* (*SE*)	Diff2 - condition (a_2) β	Diff2 - condition (a_2) b (*SE*)
Jewish Israeli	Attitudes	IG	83	3.51	2.33	2.43	-.69**	-.62 (.17)***	-.63**	-.53 (.16)**
		CG	54	2.89	2.95	2.96				
	Behavior Intentions	IG	83	2.76	1.70	1.63	-1.00[a] ***	-.91 (.14)***	-1.05[b] ***	-.98 (.15)***
		CG	54	2.57	2.60	2.61				
Arab Israeli	Attitudes	IG	83	3.59	3.53	3.46	.40[†]	.39 (.20)[†]	.31	.31 (.20)
		CG	56	3.49	3.14	3.15				
	Behavior Intentions	IG	83	3.12	3.63	3.10	.60**	.63 (.21)**	.11	.11 (.20)
		CG	56	3.16	3.00	2.99				

Note. Latent means = values between 1 and 5 with 1 signifying lowest and 5 highest agreement; IG = intervention group; CG = control group; t1 = pretest (before the first intergroup meeting); t2 = posttest (exactly at the end of the last intergroup meeting); t3 = follow-up (about weeks after the last intergroup meeting); Diff1 = latent difference score between t1 and t2; Diff2 = latent difference score between t1 and t3; a_1 = effect of the condition on Diff1, see Figure 1; a_2 = effect of the condition on Diff2, see Figure 1.

[a] The standardized coefficient β can be greater 1 according to the fact that Diff scores and dependent variables at t1 can correlate within the latent change model (Deegan, 1978).

[†] $p < .10$. ** $p < .01$. *** $p < .001$.

Table 2

Relations Between Personal Experiences Reported at t2 and Short- and Long-Term Attitudes and Behavior Intentions for Minority and Majority Students

Sample	DV		*B*	*SE B*	β	DV		*B*	*SE B*	β
Jewish Israeli	Attitudes t2	Constant	.54	.61		Attitudes t3	Constant	2.41	.71	
		Attitudes t1	.68	.17	.56***		Attitudes t1	.25	.19	.27
		Experiences t2	.44	.16	.38*		Experiences t2	.37	.19	.41†
	Behavior intentions t2	Constant	.18	.44		Behavior intentions t3	Constant	.53	.62	
		Behavior t1	.66	.14	.73***		Behavior t1	.59	.19	.61**
		Experiences t2	.12	.17	.11		Experiences t2	.06	.24	.05
Arab Israeli	Attitudes t2	Constant	2.34	.58		Attitudes t3	Constant	1.70	.43	
		Attitudes t1	.33	.15	.35*		Attitudes t1	.50	.12	.56***
		Experiences t2	.13	.43	.05		Experiences t2	.21	.29	.09
	Behavior intentions t2	Constant	2.06	.49		Behavior intentions t3	Constant	1.71	.51	
		Behavior t1	.36	.13	.36**		Behavior t1	.43	.14	.40**
		Experiences t2	.64	.34	.25†		Experiences t2	.64	.36	.24†

Note. DV = dependent variable; t1 = pretest (before the first intergroup meeting); t2 = posttest (exactly at the end of the last intergroup meeting); t3 = follow-up (about nine weeks after the last intergroup meeting).
† $p < .10$. * > .05. ** $p < .01$. *** $p < .001$.

III. ANCILLARY ANALYSIS

Additional analyses that are related to the research questions and that have not been covered by the two submitted manuscripts were carried out and are presented in the following chapter. First, the contextual situation that was present during the course of the gathering of the data was investigated (Israel-Gaza Conflict 2012; Chapter III.1). Furthermore, the examination of the research questions for all samples included in this thesis concerning students' personal experience was completed (Chapter III.2). Third, further measurement invariance and noninvariance analyses were conducted (Chapter III.3). Fourth, it was examined whether the used missing data technique (maximum likelihood estimation) revealed comparable results to an alternative missing data technique (multiple imputation technique; Chapter III.4). Finally, gender differences were explored (Chapter III.5).

1. Assessment of the Israel-Gaza Conflict 2012

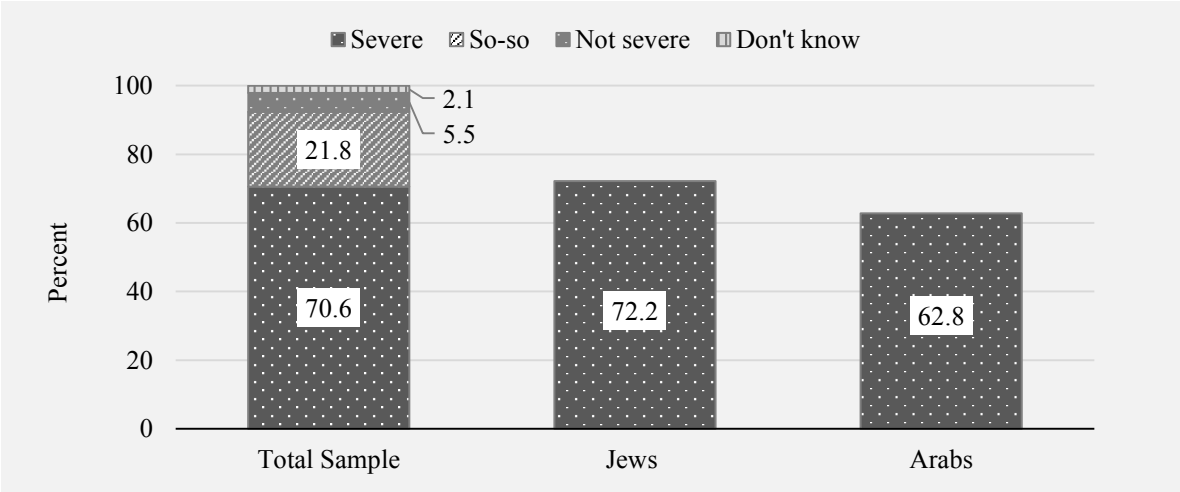
The Israel-Gaza Conflict 2012 occurred from November 14th-21st, 2012, in between each of the two analyzed interventions (both lasting from October until December 2012). The conflict included rocket attacks and air strikes between Israel and Gaza. Within the Israel-Gaza Conflict 2012 more than 140 individuals - including civilians - were killed (Rohde, 2012). We decided to investigate external data related to this period in order to explore how both conflicting groups (Jewish Israeli and Arab Israeli) perceived the Israel-Gaza Conflict 2012 and whether it might have had an impact on our results. In the scope of two surveys the Israeli Democracy Institute collected data: 1. the Israeli Democracy Index and 2. the Peace Index. Both surveys questioned Jews and Arabs in Israel either about their attitudes toward Jewish-Arab-Tensions or peace negotiations between Israel and Palestine. The surveys were conducted repeatedly; however, new individuals were surveyed each time. In sum, both surveys illustrated that Jews and Arabs responded to the Israel-Gaza Conflict 2012. It was also shown that both groups had deviating reactions: Jews being more pessimistic than Arabs after the Israel-Gaza Conflict 2012. In the following, the findings of the two surveys are presented in more detail.

1. Israeli Democracy Index. This index "provides statistical information that gives a reliable and comprehensive picture on the quality and functioning of democracy, as well as the way it is perceived by the public" in Israel (Israeli Democracy Institute [IDI], 2015a,

para. 1). It is based on a yearly survey, which is predominantly conducted in April and May of each year. Due to the fact that the Israel-Gaza Conflict 2012 took place in November 2012, we looked at the survey data in the years 2012 and 2013 to investigate whether and how the opinion of Jews and Arabs changed within this time frame (i.e., before and after this specific conflict). In 2012, the survey investigated a representative sample of N = 1,025 participants (n = 834 Jews, n = 191 Arabs); in 2013 a sample of N = 1,000 participants was surveyed (n = 852 Jews, n = 148 Arabs). Both times adults at the age of 18 and older were questioned (Hermann, Atmor, Heller, & Lebel, 2012, p. 20-21; Hermann et al., 2013a, p. 18).

In the year 2012, Jews and Arabs perceived tensions between the two groups differently with ca. 10% more Jews than Arabs stating to perceive the tensions as severe (Hermann et al., 2012; Figure 1).

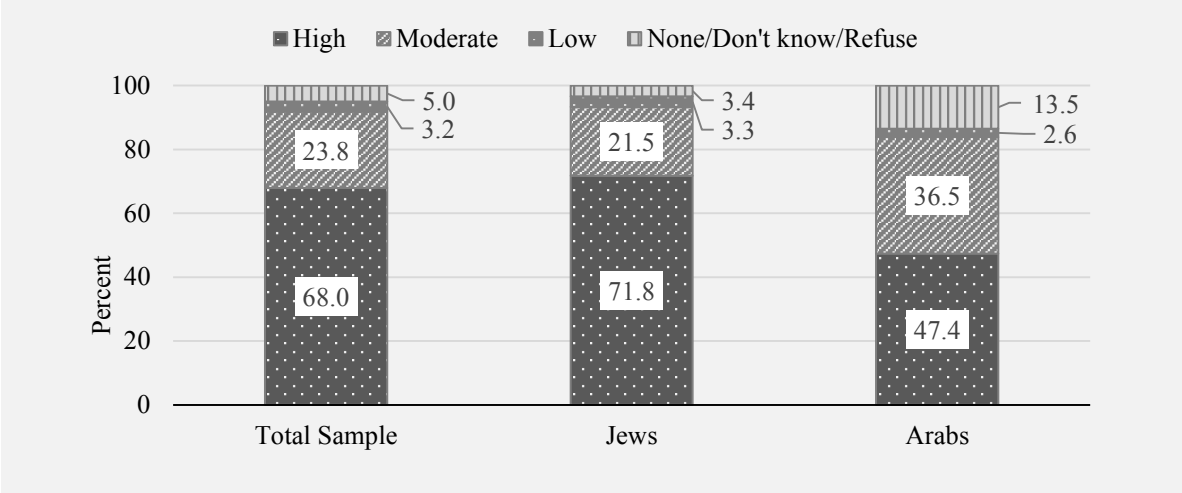
Figure 1: Assessment of Tensions Between Jews and Arabs in 2012



Note. N_{total} = 1025 (n_{Jews} = 834; n_{Arabs} = 191); no information of the Jewish and Arab answers concerning other answer possibilities (i.e., so-so, not severe, and don't know) were illustrated (Hermann et al., 2012, p. 60).

In 2013, 71.8% of the Jews in comparison to 47.4% of the Arabs perceived the level of Jewish-Arab tensions as high, illustrating a difference of almost 25% (Hermann et al., 2013a, p. 85, Figure 2). The findings illustrate a strong difference between the opinion of Jews and Arabs in their perception about the tensions between the two groups in 2013 and a smaller difference in 2012.

Figure 2: Assessment of Tensions Between Jews and Arabs in 2013



Note. N_{total} = 1000 (n_{Jews} = 852; n_{Arabs} = 148; Hermann et al., 2013a, p. 85).

The items presented to the participants differed between the two years. In 2012 participants were asked: "How severe is the tension in each of these areas?" with one category being Jews/Arabs and the answering options being *severe*, *so-so*, *not severe*, and *don't know/refuse* (Hermann et al., 2012, p. 60) whereas in 2013, participants were asked to rank the "Level of Jewish-Arab tensions" within the answering scheme *high*, *moderate*, *low*, and *none/don't know/refuse* (Hermann et al., 2013a, p. 85). Due to the different items used in each year and because of the deviating population, comparisons between the two years can only be regarded as indicator for a possible change within the opinion of Jews and Arabs, but not as a definite change within their respective perception.

Two further interesting results were reported. In 2012, participants were asked whether in their opinion "Arab citizens of Israel are discriminated against as compared to Jews". 13.7% of the Jewish in comparison to 46.6% of the Arab participants agreed strongly to this statement. This showed how different discrimination was perceived in Israel in 2012 (Hermann et al., 2012, p. 63). In 2013, 47.6% of the Jewish participants also reported the greatest aversions toward an outgroup neighbor, whereas 41.9% of the Arab participants responded that way (Hermann et al., 2013b, p. 6). Results showed that there are Jewish-Arab frictions in Israel within both years: discrimination against Arab Israeli participants was stronger felt by themselves then by Jewish Israeli participants and Jewish Israeli participants had a greater aversion to be a neighbor of an Arab Israeli than vice versa.

2. Peace Index. The Peace Index is a monthly survey about "trends in Israeli public opinion regarding the Israeli-Palestinian conflict and relations between Jews and Arabs in

Israel and their impact on Israeli society" (IDI, 2015b, para. 1). Due to its monthly survey, more direct conclusions in regard to the Israel-Gaza Conflict 2012 can be drawn from the Peace Index than from the Israeli Democracy Index because we could look at the data shortly before and after the Israel-Gaza Conflict 2012. *N* = 601 participants were surveyed from October 22nd-24th, 2012 (October) and *N* = 598 participants from November 28th until December 2nd, 2012 (November; IDI, 2012a, 2012b). In Table 6, two questions are presented that were asked at both time points.

Table 6: Israeli Public Opinion Before and After the Israel-Gaza Conflict 2012

Question	Jews		Arabs	
	October	November	October	November
What is your position on holding peace negotiations between Israel and the Palestinian Authority?				
Strongly in favor	24.9%	27.1%	53.8%	54.6%
Somewhat in favor	40.8%	31.0%	8.5%	25.3%
Somewhat opposed	16.6%	11.2%	12.6%	5.4%
Strongly opposed	*12.0%*	*23.5%*	*22.3%*	*12.0%*
Don't know/refuse to answer	5.6%	7.2%	2.8%	2.7%
Do you believe or not believe that negotiations between Israel and the Palestinian Authority will lead to peace between Israel and the Palestinians in the coming years?				
Strongly believe	3.7%	6.1%	23.1%	20.9%
Somewhat believe	20.7%	14.3%	14.1%	27.1%
Somewhat don't believe	30.5%	27.3%	12.5%	25.3%
Don't believe at all	*43.6%*	*50.7%*	*50.3%*	*21.2%*
Don't know/refuse to answer	1.4%	1.5%	0.0%	5.5%

Note. $N_{\text{total October}}$ = 601; $N_{\text{total November}}$ = 598 (IDI; 2012a, 2012b).

Herein, results indicate that Jewish participants' strong opposition against peace negotiations increased within one month (October 12.0%; November 23.5%). In contrast, Arabs indicated a strong decrease in their opposition against peace negotiations (October 22.3%; November 12.0%). Additionally, answers between the two groups differed concerning the belief that negotiations would lead to peace: 43.6% of the Jews did not believe at all in negotiations in October and even in November about 50.7% did so. In comparison, in October 50.3% of the Arabs did not believe in negotiations at all whereas

this percentage decreased to 21.2% in the November survey (IDI, 2012a, 2012b). As in the Peace Index, findings illustrated opposite developments between the perceptions of the two groups: Jews were more opposed against holding peace negotiations and believed less in peace negotiations to lead to peace shortly after the Israel-Gaza Conflict 2012 than shortly before and Arabs illustrated less opposition to peace negotiations and believed less that negotiations do not lead to peace at all after the conflict than before.

2. Analysis of the Relevance of Students' Intergroup Experience

As presented in the research questions (Chapter I.4), we were interested whether the valence of student's intergroup experience was related to short- and long-term intergroup contact intervention effects within a conflict area. The findings from four samples were already presented in our two submitted manuscripts (Manuscript #1: Jewish Israeli 4th grade students and Arab Israeli 4th grade students questioned three times; Manuscript #2: Jewish Israeli 6th grade students and Arab Israeli 6th grade students). In the following, the two remaining samples were analyzed (1. Jewish Israeli 4th grade students questioned five times and 2. Arab Israeli 4th grade students questioned five times; Chapter III.2.1). Additionally, we present an overview of the findings concerning all six samples (Chapter III.2.2).

2.1. Analysis of the Relevance of Students' Intergroup Experience of the Remaining Samples

The two samples, which were not examined in Manuscript #1 or Manuscript #2, were the Jewish Israeli 4th grade cohort and the Arab Israeli 4th grade cohort questioned five times.

Jewish Israeli 4th grade students questioned five times. The sample consisted of $N = 33$ students (48% female). Students were surveyed five times: one week ahead of the intervention (t1); one week after the first meeting (t2); ca. four weeks later, before the second meeting (t3); after all four meetings (t4); and ca. eleven weeks after the fourth meeting (t5). Analogical to the research procedure in Manuscript #1, two independent raters classified student's open-ended answers to the question "Did you tell your parents about the meetings? If yes, what did you tell them?" into the categories positive, negative, and neutral comments using the software MAXQDA (Version 11; Kuckartz, 1995-2015). In three cases, comments were double coded (i.e., included a positive and a negative aspect in one comment). These cases were excluded from further analysis. Interrater reliability was good

(Krippendorff's α = .85), pointing to a good consensus between two raters (Hayes & Krippendorff, 2007). We continued to use the categorization of the first rater. At the posttest (t4), 85% of the IG students, who attended the questioning, wrote a comment (*N* = 23 comments). The seven positive comments included aspects about *positive emotions* (e.g., "I told them [parents] I was very curious to know what we will do with them"), *pleasant interactions* (e.g., "I said we met Israeli Arab children (…) and it was really nice with them"), and *friendship* (e.g., "I told them I found friendship and had a lot of fun and I want to meet them again"). There were also two negative and fourteen neutral comments. As in Manuscript #1, we analyzed the interrelation between written answers and quantitative findings (Moran-Ellis, 2006). We generated a new variable, in which we coded the comments at t4 (-1 = negative comments; 0 = neutral comments; 1 = positive comments). We calculated a regression analyses between t4 comments and the quantitative results at t4 and t5 (while controlling for the quantitative values at the pretest). Results showed that expressed positive experience was related to more positive short- and long-term attitudes (t4: β = .41, *p* = .094 [marginal significance]; t5: β = .53, *p* = .024) as well as more positive short- but not long-term behavior intentions (t4: β = .53, *p* = .043; t5: β = .41, *p* = .131). Thus, data from Jewish Israeli 4th grade students questioned five times supported the research hypothesis, showing that expressed positive experience (i.e., positive emotions, pleasant interactions, finding friends) was related to positive short-term intervention effects in both dependent variables. The prediction of long-term intervention effects was also confirmed for the attitude variable.

Arab Israeli 4th grade students questioned five times. The Arab sample consisted of *N* = 30 students (53% females). We now tested the Arab Israeli 4th grade sample analogously to the Jewish Israeli 4th grade questioned five times sample. Arab Israeli students were questioned at about the same time as Jewish Israeli. Interrater agreement was satisfactory (Krippendorff's α = .76; Krippendorff, 2004). Therefore, we kept the categorization of the first rater. At the posttest (t4) 96% of the students, who attended the questionings, reported their experience (*N* = 24 comments; 20 positive comments, 4 neutral comments). Due to the fact that no longitudinal measurement invariance was given for attitudes within the Arab Israeli sample questioned three times, we only tested the influence of student's intergroup experience concerning the behavior intention variable. The regression analysis did not show a meaningful relation between the valence of student's intergroup experience and positive short- and long-term behavior intentions outcomes

(t4: $\beta = -.16$, $p = .429$; t5: $\beta = .38$, $p = .155$).[1] Therefore, data from Arab Israeli 4th grade students questioned five times sample did not confirm the research hypothesis and positive intergroup experience were not related to positive short- and long-term intergroup contact intervention effects.

2.2. Overview of all Results Regarding the Relevance of Students' Intergroup Experience

In Table 7, results of all six samples concerning the respective relation of student's intergroup experience, which were reported at the posttest (t2 or t4),[2] and short- as well as long-term intergroup contact intervention effects are presented.

[1] Listwise deletion was used for the regression analysis; thus, only two students were in the neutral category.

[2] Both time points indicate the questioning at the end of all meetings; the different numbers are related to the fact that some students were questioned three and other students five times.

Table 7: Overview of the Analyses of Students' Intergroup Experiences

Sample	DV Short-Term		*B*	*SE B*	β	DV Long-Term		*B*	SE *B*	β
Jews 4th (3 POM)	Attitudes t2	Constant	2.58	.31		Attitudes t3	Constant	2.29	.42	
		Attitudes t1	.40	.10	.50***		Attitudes t1	.37	.13	.37**
		Experiences t2	.50	.25	.25†		Experiences t2	.95	.38	.33*
	Behavior intentions t2	Constant	1.28	.39		Behavior intentions t3	Constant	1.81	.53	
		Behavior t1	.55	.11	.54***		Behavior t1	.27	.15	.20
		Experiences t2	.74	.30	.28*		Experiences t2	1.67	.44	.48***
Jews 4th (5 POM)	Attitudes t4	Constant	3.07	.79		Attitudes t5	Constant	1.66	1.13	
		Attitudes t1	.32	.24	.31		Attitudes t1	.66	.35	.39†
		Experiences t4	.46	.26	.41†		Experiences t4	1.06	.42	.53*
	Behavior intentions t4	Constant	4.09	1.09		Behavior intentions t5	Constant	2.88	1.26	
		Behavior t1	-.05	.31	-.04		Behavior t1	.15	.37	.10
		Experiences t4	.89	.40	.53*		Experiences t4	1.06	.66	.41
Arabs 4th (3 POM)	Behavior intentions t2	Constant	4.32	.41		Behavior intentions t3	Constant	3.75	.55	
		Behavior t1	-.22	.10	-.37*		Behavior t1	.02	.14	.03
		Experiences t2	.65	.36	.31†		Experiences t2	.15	.48	.07
Arabs 4th (5 POM)	Behavior intentions t4	Constant	3.12	.69		Behavior intentions t5	Constant	2.44	1.16	
		Behavior t1	.40	.14	.57*		Behavior t1	.15	.24	.16
		Experiences t4	-.33	.41	-.16		Experiences t4	1.22	.81	.38
Jews 6th (3 POM)	Attitudes t2	Constant	.54	.61		Attitudes t3	Constant	2.41	.71	
		Attitudes t1	.68	.17	.56***		Attitudes t1	.25	.19	.27
		Experiences t2	.44	.16	.38*		Experiences t2	.37	.19	.41†
	Behavior intentions t2	Constant	.18	.44		Behavior intentions t3	Constant	.53	.62	
		Behavior t1	.66	.14	.73***		Behavior t1	.59	.19	.61**
		Experiences t2	.12	.17	.11		Experiences t2	.06	.24	.05
Arabs 6th (3 POM)	Attitudes t2	Constant	2.34	.58		Attitudes t3	Constant	1.70	.43	
		Attitudes t1	.33	.15	.35*		Attitudes t1	.50	.12	.56***
		Experiences t2	.13	.43	.05		Experiences t2	.21	.29	.09
	Behavior intentions t2	Constant	2.06	.49		Behavior intentions t3	Constant	1.71	.51	
		Behavior t1	.36	.13	.36**		Behavior t1	.43	.14	.40**
		Experiences t2	.64	.34	.25†		Experiences t2	.64	.36	.24†

Note. DV = dependent variable; POM = points of measurement; t1 = pretest (one week before the first meeting); t2 = posttest (exactly at the end of all meetings for samples questioned three times); t3 = follow-up (about nine to eleven weeks after the end of all meetings for samples questioned three times); t4 = posttest (exactly at the end of all meetings for samples questioned five times, i.e., t2 in samples questioned three times); t5 = follow-up (about nine to eleven weeks after the end of all meetings for samples questioned five times, i.e., t3 in samples questioned three times).

As shown, in all three Jewish Israeli samples the reported valence of the intergroup experience was related (at least marginally) to short- and long-term attitudes toward Arab Israeli students.[3] Students expressing positive experiences had more positive attitudes toward Arbs than students who reported negative or neutral comments. This is especially interesting considering the fact that in the Jewish Israeli 4th grade cohort the expressed valence of the experience was predominantly positive whereas in the Jewish Israeli 6th grade cohort comments were mainly negative. Thus, Jewish Israeli student's comments regarding their intergroup experience can be used as an indicator for short- and long-term attitudes toward Arab Israeli students. Moreover, student's experience were related to short-term behavior intentions in the Jewish Israeli 4th grade cohort questioned five times while they predicted short- and long-term behavior intentions in the Jewish Israeli 4th grade cohort questioned three times. No relations were found in the 6th grade cohort. Student's intergroup experience were also related to short-term behavior intentions in the Arab Israeli 4th grade cohort questioned three times, and to short- and long-term behavior intentions in the Arab Israeli 6th grade cohort (marginal effect). Overall, the valence of student's comments was less related to contact intervention effects for the Arab Israeli than for the Jewish Israeli cohorts.

3. Measurement Invariance and Noninvariance Analysis

As mentioned in Chapter I.3.3. measurement invariance (MI) calculations have rarely, if at all, been conducted in longitudinal (quasi-)experimental studies related to intergroup contact interventions. In fact, they are not often realized in contact literature in general yet (e.g., Swart, Hewstone, Christ, & Voci, 2011). We used our data, meaning the four samples, which were questioned three times (Table 4; Chapter I.4.) to compute not only MI between the two ethnicities but also further cross-group and longitudinal MI analyses (e.g., age differences and measurement noninvariance analysis). In Chapter III.3.1. we tested whether MI was given between different age cohorts within each of the ethnic groups. Chapter III.3.2. investigated the sample in which measurement noninvariance was revealed within the attitude variable in more detail. Chapter III.3.3. presents an overview of

[3] Within all samples comments were categorized as positive, neutral, or negative. Nevertheless, analyses of students' intergroup experience differ slightly between the samples as in some samples no negative comments appeared and in others the neutral category was excluded from the analysis to enable insights into the different effects of students with a positive and a negative intergroup experience.

all measurement invariance and noninvariance analysis of all four samples of this thesis, which were analyzed using latent change modeling.

3.1. Measurement Invariance Analysis

We tested whether students within the same ethnicity/religion (i.e., Jews or Arabs) but from different age cohorts (i.e., 4th and 6th grade) had an equal understanding of the measured constructs. That is, we analyzed whether the understanding of the measured constructs differed between the two age cohorts.

Age differences with regard to Jewish Israeli 4th and 6th grade students. We examined whether Jewish Israeli 4th (N = 113 students) and 6th grade students (N = 137 students) understood the two constructs attitudes and behavior intentions equally. Again, MI across groups (4th and 6th graders) was investigated at three times of measurement. At least cross-group partial metric partial scalar MI for attitudes (pretest: χ^2_{corr} = 3.67, df = 4, p = .454, CFI = 1.00, RMSEA = .00, SRMR = .07; posttest: χ^2_{corr} = 6.15, df = 5, p = .292, CFI = .99, RMSEA = .05, SRMR = .07; follow-up test: χ^2_{corr} = 2.68, df = 4, p = .613, CFI = 1.00, RMSEA = .00, SRMR = .07) and scalar MI for behavior intentions were given at each occasion of time (pretest: χ^2_{corr} = 4.98, df = 5, p = .418, CFI = 1.00, RMSEA = .00, SRMR = .05; posttest: χ^2_{corr} = 4.54, df = 5, p = .474, CFI = 1.00, RMSEA = .00, SRMR = .07; follow-up test: χ^2_{corr} = 3.24, df = 5, p = .664, CFI = 1.00, RMSEA = .00, SRMR = .07).[4] These findings indicated that in the Jewish Israeli samples no difference in the understanding of the measured constructs appeared between the age cohorts, for example, due to maturation.

Age differences with regard to Arab Israeli 4th and 6th grade students. Analogously, we explored whether data were comparable from Arab Israeli students at different age cohorts (i.e., 4th grade [N = 108 students] and 6th grade [N = 139 students]). We investigated whether cross-group MI between both age cohorts was given within the behavior intention variable at each point in time. Solely MI within the behavior intentions variable was investigated, because in the Arab Israeli 4th grade cohort no longitudinal MI within the attitude variable was found. The attitude variable was examined in more detail in Chapter III.3.2. (measurement noninvariance analysis). Within this data set, at least cross-group

[4] Longitudinal MI across all three times of measurement (from October 2012 till February 2013) was tested. Longitudinal partial strict scalar MI was reached for attitudes (χ^2_{corr} = 29.91; df = 27; p = .318; CFI = 1.00; RMSEA = .02; SRMR = .05) and strict scalar MI for behavior intentions (χ^2_{corr} = 43.54; df = 29; p = .041; CFI = .98; RMSEA = .05; SRMR = .04). Both age cohorts were merged for the analysis.

partial metric partial scalar MI was given at all three occasions of measurement for the behavior intention variable (pretest: χ^2_{corr}[5] = 2.29,*df* = 4, *p* = .682, CFI = 1.00, RMSEA = .00, SRMR = .07[6]; posttest: χ^2_{corr} = 5.59, *df* = 4, *p* = .232, CFI = .99, RMSEA = .06, SRMR = .09; follow-up test: χ^2_{corr} = 3.82, *df* = 5, *p* = .575, CFI = 1.00, RMSEA = .00, SRMR = .04).[7] Thus, no deviating understanding, for example, based on maturation, was found for the behavior intention variable within the two Arab Israeli age cohorts. Results indicated that there are no differences in the understanding of the attitude and behavior intention variable across age cohorts in both ethnicities.

3.2. Measurement Noninvariance Analysis

Due to the fact that longitudinal MI was not found within the attitude variable in the Arab Israeli 4th grade cohort questioned three times, we examined measurement noninvariance findings in more detail.

Arab Israeli 4th grade students questioned three times. As mentioned, no longitudinal MI was achieved for the attitude variable. Due to this fact, we explored the different steps of the measurement invariance analysis to enable a deeper understanding of the occurrence of measurement noninvariance over time (i.e., response shifts). When looking at meaningful influences that might have caused a response shift, literature mentions three explanations are found in literature (see Chapter I.3.3.): maturation, external events, or participation in an intervention (Millsap & Hartog, 1988).

Maturation. We ruled out maturation as a possible explanation, because no transition between two developmental stages are postulated for 4th grade students (10-11 years-old), for example, in Kohlberg's stages of moral development (Garz, 2006). Thus, we did not assume that the understanding of the attitude variable changed within the five months period (October 2012 to February 2013), in which students were questioned. Besides, no measurement noninvariance was found within the Jewish Israeli 4th grade cohort for the attitude variable, indicating no changes according to developmental changes.

---

5 Due to the use of a MLR the χ^2 value was corrected (corr; Muthén & Muthén, 1998-2012).

6 Model fit values were presumed as satisfactory in case χ^2 *p*-value ≥ .01, CFI ≥ .95, RMSEA ≤ .08, and SRMR ≤ .10 (Schermelleh-Engel, Moosbrugger, & Müller, 2003, p. 52). Partial measurement invariance was given if at least two factor loading and two intercepts remained constrained (Steenkamp & Baumgartner, 1998).

7 Longitudinal strict scalar MI was also obtained (χ^2_{corr} = 28.79; *df* = 29; *p* = .476; CFI = 1.00; RMSEA = .00; SRMR = .05). In other words, MI was discovered in the course of time (from October 2012 until February 2013). Thereby, both age cohorts were treated as one data set.

External event. As mentioned above (Chapter III.1), an external incident occurred in between the pre- and the posttest. The Israel-Gaza Conflict 2012 including "a series of rocket attacks from both sides" (Gaza and Israel; Larsen, 2012) took place from November 14th-21st, 2012. In case the external incident would have been the reason for the revealed measurement noninvariance, cross-group MI between IG and CG would be given at each point in time, because the external impact would have influenced IG and CG equally. However, no longitudinal MI would be given, because the understanding of the construct would have differed between the pre- and the posttest.

Participation in an intervention. As found in previous research, participation in an intervention could also have led IG students to understand items within the questionnaire at the posttest in a different way than at the pretest (e.g., Fokkema et al., 2013). In case a response shift occurred due to IG student's participation cross-group MI between IG and CG would be given at the pretest, but not at the posttest (after the IG had participated in the intergroup meetings). No longitudinal MI would be found.

In order to investigate whether one of these explanations fitted to the data, we analyzed longitudinal and cross-group MI for the attitude variable in the Arab Israeli 4th grade sample. As illustrated in Manuscript #1, in the Arab Israeli 4th grade cohort $N = 108$ students were surveyed: 60 IG students (52% females) and 48 CG students (46% females).[8] The attitude variable was measured by three items based on an adaption of scales used by Lanphen (2011): "Arab Israeli students are … nice; ... friendly; … kind." Each item could be answered on a five-point scale (1 = *not very much* to 5 = *very much*). Thus, higher scores indicated attitudes that are more positive. The reliability of the variable was satisfactory at each occasion of measurement ($\omega_{t1} = .73$, $\omega_{t2} = .76$, $\omega_{t3} = .79$). As mentioned above, we discovered that longitudinal partial metric partial scalar MI across all three points in time was not achieved within the attitude variable (baseline model: $\chi^2_{corr} = 27.14$, $df = 15$, $p = .028$, CFI = .93, RMSEA = .09, SRMR = .08). In addition to the low CFI value, there were problems with one residual variance, which could not be solved without changing the overall data structure. To test the possible explanations of measurement noninvariance ex post facto, cross-group MI calculations at each occasion of measurement and longitudinal MI analysis including computations with only two occasions of measurement were conducted. Data showed at least cross-group partial metric partial scalar MI at each point of

[8] Due to the small sample size ($N = 30$), measurement noninvariance analysis were not conducted for Arab Israeli 4th grade students, who were questioned five times.

measurement (pretest: χ^2_{corr} = 2.47, *df* = 4, *p* = .650, CFI = 1.00, RMSEA = .00, SRMR = .08; posttest: χ^2_{corr} = 3.36, *df* = 6, *p* = .762, CFI = 1.00, RMSEA = .00, SRMR = .10, follow-up test: χ^2_{corr} = 3.63, *df* = 7, *p* = .821, CFI = 1.00, RMSEA = .00, SRMR = .07). Moreover, data revealed no longitudinal MI between the pre- and posttest (metric model: χ^2_{corr} = 15.18, *df* = 7, *p* = .034, CFI = .91, RMSEA = .11, SRMR = .09)[9] and the post- and follow-up test (baseline model: χ^2_{corr} = 18.16, *df* = 6, *p* = .006, CFI = .89, RMSEA = .14, SRMR = .08). However, a strict scalar MI between the pre- and the follow-up test was achieved (χ^2_{corr} = 11.39, *df* = 9, *p* = .250, CFI = .98, RMSEA = .05, SRMR = .06).

Looking at the findings, no longitudinal MI across all three points in time and no longitudinal MI between pre- and posttest was achieved. However, cross-group MI between the IG and CG occurred at each point in time. Thus, findings indicated that the participation in the intergroup intervention could be excluded as a possible explanation for the response shift, because IG and CG students did not differ in their understanding of the attitude variable at the posttest. Instead, the found measurement noninvariance might have been related to the external incident (e.g., Israel-Gaza Conflict 2012), which occurred between the pre- and the posttest measurement and which affected IG and CG students equally. According to the fact that not even the baseline model of the longitudinal MI analysis across all three points in time presented a satisfactory model fit, data indicate that a reconceptualization occurred. A reconceptualization means that students' definition of the measured construct (i.e., attitudes toward the outgroup) was different between pre- and posttest. Thus, a complete change in the understanding of the underlying construct occurred.[10] Looking at research literature, there is one article that discovered measurement noninvariance (response shifts) based on the experience of a traumatic event (Lommen, van de Schoot, & Engelhard, 2014). There are further aspects, which could explain the found measurement noninvariance. One possible explanation could be the appearance of methodical problems, more explicitly communicational problems (usually referred to as method bias in cross-group MI analysis). However, due to the findings described above, communication problems seem not a possible explanation, because IG students were questioned by different instructors than CG students. Thus, communication problems would

[9] Regarding the longitudinal MI analysis, data from IG and CG were merged.

[10] In general, no assumptions about the longitudinal MI between the post-and the follow-up test existed. However, findings showed that the changed understanding at the posttest was not long lasting but returned to the original understanding until the follow-up measurement. This indicated that all students understood the measured constructs in the pre- and follow-up questionings equally, but differently at the posttest.

have appeared only in one group and as shown in cross-group MI results, IG and CG students understood the variable equally at each point of measurement. However, other reasons, which might explain the occurrence of the response shift could not be ruled out (e.g., interaction effects: Variable x Time x Age).

To complete the measurement noninvariance analysis within the attitude variable, we also investigated longitudinal measurement noninvariance findings within two further samples: first, age differences between Arab Israeli 4th and 6th grade students; second, ethnic differences related to Arab and Jewish 4th grade students. Results referring to the behavior intention variable were already presented before (Arab Israeli 4th and 6th grade students: Chapter III.3.1; Arab Israeli and Jewish Israeli 4th grade students: Manuscript #1).

Age differences with regard to Arab Israeli 4th and 6th grade students. Within the data set, no longitudinal MI for the attitude variable was found (baseline model: χ^2_{corr} = 38.55, *df* = 15, *p* = .001, CFI = .94, RMSEA = .08, SRMR = .05). Thus, within the baseline model a low CFI was given. Additionally, the χ^2-difference test indicated that the baseline model was different to the metric model; the provided modification indices did not fit to the overall data structure and could not be applied. Cross-group partial metric partial scalar MI was found at the pretest (χ^2_{corr} = 2.13, *df* = 3, *p* = .546, CFI = 1.00, RMSEA = .00, SRMR = .05). No cross-group MI was given at the posttest (baseline model: χ^2_{corr} = 9.92, *df* = 2, *p* = .007, CFI = .90, RMSEA = .20, SRMR = .06) and the follow-up test (baseline model: χ^2_{corr} = 7.95, *df* = 2, *p* = .019, CFI = .94, RMSEA = .17, SRMR = .04).[11] Thus, no clear results were found.[12]

Ethnic differences in regard to Arab Israeli and Jewish Israeli 4th grade students. Data revealed no longitudinal MI. On the one hand the residual variance of one indicator could not be fixed adequately; on the other hand, the χ^2-difference test illustrated differences between the baseline and the metric model. Because the provided modification indices did not fit to the overall data structure, they were not implemented (χ^2_{corr} = 22.67, *df* = 15, *p* = .091, CFI = .98, RMSEA = .05, SRMR = .05). Cross-group MI between the two ethnicities was found at the pre- and posttest (pretest: χ^2_{corr} = 4.58, *df* = 4, *p* = 333, CFI = .99, RMSEA = .04, SRMR = .06; posttest: χ^2_{corr} = 5.30, *df* = 6, *p* = .506, CFI = 1.00, RMSEA = .00, SRMR = .10). No cross-group MI was discovered at the follow-up test

[11] Both age cohorts were merged for the longitudinal MI calculations.

[12] Due to the focus of this thesis in regard to sustainable effects, cross-group measurement noninvariance findings were not explored in more detail.

(partial metric partial scalar model: $\chi^2_{corr} = 10.23$, $df = 5$, $p = .069$, CFI = .96, RMSEA = .11, SRMR = .12).[13]

3.3. Overview of all Measurement Invariance and Noninvariance Analysis

In Table 8, the results of all MI analyses that were conducted within this thesis are presented. Due to the complex structure of the table, we shortly introduce its components. MI analyses were examined for three different longitudinal data sets: 1. Data sets from one ethnicity separated according to their age group (i. Jewish Israeli 4th grade students ii. Jewish Israeli 6th grade students iii. Arab Israeli 4th grade students iv. Arab Israeli 6th grade students) 2. Data sets from one ethnicity including students from different age groups (i. Jewish Israeli 4th and 6th grade students ii. Arab Israeli 4th and 6th grade students) 3. Data sets with both ethnicities restricted to one age group (1. Jewish Israeli and Arab Israeli 4th grade students 2. Jewish Israeli and Arab Israeli 6th grade students). Respectively, cross-group MI was related to three different aspects: A. Data sets that examined students from one ethnicity (data sets related to 1.) analyzed cross-group differences between the conditions (IG, CG). B. Data sets including students from one ethnicity coming from different age groups (data sets related to 2.) investigated an equal understanding of the underlying constructs with regard to age groups (4th and 6th grade students). C. Data sets including students from both ethnic backgrounds and one age group (data sets related to 3.) tested different understandings of the constructs regarding ethnicity (Jewish Israeli and Arab Israeli students). As before, longitudinal and cross-group MI were examined for both dependent variables separately.

[13] Again, according to our focus on sustainable effects, cross-group measurement noninvariance findings were not explored in more detail.

Table 8: Overview of Measurement Invariance and Noninvariance Analyses

Cohort	Attitudes				Behavior Intentions			
	Jews 4th grade	Jews 6th grade	Arabs 4th grade	Arabs 6th grade	Jews 4th grade	Jews 6th grade	Arabs 4th grade	Arabs 6th grade
Jews 4th grade	✓	✓	x[a]	-	✓	✓	✓	-
Jews 6th grade		✓	-	✓		✓[b]	-	✓
Arabs 4th grade			x	x[c]			✓[d]	✓
Arabs 6th grade				✓[e]				✓

Note. The calculations that this table is based on are displayed within Manuscript #1, Manuscript #2, or within the ancillary analysis of the synopsis (Chapter III.3.). The following symbols display results about longitudinal MI: ✓ = longitudinal partial metric partial scalar MI was achieved; x = longitudinal MI was not achieved; - = longitudinal MI was not investigated, due to the fact that no relevant research question was given; cross-group MI was investigated whenever longitudinal MI was analyzed at all three occasions of measurement; in case it was not achieved at one occasion of measurement it was explicitly referred to with superscript letters. [a] No cross-group MI between Jewish and Arab Israeli 4th grade students at t3 (ethnicity). [b] No cross-group MI between IG and CG at t3 (condition). [c] No cross-group MI between Arab Israeli 4th and 6th grade students at t2 and t3 (age). [d] No cross-group MI between IG and CG at t2 (condition). [e] No cross-group MI between IG and CG at t1 (condition).

As can be seen in Table 8 longitudinal MI was given at all occasions of measurement except in cases in which the attitude variable of the Arab Israeli 4th grade cohort was involved. With the exception of this variable in the Arab Israeli 4th grade cohort, no differences within the understanding of the underlying constructs across time were found for data sets including different conditions, age groups, or ethnicities. No cross-group MI was detected in six cases (1. Jewish and Arab Israeli 4th grade students at t3 within the attitude variable 2. IG and CG at t3 for Jewish Israeli 6th grade within the behavior intention variable 3+4. Arab Israeli 4th and 6th grade students at t2 and t3 within the attitude variable 5. IG and CG at t2 for Arabs 4th grade students within the behavior intention variable 6. IG and CG at t1 for Arabs 6th grade students within the attitude variable).

4. Multiple Imputation as Alternative Missing Data Technique

As described above (Chapter I.3.5.) state of the art to handle missing values is either to calculate data using maximum likelihood estimation (ML) or to impute and analyze data via multiple imputations (MIM; Baraldi & Enders, 2010; Enders, 2013; Feng et al., 2012; Schafer & Graham, 2002). ML was applied in all four samples that were analyzed using latent change modeling. Due to the small sample sizes of the two samples that were questioned five times, we calculated data on the basis of manifest variables and missing data with the MIM technique. Although, only minor differences within the two missing data techniques are expected in case no auxiliary variables are included, missing data within the four larger samples were also calculated using MIM (Peugh & Enders, 2004). Due to the fact that both types of missing data techniques are possible and no empirical experiences and comparisons concerning missing data techniques within contact intervention literature exist, we were interested in ascertaining that the results are stable even when using the MIM missing data technique. The analysis of short-, long-term, and fading effects within the latent change models were additionally conducted using MIM in the four larger samples. For MIM calculations, ten data sets were generated (Collins et al., 2001). In M*plus* "parameter estimates are averaged over the set of analyses, and standard errors are computed using the average of the standard errors over the set of analyses and the between analysis parameter estimate variation" (Muthén & Muthén, 1998-2012, p. 388). Table 9 presents the results for both missing data techniques.

Table 9: Comparison of the Study Results Regarding two Missing Data Techniques

			Latent Means t1	t2	t3	Diff1-condition β	Diff1-condition *b* (*SE*)	Diff2-condition β	Diff2-condition *b* (*SE*)	Diff3-condition β	Diff3-condition *b* (*SE*)
								Attitudes			
Jews 4th	ML	IG	3.07	4.01	3.66	.59***	.49*** (.08)	.23*	.25* (.10)	-.24**	-.24* (.10)
		CG	3.15	3.03	3.15						
	MIM	IG	3.11	3.90	3.61	.43***	.36*** (.09)	.20*	.22* (.09)	-.21	-.14 (.10)
		CG	3.14	3.18	3.17						
								Behavior Intentions			
	ML	IG	3.19	2.99	2.69	.11	.10 (.13)	.02	.02 (.14)	-.09	-.08 (.13)
		CG	2.83	2.78	2.65						
	MIM	IG	3.17	2.95	2.70	.10	-.09 (.11)	.02	.03 (.13)	-.07	-.06 (.11)
		CG	2.84	2.76	2.65						
								Behavior Intentions[a]			
Arabs 4th	ML	IG	2.47	3.62	3.37	.19**	.36** (.14)	.11	.20 (.12)	-.13	-.17 (.17)
		CG	2.70	2.89	2.98						
	MIM	IG	2.53	3.47	3.36	.12†	.24† (.13)	.09	.16 (.12)	-.07	-.08 (.16)
		CG	2.73	3.00	3.05						
								Attitudes			
Jews 6th	ML	IG	3.51	2.33	2.43	-.69**	-.62*** (.17)	-.63**	-.53** (.16)	.14	.09 (.18)
		CG	2.89	2.95	2.96						
	MIM[b]	IG	3.47	2.40	2.48	-.59**	-.53** (.17)	-.55**	-.46** (.16)	.11	.07 (.18)
		CG	2.89	2.93	2.94						
								Behavior Intentions			
	ML	IG	2.76	1.70	1.63	-1.00[c]***	-.91*** (.14)	-1.05[c]***	-.98*** (.15)	-.11	-.07 (.16)
		CG	2.57	2.60	2.61						
	MIM	IG	2.69	1.75	1.66	-.90***	-.81*** (.14)	-.95 ***	-.89*** (.16)	-.13	-.09 (.16)
		CG	2.58	2.55	2.56						
								Attitudes			
Arabs 6th	ML	IG	3.59	3.53	3.46	.40†	. 39† (.20)	.31	.31 (.20)	-.08	-.08 (.23)
		CG	3.49	3.14	3.15						
	MIM[d]	IG	3.60	3.49	3.44	.34	.32 (.21)	.28	.27 (.20)	-.05	-.05 (.25)
		CG	3.49	3.17	3.17						
								Behavior Intentions			
	ML	IG	3.12	3.63	3.10	.60 **	.63** (.21)	.11	.11 (.20)	-.58*	-.53* (.21)
		CG	3.16	3.00	2.99						
	MIM	IG	3.11	3.56	3.14	.61**	.64** (.21)	.15	.15 (.21)	-.54*	-.49* (.24)
		CG	3.18	2.91	2.99						

Note. Jewish Israeli 4th grade students: $N = 113$ for positive attitudes and $N = 112$ for positive behavior intentions; IG ($n = 60$); CG ($n_{att} = 53$; $n_{beh} = 52$); Arab Israeli 4th grade students: $N = 108$; IG ($n = 60$); CG ($n = 48$); Jewish Israeli 6th grade students: $N = 137$; IG ($n = 83$); CG ($n = 54$); Arab Israeli 6th grade students: $N = 139$; IG ($n = 83$); CG ($n = 56$); latent means = values between 1 and 5 with 1 signifying lowest and 5 highest agreement; t1 = pretest (one week before the first intergroup meeting); t2 = posttest (exactly at the end of all intergroup meetings); t3 = follow-up (depending on the samples about nine to eleven weeks after the end of all intergroup meetings); Diff1 = latent difference score between t1 and t2; Diff2 = latent difference score between t1 and t3; Diff3 = latent difference score between t2 and t3.
[a] Only results for positive behavior intentions are displayed because no longitudinal measurement invariance was given for positive attitudes. [b] When calculating the latent difference scores using MIM, the model fit indices were satisfactory except the CFI, which was ≥ .93. [c] The standardized coefficient β can be greater 1 according to the fact that Diff scores and dependent variables at t1 can correlate within the latent change model (Deegan, 1978).[d] When calculating the latent difference scores using MIM, the model fit indices were satisfactory except the CFI, which was ≥ .88.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

As shown in Table 9, small differences in the results were found when using the two missing data techniques. Within the Jewish Israeli 4th grade cohort, results were similar. However, the effect of the intervention on Diff3 (fading effect between t2 and t3) concerning attitudes was not significant when using MIM calculations (β = -.21, $p = .111$), as opposed to ML calculations (β = -.24, $p = .002$). Similar differences occurred in the respective *b* coefficients. There were only minor differences in the results of the Arab Israeli 4th grade cohort. As illustrated the effect of the intervention on Diff1 (short-term effect that was measured directly after the end of the four intergroup meetings) within the behavior intention variable was significant when using ML (β = .19, $p = .008$) whereas it was only marginally significant when using MIM (β = .12, $p = .078$). The same differences were given within the *b* coefficients. In the Jewish Israeli 6th grade cohort only the magnitude regarding the coefficient *b* of Diff1 was deviating showing the effect of the intervention on the attitude variable. Finally, the results of the Arab Israeli 6th grade cohort showed that the marginally significant effect of the intervention on Diff1 within the attitude variable was given when calculating data via ML (β = .40, $p = .056$) but not when using the MIM technique (β = .34, $p = .128$).

Overall, the comparison of the two missing data techniques demonstrated that outcomes using MIM were most of the times slightly smaller (more conservative) compared to results revealed when ML was used. The structure of the data remained similar.

5. Analysis of Gender Differences

Main effects of gender on prejudice have already been discovered by other authors, illustrating that females have less explicit prejudice than males (Ekehammar, Akrami, & Araya, 2003; Hoxter & Lester, 1994). However, within their meta-analysis Pettigrew and Tropp (2006) did not find any differences on the contact-prejudice relation regarding participants' gender ($Q_B[1] = 0.70$, $p = .40$). Therefore, no moderating effects of gender were assumed ahead of the analysis. Nevertheless, Pettigrew and Tropp stated, "participants' gender proves to be a minor factor in interpreting contact-prejudice effects" (2006, p. 764). In their data female participants showed a larger contact-prejudice effect ($r = -.21$, CI [-.26, -.17], $k = 63$ samples) than males ($r = -.19$, CI [-.23, -.14], $k = 59$ samples).

According to the fact that the occurrence of a gender effect could not be ruled out for intractable conflict contexts it was explored whether gender was related to the contact intervention-prejudice effect over time ex post facto. To test gender differences the variable *Gender* (male = 1; female = 2) and the interaction variable *Gender x Condition* were included in the latent change models to explain the three latent difference scores Diff1 (pretest-posttest), Diff2 (pretest-follow-up test), and Diff3 (posttest-follow-up test). For Jewish Israeli 4th grade students no gender effects with regard to both dependent variables were discovered. In addition, no gender effects were discovered within the behavior intention variable in the Arab Israeli 4th grade sample.[14] Within the Jewish Israeli 6th grade sample an effect with regard to Diff2 (pretest-follow-up test) was discovered indicating that females had a less negative attitude change than males from the pretest to the follow-up test ($\beta = .15$, $p = .050$). The model fit of the LCM for attitudes ($\chi^2_{corr} = 98.82$; $df = 69$; $p = .011$; CFI = .95; RMSEA = .06; SRMR = .08) remained satisfactory even after the inclusion of the two variables. Also, group differences (IG; CG) remained for short-term (Diff1: $\beta = -.33$, $p = .001$) and long-term contact intervention effects (Diff2: $\beta = -.29$, $p = .003$). In the Arab Israeli 6th grade sample, no gender effects were discovered.[15]

In sum, negative contact led to a weaker decrease of long-term attitudes for females in comparison to males between the pre- and the follow-up test in the Jewish Israeli 6th grade sample. No overall gender effect appeared across all four samples.

[14] Due to the missing longitudinal measurement invariance in the attitude variable we tested gender effects only for the behavior intention variable.

[15] According to our small sample size in the Jewish and Arab Israeli 4th grade cohorts that were questioned five times no gender effects were analyzed in these two samples.

IV. CONCLUSION AND DISCUSSION

1. Conclusion

This thesis focused on face-to-face contact interventions with repeated intergroup meetings in intractable conflicts. Intergroup contact interventions were chosen as a mean to achieve improved intergroup relations, that is attitudes and behavior intentions toward the outgroup. The focus was on intractable conflicts to show the effectiveness of intergroup contact interventions in challenging situations. The following aspects have been explored in particular: systematic investigation of contact intervention effects over time (i.e., short- and long-term, as well as fading effects of intergroup contact interventions), characteristics of interventions and their participants that influence intergroup contact effects (e.g., group status, repetition of intergroup meetings, valence of reported personal experience within the contact interaction), and state of the art methodical approaches to analyze quasi-experimental data (i.e., latent variable modeling, longitudinal and cross-group measurement invariance analysis, control for hierarchical effects, missing data analysis). In order to investigate the above-mentioned aspects, six samples were examined.[1] Each sample was analyzed regarding two dependent variables –attitudes and behavior intentions. The intervention groups of these samples took part in two distinct intergroup contact interventions in Israel. The combined results of all samples are described in the following paragraphs.

Short-, long-term, and fading effects. The research question as to whether repeated intergroup contact meetings lead to short-term attitude and behavior intention changes can be affirmed in almost all samples and variables: Short-term contact intervention effects were revealed within all four samples regarding positive attitudes[2] and in four out of six samples in regard to positive behavior intentions. The research question regarding repeated contact meetings leading to long-term attitude and behavior intention changes is supported only partly: long-term effects were found within three samples investigating positive attitudes and within one sample examining behavior intentions. As mentioned, in contrast to our expectations, in one sample, negative short- and long-term effects in attitude and behavior changes appeared. We also investigated whether fading effects (i.e., the decrease

[1] Jewish Israeli 4th grade students – questioned three times, Jewish Israeli 4th grade students – questioned five times, Arab Israeli 4th grade students – questioned three times, Arab Israeli 4th grade students – questioned five times, Jewish Israeli 6th grade students – questioned three times, and Arab Israeli 6th grade students – questioned three times (see also Table 4).

[2] Two out of six samples were not analyzed regarding attitudes due to missing longitudinal measurement invariance in the attitude variable.

of positive attitudes or positive behavior by the time of the follow-up measurement) existed. Fading effects were discovered in two samples concerning attitudes and in two samples with regard to behavior intentions.

Status differences. We were also interested in the occurrence of status differences in the findings. Therefore, we tested whether the Ethnicity x Group interaction had a significant impact on attitude and behavior intention changes between pre- and posttest as well as pre- and follow-up test. Status differences were found for the 6th grade cohorts referring to short- and long-term effects (Manuscript #2).[3] Upon a closer look at minority and majority group's results, no long-term contact intervention effects were revealed within any of the minority sample (Arab Israeli participants). Additionally, it was observed that the two status groups assessed the exact same intervention differently: The Arab Israeli 6th grade cohort experienced the intervention positively whereas the Jewish Israeli 6th grade cohort evaluated it partly negatively. This finding underlines that there are different perceptions of the contact interventions by both groups. A further difference between the status groups was their reaction to the Israel-Gaza Conflict 2012, the data from two Israeli surveys showed Jews being more and Arabs less pessimistic toward the conflict after the external event. Thus, majority and minority differed concerning the effects of the intergroup contact intervention, in their perception of the intervention as well as in their reaction to the external event.

The role of repetition. In accordance with previous research, we were interested in learning whether the repetition of contact enables prolonged attitude and behavior intention changes. Therefore, we investigated the role of repeated contact within two samples more specifically. Concerning the attitude variable in the majority sample data showed that participation in more contact meetings led to stronger short-term findings and to positive long-term effects. Four intergroup meetings led to stronger short-term positive behavior intentions than a single one. In the minority sample, no effects of the repetition were found; nevertheless, means pointed in the expected direction. In conclusion, repeated contact interventions were relevant especially in the majority sample.

The role of the reported personal experience. We were also interested in the valence of students' personal experience, which was reported at the end of the contact intervention (posttest). We explored whether the reported experiences was a predictor for short- and

[3] Status differences in regard to short- and long-term effects were also calculated for 4th grade students (Manuscript #1; Study 1). No differences were discovered with regard to the behavior intention variable. Again, the attitude variable was not examined due to missing longitudinal measurement invariance.

long-term findings; that is, whether or not positive experiences were associated with improved attitudes and behavior intentions. We discovered that the valence of students' intergroup experience was related to short-term effects on attitudes in three out of four samples and short-term effects on behavior intentions in four out of six samples. Long-term effects were revealed in three out of four samples with respect to attitude changes[4] and in two out of six samples within the behavior intentions variable. Thus, results are inconclusive regarding long-term effects. However, findings suggest that reported intergroup experience may be a good indicator of contact intervention effects for the majority sample as most of the effects were discovered within the this status group.

Methodical aspects. Latent variable modeling was applied in all samples that were questioned three times; the two remaining samples were investigated using manifest variables due to their small sample size. The hypotheses were tested with two latent change models (i.e., specific types of structural equation models), which were extended to analyze short- and long-term as well as fading effects for quasi-experimental data. Thereby, advantages that are linked to latent variable modeling (e.g., inclusion of measurement error, illustration of model fit indices, measurement invariance analysis) were integrated in the analysis. Longitudinal measurement invariance (MI) was computed for each sample, for ethnicity including both age groups, and age groups including both ethnicities: All of the analyses were conducted for both dependent variables. Only one sample violated the assumption of measurement equivalence.[5] Cross-group MI was calculated concerning the experimental condition (IG, CG), ethnicity (Jewish, Arab), as well as age group (4th grade, 6th grade) at each of the three points of measurement (pre-, post-, and follow-up-test). The findings revealed invariance between IG and CG, the two ethnicities, as well as the two age groups was ascertained within the results. Thus, groups understood the underlying constructs equally across time. In only one sample, no longitudinal MI was found for the attitude variable and in four samples, no cross-group MI was discovered (see Table 8). In the analyses, we were able to show that a more detailed investigation of cases showing measurement noninvariance results can lead to a deepened understanding of the data. For example, external events might influence participant's understanding of the measured constructs. To control for hierarchical effects within the latent change models we applied an appropriate method. Thus, we included variables for each level-2-unit in our model to

[4] Lesser samples were examined in regard to attitudes due to the missing longitudinal measurement invariance in the Arab Israeli 4th grade cohort.

[5] Samples that were combined with this sample did also lack longitudinal MI (see Chapter III.3.3.).

implement a fixed effect approach. Finally, we illustrated that two missing data techniques (maximum likelihood estimation and multiple imputation) lead to similar results, meaning the direction of the results are equal between both techniques. Nevertheless, in general the multiple imputation technique showed slightly smaller effects than the maximum likelihood method, meaning it can be regarded as the more conservative estimation in this thesis.

Conflict context. Due to the fact that the Israel-Gaza Conflict 2012 occurred during the data collection, we also investigated the impact of this conflict using external data and found that the Israeli population showed different reactions depending on their status. Thus, Jewish Israeli increased in their strong opposition on holding peace negotiations and increased in their agreement to the item "don't believe at all" that negotiations lead to peace. In contrast, Arab Israeli showed the opposite reactions, they decreased in their strong opposition on holding peace negotiations; for the item stating that negotiations lead to peace the agreement to "don't believe at all" also decreased.

Gender differences. Although no gender effects were assumed ahead of the analyses, we explored whether gender had an impact on the results. No general gender differences were discovered. It has to be considered that the power within these analyses was low.[6]

2. Discussion

According to a representative public opinion survey, by now almost 16% of the Jewish Israeli and 6% of the Arab Israeli population have already taken part in an intergroup encounter. Given these rates and the conceivable potential of intergroup contact interventions in intractable conflict areas, a meaningful investigation of these interventions is necessary (Maoz, 2011).

Short-, long-term, and fading effects. The short-term changes found in this thesis corresponded with the results shown in the meta-analysis by Lemmer and Wagner (2015). Within conflict areas, short-term changes can be expected after participation in an intergroup contact intervention. Concerning long-term changes after intergroup contact in a conflict area, we only discovered effects for the majority. As to the minority, no clear conclusion can be drawn. On the one hand, Jayusi (2009) also discovered no long-term effects for minority samples; on the other hand, Lemmer and Wagner (2015) listed some

[6] We found effects only within the Jewish Israeli 6th grade sample indicating females had less negative changes within their attitudes between the pre- and the posttest than males.

long-term effects for minority samples in conflict areas in their meta-analysis (e.g., Kropiunigg & Pabst, 2007). These contradicting findings indicate that until now there has been no conclusive understanding under which conditions long-term effects occur for the minority.

We also revealed negative long-term contact intervention effects for the majority. As shown in the analysis, this negative outcome was associated with negative personal experiences. The negative contact situation existed only present in 6th grade and not in 4th grade. Therefore, it might have been related to the age of the students, who were in a developmental phase where bullying occurs more often (e.g., Aboud & Brown, 2013). The negative contact experience might also have been associated with the external event (Israel-Gaza Conflict 2012), which occurred during the intervention. In previous contact literature negative short-term effects have already been revealed, negative long-term contact intervention effects on the other hand have not been examined systematically until today (e.g., Graf et al., 2014).

Fading-effects existed in three samples that showed positive short-term effects. This can be seen as an affirmation of previous findings, which indicated already that intervention effects are not sustainable in intractable conflict areas (e.g., Shani, 2015). In conclusion, we assume that positive short-term contact intervention effects decrease most often after the end of the intervention in intractable conflict areas. According to our findings, the time span for these fading effects is associated with different aspects, for example, characteristics of the intervention (e.g., repetition of the intergroup meetings) and status (minority or majority) of the participants.

Explanations for the deviating outcomes for attitude and behavior intentions might be related to student's age. As Farhan (2008) stated, "consistencies between attitudes and behavior are less common in children because of a higher susceptibility for concrete situational aspects and the lower sophistication of cognitive and attitudinal systems" (p. 16). Another explanation might be that minority and majority groups differed in their sensitivity to the dependent variables.

Status differences. First, one has to acknowledge that in this thesis status differences mean a conglomerate of the minority-majority proportion in terms of the number of people living in the population, as well as ethnical and cultural differences. Due to the intermingling of these influences, differences could only be examined for all of these aspects. We revealed long-term contact intervention effects solely for majority group

members. Repeated participation in intergroup meetings enabled long-term effects, again only for majority group members. In contrast, the valence of students' personal experience as either positive, neutral, or negative was relevant for both status groups in regard to their short- and long-term changes; however, in the minority group solely the behavior intention variable was affected. This indicates that for minority members behavior intentions might be the more sensitive dependent variable than the attitude variable. This aspect has also been described in past literature, showing that for minority members intergroup contact might affect distance rather than outgroup evaluation (Bastian et al., 2012). As mentioned in the introduction, differences in contact intervention effects might suggest that minority and majority members follow different needs (Salomon, 2011; Shnabel & Nadler, 2008; Tropp & Pettigrew, 2005). A further explanation from the theoretical and empirical background is a possible ceiling effect for the minority, which implies that at the time of the pretest minority members already had a more positive score on the dependent variables so that they could not improve as much as the majority members. In the data of this thesis, Arab Israeli students did in fact show higher positive pretest-values. Except for one sample, on average these values were not above 3.5 on a 5 point scale, so that we do not assume the ceiling effect to explain the data. Recognizing that the exact same contact intervention can be regarded as positive for the Arab Israeli and as negative for the Jewish Israeli students, we assume that change for both groups is even less related to each other than has been expected until now.

The role of repetition. Lemmer and Wagner (2015) discovered in their meta-analysis that the duration of interventions had no impact on short-term outcomes; data however indicate that the structure of the intervention has an impact on the sustainability of contact effects in intractable conflict areas. Thus, our results as well as previous findings illustrate that at least for majority members contact repetition in conflict areas has an impact (e.g., Jayusi, 2009). Previous findings indicated that an additional intervention, which is different from the first intervention, at some point after the end of the first intervention supports the prolongation of the contact intervention effects, whereas we were able to show that also one intervention, which consists of several meetings, facilitates long-term contact intervention effects. We can only speculate that the impact of these repetitions might follow similar rules as learning processes within operant conditioning: "the shorter the intervening interval, the speedier and more marked the conditioning" (e.g., Skinner, 1948, p. 169). The results of this thesis indicate that after repeated negative intergroup contact negative long-

term contact intervention effects existed. However, we cannot prove, whether these negative long-term effects were related to the repetition of negative contact situations, the appearance of one negative contact meeting, or interaction of the intervention with the specific conflict context.

The role of reported personal experience. Reported personal experience seems to be a relevant indicator for long-term effects. However, the relevance differs between minority and majority members. As shown in former research, this might depend on the fact that contact quality and contact quantity has a deviating relevance for majority members. In a longitudinal study Vezzali, Giovannini, and Capozza (2010) discovered that contact quality is more relevant for majority group members than for minority members and that both groups are effected by contact quantity (see also Feddes, Noack, & Rutland, 2009). Given that contact interventions strive to enable a positive contact situation for all participants, most of the time it focuses more on positive contact than on contact quantity; therefore, contact interventions might have a stronger long-term impact on majority members.

Methodical aspects. In the light of the fact that measurement invariance (MI) and measurement noninvariance are infrequently analyzed within contact intervention research, the analyses were explorative and not based on any presuppositions; thus, we illustrated that it is possible to achieve MI between IG and CG, status groups, and age cohorts. However, MI analyses are necessary and have to be confirmed in future work because MI is not certainly given, as can be seen in one of our samples. Our findings only let us speculate why longitudinal MI was not achieved for attitudes but for behavior intentions in one of our samples. One possible explanation is that the positive attitude variable was more sensitive to the external influence than the positive behavior intention variable and therefore a response shift occurred. Findings related to response shifts after interventions in other fields of research (e.g., organizational interventions; Millsap & Hartog, 1988) also indicate that the verification of MI is essential before interpreting latent means over time and across groups. Additionally, the analysis of measurement noninvariance can lead to meaningful insights.

Conflict context. The Israeli Democracy Index as well as The Peace Index indicated that the Israel-Gaza Conflict 2012 was perceived differently within the Jewish Israeli and the Arab Israeli population. These two surveys interviewed adults, which leads to the conclusion that findings can only hint at the influence of the conflict event on the public opinion and should not be freely transferred to students' opinions (age 10 to 13 years). Although only

very few students referred to the specific conflict situation (e.g., “I'm a little scared of Arabs at the moment”), a general impact due to this external incident might be assumed based on the external data.

Critical remarks. The investigation and promotion of intergroup contact interventions has been subject to a critical discussion among scientists in recent years. To embed this thesis in this debate, we hereafter shortly outline its content. The controversy of whether or not intergroup contact achieves its overall goals of improving intergroup relations and leading to a more just society is the focus of the discussion. Two main aspects have been especially emphasized: First, critics have stated that intergroup contact does not necessarily bring majority members to support political policies aiming to reduce social inequalities (e.g., Dixon, Durrheim, & Tredoux, 2005); second, intergroup contact has been found to reduce minority members’ motivation to take action against social inequality (the Reicher effect; Reicher, 2007; see also Dixon, Tropp, Durrheim, & Tredoux, 2010; Tropp, Hawi, van Laar, & Levin, 2012). In their book, Pettigrew and Tropp (2011) extensively respond to these critics by providing evidence that intergroup contact is typically improving “attitudes towards policy changes beneficial for the less-powerful outgroup” (p. 172) within the majority and stating that the distinction between collective action participation and prejudice reduction approaches is too sharply defined: “some contact outcomes further mobilization, whereas others counter it. Mobilization itself will in turn influence intergroup contact – increasing it with outgroup allies and decreasing it with outgroup opponents” (p. 178). Besides, Becker, Wright, Lubensky, and Zhou (2013) discovered that collective action of minority members could be retained when majority members communicate that they believe status differences between the groups to be illegitimate. In the light of these results, the reduction of prejudice through intergroup contact is not mutually exclusive to minorities’ collective action tendencies to end societal inequalities. Therefore, one can regard intergroup contact intervention research as capable of improving intergroup relations and leading to a more just society.

3. Limitations

The major limitation of this thesis is the implementation of a quasi-experimental instead of an experimental design (i.e., missing randomization of IG and CG) so that causal inferences are limited. We used this design out of practical reasons that is we investigated an established intergroup program and did not implement a new program ourselves. This

procedure was chosen due to missing relations to schools that are located in conflict areas, limited capacities, and language barriers. Therefore, it was not possible to randomize IG and CG. We are aware that a randomization in general is very difficult to implement in conflict areas (e.g., little overall interest in contact with the outgroup) and that many other studies did not even include a non-randomized CG in their research studies (see Chapter I.2.2). Nevertheless, we regard the quasi-experimental design as our major limitation. Based on this design, results have to be assessed with some caution in order to enable a more accurate view on the threats to the internal validity of the results. We discussed to which extent they apply to this research within the respective manuscripts.

A further relevant limitation is related to the external validity of the results. With this thesis, we extend the external validity of contact intervention research in intractable conflicts regarding different aspects. Effects from previous studies could be confirmed with regard to other units (i.e., younger students)[7], other treatments (i.e., another intergroup contact intervention), and different outcome operationalization (i.e., dependent variables were not exactly the same). Nevertheless, one main external validity limitation remains: It is uncertain whether our results are transferable to other settings. Although it might be implicitly assumed that different intractable conflicts are alike, (because they consist of the same components; e.g., long lasting, violent, not solvable, existential meaning for the population; Bar-Tal, 1998; Kriesberg, 1998), it is not proven that the situations within various intractable conflicts have a similar impact on contact intervention results. A further similar limitation is due to the missing explicit measurement of the perception of the contextual incident (i.e., Israel-Gaza Conflict 2012). Although a quasi-experimental control group design was used, an explicit measurement of participants' perception of the contextual event would have been more informative and convincing. Especially in intractable conflict areas, where external incidents occur on a regular basis, the inclusion of such a measurement would have been meaningful to gather more information about possible contextual effects.

One other limitation is that the open-ended question ("Did you tell your parents about the meetings? If yes, what did you tell them?") was not explicitly validated ahead of this research. Discussions with a group of local organizers who supervised the school program in Israel and had years of experience in leading these interventions led us to the

[7] We investigated 4th and 6th grade students whereas other interventions predominantly focus on older students (i.e., 10th to 12th grade; Arnon, 2010; Jauysi, 2009; Schroeder & Risen, 2014; Shani, 2015).

inclusion of this question. Nevertheless, the operationalization was newly created and therefore lacks evidence of its construct validity, so we can only suspect that students responded to the question in line with their personal experience and not based on other aspects (e.g., social desirability).

Finally, we followed a relatively strict methodical approach. However, when more knowledge of auxiliary variables had been given, these could have been included in the multiple imputation techniques and more precise analyses of missing data would have been possible. In addition, more participants would have enabled an analysis of all samples with latent variable modeling. Practical reasons inhibited this approach.

4. Implications

The implications of this thesis are straightforward. Intergroup contact interventions improve intergroup relations in the short run even in conflict areas. More sustainable effects are achievable under certain conditions. According to this thesis, repetitions are a meaningful component to prolong contact effects in intractable conflict areas for the majority group. This implies that the intervention should be constructed in a way which enables a continued involvement of students with outgroup members. It is also conceivable that virtual contact follow-ups via social networks (Tavakoli, Hatami, & Thorngate, 2010) could enable a prolongation of contact effects. Rosen and Perkins (2013) name further proposals for continued contact for example mutual projects (i.e., sports), involving family and community, mutual ceremonies, and perspective taking activities. A further consideration would be to use other interventions to prolong intergroup contact effects: for example (indirect) emotional regulation sessions (Halperin, Cohen-Chen, & Goldenberg, 2014). An important factor for the implementation of such projects is the feasibility of the solution since financial resources are often limited. The findings furthermore implicate that the facilitation of more sustainable contact intervention effects can be attained by ensuring a positive intergroup contact experience for participants. Definite suggestions for the minority group are not possible so far. More research needs to be conducted to explore the needs of members of the minority since long-term effects occur exclusively for majority group members. It also seems crucial to avert the occurrence of a negative intergroup contact situation within an intractable conflict area (i.e., a contact situation including anxiety, fear, and physical violence) as this contact can worsen short- and even long-term intergroup relations. To prevent these situations, it is essential to inform and train

instructors, guides, and organizers of intergroup interventions to hinder the development of negative contact situations and intervene if they occur nonetheless.

Theoretically, a focus on longitudinal studies is the strongest implication of this thesis. Given that intervention effects are known to fade in conflict areas, it is crucial to know the processes that cause the reverse changes and thus to enable contact effects having an impact in the long run.

Methodically the thesis illustrates the importance of up-to-date methodical approaches: by using methods such as MI analysis, adequate missing data techniques, and control for hierarchical data we identified results that are more trustworthy (i.e., statistically valid). However, the findings also illustrate that the time intervals within the research design have to be chosen reasonably to achieve meaningful effects (i.e., follow-up measurements should be placed at some time after the intervention when a change in the dependent variable is expected).

5. Future Research

Future research should examine more moderating and mediating variables that influence long-term effects in non-conflict and conflict areas. There are many possibilities to discover meaningful variables. There is a tremendous amount of qualitative research about intergroup interventions in intractable conflict areas. It seems plausible to examine these articles systematically to discover possible moderating or mediating effects for improved long-term relations between opposing groups in intractable conflicts. The subsequent step would be to integrate these established variables into quantitative research. Wagner (personal communication, February 25, 2015) proposed to investigate *communication networks* (direct communication, exchange of views between individuals, and information distributed by mass media) as a variable, which might serve as either a moderating or mediating influence for long-term contact intervention effects. Thus, communication networks could be regarded as a type of continued engagement with intergroup relations and – under ideal conditions – enable a positive debate and a continuation of contact effects. In line with discovering possible moderating and mediating variables, there are also many quantitative studies revealing variables, which are related to positive short-term outcomes after participation in intergroup interventions in (non-)conflict areas. There is little research investigating these variables systematically concerning long-term effects. One of few exceptions in which intergroup contact in conflict areas was

investigated (not in the context of intergroup interventions) is an article by Swart et al. (2011) that revealed a full longitudinal mediation of contact via affective empathy on outgroup attitudes.

A further extension of this research would be to conduct laboratory experimental studies to test variables that are considered to have a meaningful effect on short- and long-term contact intervention effects in conflict areas. Due to the fact that there is yet little knowledge about contextual influences regarding the contact intervention - prejudice relation experiments could manipulate the strength of a conflict (context). Thereby, it could be investigated whether strong conflicts have a specific (moderating) impact on long-term contact effects in comparison to less strong conflicts. Participants could be divided into groups using a minimal-group paradigm (or existing group identifications) and conflict settings could be manipulated with vignettes that describe or use existing contextual situations including specific collective memories, collective emotional orientation, and an ethos of conflict. In these scenarios, short- and long-term intergroup relations could be investigated as outcomes. A further laboratory experiment could focus on moderators that enable long-term contact intervention effects such as the repetition of intergroup contact. Participants could meet face-to-face first and afterwards the contact situation would be refreshed by virtual meetings (e.g., via internet based communication; Yablon & Katz, 2001; see also Ellis & Maoz, 2007).

Moreover, an even greater focus should be placed on state-of-the-art statistical methods; for example, multi-level modeling should be considered more extensively in contact intervention research. In future research latent variable modeling and measurement invariance analysis should be applied whenever possible to ascertain comparability and develop assumptions of latent means across groups and time. A further research area could be to investigate curvilinear effects to ensure a broader knowledge of long-term contact intervention effects. There is hardly any knowledge about the exact development of long-term as well as fading effects after the participation in an intergroup intervention, so that a clear interpretation of the data is difficult (Kline, 2011). If possible, future studies should include multiple follow-up questionings after meaningful time spans.

This thesis began with a quotation by Tropp (2012) “conflicts based in ethnic, religious, and racial differences continue to erupt around the world, despite decades of interventions and scholarly research” (p. 3). Although this thesis cannot solve conflicts, it

contributes to a better understanding of short-term and sustainable effects of intergroup contact interventions in intractable conflict areas and the understanding of methodical research approaches to measure these effects. In conclusion, one can consider it as a small part of a greater whole, which enables future research to continue on a slightly advanced knowledge base about intergroup contact interventions in intractable conflict areas and their effects on intergroup relations.

V. REFERENCES

Aboud, F. E., & Brown, C. S. (2013). Positive and negative intergroup contact among children and its effect on attitudes. In G. Hodson and M. Hewstone (Eds.), *Advances in intergroup contact* (pp. 176–199). Hove, UK: Psychology Press.

Abu-Nimer, M. (2004). Education for coexistence and Arab-Jewish encounters in Israel: Potential and challenges. *Journal of Social Issues, 60*(2), 405–422. doi:10.1111/j.0022-4537.2004.00121.x

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086–1120. doi:10.1016/j.leaqua.2010.10.010

Arnon, M. (2010). *Reflection on the activities of a peace education program as a tool for reinforcing/rehabilitating its long term effects* (Unpublished doctoral dissertation). University of Haifa, Haifa, Israel. [Hebrew]

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495–508. doi:10.1080/10705511.2014.919210

Azar, E. E. (1990). *The management of protracted social conflict: Theory and cases*. Hampshire, UK: Dartmouth Publishing Company Limited.

Azar, E. E., Jureidini, P., & McLaurin, R. (1978). Protracted social conflict: Theory and practice in the Middle East. *Journal of Palestine Studies*, *8*(1), 41–60. doi:10.2307/2536101

Bagozzi, R. P., & Yi, Y. (1989). On the use of structural equation models in experimental designs. *Journal of Marketing Research*, *26*(3), 271–284. doi:10.2307/3172900

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5–37. doi:10.1016/j.jsp.2009.10.001

Barclay, R., & Tate, R. B. (2014). Response shift recalibration and reprioritization in health-related quality of life was identified prospectively in older men with and without stroke. *Journal of Clinical Epidemiology, 67*(5), 500–507. doi:10.1016/j.jclinepi.2013.12.003

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R. M., Harwood, J., … Sibley, C. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, *38*(12), 1629–1643. doi:10.1177/0146167212457953

Bar-Natan, I., Rosen, Y., Salomon, G. (2010). *Does friendship between adversaries generalize?* Unpublished manuscript, University of Haifa, Haifa, Israel.

Bar-Tal, D. (1990). Israeli-Palestinian conflict: A cognitive analysis. *International Journal of Intercultural Relations*, *14*(1), 7–29. doi:10.1016/0147-1767(90)90045-X

Bar-Tal, D. (1996). Development of social categories and stereotypes in early childhood: The case of "the Arab" concept formation, stereotype and attitudes by Jewish children in Israel. *International Journal of Intercultural Relations*, *20*(3-4), 341–370. doi:10.1016/0147-1767(96)00023-5

Bar-Tal, D. (1998). Societal beliefs in times of intractable conflict: The Israeli case. *The International Journal of Conflict Management*, *9*(1), 22–50. doi:10.1108/eb022803

Bar-Tal, D. (2000). From intractable conflict through conflict resolution to reconciliation: Psychological analysis. *Political Psychology*, *21*(2), 531–365. doi:10.1111/0162-895X.00192

Bar-Tal, D. (2004). The necessity of observing real life situations: Palestinian-Israeli violence as laboratory for learning about social behaviour. *European Journal of Social Psychology, 34*(6), 677–701. doi:10.1002/ejsp.224

Bar-Tal, D. (2007). Sociopsychological foundations of intractable conflicts. *American Behavioral Scientist*, *50*(11), 1430–1453. doi:10.1177/0002764207302462

Bar-Tal, D. (2013). Intractable conflicts: Socio-psychological foundations and dynamics. Cambridge, UK: Cambridge University Press.

Bar-Tal, D., Halperin, E., & de Rivera, J. (2007). Collective emotions in conflict situations: Societal implications. *Journal of Social Issues, 63*(2), 441–460. doi:10.1111/j.1540-4560.2007.00518.x

Bar-Tal, D., & Labin, D. (2001). The effect of a major event on stereotyping: Terrorist attacks in Israel and Israeli adolescents' perceptions of Palestinians, Jordanians and Arabs. *European Journal of Social Psychology, 31*(3), 265–280. doi:10.1002/ejsp.43

Bar-Tal, D., Raviv, A., Raviv, A., & Dgani-Hirsh, A. (2009). The influence of the ethos of conflict on Israeli Jews' interpretation of Jewish Palestinian encounters. *Journal of Conflict Resolution*, *53*(1), 94–118. doi:10.1177/0022002708325942

Bar-Tal, D., Rosen, Y., & Nets-Zehngut, R. (2010). Peace education in societies in intractable conflicts: Goals, conditions, and directions. In G. Salomon (Ed.), *Handbook on peace and conflict* (pp. 21–43). New York, NY: Psychology Press.

Bar-Tal, D., & Teichman, Y. (2005). Stereotypes and prejudice in conflict: Representations of Arabs in Israeli Jewish society. Cambridge, UK: Cambridge University Press.

Bastian, B., Lusher, D., & Ata, A. (2012). Contact, evaluation and social distance: Differentiating majority and minority effects. *International Journal of Intercultural Relations, 36*(1), 100–107. doi:10.1016/j.ijintrel.2011.02.005

Baumeister, R. F., Bratlavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. doi:10.1037/1089-2680.5.4.323

Becker, J. C., Wright, S. C., Lubensky, M. E., & Zhou, S. (2013). Friend or ally: Whether cross-group contact undermines collective action depends on what advantaged group members say (or don't say). *Personality and Social Psychology Bulletin*, *39*(4), 442–455. doi:10.1177/0146167213477155

Beelmann, A., & Heinemann, K. S. (2014). Preventing prejudice and improving intergroup attitudes: A meta-analysis of child and adolescent training program. *Journal of Applied Developmental Psychology, 35*(1), 10–25. doi:10.1016/j.appdev.2013.11.002

Bekerman, Z., & Horenczyk, G. (2004). Arab-Jewish bilingual coeducation in Israel: A long-term approach to intergroup conflict resolution. *Journal of Social Issues, 60*(2), 389–404. doi:10.1111/j.0022-4537.2004.00120.x

Bekhuis, H., Ruiter, S., & Coenders, M. (2013). Xenophobia among youngsters: The effect of inter-ethnic contact. *European Sociological Review*, *29*(2), 229–242. doi:10.1093/esr/jcr057

Ben-Ari, R., & Amir, Y. (1986). Contact between Arab and Jewish youth in Israel: Reality and potential. In M. Hewstone & R. Brown (Eds.), *Contact and conflict in intergroup encounters* (pp. 45–58). Oxford, UK: Basil Blackwell.

Benhaida, S. (2015, January 6). Oslo process buried, analysts say, as Israeli-Palestinian conflict enters new phase. *Your Middle East*. Retrieved from http://www.yourmiddleeast.com/news/oslo-process-buried-analysts-say-as-israelipalestinian-conflict-enters-new-phase_28971

Bergen, T. J., Jr. (2001). The development of prejudice in children. *Education, 122*(1), 154–162.

Berger, R., Benatov, J, Abu-Raiya, H., & Tadmor, T. C. (2016). *Reducing prejudice and promoting positive intergroup attitudes among elementary-school children in the context of the Israeli-Palestinian conflict.* Manuscript submitted for publication.

Bontempo, D., Grouzet, F. M. E., & Hofer, S. M. (2012). Measurement issues in the analysis of within-person change. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences* (pp. 97–142). New York, NY: Routledge.

Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 149–182). Mahwah, NJ: Lawrence Erlbaum Associates.

Brown, R. (2010). *Prejudice: Its social psychology* (2nd ed.). Oxford, UK: Wiley-Blackwell.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

Burton, J. W. (1987). *Resolving deep-rooted conflict: A handbook*. Lanham, MD: University Press of America.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. doi:10.1080/10705510701301834

Christ, O., Schmid, K., Lolliot, S., Swart, H., Stolle, D., Tausch, N., … Hewstone, M. (2014). Contextual effect of positive intergroup contact on outgroup prejudice. *Proceedings of the National Academy of Sciences*, *111*(11), 3996–4000. doi:10.1073/pnas.1320901111

Christ, O., Ulrich, J., & Wagner, U. (2008, July). *The joint effects of positive and negative intergroup contact on attitudes and attitude strengths*. Paper presented at the general meeting of the European Association of Experimental Social Psychology, Opatija, Croatia.

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, *5,* 982. doi:10.3389/fpsyg.2014.00982

Cohen, S. P., & Azar, E. E. (1981). From war to peace: The transition between Egypt and Israel. *The Journal of Conflict Resolution*, *25*(1), 87–114. doi:10.1177/002200278102500104

Coleman, P. T. (2003). Characteristics of protracted, intractable conflict: Toward the development of a metaframework-I. *Peace and Conflict: Journal of Peace Psychology*, *9*(1), 1–37. doi:10.1207/S15327949PAC0901_01

Coleman, P. T. (2006). Intractable Conflict. In M. Deutsch, P. T. Coleman, & E. D. Marcus (Eds.), *The handbook of conflict resolution: Theory and practice* (7th ed., pp. 533–559). San Francisco, CA: Jossey-Bass.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, 57*, 505–528. doi:10.1146/annurev.psych.57.102904.190146

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351. doi:10.1037/1082-989X.6.4.330

Connolly, E. (1992). *The Edu-Train project and the reduction of prejudice* (Unpublished master thesis). Rand Afrikaans University, Johannesburg, South Africa.

Council of Religious Institutions of the Holy Land. (2013). *Victims of our own narratives? Portrayal of the "other" in Israeli and Palestinian school books* (Study Report). Retrieved from http://d7hj1xx5r7f3h.cloudfront.net/Israeli-Palestinian_School_Book_Study_Report-English.pdf

Crocker, C. A., Hampson, F. O., & Aall, P. (2005). *Grasping the nettle: Analyzing cases of intractable conflict*. Washington, DC: United States Institute of Peace.

Cuhadar, E., Genc, O. G., & Kotelis, A. (2015). A Greek-Turkish peace project: Assessing the effectiveness of interactive conflict resolution. *Southeast European and Black Sea Studies*. Advance online publication. doi:10.1080/14683857.2015.1020141

Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. *Survey Research Methods, 2*(1), 33–46. doi:10.18148/srm/2008.v2i1.365

Davidov, E., & De Beuckelaer, A. (2010). How harmful are survey translations? A test with Schwartz's Human Values Instrument. *International Journal of Public Opinion Research*, *22*(4), 485–510. doi:10.1093/ijpor/edq030

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75. doi:10.1146/annurev-soc-071913-043137

Davies, K., Tropp, L. R., Aron, A., Pettigrew, T. F., & Wright, S. C. (2011). Cross-group friendships and intergroup attitudes: A meta-analytic review. *Personality and Social Psychology Review, 15*(4), 332–351. doi:10.1177/1088868311411103

De Beuckelaer, A., & Swinnen, G. (2011). Biased latent variable mean comparison due to measurement noninvariance: A simulation study. In E. Davidov, P. Schmidt & J. Billiet (Eds.), *Cross-Cultural Analysis: Methods and Applications* (pp. 117–147). New York, NY: Routledge.

Deegan, J., Jr. (1978). On the occurrence of standardized regression coefficients greater than one. *Educational and Psychological Measurement*, *38*(4), 873–888. doi:10.1177/001316447803800404

Deutsche Gesellschaft für die Vereinten Nationen. (2011). *UN Basis-Informationen 44: Die Vereinten Nationen und der Nahost-Konflikt* [UN basic information 44: The United Nations and the Middle East Conflict]. Retrieved from the United Nations Associations of Germany http://www.dgvn.de/fileadmin/user_upload/PUBLIKATIONEN/Basis_Informationen/Basisinfo44_Naher_Osten.pdf [German]

Dixon, J., & Durrheim, K. (2003). Contact and the ecology of racial division: Some varieties of informal segregation. *British Journal of Social Psychology*, *42*(1), 1–23. doi:10.1348/014466603763276090

Dixon, J., Durrheim, K., & Tredoux, C. (2005). Beyond the optimal contact strategy: A reality check for the contact hypothesis. *American Psychologist, 60*(7), 697–711. doi:10.1037/0003-066X.60.7.697

Dixon, J., Tropp, L. R., Durrheim, K., & Tredoux, C. (2010). "Let them eat harmony": Prejudice-reduction strategies and attitudes of historically disadvantaged groups. *Current Directions in Psychological Science*, *19*(2), 76–80. doi:10.1177/0963721410363366

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. doi:10.1111/bjop.12046

Ekehammar, B., Akrami, N., & Araya, T. (2003). Gender differences in implicit prejudice. *Personality and Individual Differences*, *34*(8), 1509–1523. doi:10.1016/S0191-8869(02)00132-0

Elcheroth, G., Doise, W., & Reicher, S. (2011). On the knowledge of politics and the politics of knowledge: How a social representations approach helps us rethink the subject of political psychology. *Political Psychology*, *32*(5), 729–758. doi:10.1111/j.1467-9221.2011.00834.x

Ellis, D. G., & Maoz, I. (2002). Cross-cultural argument interactions between Israeli-Jews and Palestinians. *Journal of Applied Communication Research, 30*(3), 181–194. doi:10.1080/00909880216583

Ellis, D. G., & Maoz, I. (2007). Online argument between Israeli Jews and Palestinians. *Human Communication Research*, *33*, 291–309. doi:10.1111/j.1468-2958.2007.00300.x

Enders, C. K. (2013). Analyzing structural equation modeling with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A second course* (pp. 493–519). Charlotte, NC: Information Age Publishing Inc.

Eshel, Y. (1999). Effects of in-group bias on planned encounters of Jewish and Arab youth. *The Journal of Social Psychology*, *139*(6), 768–783. doi:10.1080/00224549909598256

Eshel, Y., & Dicker, R. (1995). Congruence and incongruence in perceived ethnic acceptance among Israeli students. *The Journal of Social Psychology*, *135*(2), 251–262. doi:10.1080/00224545.1995.9711429

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science, 326*(5952), 535–538. doi:10.1126/science.1168244

Farhan, T. (2008). *If you don't like them by now – what makes you like them next year? Social-cognitive and social predictors of prejudice in school children in a two-wave longitudinal study* (Unpublished doctoral dissertation). Philipps-University Marburg, Marburg, Germany.

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (14th ed., pp. 161–202). New York, NY: Academic Press.

Feddes, A.R., Noack, P., & Rutland, A. (2009). Direct and extended friendship effects on minority and majority children's interethnic attitudes: A longitudinal study. *Child Development, 80*(2), 377–390. doi:10.1111/j.1467-8624.2009.01266.x

Fell, B. (2014, July). *Buffer or poison? The interaction of positive and negative contact*. Paper session presented at the general meeting of the European Association of Social Psychology, Amsterdam, Netherlands.

Feng, D., Cong, Z., & Silverstein, M. (2012). Missing data and attrition. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences* (pp. 71–96). New York, NY: Routledge.

Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: An illustration of longitudinal measurement invariance. *Psychological Assessment*, *25*(2), 520–531. doi:10.1037/a0031669

Frijda N. H., & Mesquita B. (1994). The social roles and functions of emotions. In S. Kitayama & M. Hazel (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51–87). Washington, DC: American Psychological Association.

GA Res 70/1, Transforming our world: the 2030 Agenda for Sustainable Development, UN Doc A/RES/70/1 (21 October 2015). Retrieved from http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

Garz, D. (2006). *Sozialpsychologische Entwicklungstheorien: Von Mead, Piaget und Kohlberg bis zu Gegenwart* [Social psychological theories of development: From Mead, Piaget, and Kohlberg to the present] (3rd ed.). Wiesbaden, Germany: VS Verlag. [German]

Geiser, C. (2010). *Datenanalyse mit Mplus: Eine anwendungsorientierte Einführung* [Data Analysis with Mplus: An application-oriented introduction]. Wiesbaden, Germany: Verlag für Sozialwissenschaften. [German]

Goertz, G., & Diehl, P. F. (1993). Enduring rivalries: Theoretical constructs and empirical patterns. *International Studies Quarterly*, *37*(2), 147–171. doi:10.2307/2600766

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *The Journal of Applied Behavioral Science*, *12*(2), 133–157. doi:10.1177/002188637601200201

Gollwitzer, M., Christ, O., & Lemmer, G. (2014). Individual differences make a difference: On the use and the psychometric properties of difference scores in social psychology. *European Journal of Social Psychology*, *44*(7), 673–682. doi:10.1002/ejsp.2042

Gollwitzer, M., & Jäger, R. S. (2014). *Evaluation kompakt* [Evaluation compact]. Weinheim, Germany: Beltz. [German]

Graf, S., Paolini, S., & Rubin, M. (2014). Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five Central European countries. *European Journal of Social Psychology*, *44*(6), 536–547. doi:10.1002/ejsp.2052

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10*(1), 80–100. doi:10.1207/S15328007SEM1001_4

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530

Guffler, K., & Wagner, U. (2016). *Backfire of good intentions: Unexpected long-term contact intervention effects in an intractable conflict area.* Manuscript submitted.

Guffler, K., & Wagner, U. (2016). *Long-term contact intervention effects in a conflict area: The role of personal experiences and the repetition of intergroup contact meeting*. Manuscript submitted.

Halabi, R., & Zak, M. (2006). *Palestinian-Jewish youth encounters at the School for Peace*. Wahat al-Salam/ Neve Shalom: School for Peace Research Center.

Halperin, E., Bar-Tal, D., Nets-Zehngut, R., & Drori, E. (2008). Emotions in conflict: Correlates of fear and hope in the Israeli-Jewish society. *Peace and Conflict, 14*(3), 233–258. doi:10.1080/10781910802229157

Halperin, E., Cohen-Chen, S., & Goldenberg, A. (2014). Indirect emotion regulation in intractable conflicts: A new approach to conflict resolution. *European Review of Social Psychology, 25*(1), 1–31. doi:10.1080/10463283.2014.923155

Hammack, P. L. (2006). Identity, conflict, and coexistence: Life stories of Israeli and Palestinian adolescents. *Journal of Adolescent Research, 21*(4), 323–369. doi:10.1177/0743558406289745

Hammack, P. L. (2010). Narrating hyphenated selves: Intergroup contact and configurations of identity among young Palestinian citizens of Israel. *International*

Journal of Intercultural Relations, 34(4), 368–385. doi:10.1016/j.ijintrel.2010.03.002

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. doi:10.1080/19312450709336664

Hayes, B. C., McAllister, I., & Dowds, L. (2007). Integrated education, intergroup relations, and political identities in Northern Ireland. *Social Problems*, *54*(4), 454–482. doi:10.1525/sp.2007.54.4.454

Heidelberg Institute for International Conflict Research ([HIIK] 2015a). *Conflict Barometer 2014: No. 23.* Retrieved from http://www.hiik.de/en/konfliktbarometer/pdf/ConflictBarometer_2014.pdf

Heidelberg Institute for International Conflict Research ([HIIK] 2015b). *Methodological approach: Conflict definition*. Retrieved from http://www.hiik.de/en/methodik/index.html

Hermann, T., Atmor, N., Heller, E., & Lebel, Y. (2012). *The Israeli democracy index 2012.* Retrieved from http://en.idi.org.il/media/1365574/Index2012%20-%20Eng.pdf

Hermann, T., Heller, E., Atmor, N., & Lebel, Y. (2013a). *The Israeli democracy index 2013* Retrieved from http://en.idi.org.il/media/2720081/Democracy%20Index%20English%202013.pdf

Hermann, T., Heller, E., Atmor, N., & Lebel, Y. (2013b). *The Israeli democracy index 2013. Main findings.* Retrieved from http://en.idi.org.il/media/2726731/2013-Democracy-Index-Main-Findings.pdf

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117–144. doi:10.1080/03610739208253916

Hox, J. J. (2010). Multilevel analysis: Techniques and applications. New York, NY: Routledge.

Hoxter, A. L., & Lester, D. (1994). Gender differences in prejudice. *Perceptual and Motor Skills*, *79*(3f), 1666–1666. doi:10.2466/pms.1994.79.3f.1666

Hughes, J., Lolliot, S., Hewstone, M., Schmid, K., & Carlisle, K. (2012). Sharing classes between separate schools: A mechanism for improving inter-group relations in Northern Ireland? *Policy Futures in Education*, *10*(5), 528–539. doi:10.2304/pfie.2012.10.5.528

IBM Corp. (2012). IBM SPSS Statistics for Windows [Computer software]. Armonk, NY: IBM Corp.

Intergovernmental Panel on Climate Change (2014). *Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change* (Core Writing Team, R. K. Pachauri and L. A. Meyer [Eds.]). Geneva, Switzerland: IPCC. Retrieved from http://ar5-syr.ipcc.ch/ipcc/ipcc/resources/pdf/IPCC_SynthesisReport.pdf

Israeli Democracy Institute ([IDI] 2012a). *The peace index – November 2012.* Retrieved from http://en.idi.org.il/media/1984056/The%20Peace%20Index%20Data%20-%20November%202012.pdf

Israeli Democracy Institute ([IDI] 2012b). *The peace index – October 2012.* Retrieved from http://en.idi.org.il/media/1838541/The%20Peace%20Index%20Data%20-%20October%202012.pdf

Israeli Democracy Institute ([IDI] 2015a). *The Israeli democracy index.* Retrieved from http://en.idi.org.il/tools-and-data/guttman-center-for-surveys/the-israeli-democracy-index/

Israeli Democracy Institute ([IDI] 2015b). *The peace index.* Retrieved from http://en.idi.org.il/tools-and-data/guttman-center-for-surveys/the-peace-index/

Issmer, C. (2012). *Explaining aggressive and delinquent behaviors of disadvantaged adolescents: The impact of negative metastereotypes* (Unpublished doctoral dissertation). Philipps-University Marburg, Marburg, Germany.

Jayusi, W. (2009). *Restoring the attitudes of peace education participants through peer-tutoring* (Unpublished doctoral dissertation). University of Haifa, Haifa, Israel. [Hebrew]

Kamfer, L. & Venter, D. J. L. (1994). First evaluation of a stereotype reduction workshop. *South African Journal of Psychology, 24*(1), 13–20. doi:10.1177/008124639402400102

Kano, Y. (2001). Structural equation modeling for experimental data. In R. Cudeck, K. G. Jöreskog & D. Sörbom (Eds.), *Structural equation modeling: Present and future. A Festschrift in honor of Karl Jöreskog* (pp. 381–402). Lincolnwood, IL: Scientific Software International Incorporation.

Kashmir tensions: There are many bloody milestones along the way - expert. (2013, August 8). *KashmirWatch*. Retrieved from http://kashmirwatch.com/interviews.php/2013/08/08/kashmir-tensions-there-are-many-bloody-milestones-along-the-way-expert.html

3-front Kashmir struggle unstoppable. (2015, January 5). *KashmirWatch*. Retrieved from http://kashmirwatch.com/news.php/2015/01/05/3-front-kashmir-struggle-unstoppable.html

Kenward, M. G., & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, *13*(3), 236–247.

Khattab, N., Miaari, S., Manor, A., Nabwani, O., & Kagya, S. (2013). *Social justice in Jewish-Arab Relations in Israel*. Retrieved from The Israeli Democracy Institute http://www.en.idi.org.il/analysis/articles/social-justice-in-jewish-arab-relations-in-israel/

Kilpatrick, R., & Leitch, R. (2004). Teachers' and pupils' educational experiences and school-based responses to the conflict in Northern Ireland. *Journal of Social Issues*, *60*(3), 653–586. doi:10.1111/j.0022-4537.2004.00372.x

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: The Guilford Press.

Kriesberg, L. (1998). Intractable conflicts. In E. Weiner (Ed.), *The handbook of interethnic coexistence* (pp. 332–342). New York, NY: Continuum Books.

Kriesberg, L. (2005). Nature, dynamics, and phases of intractability. In C. A. Crocker, F. O. Hampson, & P. Aall (Eds.), *Grasping the nettle: Analyzing cases of intractable conflict* (pp. 65–97). Washington, DC: United States Institute of Peace.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30*(3), 411–433. doi:10.1111/j.1468-2958.2004.tb00738.x

Kropiunigg, U., & Pabst, B. (2007). Selbstwertsteigerung und Vorurteilsreduktion bei Jugendlichen eines multiethnischen Peace Camps [Improvement of self-esteem and reduction of prejudice by multiethnic peace camps for juveniles]. *Psychotherapie Forum*, *15*(2), 63–72. doi:10.1007/s00729-007-0194-1 [German]

Kuckartz, U. (1995-2015). MAXQDA [Computer software]. Berlin, Germany: VERBI GmbH.

Kumkale, G. T., & Albarracín, D. (2004). The sleeper effect in persuasion: A meta-analytic review. *Psychological Bulletin*, *130*(1), 143–172. doi:10.1037/0033-2909.130.1.143

Kupermintz, H., & Salomon, G. (2005). Lessons to be learned from research on peace education in the context of intractable conflict. *Theory Into Practice, 44*(4), 293–302. doi:10.1207/s15430421tip4404_3

Lanphen, J. (2011). *Kooperatives Lernen und Integrationsförderung: Eine theoriegeleitete Intervention in ethnisch heterogenen Schulklassen* [Cooperative learning and promoting integration: A theory-based intervention in ethnically heterogeneous classrooms]. Münster, Germany: Waxmann. [German]

Larsen, S. (2012, November 27). *Gaza Conflict* [Video file]. Retrieved from http://www.abc.net.au/btn/story/s3639690.htm

Lazarus, N. (2011). *Evaluating peace education in the Oslo-Intifada generation: A long-term impact study of Seeds of Peace 1993-2010* (Unpublished doctoral dissertation). American University, Washington, DC, USA.

Lazovsky, R. (2007). Educating Jewish and Arab children for tolerance and coexistence in a situation of ongoing conflict: An encounter program. C*ambridge Journal of Education, 37*(3), 391–408. doi:10.1080/03057640701546706

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*(2), 152–168. doi:10.1002/ejsp.2079

Leonardi, A., Magos, K., & Oikonomou, A. (2014). From the fear of death to the fear of 'dissimilar other': A research in elementary school classes. *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, *42*(1), 77–88. doi:10.1080/03004279.2011.652141

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. doi:10.2307/2290157

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: The Guilford Press.

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, *31*(4), 357–365. doi:10.1177/0165025407077757

Lommen, M. J. J., van de Schoot, R., & Engelhard, I. M. (2014). The experience of traumatic events disrupts the measurement invariance of a posttraumatic stress scale. *Frontiers in Psychology, 5,* 1304. doi:10.3389/fpsyg.2014.01304

Luiz, D., & Krige, P. (1985). The effect of social contact between South African White and Colored adolescent girls: A follow-up study. *Journal of School Psychology*, *125*(3), 407–408. doi:10.1080/00224545.1985.9922909

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. doi:10.1027/1614-2241.1.3.86

Maede, A. W., Lautenschlaeger, G. J., & Hecht, J. E. (2005). Establishing measurement equivalence and invariance in longitudinal data with item response theory. *International Journal of Testing*, *5*(3), 279–300. doi:10.1207/s15327574ijt0503_6

Maiese, M. (2003). *Causes of disputes and conflicts.* Retrieved from http://www.beyondintractability.org/essay/underlying-causes

Malhotra, D., & Liyanage, S. (2005). Long-term effects of peace workshops in protracted conflicts. *Journal of Conflict Resolution*, *49*(6), 908–924. doi:10.1177/0022002705281153

Maoz, I. (2000). An experiment in peace: Reconciliation-aimed workshops of Jewish-Israeli and Palestinian Youth. *Journal of Peace Research, 37*(6), 721–736. doi:10.1177/0022343300037006004

Maoz, I. (2002). Conceptual mapping and evaluation of peace education programs: The case of education for coexistence through intergroup encounters between Jews and Arabs in Israel. In G. Salomon & B. Nevo (Eds.), *Peace education: The concept, principles, and practices around the world* (pp. 259–269). Mahwah, NJ: Lawrence Erlbaum Associates.

Maoz, I. (2010). Educating for peace through planned encounters between Jews and Arbs in Israel: A reappraisal for effectiveness. In G. Salomon & E. Cairns (Eds.), *Handbook of peace education* (pp. 303–314). New York, NY: Psychology Press.

Maoz, I. (2011). Does contact work in protracted asymmetrical conflict? Appraising 20 years of reconciliation-aimed encounters between Israeli Jews and Palestinians. *Journal of Peace Research, 8*(1), 115–125. doi:10.1177/0022343310389506

Maoz, I., Bar-On, D., Bekerman, Z., & Jaber-Massarwa, S. (2004). Learning about 'good enough' through 'bad enough': A story of a planned dialogue between Israeli Jews and Palestinians. *Human Relations, 57*(9), 1075–1101. doi:10.1177/0018726704047139

Maoz, I., Bekerman, Z., & Sheftel, M. (2007). Can talking to each other really make a difference? Perspectives on reconciliation-aimed dialogues in the conflict between Israeli-Jews and Palestinians. In J. Kuriansky (Ed.), *Beyond bullets and bombs: Grassroots peacebuilding between Israelis and Palestinians* (pp. 37–49). Westport, CT: Praeger.

Maoz, I., Steinberg, S., Bar-On, D., & Fakhereldeen, M. (2002). The dialogue between the 'Self' and the 'Other': A process analysis of Palestinian–Jewish encounters in Israel. *Human Relations, 55*(8), 931–962. doi:10.1177/0018726702055008178

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*, 577–605. doi:10.1146/annurev.psych.60.110707.163612

McArdle, J. J., & Prindle, J.J. (2008). A latent change score analysis of a randomized clinical trial in reasoning training. *Psychology and Aging, 23*(4), 702–719. doi:10.1037/a0014349

MacCallum, R., & Austin, J. T. (2000). Application of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226. doi:10.1146/annurev.psych.51.1.201

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*(1), 40–48. doi:10.1037//0021-843X.110.1.40

Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, *73*(3), 574–584. doi:10.1037/0021-9010.73.3.574

Moltmann, B. (2014). Nordirland [Northern Ireland]. In *Dossier innerstaatliche Konflikte (Konflikportraits) [Dossier innerstate conflicts (conflict portraits)]*. Retrieved from the German Federal Agency for Civic Education http://www.bpb.de/internationales/weltweit/innerstaatliche-konflikte/54664/nordirland [German]

Moran-Ellis, J. (2006). Triangulation and integration: Processes, claims and implications. *Qualitative Research*, *6*(1), 45–59. doi:10.1177/1468794106058870

Muldoon, O. T., & Trew, K. (2000). Children's experience and adjustment to political conflict in Northern Ireland. *Peace and Conflict: Journal of Peace Psychology*, *6*(2), 157–176. doi:10.1207/S15327949PAC0602_4

Muthén, L. (2012, April 23). *Type = Complex* [Online forum comment]. Retrieved from http://www.statmodel.com/discussion/messages/12/7975.html?1360965134

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2014). M*plus* (Version 7.2) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Narratives of 1948. (2002) [Roundtable debate]. *Palestine-Israel Journal of Politics, Economics, and Culture, 9*(4). Retrieved from http://www.pij.org/details.php?id=105

Nevo, B., & Brem, I. (2002). Peace education programs and the evaluation of their effectiveness. In G. Salomon & B. Nevo (Eds.), *Peace education: The concept, principles, and practices around the world* (pp. 271–282). Mahwah, NJ: Lawrence Erlbaum Associates.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, *6*(3), 328–362. doi:10.1177/1094428103254673

Niedenthal, P. M., & Brauer, M. (2012). Social functionality of human emotion. *Annual Review of Psychology, 63*, 259–285. doi:10.1146/annurev.psych.121208.131605

Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, *14*(3), 587–598. doi:10.1007/s11136-004-0830-y

Paluck, E. L (2012). Interventions aimed at the reduction of prejudice and conflict. In L. R. Tropp (Ed.), *The Oxford Handbook of Intergroup Conflict* (pp. 179–192). New York, NY: Oxford University Press.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*(1), 339–367. doi:10.1146/annurev.psych.60.110707.163607

Paolini, S., Harwood, J., & Rubin, M. (2010). Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin*, *36*(12), 1723–1738. doi:10.1177/0146167210388667

Paolini, S., Harwood, J., Rubin, M., Husnu, S., Joyce, N., & Hewstone, M. (2014). Positive and extensive intergroup contact in the past buffers against the disproportionate impact of negative contact in the present. *European Journal of Social Psychology*, *44*(6), 548–562. doi:10.1002/ejsp.2029

Peace Research Institute in the Middle East (2006). *Learning each other's historical narrative Palestinians and Israelis*. Retrieved from the Peace Research Institute in the Middle East http://vispo.com/PRIME/narrative.pdf

Pettigrew, T. F. (1971). *Racially separate or together?* New York, NY: McGraw-Hill.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85. doi:10.1146/annurev.psych.49.1.65

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. doi:10.1037/0022-3514.90.5.751

Pettigrew, T. F., & Tropp, L. R. (2011). *When groups meet: The dynamics of intergroup contact.* New York, NY: Psychology Press.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525–556. doi:10.3102/00346543074004525

Pilecki, A., & Hammack, P. L. (2014). Negotiating the past, imagining the future: Israeli and Palestinian narratives in intergroup dialogue. *International Journal of Intercultural Relations, 43*, 100–113. doi:10.1016/j.ijintrel.2014.08.019

Pope, H. (2014, November 12). A precious chance to end Turkey's conflict with the Kurds. *Financial Times*. Retrieved from http://www.ft.com/cms/s/0/bea838e4-698b-11e4-8f4f-00144feabdc0.html#axzz3Ob9H4IhG

Pugesek, B. H., Tomer, A., & von Eye, A. (2009). *Structural equation modeling: Application in ecological and evolutionary biology*. Cambridge, UK: Cambridge University Press.

Raabe, T., & Beelmann, A. (2011). Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences. *Child Development, 82*(6), 1715–1737. doi:10.1111/j.1467-8624.2011.01668.x

Reicher, S. (2007). Rethinking the paradigm of prejudice. *South Africa Journal of Psychology*, *35*(4), 412–432. doi:10.1177/008124630703700410

Rohde, D. (2012, November 22). After the ceasefire [Online newspaper article]. *Reuters*. Retrieved from http://blogs.reuters.com/david-rohde/2012/11/22/after-the-ceasefire/

Ron, Y., & Maoz, I. (2013). Dangerous stories: Encountering narratives of the other in the Israeli–Palestinian conflict. *Peace and Conflict: Journal of Peace Psychology*, *19*(3), 281–294. doi:10.1037/a0033686

Ron, Y., Maoz, I., & Bekerman, Z. (2010). Dialogue and ideology: The effect of continuous involvement in Jewish-Arab dialogue encounters on the ideological perspectives of Israeli-Jews. *International Journal of Intercultural Relations, 34*(6), 571–579. doi:10.1016/j.ijintrel.2010.07.001

Rosen, K. (2015). *Ist Wegrennen wünschenswert? Veränderung der Bewertung von Emotionsregulationsstrategien aufgrund des Präventionsprogramms PiKS* [Is running away desirable? Changing the assessment of emotion regulation strategies due to the prevention program PiKS] (Unpublished diploma thesis). Philipps-University Marburg, Marburg, Germany. [German]

Rosen, Y., & Perkins, D. (2013). Shallow roots require constant watering: The challenge of sustained impact in educational programs. *International Journal of Higher Education, 2*(4), 91–100. doi:10.5430/ijhe.v2n4p91

Rosen, Y., & Salomon, G. (2011). Durability of peace education effects in the shadow of conflict. *Social Psychology of Education, 14*(1), 135–147. doi:10.1007/s11218-010-9134-y

Ross, K. (2016). Peacebuilding through inter-group encounters: rethinking the contributions of 'mainstream' and 'politicised' approaches. *Peacebuilding*. Advance online publication. doi:10.1080/21647259.2016.1168086

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. doi:10.1093/biomet/63.3.581

Russell, D. W., Kahn, J. H., Altmeier, E. M., & Spoth, R. (1998). Analyzing data from experimental studies: A latent variable structural equation modeling approach. *Journal of Counseling Psychology*, *45*(1), 18–29.

Saguy, T., Dovidio, J. F., & Pratto, F. (2008). Beyond contact: Intergroup contact in the context of power relations. *Personality and Social Psychology Bulletin*, *34*(3), 432–445. doi:10.1177/0146167207311200

Sagy, S. (2002). Intergroup encounters between Jewish and Arab students in Israel: Towards an interactionist approach. *Intercultural Education, 13*(3), 259–274. doi:10.1080/1467598022000008341

Sak, G. (2013, January 12). The ethnic conflict in Cyprus is turning into a greater game [Online newspaper article]. *Hurriyet Daily News*. Retrieved from

http://www.hurriyetdailynews.com/the-ethnic-conflict-in-cyprus-is-turning-into-a-greater-game.aspx?pageID=238&nID=38891&NewsCatID=403

Salomon, G. (2002). The nature of peace education: Not all programs are created equal. In G. Salomon & B. Nevo (Eds.), *Peace education: The concept, principles, and practices around the world* (pp. 3–15). Mahwah, NJ: Lawrence Erlbaum Associates.

Salomon, G. (2004). Does peace education make a difference in the context of an intractable conflict? *Peace and Conflict: Journal of Peace Psychology*, *10*(3), 257–274. doi:10.1207/s15327949pac1003_3

Salomon, G. (2006). Does peace education really make a difference? *Peace and Conflict: Journal of Peace Psychology*, *12*(1), 37–48. doi:10.1207/s15327949pac1201_3

Salomon, G. (2009a). Peace education: Its nature, nurture and the challenges it faces. In J. de Rivera (Ed.), *Handbook on Building Cultures of Peace* (pp. 107–121). New York, NY: Springer.

Salomon, G. (2009b). *Recent research findings by the center for research on peace education*. Retrieved from University of Haifa, Center for Research on Peace Education website: http://peach.haifa.ac.il/images/f/f9/Summary_of_our_findings1.pdf

Salomon, G. (2011). Four major challenges facing peace education in regions of intractable conflict. *Peace and Conflict: Journal of Peace Psychology*, *17*(1), 46–59. doi:10.1080/10781919.2010.495001

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. doi:10.1007/BF02296192

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. doi:10.1037//1082-989X.7.2.147

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schleien, S. M. (2007). *Confronting the intractable: An evaluation of the Seeds of Peace experience* (Unpublished doctoral dissertation). University of Waterloo, Waterloo, Canada.

Schroeder, J., & Risen, J. L. (2016). Befriending the enemy: Outgroup friendship longitudinally predicts intergroup attitudes in a coexistence program for Israelis and Palestinians. *Group Processes and Intergroup Relations, 19*(1), 72–93. doi:10.1177/1368430214542257

Schwartz, C. E., & Sprangers, M. A. G. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine*, *48*(11)*,* 1531–1548. doi:10.1016/S0277-9536(99)00047-7

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shani, M. (2015). *A theory and practice of coexistence: Improving coexistence orientation through mixed-model encounters between Jews and Palestinians in Israel* (Unpublished doctoral dissertation). University of Bremen, Bremen, Germany.

Sharing Education Programme (2012). *A Report on the Sharing Education Learning Forum Conference 2012*. Retrieved from http://www.schoolsworkingtogether.com/documents/SELF%20Conference%20Report%20December%202012.pdf

Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting

reconciliation. *Journal of Personality and Social Psychology, 94*(1), 116–132. doi:10.1037/0022-3514.94.1.116

Skinner, B. F. (1948). Superstition' in the pigeon. *Journal of Experimental Psychology, 38*, 168–172. doi:10.1037/h0055873

Smith, H. L. (2013). Research design: Toward a realistic role of causal analysis. In S. L Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 45–73). Dordrecht, Netherlands: Springer.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage Publications Ltd.

Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health related quality of life research: A theoretical model. *Social Science & Medicine*, *48*(11), 1507–1515. doi:10.1016/S0277-9536(99)00045-3

Stark, T. H., Flache, A., & Veenstra, R. (2013). Generalization of positive and negative attitudes toward individuals to outgroup attitudes. *Personality and Social Psychology Bulletin*, *39*(5), 608–622. doi:10.1177/0146167213480890

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90. doi:10.1086/209528

Stephan, C. W., Hertz-Lazarowitz, R., Zelniker, T., & Stephpan, W. G. (2004). Introduction to improving Arab-Jewish relations in Israel: Theory and practice in coexistence educational programs. *Journal of Social Issues, 60*(2), 237–252. doi:10.1111/j.0022-4537.2004.00109.x

Stephan, W. G., Stephan, C. W., & Gudykunst, W. B. (1999). Anxiety in intergroup relations: A comparison of anxiety/uncertainty management theory and integrated threat theory. *International Journal of Intercultural Relations*, *23*(4), 613–628. doi:10.1016/S0147-1767(99)00012-7

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research Online*, *2*(1), 21–33.

Swart, H., Hewstone, M., Christ, O., & Voci, A. (2011). Affective mediators of intergroup contact: A three-wave longitudinal study in South Africa. *Journal of Personality and Social Psychology*, *101*(6), 1221–1238. doi:10.1037/a0024450

Tavakoli, M., Hatami, J., & Thorngate, W. (2010). Changing stereotypes in Iran and Canada using computer mediated communication. *Journal of Intercultural Communication*, *23*(5).

Techakesari, P., Barlow, F. K., Hornsey, M. J., Sung, B., Thai, M., & Chak, J. L. Y. (2015). An investigation of positive and negative contact as predictors of intergroup attitudes in the United States, Hong Kong, and Thailand. *Journal of Cross-Cultural Psychology*, *46*(3), 454–468. doi:10.1177/0022022115570313

Teichman, Y., Bar-Tal, D., & Abdolrazeq, Y. (2007). Intergroup biases in conflict: Reexamination with Arab pre-adolescents and adolescents. *International Journal of Behavioral Development*, *31*(5), 423–432. doi:10.1177/0165025407081470

The Jerusalem Institute for Israel Studies (2015a). Extent of poverty among families, persons and children, in Israel and Jerusalem, by population group, age and level of education, 2012, 2013. *Statistical Yearbook. 2015 Edition*. Retrieved from http://www.jiis.org.il/.upload/yearbook/2015/shnaton_F0215.pdf

The Jerusalem Institute for Israel Studies (2015b). Population aged 15 and over, and aged 25-54 (prime working-age group) in Israel and Jerusalem, by labor force characteristics, population group and gender, 2013. *Statistical Yearbook. 2015*

Edition. Retrieved from http://www.jiis.org.il/.upload/yearbook/2015/shnaton_G0115.pdf

The Jerusalem Institute for Israel Studies (2015c). Population of Israel and Jerusalem, by population group, 1922-2013. *Statistical Yearbook. 2015 Edition*. Retrieved from http://www.jiis.org.il/.upload/yearbook/2015/shnaton_C0115.pdf

The Jerusalem Institute for Israel Studies (2015d). Students in universities in Israel, by institution, degree and population Ggroup, 2012/13. *Statistical Yearbook. 2015 Edition*. Retrieved from http://www.jiis.org.il/.upload/yearbook/2015/shnaton_M2115.pdf

Thies, C. G. (2001). A social psychological approach to enduring rivalries. *Political Psychology*, *22*(4), 693–725. doi:10.1111/0162-895X.00259

Thompson, M. S., & Green, S. B. (2013). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A second course* (2nd ed., pp. 163–218). Charlotte, NC: Information Age Publishing, Inc.

Tropp, L. R. (2012). Understanding and responding to intergroup conflict: Toward an integrated analysis. In L. R. Tropp (Ed.), *The Oxford Handbook of Intergroup Conflict* (pp. 3–10). New York, NY: Oxford University Press.

Tropp, L. R., Hawi, D. R., van Laar, C., & Levin, S. (2012). Cross-ethnic friendships, perceived discrimination, and their effects on ethnic activism over time: A longitudinal investigation of three ethnic minority groups. *British Journal of Social Psychology*, *51*(2), 257–272. doi:10.1111/j.2044-8309.2011.02050.x

Tropp, L. R., & Pettigrew, T. F. (2005). Relationships between intergroup contact and prejudice among minority and majority status groups. *Psychological Science, 16*(12), 651–657. doi:10.1111/j.1467-9280.2005.01643.x

Turner, R. N., Crisp, R. J., & Lambert, E. (2007). Imagining intergroup contact can improve intergroup attitudes. *Group Processes and Intergroup Relations, 10*(4), 427–441. doi:10.1177/1368430207081533

Ullman, J. B. (2006). Structural equation modeling: Reviewing the basics and moving forward. *Journal of Personality Assessment*, *87*(1), 35–50. doi:10.1207/s15327752jpa8701_03

UN says evidence on ethnic massacres in S. Sudan may be sufficient for prosecutions. (2015, January 15). *Radio Tamazuj*. Retrieved from https://radiotamazuj.org/en/article/un-says-evidence-ethnic-massacres-s-sudan-may-be-sufficient-prosecutions

van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: Chapman and Hall/CRC.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–69. doi:10.1177/109442810031002

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4,* 770. doi:10.3389/fpsyg.2013.00770

van de Vijver, F. J. R. (1998). Towards a theory of bias and equivalence. In Zuma (Ed.), *ZUMA Nachrichten Spezial Nr. 3* (pp. 41–65). Retrieved from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial3.pdf

van de Vijver, F. J. R. (2011). Capturing bias in structural equation modeling. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 3–34). New York, NY: Routledge.

Vered, S., & Bar-Tal, D. (2014). Routinization of the Israeli-Arab conflict: The perspective of outsiders. *Israel Studies Review*, *29*(1), 41–61. doi:10.3167/isr.2014.290104

Vezzali, L., Giovannini, D., & Capozza, D. (2010). Longitudinal effects of contact on intergroup relations: The role of majority and minority group membership and intergroup emotions. *Journal of Community and Applied Social Psychology, 20*(6), 426–479. doi:10.1002/casp.1058

Werner, C., & Schermelleh-Engel, K. (2009). *Strukturgleichungsmodelle: Vorteile, Herausforderungen und Probleme [Structural equation modeling: Advantages, challenges, and problems]*. Retrieved from https://www.yumpu.com/de/document/view/24641708/strukturgleichungsmodelle-vorteile-herausforderungen-und- [German]

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604. doi:10.1037/0003-066X.54.8.594

Wright, S., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*, *73*(1), 73–90. doi:10.1037/0022-3514.73.1.73

Yablon, Y. B., & Katz, Y. J. (2001). Internet-based group relations: A high school peace education project in Israel. *Educational Media International, 38*(2–3), 175–182. doi:10.1080/09523980110043591

Zercher, F., Schmidt, P., Cieciuch, J., & Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: Exact vs. approximate measurement invariance. *Frontiers in Psychology, 6*, 733. doi:10.3389/fpsyg.2015.00733

ZUSAMMENFASSUNG

Contact Interventions in Intractable Conflicts:
Long-Term Attitude and Behavior Intention Changes in Israel

Trotz jahrzehntelanger Interventionen und wissenschaftlicher Forschung brechen weltweit immer wieder Konflikte aus, die auf ethnischen, religiösen und rassistischen Unterschieden basieren (Tropp, 2012, p. 3). Alleine das Heidelberger Institut für Internationale Konfliktforschung (HIIK; 2015a) nennt für das Jahr 2014 in seiner jährlichen Analyse zum globalen Konfliktgeschehen 21 Kriege und 25 begrenzte Kriege. Die Konsequenzen dieser Konflikte für die betroffene Bevölkerung sind gravierend und reichen von der Beeinträchtigung der körperlichen Unversehrtheit bis zur Sicherstellung grundlegender menschlicher Bedürfnisse wie zum Beispiel Essen, Wasser, Schutz, Sicherheit und Selbstverwirklichung (Maiese, 2003, Segment 5).

Eine Möglichkeit, Konflikte zwischen Gruppen zu reduzieren, wurde 1954 von Gordon Allport vorgestellt. In seinem Buch *The Nature of Prejudice* zeigte Allport auf, dass Kontakt zwischen konfligierenden Gruppen unter optimalen Bedingungen Vorurteile zwischen den Gruppen reduzieren kann. Seine damaligen Überlegungen wurden bis heute vielfach empirisch von Wissenschaftlerinnen und Wissenschaftlern bestätigt (z. B. Pettigrew & Tropp, 2006). Dabei zeigte sich auch, dass Kontakt, im Rahmen von strukturierten Programmen (Interventionen), die sich das Ziel setzen, die Beziehungen zwischen konfligierenden Gruppen zu verbessern, ebenfalls positive Wirkungen entfaltet (z. B. Lemmer & Wagner, 2015). Intergruppenkontakt im Rahmen von Interventionen erzielt dabei nicht nur positive kurzfristige Effekte, sondern auch langfristige Effekte. Im Speziellen fanden Wissenschaftlerinnen und Wissenschaftler heraus, dass positive kurzfristige Effekte auch in Konfliktregionen realisiert werden können (z. B. Lemmer & Wagner, 2015). Ausgehend davon, dass eine langfristige, das heißt nachhaltige Verbesserung der Intergruppenbeziehungen notwendig ist, um Konflikte dauerhaft zu reduzieren oder zu beenden, stellt sich die Frage: Wie sollten Intergruppeninterventionen in Konfliktgebieten umgesetzt werden, um auch dort nachhaltige Effekte zu ermöglichen?

Ziel dieser Doktorarbeit war es daher Kontaktprogramme in einem Konfliktgebiet systematisch auf kurz- und langfristige Effekte zu untersuchen. Im Rahmen der Arbeit wurden zwei Kontaktinterventionen zwischen arabischen (Minorität) und jüdischen Israelis

(Majorität) in Israel analysiert (4. und 6. Schuljahr). Dabei wurden sechs unterschiedliche Stichproben ausgewertet: drei Stichproben der Minorität und drei der Majorität. Ausgehend von der aktuellen Forschungslage wurde getestet, welche Rolle der Status der Teilnehmerinnen und Teilnehmer einnahm (z. B. Mitglied der Mehrheit oder der Minderheit), welchen Einfluss die Wiederholung von Intergruppenkontakt aufwies und in wie weit die individuelle Bewertung des Interguppenkontakts mit den kurz- und langfristigen Effekten der Intervention in Verbindung stand. Die Effekte wurden mit Hilfe der abhängigen Variablen positive Einstellung und positive Verhaltensintention gemessen. Teilnahme beziehungsweise Nicht-Teilnahme an der Kontaktintervention diente als unabhängige Variable. Insbesondere sollte durch die Analysen sowohl das theoretische Wissen im Bereich der Kontakttheorie erweitert als auch praktische Empfehlungen für die Umsetzung von Kontaktinterventionen in Konfliktgebieten identifiziert werden. Ein besonderer Schwerpunkt der Arbeit lag dabei auf der methodischen Auswertung, die sich am neuesten Stand der Forschung im Bereich der Strukturgleichungsmodelle und der Veränderungsmessung bei experimentellen Forschungsdesigns orientierte.

Manuskript #1 umfasst die Ergebnisse der Kontaktintervention mit arabischen und jüdischen Israelis der 4. Schulklasse. Es wurden vier Stichproben untersucht (zwei jüdische und zwei arabische; in beiden Ethnien wurde eine Stichprobe dreimal und die andere fünfmal befragt). Dabei wurden bei den jüdischen Stichproben positive kurz- und langfristige Effekte im Bereich der Einstellungsänderung (aber nicht der Verhaltensabsichten) gefunden. Des Weiteren wurde festgestellt, dass die wiederholte Teilnahme an Intergruppenkontakten zu einer Ausdehnung der Langfristigkeit der Effekte führte. Ebenfalls zeigte sich, dass die individuelle Bewertung der Kontaktintervention einen Einfluss auf kurz- und langfristige Effekte hatte. Das heißt, wenn der Kontakt durch die Schülerinnen und Schüler positiv beschrieben wurde, waren die kurz- und langfristigen Effekte positiver als bei einer negativen oder neutralen Bewertung des Kontaktes. In den beiden arabischen Stichproben wurde nur die Verhaltensintention untersucht, da Voranalysen ergaben, dass keine längsschnittliche skalare Messinvarianz für die Einstellungsvariable gegeben war und daraufhin keine Längsschnittanalysen durchgeführt werden konnten. Bei der Minorität konnten kurzfristige aber keine langfristigen Effekte gefunden werden. Zudem hatte die wiederholte Teilnahme an Intergruppenkontakt als auch die Bewertung der Kontaktsituation keinen Einfluss auf die Nachhaltigkeit.

Manuskript #2 untersuchte die Schülerinnen und Schüler der 6. Schulklasse. Entgegen der theoretisch abgeleiteten Annahme wurde bei der jüdischen Stichprobe der Kontakt als überwiegend negativ wahrgenommen. Entsprechend wurden sowohl kurz- als auch langfristige negative Kontaktinterventionseffekte identifiziert. Inwieweit hierfür die Wiederholung des Intergruppenkontaktes relevant war, konnte nicht untersucht werden. Die Bewertung der Kontaktintervention hatte zudem einen Effekt auf die kurz- und langfristigen Ergebnisse, das heißt, wenn der Kontakt von den Schülerinnen und Schülern negativ wahrgenommen wurde, waren die kurz- und langfristigen Kontakteffekte schlechter als nach einer positiven oder neutralen Wahrnehmung. Ebenso wie in Manuskript #1 zeigten sich bei der arabischen Stichprobe lediglich kurzfristige Effekte. Interessanterweise waren diese aber im Vergleich zur Majorität positiv, so dass dieselbe Kontaktsituation von Minorität und Majorität unterschiedlich wahrgenommen wurde. Die Bewertung der Kontaktsituation zeigte bei der arabischen Stichprobe ebenfalls einen positiven Einfluss auf die kurz- und langfristigen Verhaltensintentionen.

Die abschließende Synopsis führt zusätzliche Analysen auf, deren Ergebnisse noch nicht in einem der beiden Manuskripte abgebildet wurden. Dabei konnte unter Zuhilfenahme externer Daten gezeigt werden, dass ein kontextuelles Ereignis (Israel-Gaza Konflikt 2012: Gewaltausbruch zwischen Juden und Palästinensern im November 2012, d. h. ungefähr in der Mitte von beiden Kontaktinterventionen) sowohl von der jüdischen als auch von der arabisch israelischen Bevölkerung wahrgenommen wurde, aber gegensätzliche Reaktionen hervorrief (Juden: Verstärkung der Opposition zu Friedensverhandlungen; Araber: Verringerung der Opposition; Israeli Democracy Institute, 2012a, 2012b). Des Weiteren konnte bis auf einen Fall (Einstellungsvariable bei den arabisch israelischen Schülerinnen und Schülern der 4. Klasse) bei allen Stichproben und jeweils beiden abhängigen Variablen über die Zeit hinweg längsschnittliche skalare Messinvarianz festgestellt werden. Anschließend konnte anhand der zusätzlichen Analysen aufgezeigt werden, dass die Berechnung der fehlenden Werte mit Hilfe der multiplen Imputation zu leicht konservativeren Ergebnissen führt als die Berechnung mit dem maximum likelihood Schätzer. Systematische Geschlechtsunterschiede wurden nicht gefunden.

Insgesamt leisten die Ergebnisse sowohl einen Beitrag für die zukünftige Forschung als auch für die Praxis. Für die Forschung weist die Arbeit vor allem darauf hin, dass Minorität und Majorität unterschiedlich langfristige Effekte nach der Teilnahme an einer Kontaktintervention in Konfliktregionen aufzeigen; ebenso dass die Einflussfaktoren, die

auf diese Effekte wirken (z. B. Wiederholung der Kontaktsituationen, Bewertung der Kontaktsituation) zwischen den Gruppen variieren. Daher sollte zukünftig vor allem genauer überprüft werden, worin die Unterschiede zwischen Minorität und Majorität liegen und welche Variablen langfristige Effekte von Kontaktinterventionen in Konfliktgebieten erklären können. Für die Praxis zeigt sich, dass bei der Umsetzung der Interventionen Intergruppenkontakte wiederholt werden sollten, damit zumindest bei der Majorität nachhaltigere Effekte ermöglicht werden können. Zudem sollten Praktiker intensiv geschult werden, um negative Kontaktsituationen zwischen den Gruppen und damit nachhaltige negative Intergruppenbeziehungen zu erkennen und zu verhindern.

Literatur

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Israeli Democracy Institute. (2012a). *The peace index – November 2012.* Retrieved from http://en.idi.org.il/media/1984056/The%20Peace%20Index%20Data%20-%20November%202012.pdf

Israeli Democracy Institute. (2012b). *The peace index – October 2012.* Retrieved from http://en.idi.org.il/media/1838541/The%20Peace%20Index%20Data%20-%20October%202012.pdf

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*(2), 152–168. doi:10.1002/ejsp.2079

Maiese, M. (2003). *Causes of disputes and conflicts.* Retrieved from http://www.beyondintractability.org/essay/underlying-causes

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. doi:10.1037/0022-3514.90.5.751

Tropp, L. R. (2012). Understanding and responding to intergroup conflict: Toward an integrated analysis. In L. R. Tropp (Ed.), *The Oxford Handbook of Intergroup Conflict* (pp. 3–10). New York, NY: Oxford University Press.

DANKSAGUNG

Als erstes möchte ich mich bei meinem Doktorvater Prof. Dr. Ulrich Wagner bedanken, der es mir ermöglicht hat, zu meinem Wunschthema zu promovieren. Für die Offenheit und das Vertrauen, welches Du mir von Beginn an entgegen gebracht hast, bedanke ich mich sehr. Die intensive und sehr hilfreiche Betreuung, wie die Tatsache, dass Du für Rückfragen immer ein offenes Ohr hattest, haben mir diese Arbeit möglich gemacht. Ebenso möchte ich mich bei Prof. Dr. Peter Schmidt bedanken, der diese Arbeit als Zweitgutachter begleitet hat und mir methodisch wie menschlich stets ein sehr angenehmer und hilfegebender Ansprechpartner war. Ein besonderer Dank gilt Prof. Dr. Gabi Salomon, der mich bei der Suche nach Kontaktprogrammen in Israel unterstützt hat. Ebenso bedanke ich mich bei Ran Shorenson und Carmela Arnon (Koordinator und Koordinatorin des israelischen Schulprogramms). Ohne ihr Engagement wäre eine Untersuchung des Kontaktprogramms in Israel nicht umsetzbar gewesen. Bei allen Wissenschaftlerinnen und Wissenschaftlern, die mich während der Promotion begleitet und unterstützt haben, möchte ich mich ebenfalls bedanken: Eure Impulse, Euer Wissen, Euer Mitdenken wie Eure Aufmunterungen haben mir sehr geholfen. Insbesondere möchte ich Lisa Gutenbrunner, Stefan Thörner und Judith Lanphen für ihre unermüdlichen methodischen und inhaltlichen Hilfestellungen danken. Ohne Euch hätte ich vermutlich noch ein paar Jahre länger an dieser Arbeit gesessen. Jan Weber und Caroline Woderich danke ich für die Korrekturen der englischsprachigen Texte. Auch gilt mein Dank der gesamten Arbeitseinheit Sozialpsychologie und explizit auch allen studentischen Hilfskräften sowie den Wissenschaftlerinnen und Wissenschaftlern des Graduiertenkollegs „Gruppenbezogene Menschenfeindlichkeit“. Prof. Dr. Thomas F. Pettigrew, den ich während einer seiner Aufenthalte in Deutschland persönlich kennenlernen durfte, danke ich für seine Anregungen und seinen Kommentare zu meiner Arbeit.

Betonen möchte ich in dieser Danksagung, dass es nicht nur das Wissenschaftliche, sondern in gleichem Maße auch das Menschliche ist, für das ich mich bei allen Personen bedanke. Die ungezwungenen Gespräche, das gemeinsame Lachen, Joggen, Spielen; all diese scheinbaren Kleinigkeiten und Momente, die geplant oder ungeplant Teil meines Alltags gewesen sind, waren mir Unterstützung, Motivation und wertvolle Begleitung in dieser Zeit. In diesem Sinne möchte ich mich auch bei meiner Familie bedanken. Auch sie hat mich begleitet, motiviert und ist für mich da gewesen. Zudem hat Sie mir immer wieder

einen Ausgleich geboten und Raum, um einmal inne zu halten. Meinem Mann Matthias gilt darüber hinaus ein ganz besonderer Dank. Er hat mich bei meinem Wunsch, mich mit diesem Thema auseinanderzusetzen und diese Promotion zu schreiben, nicht nur unterstützt, sondern auch über alle Hindernisse hinweg aktiv gefördert.

Ich möchte mich von ganzem Herzen bei Euch allen für diese wertvolle Zeit bedanken und hoffe, dass ich während der vergangenen Jahre nicht nur genommen, sondern Euch und der Forschungsgemeinschaft auch etwas zurückgegeben habe.

ANGABEN ZUR PERSON

ERKLÄRUNG DER AUTORIN

Hiermit versichere ich, dass ich meine Dissertation „Contact Interventions in Intractable Conflicts: Long-Term Attitude and Behavior Intention Changes in Israel“ selbstständig, ohne unerlaubte Hilfe angefertigt habe und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

_____ _____

Ort, Datum Kerstin Guffler