Implementation of forgiveness education to increase *agape* love in three culturally distinct areas: Belfast, Northern Ireland, Israel, and Taiwan

By

Jiahe Wang Xu

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Educational Psychology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 06/07/2023

The dissertation is approved by the following members of the Final Oral Committee:

Dr. Robert Enright, Professor, Educational Psychology

Dr. Amy Bellmore, Professor, Educational Psychology

Dr. Sarah Short, Assistant Professor, Educational Psychology

Dr. Julie Poehlmann, Professor, Human Development & Family Studies

Table of Contents

Acknowledgements iii
Abstract v
Chapter 1: Introduction to the Research Idea 1
Chapter 2: Literature Review and Hypotheses 6
Theoretical and Philosophical Perspectives on Forgiveness and *Agape* Love 6
Forgiveness Education Interventions and the Curriculum 12
Research Question 14
Hypotheses 14
Chapter 3: Methods 19
Participants Recruitment 19
Teacher Training and Fidelity Check 20
Intervention Curriculum 21
Experimental Design and Testing Procedures in the Intervention 23
Research Instruments 28
Chapter 4: Results 41
Northern Ireland Students 42
Reliability and Validity of the Initial Assessment 42
Effect of the *Agape* Love and Forgiveness Education Program 43
Israel Students 54
Reliability and Validity of the Initial Assessment 54

Effect of the *Agape* Love and Forgiveness Education Program 56
Taiwan Students 65

Reliability and Validity of the Initial Assessment 65
Effect of the *Agape* Love and Forgiveness Education Program 67
All Teachers Combined 78

Reliability and Validity of the Initial Assessment 78
Effect of the *Agape* Love and Forgiveness Education Program 80
Chapter 5: Discussion 92
General Discussion of the Results 92

Limitations of the Study 111

Directions for Future Research 114

Conclusion 116

References 117
Appendix A: Fidelity Checklist 127
Appendix B: Criteria of Problem Intervention Coding 128
Appendix C: Criteria of Enright *Agape* Love Inventory (EALI) Story Rating 129
Appendix D: Criteria of Enright Forgiveness Inventory (EFI) Story Rating 130

Acknowledgements

No words could be used to express the debt of gratitude I owe my advisor, Dr. Robert Enright, for his guidance, support, and encouragement throughout my doctoral studies, especially with the process of generating this dissertation. I would like to thank Dr. Amy Bellmore, Dr. Sarah Short, Dr. Julie Poehlmann, for their advice on my dissertation proposal. To Dr. Dan Bolt, I am grateful for his providing insightful views on the *agape* love scales' construction and validation. I also want to give thanks to Jennifer Krug, Amy Moschkau, and Luci Trinastic, who have lent great support to my journey in Educational Psychology and to our project.

I would never have been able to accomplish my doctoral studies without the support of my family. My dearest husband, Wang Xuwei, who does not only express *agape* love and forgiveness daily to me to allow me to focus on my work but also provides numerous technical supports to the project using his knowledge of computer science. My two young boys, Alpha who is nine and Nathan who is seven now, grew up to witness my life in graduate school, lent their mom to her research a lot, and also learned about *agape* love and forgiveness at a young age. I owe a debt of gratitude to my parents, Xu Liguo, and He Chuanhong, my in-laws, Wang Baode (who passed away during the pandemic) and Qian Meihua, my grandparents, Jia Qibing and He Yongjing, who have generously supported me emotionally and financially throughout my graduate study. I would like to give thanks to the team members with whom I worked together and from whom I received help: Dr. Chansoon (Danielle) Lee, Ms. Jacqueline Song, Ms. Hannah Rapp, Ms. Moon Evans, and Ms. Katrina Welborn. I would like to thank the leaders and researchers at our three research sites: Dr. Anita Gracie, Ms. Jill Magennis, Dr. Tali Gal, Ms. Laiki Saban, and Dr. Shih-Tseng, Tina Huang, as well as the educators and students who spent endless effort on this project.

Finally, I give thanks to my Redeemer and my Lord, God the Father, God the Son, God the Holy Spirit, who makes all things possible.

Abstract

This is a first study of its kind to investigate the effects of forgiveness education on the development of *agape* love with fifth graders and their teachers in three culturally distinct sites: Northern Ireland, Israel, Taiwan. *Agape* love, drawn from an ancient Greek concept, is defined as a moral virtue in which a person willingly and unconditionally offers goodness, at a cost to the giver, to another or others in need. There is a giving of the self to the other(s) that is: (a) understood (conscious awareness of what this is and so *agape* can occur toward many across time), (b)motivated (I want to do this and for the good of the other), (c) willed (I will do this, again for the good of the other), and (d) acted upon (behavior) toward other people in such a way that the actions cost the one expressing that love (Enright, Wang Xu, Rapp, Evans, & Song, 2022). In my dissertation, I show the effect of a forgiveness education intervention on the development of *agape* love in children (10 to 12 years old) and teachers, based on the analysis of survey data collected from schools in the three sites at four time points throughout the academic year of 2021-22. The following measures were repeatedly administered in the surveys: recently validated *agape* love scales, Enright Forgiveness Inventory, anger, depression, hope, self-esteem, and empathy. The study is based on the hypothesis that through educating children about forgiveness, which is a conscious and deliberate expression of *agape* love, it will foster the moral virtues of *agape* love and forgiveness in the two major developmental domains of cognitive and socio-emotional development with improvement in psychological well-being in the children and the educators.

Chapter 1: Introduction to the Research Idea

Forgiveness has developed as a powerful healing tool for adult survivors of deep, personal, unfair hurt, first by Enright and the Human Development Study Group, in over three decades (Freedman & Enright, 1996; Freedman & Zarifkar, 2015; Hebl & Enright, 1993; Reed & Enright, 2006). The strong evidence of forgiveness therapy (see the four meta-analyses: Akhtar & Barlow, 2018; Baskin & Enright, 2004; Lundahl et al., 2008; Wade et al., 2014) led to the idea of implementing forgiveness education in schools so students can cope with anger and deep hurt. The first forgiveness education to be researched was with college students (Al-Mabuk, Enright, & Cardis, 1995). This then led to the first such program with children, which occurred in Belfast, Northern Ireland (Enright et al., 2007), with other research programs to follow in the United States (Gambaro et al., 2008; Holter et al., 2008; Enright, et al., 2014), Korea (Park et al., 2013), Turkey (Taysi and Vural, 2015), Pakistan (Rahman et al., 2018), and Greece (Vassilopoulos et al., 2020).

Forgiveness education can be conducted by school counselors, psychologists, and teachers with individuals or in classrooms. It also has shown the effect of bullying prevention and intervention (Egan & Todorov, 2009; Enright, 2012a). A recent meta-analysis of research on forgiveness education has identified a significant increase of forgiveness as well as significant reductions of anger (Rapp, Wang Xu, Enright, 2022). In addition, some studies of forgiveness education interventions in schools in different world regions have found a reduction of mental health problems such as depression and the development of positive psychological characteristics such as hope and self-esteem (Hui & Chau, 2009; Shechtman, Wade, & Khoury, 2009; Taysi & Vural, 2016).

Agape love, drawn from an ancient Greek concept, is defined as a moral virtue in which a person willingly and unconditionally offers goodness, at a cost to the giver, to another or others in need. There is a giving of the self to the other(s) that is: (a) understood (conscious awareness of what this is and so *agape* can occur toward many across time), (b) motivated (I want to do this and for the good of the other), (c) willed (I will do this, again for the good of the other), and (d) acted upon (behavior) toward other people in such a way that the actions cost the one expressing that love (Enright, Wang Xu, Rapp, Evans, & Song, 2022). *Agape* differs from all other forms of love, such as *eros* (romantic love), *storge* (the natural love between parent and child), *philia* (the friendship love) and compassionate love (a concern for all of humanity). Compassionate love is a general tenderness (including affective, cognitive, and behavioral attitudes) toward all people (Sprecher & Fehr, 2005). In contrast, *agape* love requires heroic commitment and concrete actions to the betterment of others. *Agape* may aid humanity in reaching its highest level when people are deliberately, consciously, and willingly engaged in cultivating this moral virtue toward one another. Based on the definition of *agape* love as a moral virtue, it is expected that *agape* love can be developed in children as transforming self-good to serving others. Thus, compared to other positive psychological changes, *agape* love is more than a healthy emotional state as it offers kindness and service to others, across cognition, affect, and behavior. Because forgiveness as a moral virtue and in its essence (in other words, as understood on its highest possible level) concerns goodness toward offending others, its highest expression is *agape* love toward those who caused pain and suffering to the forgiver. Thus, the first expectation is that *agape* love can be developed in both students and teachers following Forgiveness Education.

The study of forgiveness development with adolescents and young adults shows that forgiveness as a conscious and deliberate expression of *agape* love reaches its higher degrees as

age increases into adulthood (Enright, Santos, & Al-Mabuk, 1989). Similarly, the developmental hypothesis of *agape* love should differ in children and adults. Based on our definition of *agape* love, it should correlate with at least two major developmental domains (socio-emotional domain and cognitive domain) in children and adolescents. In fact, when *agape* love is fostered in socio-emotional developmental and cognitive developmental domains, it is possible to be expressed in positive behaviors that may transform interpersonal relationships, including the reduction in prejudice toward groups historically in conflict with one another because two central characteristics of *agape* love are to be in service to others and to see the inherent worth of all people. To scientifically assess *agape* love, our research group, led by Dr. Enright, developed, and validated four scales measuring *agape* love and one *agape* love knowledge scale, in both adults' and children' versions, using the samples in the United States.

The *agape* love scales were administered to assess cognitions, emotions, and behaviors associated with the development of *agape* love in children and teachers following Forgiveness Education interventions with elementary students in 5th grade (US equivalency, P7 in Northern Ireland, ages 10-12) in Northern Ireland, Israel, and Taiwan. The three world sites share the fact that ethnic/religious/political groups are in conflict and at the same time each world site is distinctly different in its ethnicity, religious traditions, customs, types of social conflicts, as well as psychological resources on which to draw. The first goal is to explore the development of *agape* love in the context of Forgiveness Education within each culture; then the second goal is to compare the development of *agape* love through Forgiveness Education in distinctly different cultural sites.

Grade 5 is chosen for two reasons: (a) children in this grade are firmly in the cognitive developmental stage of concrete operations and so they will be able to understand causes and

consequences of people's actions (including unjust actions) and such positive development as *agape* love; and (b) they are just entering early adolescence in which peer interactions become more important. Thus, the development of peer harmony becomes quite essential at this grade, and at the same time, stereotypical prejudices can form toward groups that are different from their own. Forgiveness education, then, should be effective in the students' understanding of what forgiveness is and is not, the nature of injustice and the effects of being treated unjustly. At the same time, they should be able to reduce any developing prejudices against other peer groups, other pathway groups, and groups toward which their ancestors have been in conflict.

The cultures chosen will allow the examination of diverse racial and ethnic backgrounds relative to the impact of Forgiveness Education. Although Forgiveness Education has been done with varying populations of children in different cultures, none of the studies compares the cross-cultural differences in developing forgiveness and its impact on other negative and positive psychological constructs. Neither has any study ever been done to examine the impact of Forgiveness Education on developing *agape* love. As Aristotle taught, there is a difference between the essence of any moral virtue (what it is at its unchanging core) and the existence of that virtue (how it is expressed in different cultures and different situations by different people). The essence of either *agape* love or forgiveness is the same across all cultures, yet it will be very interesting to examine the differences in expressing and practicing these two moral virtues in different cultures.

In this doctoral dissertation, Chapter 2 includes a literature review and hypotheses: a. to further discuss the relatively well-developed psychological construct of forgiveness in connection with the relatively new psychological construct of *agape* love and to introduce the

agape love scales; b. to review the existing research of forgiveness education and to introduce the curriculum used in current study; c. to present the hypotheses in current study.

Chapter 3 introduces methods of the research, including recruitment of participants, teacher training, intervention curriculum, experimental design and testing procedures, as well as research instruments.

Chapter 4 presents the results of students' responses by each research site and the results of teachers' responses from three sites combined.

Chapter 5 gives a general discussion of the results, limitations of the study, and directions for future research.

Chapter 2: Literature Review and Hypotheses

Theoretical and Philosophical Perspectives on Forgiveness and *Agape* Love

The theme of forgiveness in texts could be traced as early as in the book of Genesis in Hebrew Bible -- Joseph, who had been awfully treated by his ten half-brothers at a young age, decided to forgive and extended mercy on them when his brothers came to Egypt to seek famine relief while Joseph was an adult and was the authority of Egypt at that time. Being treated unjustly is always associated with forgiveness. It is almost impossible for a person to never have experienced injustice in life. I talked with my good friend, who is now 85-year-old, about experiences of injustice. She grew up in a Christian family with relative stable social and financial status, good relationships with parents and siblings, as well as a happy and long marriage with a husband with almost no argument. She could still recall five incidents in her life in which she was unjustly treated by someone. When facing injustice in life, forgiveness is a choice. There are some other choices such as revenge or ignorance. Scholars of forgiveness in general agree on what forgiveness is not: it is not forgetting, condoning, excusing offenses, and forgiveness does not necessarily lead to reconciliation (Enright & Fitzgibbons, 2015; Exline et al., 2003). Both philosophers (Holmgren, 1993; North, 1987) and psychologists (Enright & Fitzgibbons, 2015) describe forgiveness as a moral virtue, for the one who practices it knows it is good and is motivated to effect moral goodness, first originating within the person, and then bringing this goodness to others for their good (Simon, 1986, following Aristotelian approach on identifying qualities of moral virtues). Based on the philosophical view of forgiveness as a moral virtue, it is defined as: forgiveness begins as a conscious decision to abandon one's right to resentment, judgement, and indifferent behavior toward the one who offended and involves fostering the undeserved qualities of compassion, generosity, and even love toward the one who

offended (Enright et al., 1998; see Exline et al., 2003 and Worthington, 2020 for definitional issues). There are two principles of Aristotelian philosophy that can apply to forgiveness as a moral virtue (Simon, 1986): (a) there exists an objective and true perfection (essence) of forgiveness that is not altered by the variations in the expression across historical, religious, and cultural groups; (b) most people do not reach the perfection of practicing forgiveness. Thus, the process of developing forgiveness could be viewed as a spectrum. It takes time and effort for one, who was unjustly treated, to practice and to gradually transform the emotions, cognitions, and behaviors toward the offending person. The forgiver moves from withholding the right of resentment to willfully abandoning this right and responds to the other person, who did the injustice, with some goodness, including kindness, respect, generosity, and love. Empirical studies in the field of psychology have accumulated evidence that individuals, who had forgiveness interventions, can gain relief from mental health issues such as chronic anger, anxiety, and depression and develop positive constructs such as self-esteem and hope (Wade et al., 2014).

Early empirical studies done with forgiveness interventions in the field of psychology view forgiveness as a psychotherapeutic process, taking time for the reduction in anger and the increase of positive psychological characteristics to emerge (for example, see Freedman and Enright, 1996). Wade et al. (2014) suggested that as a psychotherapeutic process, forgiveness should be promoted not simply for reducing mental health problems but for developing strengths so that the clients will move toward more positive, even optimal functioning. Enright (2012b, chapter 1), in his book <The Forgiving Life>, stated that forgiveness can be described as a practice of love at a time when it is challenging to love. Thus, forgiveness should not be viewed as a psychotherapeutic technique, in which certain behaviors are attempted in a short amount of

time. Only by viewing forgiveness as a moral virtue, which takes time, effort, and practice to develop, and practicing it with other moral virtues, such as *agape* love (will be introduced in the next paragraph), can forgiveness develop more fully. By continuous growing in the moral virtues of forgiveness and *agape* love, people might develop more optimal psychological and physical functioning.

Agape love has never been clearly defined as a term in social science. It has been referred to as "unconditional love" (Wivestad, 2008) or "altruistic love" (Post, 2002; Hendrick & Hendrick, 1986). It has been conflated with other constructs such as altruism. Post (2002) points out that *agape* love goes so far as to even include the love of one's enemies, but in the article, there is no clear exposition of how *agape* love is distinguished from altruism. Because of the unclear definition of *agape* love, it is regarded as having interchangeable or similar meaning with compassionate love, unlimited love, or altruistic love (Oman & Meyer, 2013). However, neither unlimited love nor compassionate love captures the essence of the moral virtue of love as *agape* does.

Agape love, as a moral virtue, can be clearly defined under the five categories Aristotle used to explicate a construct: **species (or essence)** of the construct; **genus** is what in common between the construct and other related concepts; **specific difference** is how the construct differs from other related concepts; **accident** is what is atypical but still within the essence of the construct; **property** is something the construct possesses but not necessary the essence.

To apply the five categories to define *agape* love, the **essence (or species)** is:

> *Agape* love is a moral virtue in which a person willingly and unconditionally offers goodness to another or others, in need, for the good of the other(s). There is a giving of the self to the other(s) that is understood (conscious awareness of what this is), motivated (I want

to do this), willed (I will do this), and acted upon (behavior) toward other people in such a way that the actions cost the one expressing that love (Enright et al., 2022).

First, the **genus** and **specific differences** can be used to compare *agape* love with other moral virtues. As a moral virtue, *agape* love shares some common qualities with all the other moral virtues, such as justice, courage, temperance, forgiveness. Moral virtues originate within the person and then are brought forth to others for good. The person expressing moral virtues is motivated and willed to deliberately effect moral goodness with moral emotions (Simon, 1986). The **specific difference** between *agape* love and other moral virtues is its giving of the self for the good of the other. Although all moral virtues are expressed for the others' good, they are not as costly as *agape* love to set forth the other's needs as the central focus. For instance, if we compare two examples: a millionaire gives $100 to a charity compared with a person who owns $200 but gives $100 to a starving friend, the former one is regarded as altruism while the latter one is an expression of *agape* love.

Second, the **genus** and **specific differences** can be used to compare *agape* love with other forms of love. *Agape* love shares common characteristics with all of the other qualities of love defined in ancient Greece (*eros*, *storge*, *philia*): all are concerned for the good of the other and view the other as a person who is special, unique, and irreplaceable; all require an investment of time given to the other; all have the quality of affection, although this is not in the essence but a property of *agape* love. The **specific difference** distinguishes *agape* love from other forms of love (*eros*, *storge*, *philia*) is that *agape* love is more costly in terms of consciously, willingly, and actively giving up energy, material possessions, comfort, and/or safety for the good of another person or other people. *Agape* love does not necessarily share the mutuality that is embedded in the other forms of love as the one who expresses *agape* love does not necessarily expect the

other to love back. The one who expresses *agape* love does so out of the other's need and not out of self-interest.

Accident in *agape* love could be when a person willingly gives one's life when expressing *agape* love to serve others. Not all who express *agape* love should be expected to give up one's life in service to the other, nor should this be regarded as the highest degree of expressing *agape* love.

The affection of *agape* love, as a quality shared with the other forms of love, is a **property** in *agape* love. The feeling of love is a part of but not the essence of *agape* love. The essence of *agape* love is that it is expressed through cognitively conscious, motivated, and willed actions. We are not necessarily in full control of our emotions, so we cannot rely on our feeling toward the other person to express *agape* love. Neither can we use our feelings to judge whether other people express *agape* love to us because *agape* love does not necessarily involve the feeling of tenderness. Instead, an action that is understood, motivated, and willed is a true and concrete expression of *agape* love, which is described in the essence of *agape* love.

How are *agape* love and forgiveness connected to each other? Forgiveness is more specifically targeted toward people who have treated us unjustly. *Agape* love can occur toward those who are fair and good to us or who are so suffering that they are not focused on us. The moral virtue of forgiving flows from the moral virtue of *agape* love. The highest form of forgiveness is when we strive, even struggle, to love the one who hurt us (Enright, 2022b). Using a concrete example, at a higher end of the "forgiveness spectrum", more *agape* love is expressed. No research to date has been conducted to examine the development of forgiveness and *agape* love. However, a study with the longitudinal design indicated that the development of forgiveness temporally precedes other-focused love (Wang Xu, Kim, Olmstead, Enright, 2021).

Based on this finding, my keen interest now is to examine the effect of forgiveness interventions on the growth of the moral virtue of *agape* love.

So, it is important to construct *agape* love measures to conceptualize its essence as a moral virtue. The scales should be constructed from the concrete definition of *agape* love and a sound theological and philosophical foundation. Both help to determine what should be included in the measure to reflect the construct as well as what should not be included in order to avoid equivocation with other constructs in item generation. The measurement is constructed to investigate the expression of *agape* love as a moral virtue in concrete actions by assessing the inner qualities of motivation, cognition, affection, and affinity of the virtue. To measure *agape* love is significant because *agape* love does not only lead to the enormous social good for those who are its recipients, but also benefits the ones who express it. This measure eventually may have wide application because there is no single human in this world who has never encountered a situation in which the person should choose whether to give up one's own needs, no matter physical or psychological, and help with another person's (or other people's) needs in a costly manner. The only difference is that, in this kind of situation, to what extent do individuals willingly and unconditionally perform self-giving actions that cost themselves? The measurement is helpful for social scientists to examine what are the benefits to these who express *agape* love, when it is assessed with other positive psychological constructs such as self-esteem and forgiveness, negative psychological traits such as anger, anxiety, and depression, as well as other mental or physical health indicators. In addition, a scale that has the capacity to measure growth is important to promote the development and implementation of meaningful interventions that target the cultivation of this moral virtue – *agape* love.

My Master's thesis presented the work of constructing *agape* love scales of adults' and children's versions and validating them using a US adults' sample as well as a US children's sample. The five scales of *agape* love are introduced in detail in Chapter 3, Method.

Forgiveness Education Interventions and the Curriculum

A recent meta-analysis done by Rapp et al. (2022) on randomized intervention studies of forgiveness education defines Forgiveness Education intervention as a group intervention that teaches children and adolescents about what forgiveness is, the benefits of forgiveness (i.e., its role in healthy relationships), as well as the process(es) of forgiveness. All forgiveness education interventions were created for children and adolescents (elementary through high school) to be administered in classrooms or groups. Forgiveness education interventions are differentiated from therapeutic forgiveness interventions by the following two points: 1) forgiveness education interventions are only designed for schools whereas therapeutic forgiveness interventions are not, although both can take place in group settings; 2) most forgiveness education interventions have the goal of first helping youth to develop the knowledge of forgiveness. Only then are they encouraged, if they choose, to develop the skill of forgiveness to achieve more social and mental well-being. The 20 studies included in the meta-analytic review involving 1,472 youth (51% female; *M*age = 11.66) from 10 countries (studies: 40% North American, 25% East Asian, 20% Middle Eastern, 15% European). Each study at least includes one forgiveness and/or one anger measure. Findings suggest that forgiveness education interventions have a significant positive effect on forgiveness (g = 0.54, 95% CI [0.36, 0.73]) and anger (g = 0.29, 95% CI [0.11, 0.47]). The other measures taken in the studies can be categorized into a) psychological measures such as aggression, anxiety, depression, empathy, hope, and self-esteem; b) behavioral measures such as bullying, delinquency, and social skills; and c) academic performance. Not all other outcomes

but depression, hope, self-esteem, and empathy are examined in the meta-analysis because too few studies adopted the same category of measures. The efficacy identified as well as the measures taken in existing empirical studies of forgiveness education interventions resonate with the goal of forgiveness education intervention: helping youth to develop the knowledge and skill of forgiveness to grow into socially and mentally healthier individuals. Thus, Rapp et al. (2022) compares forgiveness education interventions to social-emotional learning (SEL), which refers to the development of specific skills and competencies that students need to set goals, to manage behavior, to build relationships through the development of empathy, cooperation, emotion management, and so forth. In recent years, schools in the United Stated have recognized the deep connection between social emotional skills and traditional academic skills and have adopted SEL programs into the curricula (Jones & Kahn, 2018). However, forgiveness education interventions are more than teaching social emotional skills. Lin, Enright, and Klatt (2011) argued that forgiveness education should be counted into character education, which is defined by the Character Education Partnership (CEP) as "the deliberate effort by schools, families, and communities to help young people understand, care about and act upon core ethical values" (Lickona, 1996), because forgiveness itself is a moral virtue. By practicing forgiveness, a person cultivates other virtues such as compassion, generosity, and *agape* love.

"*...nothing is of more importance for the public weal, than to form and train up youth in wisdom and virtue.*" – Benjamin Franklin (Franklin, 1750)

I was reminded by our Founding Father of the United States of the importance to cultivate moral virtues in youth. This current research study, which implements the curriculum centered on teaching the two vital moral virtues of forgiveness and *agape* love in Grade 5 (P7 in Northern Ireland) in three countries echoes the vision and the call from Benjamin Franklin.

Research Question

Based on the literature review, the first goal of this study is to see if cultivating the moral virtues of *agape* love and forgiveness through the forgiveness education program will result in the growth of these two moral virtues, which can be reflected in cognition (understanding *agape* love and forgiveness), emotions (feelings associated with *agape* love and forgiveness), behaviors (practicing *agape* love and forgiveness). Then, the research will examine if such learning with development in moral virtues will impact other factors such as psychological well-being, class cooperation, as well as academic results. In this study, the investigation is with the students and the teachers, who are involved in the forgiveness education program together.

Hypotheses

Hypotheses Focused on the Students

Hypothesis I: through the forgiveness intervention (relative to those in the control group), students will grow significantly in *agape* love in the relationship in their family, school community or other social context, regardless of the culture and ethnicity to which they belong.

Hypothesis II: across all cultures, participants in the forgiveness intervention will demonstrate significant increases in forgiveness relative to those in the control group.

Hypothesis III: across all cultures, the forgiveness intervention will reduce anger relative to those in the control group.

Hypothesis IV: across all cultures, the forgiveness intervention will reduce depressive symptoms relative to those in the control group. Participants may have to be screened for clinical levels of depression at pretest and only those who show clinical compromise on this variable will be investigated.

Hypothesis V: across all cultures, the forgiveness intervention will increase the positive psychological characteristics of hope, relative to those in the control group.

Hypothesis VI: across all cultures, the forgiveness intervention will increase the positive psychological characteristics of self-esteem, relative to those in the control group.

Hypothesis VII: the forgiveness intervention will increase classroom cooperation, as rated by teachers, on the level of those whom teachers find disruptive in the forgiveness group compared to those in the control group.

Hypothesis VIII: the forgiveness intervention will increase academic grades for those in the forgiveness group compared to those in the control group. Participants may have to be screened for academic performance at pretest and only those who are low on this variable will be investigated.

The above hypotheses will be tested twice: a) between the experimental and control groups from the pre-test (T1) to the post-test (T2) in the first forgiveness intervention and b) compare the control group from the pre-test (T1) to the post-test (T2) in the first forgiveness intervention to the control-group-turned-experimental group from their first post-test (T2) to their second post-test (T3) in the second forgiveness intervention.

It is expected that the effect for the first and the second forgiveness interventions will be the same. So, across all cultures, that the gain achieved from the pre-test (T1) to the post-test (T2) in the first forgiveness intervention for the original experimental group will be statistically equal to the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second forgiveness intervention, in all the variables measured for the students.

In addition to the hypotheses, a research question is this: do cultural differences exist in the development of *agape* love through Forgiveness Education? As forgiveness is a dynamic process affected by society and culture (Ho & Fung, 2011), the results may show differences across cultures. The specifics of these differences will be more exploratory than hypothesis-testing and will be examined to explain the development of *agape* love through forgiveness in different social and cultural contexts.

Hypotheses Focused on the Teachers

Hypothesis I: through teaching Forgiveness Education curriculum (relative to those in the control group), teachers will grow significantly in *agape* love in general. Because of the smaller sample size relative to the students, teachers' data of three research sites will be combined.

Hypothesis II: across all cultures, the teachers who teach Forgiveness Education curriculum will increase significantly in forgiveness relative to those in the control group.

Hypothesis III: across all cultures, the teachers who are in the Forgiveness Education program will reduce anger relative to those in the control group.

Hypothesis IV: across all cultures, the teachers who are in the Forgiveness Education program will reduce depressive symptoms relative to those in the control group. Participants may have to be screened for clinical levels of depression at pretest and only those who show clinical compromise on this variable will be investigated.

Hypothesis V: across all cultures, the teachers who teach Forgiveness Education will increase in positive psychological characteristics of hope, relative to those in the control group.

Hypothesis VI: across all cultures, the teachers who teach Forgiveness Education will increase in positive psychological characteristics of self-esteem, relative to those in the control group.

Hypothesis VII: across all cultures, the teachers who teach Forgiveness Education will increase in positive psychological characteristics of empathy, relative to those in the control group.

The above hypotheses will be tested twice: a) between the experimental and control groups from the pre-test (T1) to the post-test (T2) in the first forgiveness intervention and b) compare the control group from the pre-test (T1) to the post-test (T2) in the no treatment period (teaching regular school curriculum) to the control-group-turned-experimental group from the first post-test (T2) to the second post-test (T3) in the second forgiveness intervention.

It is expected that the effect for the first and the second forgiveness interventions will be the same. So, across all cultures, that the gain achieved from the pre-test (T1) to the post-test (T2) in the first forgiveness intervention for the original experimental group will be statistically equal to the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second forgiveness intervention, in all the variables measured for the teachers.

Hypotheses Related to the Maintenance of Effects

Hypothesis I: across all cultures, the original experimental group's results at their post-test (T2) will be statistically similar to the scores achieved at their first follow-up test (T3) and the second follow-up test (T4) for all the variables measured for students and teachers.

Hypothesis II: across all cultures, after the second forgiveness intervention: the control-group-turned-experimental group results at their second post-test (T3) will be statistically similar to the scores achieved at their follow-up test (T4) for all variables measured for students and teachers.

Hypothesis III: across all cultures, the original experimental group's results at their second follow-up test (T4) will be statistically similar to the scores achieved by the control-group-turned-experimental group's scores at their follow-up test (T4) for all variables measured for students and teachers.

Hypothesis IV: across all cultures, the original experimental group's results at their first follow-up test (T3) will be statistically similar to the scores achieved by the control-group-turned-experimental group's scores at their follow-up test (T4) for all variables measured for students and teachers.

Chapter 3: Methods

Participants Recruitment

The procedures of recruitment, survey, and interview administration, educational psychological intervention, and data collection were approved by the ethics committees of the institutions at the three research sites: Stranmillis University College in Northern Ireland, the University of Haifa in Israel, and the National Chung Cheng University in Taiwan, as well as the University of Wisconsin-Madison's Institutional Review Board.

The researchers in each site's project team reached out to principals of local elementary schools to gain approval to conduct the education program and the experiment (surveys, interviews, and access to academic results) in schools in the academic year of 2021 to 2022. In Northern Ireland, 20 schools (eleven Maintained Schools & nine Controlled Schools) were recruited at the beginning of the study, three (two Maintained & one Controlled) dropped out due to various reasons during the study. For the remaining 17 schools, one Primary 7 (equivalent to Grade 5 in the US) class from each of 16 schools (in total 16 classes) participated from the beginning to the end of the study; one Primary 7 class from one school didn't submit the final survey at T4. In Israel, 20 schools (10 Hebrew-speaking; 10 Arabic-speaking) were recruited, with 19 schools having one Grade 5 class and one school having three Grade 5 classes (in total 22 classes) participated from the beginning to the end of the study. In Taiwan, 10 schools were recruited, one dropped out during the study. With the remaining nine schools, each school has two to three classes (in total 19 classes) participated from the beginning to the end of the study.

Parents or guardians voluntarily signed consent forms and the student him/herself signed minor assent form independently, allowing each student then to participate in the research study

(four times of survey/interview/allow access to the grades). Teachers signed consent forms to participate in the four times of surveys and/or the interviews.

Teacher Training and Fidelity Check

Before each Forgiveness Education intervention started, there was mandatory teacher training for every educator who would teach the program. The 6-hour teacher training consists of two parts: 1) a 3-hour Online Teacher Training Course with the main goal of introducing the educators to the moral virtue of forgiveness and the art of Forgiveness Education. 2) After taking the Online Teacher Training Course in a self-paced manner, the educators participated in a 3-hour interactive Teacher Training Workshop on Zoom with researchers from the United States as well as researchers from the local research team. The Online Teacher Training Course was developed by the US project team (including myself) and was translated into Arabic, Hebrew by researchers in Israel for educators from Israel, as well as Traditional Chinese by myself and two other US researchers who are fluent in Chinese for the educators in Taiwan to view. The Teacher Training Workshop conducted on Zoom with educators from Israel has two sessions, one with Hebrew interpreters presented and the other with Arabic interpreters presented. The Teacher Training Workshop conducted on Zoom with educators from Taiwan was interpreted by me if there was an English speaker or directly delivered in Mandarin by myself. The content of the workshop consists of four parts: a) introducing to the two important concepts of *agape* love and forgiveness; b) introducing to the research components of the project (surveys and interviews); c) providing guidance and instructions on how to use the Forgiveness Education curriculum; and d) introducing to the online resources for the education program.

Fidelity checks support whether an intervention program is being properly implemented and help to target training needs. We checked the fidelity of the curriculum teaching using two

mechanisms. First, in each research site, one lesson from each experimental group (the original experimental group and the control-group-turned-experimental group) was randomly selected for a fidelity check and evaluation. Some schools' policies did not allow for audio or video-recording students and teachers, so a fidelity evaluation checklist (Please see Appendix A: Fidelity Checklist) was used by a member of the site research team during an in-person observation. Fidelity checklists included core components found in the curriculum guide's lesson procedures, such as whether the major activities and the discussion questions were presented in the sessions. Second, each lesson in the curriculum guide included student & lesson evaluation questionnaires that teachers could use to reflect on whether their students actively participated, learned the concepts in a concrete way, and responded well or not well to any ideas, discussion questions, or activities from the lesson. We asked the randomly selected teachers from each site to send us the two filled evaluation forms to determine students' active participation and understanding to further support implementation fidelity.

Intervention Curriculum

The Forgiveness Education curriculum used in this current study is broadly defined as an Enright story-based curriculum (Rapp et al., 2022), published by Enright and the International Forgiveness Institute Inc., spanning pre-K through grade 12 (Enright & Knutson, 2010). The Enright story-based curriculum uses children's stories to teach about forgiveness and other related moral virtues and equips children with the knowledge of how to forgive a specific person who offended through the interactions of story characters. With the original author, Dr. Enright's permission, I revised the curriculum by adding the content that deliberately teaches about the moral virtue of *agape* love, including explaining what *agape* love is as well as how it is developed and reflected in the process of forgiveness.

Because the Forgiveness Education intervention study is with Grade 5 students (US equivalency, P7 in Northern Ireland), the stories chosen in the curriculum should appeal to children ages 10 through 12, as the stories are based on the children's developmental level. This curriculum focuses on a tender-hearted effort with expressing *agape* love in forgiving, seeking forgiveness, and receiving forgiveness. *Agape* love is emphasized throughout the lessons that as people forgive, they are lavishing love onto the offending person. *Agape* love is taught as the central moral principle in forgiving, which could also be applied outside the context of forgiving regarding how we generally treat other people.

This curriculum for 10-12-year-old students comprises 14 lessons. In lessons 1 through 5 the students study: what forgiveness is and is not; inherent worth and *agape* love within the context of giving, seeking, and receiving forgiveness; and the importance of achieving a balance in applying *agape* love and forgiveness. In lessons 6 through 13, the students learn the actual process of how to forgive. In lesson 14, the students learn the positive impacts of forgiveness and *agape* love in wider communities.

Because the curriculum was used in Taiwan, as well as in Israel in both Hebrew-speaking and Arabic-speaking schools, the curriculum was translated from English to traditional Chinese, Hebrew, and Arabic. The translation was done by professional translators and by researchers in the project teams. I helped with editing the traditional Chinese version of the curriculum. We had discussions with the local researchers and within our own research group to make sure that the words "*agape* love" and forgiveness were translated into the words in the native languages with the closest and comparable meaning as we understand them in English.

Experimental Design and Testing Procedures in the Intervention

Overall, the experimental design for the study is a multi-site, multi-level cluster-randomized trial with pre-, post-tests, and follow-ups. In each research site, students and teachers are nested within classes and classes are nested within schools to receive the Forgiveness Education; schools are randomized and nested within the treatment and control groups. The experimental design is the same across the three research sites (Northern Ireland, Israel, Taiwan). All of the procedures were completed in the same academic year of 2021-22 in three sites.

A crossover experimental design was applied: the experimental group had the first 12-weeks forgiveness intervention according to the Forgiveness Curriculum. The control group taught a regular school curriculum. Once the experimental group completed the intervention, the control group became a new experimental group and began the second 12-week forgiveness intervention, while the original experimental group taught a regular school curriculum. Thus, there is a replication of the forgiveness intervention (original experimental group and the control-group-turned-experimental group) within each research site.

Survey data were collected at four time points in the experiment: before the first 12-week Forgiveness Education intervention, after the first 12 weeks of intervention, after the second 12-week Forgiveness Education intervention, and four weeks after the second 12-week Forgiveness Education intervention ends. To incorporate the data collection into the experimental design, the procedure is: at the beginning of the academic year before any instruction for this intervention starts, the first measurement (T1) was the pre-test data collected from both the experimental group and control group. After collecting the pre-test data, the experimental group began the first 12-week Forgiveness Education intervention. The control group had instruction as usual (with no Forgiveness Education intervention). After the 12-week forgiveness intervention, the second

measurement (T2) was the post-test data collection for the experimental group, and the first post-test data collection for the control group. Then, after the school break, the previous control group turned into a new experimental group. The control-group-turned-experimental group then went through the 12-week forgiveness intervention. The original group had instruction as usual (with no forgiveness intervention). After this second 12-week forgiveness intervention, the third measurement (T3), again for both groups, includes: the second post-test data set for the control-group-turned-experimental group and the first follow-up data set for the original experimental group. After another four weeks, a fourth measurement (T4) collected a follow-up data set from the control-group-turned-experimental group and a second follow-up data set from the original experimental group. The timing of the intervention and the waves of data collection are shown in Table 1.

Parents read and signed consent forms and students had an assent form prior to participating in the study. Teachers also signed a consent form so that we could give them assessments. Participants were free to withdraw at any time without consequence. In Northern Ireland and Israel, students and teachers took the electronic surveys in Qualtrics. On the survey administration day, participants of the class used a computer room to complete the survey, with at least one researcher from the local project team present. In Taiwan, students and teachers took paper-and-pencil surveys in classrooms, with at least one researcher from the local project team present. Because each survey administration was scheduled between the researcher and the class at a particular time on a school day, if the student was absent from the school, the class teacher sent a reminder for the absentee to complete the questionnaire afterwards. In the Spring of 2022, the Coronavirus disease (COVID-19) pandemic was severely impacted by the Omicron variant.

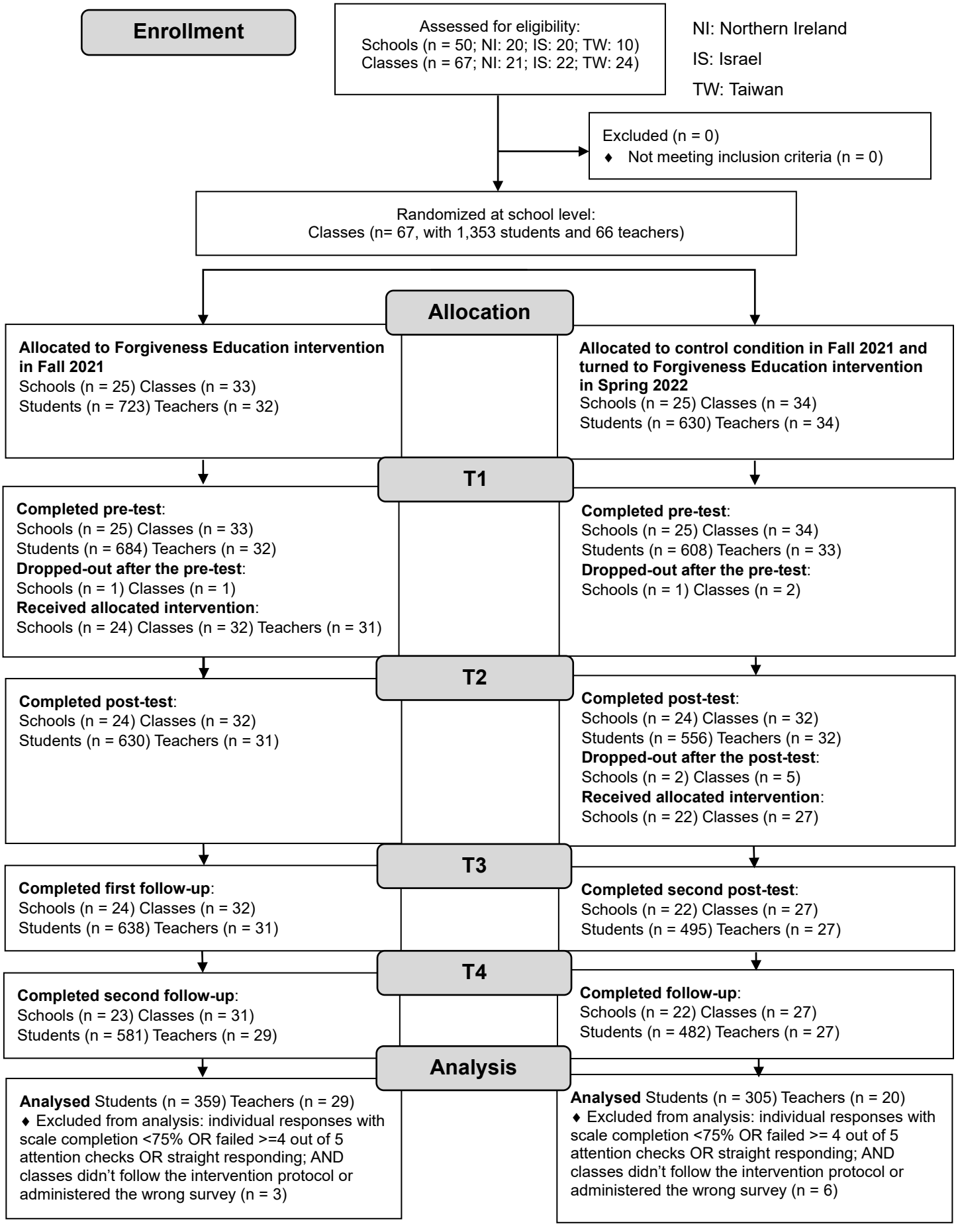
More school absentees were observed and some of the survey administrations were shifted online instead of in-person.

Table 1

Time		Group	Task
Fall of 2021	T1: <= one week before the first intervention	Experimental group	Pre-test
		Control group	
	12 weeks	Experimental group	First Forgiveness education intervention
		Control group	Introduction as usual (No intervention)
	T2: <= one week before the first intervention	Experimental group	Post-test
		Control group	
School break			
Spring of 2022	12 weeks	Experimental group	Introduction as usual (No intervention)
		Control-turned-experimental group	Second Forgiveness education intervention
	T3: <= one week after the second intervention	Experimental group	First follow-up test
		Control-turned-experimental group	Second post-test
	T4: Four weeks after the second intervention	Experimental group	Second follow-up test
		Control-turned-experimental group	Follow-up test

Timing of the intervention and the waves of data collection

The following consort table outlines the testing procedures:


Enrollment
Assessed for eligibility:
Schools (n = 50; NI: 20; IS: 20; TW: 10)
Classes (n = 67; NI: 21; IS: 22; TW: 24)
NI: Northern Ireland
IS: Israel
TW: Taiwan
Excluded (n = 0)
♦ Not meeting inclusion criteria (n = 0)
Randomized at school level:
Classes (n= 67, with 1,353 students and 66 teachers)
Allocation
Allocated to Forgiveness Education intervention in Fall 2021
Schools (n = 25) Classes (n = 33)
Students (n = 723) Teachers (n = 32)
Allocated to control condition in Fall 2021 and turned to Forgiveness Education intervention in Spring 2022
Schools (n = 25) Classes (n = 34)
Students (n = 630) Teachers (n = 34)
T1
Completed pre-test:
Schools (n = 25) Classes (n = 33)
Students (n = 684) Teachers (n = 32)
Dropped-out after the pre-test:
Schools (n = 1) Classes (n = 1)
Received allocated intervention:
Schools (n = 24) Classes (n = 32) Teachers (n = 31)
Completed pre-test:
Schools (n = 25) Classes (n = 34)
Students (n = 608) Teachers (n = 33)
Dropped-out after the pre-test:
Schools (n = 1) Classes (n = 2)
T2
Completed post-test:
Schools (n = 24) Classes (n = 32)
Students (n = 630) Teachers (n = 31)
Completed post-test:
Schools (n = 24) Classes (n = 32)
Students (n = 556) Teachers (n = 32)
Dropped-out after the post-test:
Schools (n = 2) Classes (n = 5)
Received allocated intervention:
Schools (n = 22) Classes (n = 27)
T3
Completed first follow-up:
Schools (n = 24) Classes (n = 32)
Students (n = 638) Teachers (n = 31)
Completed second post-test:
Schools (n = 22) Classes (n = 27)
Students (n = 495) Teachers (n = 27)
T4
Completed second follow-up:
Schools (n = 23) Classes (n = 31)
Students (n = 581) Teachers (n = 29)
Completed follow-up:
Schools (n = 22) Classes (n = 27)
Students (n = 482) Teachers (n = 27)
Analysis
Analysed Students (n = 359) Teachers (n = 29)
♦ Excluded from analysis: individual responses with scale completion <75% OR failed >=4 out of 5 attention checks OR straight responding; AND classes didn't follow the intervention protocol or administered the wrong survey (n = 3)
Analysed Students (n = 305) Teachers (n = 20)
♦ Excluded from analysis: individual responses with scale completion <75% OR failed >= 4 out of 5 attention checks OR straight responding; AND classes didn't follow the intervention protocol or administered the wrong survey (n = 6)

For participants who took the electronic surveys in Qualtrics, their answers from the pre-test (T1) on the first five questions of Enright *Agape* Love Inventory (EALI) and first five (or six in the children's version) questions of Enright Forgiveness Inventory (EFI-30) were recorded and linked to the following three surveys at T2, T3, T4, so their answer-input at T1 showed up under the same questions at T2, T3, T4. For participants who took paper-and-pencil surveys, their answers on the first five questions of Enright *Agape* Love Inventory (EALI) and first five (or six in the children's version) questions of Enright Forgiveness Inventory (EFI-30) were photocopied for the pre-test and inserted in the paper questionnaires administered at T2, T3, T4. This is to ensure that each participant focused on the same event with the same person involved for EALI and for EFI separately at all testing phases.

Research Instruments

Enright *Agape* Love Inventory. The Enright *Agape* Love Inventory (Enright, 2020), as with the Enright Forgiveness Inventory, assesses **the degree** of *agape* love toward a particular person in a particular situation. The EALI is composed of two sections: The first five questions ask information about a specific situation of helping another person followed by the inventory with 36 items. There is a set of instructions before the first five questions: three examples of expressing *agape* love are provided. The first example is "Suppose a very rich person gives $100 to a charity. This is helping, but not likely to be painful because of the giver's wealth. Yet, if someone has a savings of only $200 and gives $100 to a starving friend, this costly gesture likely will involve pain for the giver because of the hardship now created for the self.". The second example is "An extremely exhausted mother, who is up all night with a sick child, is feeling the pain of fatigue." The third example is "After deep conflict with a mother-in-law, a son-in-law who answers the phone, also is helping the other by taking the call and it can be emotionally

painful for him." Then we asked the respondent "to think of one experience where your helping another person(s) caused you pain, either emotionally or physically, or in some other way. Take a moment to visualize the event of that interaction. Try to see the person and try to experience what happened.

Following the instructions, the five questions ask for information about the event the respondent helped another person(s) with pain. The EALI has twelve subscales as follows:

1. Current feelings (6 items). This subscale assesses the respondent's current feeling about the situation in the past. This subscale is used for accessing the change on the degree of *agape* love over time.
2. Past feelings (2 items). This subscale assesses the respondent's recalling of the feeling at the time of the event.
3. Attitude toward the pain (2 items). This subscale assesses whether the respondent willingly embrace the pain or resist that pain.
4. Purpose of enduring the pain, part I (2 items). This first subscale of purpose assesses whether the respondent focuses on utilitarian purposes or wanting something good to occur for the one helped as the reason for enduring the pain.
5. Purpose of enduring the pain, part II (2 items). This second subscale of purpose assesses whether the respondent focuses on who the served-one is as a person as the reason for enduring the pain.
6. My identity as an *agape* love servant (6 items). This subscale assesses who the respondent is as a person relative to the *agape* love offered to another. This subscale is used for assessing the change in the degree of *agape* love over time.

7. Purpose of my actions (as distinct from the purpose for enduring the pain), part I (2 items). This subscale assesses whether the respondent wanted the action to achieve something good for the one helped as the purpose of the action.
8. Purpose of my actions (as distinct from the purpose for enduring the pain), part II (2 items). This subscale assesses whether the respondent focuses on who the served-one is as a person as the purpose of the action.
9. Willingness to bear pain (even before the incident occurs) subscale (2 items). This subscale assesses the respondent's readiness to engage in *agape* love.
10. Self-centeredness subscale (2 items). This subscale assesses the vice of self-centeredness, or who comes first as a motivation to action, the other or the self.
11. Vice of self-exhaustion by meeting wants not needs subscale (2 items). This subscale assesses whether the respondent thinks the helping was overdone to the detriment to the self.
12. Finding meaning or Outcomes of this action subscale (6 items). This subscale assesses the meaning the respondent found in that action, which focuses on what is learned, in an internal sense. This subscale is used for assessing the change in the degree of *agape* love over time.

Each item represents a specific situation that might occur under the given condition. For example, "I am not quite sure why I helped the person" is an item that assesses the presence of negative anthropological purpose of the action. Respondents focused on a particular experience of helping someone that involves physical or psychological pain and responded to each statement on a 6-point Likert scale of agreement-disagreement. The score range is 36 to 216 with a larger value indicating a response with more *agape* love. The EALI has been validated using a sample

of 709 adults in the US with a strong reliability of 0.93. In addition, EALI has a pseudo-*agape* love scale for internal validation of the EALI, which asks participants to evaluate whether the incident was truly painful by helping another person(s) who needs that help. Participants responded on a six-point Likert scale ranging from "strongly disagree" to "strongly agree" to items such as "I actually did not have much pain." Total pseudo-*agape* love scores range from 6 to 30. Participants from the initial data collection with pseudo-*agape* love scores higher than 20 were eliminated from data analysis. The 1-item *Agape* love scale that, which is an independent measure for content validity. Three subscales (EALI-18), #1 current feelings, #6 one's current identity as an *agape* love servant, and #12 finding meaning, are used to measure the change in *agape* love across time.

Enright *Agape* Love Inventory for Children (EALI-C). The 24-item short form of Enright *Agape* Love Inventory was adapted into children's language. The score ranges from 24 to 144 with a higher score corresponding to a larger degree of *agape* love. Total pseudo-*agape* love scores range from 6 to 30. Participants from the initial data collection with pseudo-*agape* love scores higher than 26 were eliminated from data analysis. Three subscales (EALI-6), current feelings, current willingness, and one's current identity as an *agape* love servant, are used to measure the change in *agape* love across time.

***Agape* Love Scale Comparing the Different Loves (AL-CDL)**. AL-CDL asks the respondents to consider sentences that reflect five different kinds of love and one theme of respect by choosing one statement over the other in a pair that they think represents the higher form of love. AL-CDL consists of 25 items: 15 items assess the degree to which respondents choose the statement of *agape* love when comparing with another statement that reflects one of the four different kinds of love (*eros*, *philia*, *storge*, compassionate love) and one theme of

respect (e.g., “I strive to be friends with others so we get along well” comparing with “I willingly put others’ needs above my own needs”). The other 10 items consist pairs of comparing two statements among the four different kinds of love and one theme of respect. Only the responses to the 15 items with the statement of *agape* love are scored, and the score range is 0 to 15.

***Agape* Love Scale Comparing the Different Loves: Children’s Scale (AL-CDL-C).** The 25-item *Agape* Love Scale Comparing the Different Loves was revised to an 18-item children’s scale. Out of the 18 items, 12 assess the degree to which participants choose the statement of *agape* love when comparing with another statement that reflects one of the four different kinds of love (*eros*, *philia*, *storge*, compassionate love) and one theme of respect. The score ranges from 0 to 12 with a higher score corresponding to a larger degree of agreement on *agape* love as a higher form of humanity.

***Agape* Love Mature/Dispositional Scale (AL-DS)**. AL-DS measures the degree of how the respondents typically respond to different groups of others (loved ones, friends, people you do not know, those outside of your family with whom you are in conflict) when there are tensions caused by the different needs (e.g., “when your needs are in tension with loved one’s needs to what degree do you: want to defer to the others’ needs for that person’s (or people’s) sake?”). The score range, on a 6-point Likert scale, is 9 to 54 with a higher score indicating a higher level of *agape* love expression.

***Agape* Love Mature/Dispositional Scale for Children (AL-DS-C).** The 9-item *Agape* Love Mature/Dispositional Scale was adapted into children’s language. The score ranges from 9 to 54 with a higher score corresponding to a higher level of *agape* love expression.

***Agape* Love Willingness to Engage in *Agape* Love Scale (AL-WS)**. AL-WS measures the degree of willingness to engage in *Agape* Love, indicating the level of potential to express the

love. Four hypothetical examples of *agape* love are given in 20 items for respondents to rate the degree (on a 6-point Likert scale) they would react toward a person when there are tensions caused by different needs (e.g., "lose sleep comforting: loved one other than a romantic partner, romantic partner, friend, stranger, person with who you are in conflict"). The score range, on a 6-point Likert scale, is 20 to 120.

***Agape* Love Willingness to Engage in *Agape* Love Scale: Children's Scale (AL-WS-C).** The Willingness to Engage in *Agape* Love Scale was adapted into children's language. The hypothetical examples are the same as the ones in the adult's scale. The options for reacting towards a particular person are changed into "father, mother, brothers and sisters, friend, person who has laughed at you, embarrassing you, someone your age whom you do not know." The score ranges from 24 to 144 with a higher score corresponding to a higher level of potential to express *agape* love.

***Agape* Love Knowledge Scale (AL-KS)**. AL-KS (14 items) measures one's knowledge of *agape* love as a moral virtue on a true-false scale--either represent the correct or the incorrect understanding of *agape* love (e.g., "*agape* love does not mean to ignore injustice."). The score range is 0 to 14.

***Agape* Love Knowledge Scale for Children (AL-KS-C)**. Eight items were selected from the adult's scale and adapted into children's language to measure a child's knowledge of the *agape* love as a moral virtue on a true-false scale--either the statements represent *agape* love or are incorrect about it (e.g., "When I give *agape* love to the person who hurt me, I should also ask the person to behave better."). The score ranges from 0 to 8. Higher scores corresponded to a higher level of understanding *agape* love.

Enright Forgiveness Inventory Short Form (EFI-30, Subkoviak et al., 1995). EFI-30 is a 30-item self-report measure of interpersonal forgiveness that includes six subscales: Positive and Negative Affect; Positive and Negative Cognition; and Positive and Negative Behavior (Enright & Fitzgibbons, 2015). The score ranges from 30 to 180 (6-point Likert scale) with a higher score corresponds to a higher level of forgiveness. Five pseudo-forgiveness questions are included at the end of the EFI-30, which ask participants to evaluate whether the incident was truly hurtful. Participants responded on a six-point Likert scale ranging from "strongly disagree" to "strongly agree" to items such as "I was never bothered by what happened." Total pseudo-forgiveness scores range from 6 to 30. Participants from the initial data collection with pseudo-forgiveness scores higher than 20 were eliminated from data analysis. EFI-30 has demonstrated good internal consistency across various cultures (Enright et al., 2021).

Enright Forgiveness Inventory for Children (EFI-C, Enright, 2000). The EFI-C is a 30-item children's version of the Enright Forgiveness Inventory. Children describe an incident in which they were unjustly hurt by someone. EFI-C includes six subscales: Positive and Negative Affect; Positive and Negative Cognition; and Positive and Negative Behavior. The score ranges from 30 to 120, on a 4-point Likert scale, with a higher score corresponding to a higher level of forgiveness. Three pseudo-forgiveness questions are included at the end of the EFI-C, which ask participants to evaluate whether the incident was truly hurtful. Participants responded on a four-point Likert scale ranging from "No!" "a little bit no" "a little bit yes" to "Yes!" to items such as "Were your feelings hurt?" Total pseudo-forgiveness scores range from 4 to 12. Participants from the initial data collection with pseudo-forgiveness scores equal to 12 were eliminated from data analysis. Reliability and validity in Forgiveness Education interventions and the relationship to school-related behaviors are good (see, for example, Gambaro et al., 2008).

Clinical Anger Scale (CAS; Snell et al., 1995). The CAS is a 21-item self-report scale designed to assess severity of clinical anger (affective, cognitive, physiological, social, and behavioral symptoms due to anger). In this study, item 21 (0 = "I don't feel so angry that it interferes with my interest in sex.") was excluded from the original scale because people from certain cultures are not comfortable answering this item. Respondents are asked to rate different dimensions of anger symptoms on a scale of 0-3 by selecting the statement that best reflects their feelings (e.g., 0 = I am not all that angry about things; 1 = I am becoming more hostile about things than I used to be; 2 = I am pretty angry about things these days; or 3 = I am angry and hostile about everything). The total scores in this study range from 0 to 60. Higher total scores indicate greater clinical anger. Previous research has demonstrated that the scale correlates with state and trait anger as measured by the State-Trait Anger Scale (Spielberger et al., 1983), r's = 0.55 - 0.61, $p < 0.001$ (Snell et al., 1995). The scale has been used in previous research with evidence of adequate psychometric properties as a measure of clinical anger (Reyes and Hicklin, 2005).

The Anger Expression Scale for Children (AESC; Steele, Legerski, Nelson, & Phipps, 2009). AESC is a 26-item self-report measure. It has a four-point Likert-type response format (1: Almost never, 2: Sometimes, 3: Often, and 4: Almost always) with higher values keyed to greater agreement with the items. Cumming, Poling, and Smith (2020) examined the factor structure of the scale with 2,020 fourth and fifth graders and found a modified two-factor model of Anger Expression factor and Anger Control factor. One administered item ("I stay mad at people but keep it secret.") was excluded from analysis because it is not in the two-factor model reported by Cumming et al. (2020). Anger Expression (16 items) was positively associated with aggression and externalizing problems. The score of Anger Expression ranges from 16 to 64,

with a higher score indicating greater aggression. Anger Control (9 items) was related to emotion control and less internalizing behaviors. The score of Anger Control ranges from 9 to 36, with a higher score indicating less aggression.

Patient Health Questionnaire Eight-item Depression Scale (PHQ-8, Kroenke et al., 2009). The PHQ-8 consists of eight of the nine criteria on which the DSM-IV diagnosis of depressive disorders is based (American Psychiatric Association, 1994). The ninth question in the DSM-IV assesses suicidal or self-injurious thoughts. Research indicates that the deletion of this question has only a minor effect on scoring because thoughts of self-harm are fairly uncommon in the general population, and the ninth item is by far the least frequently endorsed item on the PHQ-9 (Huang et al., 2006a, Kroenke and Spitzer, 2002, Lee et al., 2007, Rief et al., 2004). Indeed, the two original validation studies of the PHQ totaling 6000 patients established that identical scoring thresholds for depression severity could be used for the PHQ-9 and PHQ-8 (Kroenke and Spitzer, 2002). The PHQ-8 response set was standardized by asking the number of days in the past 2 weeks the respondent had experienced a particular depressive symptom. The modified response set was converted back to the original response set: 0 to 1 day = “not at all,” 2 to 6 days = “several days,” 7 to 11 days = “more than half the days,” and 12 to 14 days = “nearly every day,” with points (0 to 3) assigned to each category, respectively. The scores for each item are summed to produce a total score between 0 and 24 points. A total score of 0 to 4 represents no significant depressive symptoms. A total score of 5 to 9 represents mild depressive symptoms; 10 to 14, moderate; 15 to 19, moderately severe; and 20 to 24, severe. (Kroenke et al., 2001) Current depression was defined in two ways: 1) a PHQ-8 algorithm diagnosis of major depression (this requires either the first or second item [depressed mood or anhedonia] to be present “more than half the days” and at least 5 of the 8 symptoms to be present “more than half

the days") or other depression (2 to 4 symptoms, including depressed mood or anhedonia, are required to be present "more than half the days"); 2) a PHQ-8 score of ≥ 10, which has an 88% sensitivity and 88% specificity for major depression (Kroenke and Spitzer, 2002) and, regardless of diagnostic status, typically represents clinically significant depression (Corson et al., 2004, Kroenke et al., 2001).

Patient Health Questionnaire (PHQ) Screener for Depression: PHQ-8 Modified for Adolescents (PHQ-8A). Modified with permission from the PHQ (Spitzer, Williams & Kroenke, 1999) by J. Johnson (Johnson et al., 2002). The items of PHQ-8 are rated on a scale from "not at all" (0) to "nearly every day" (3), with higher scores indicating greater frequency of depressive symptoms. Individuals are asked to report how bothered they have been over the past 2 weeks for problems such as feeling down, tired, or feeling bad about themselves. The sum of the items indicates the overall level of depressive symptoms indicating mild (5-9), moderate (10-14), moderately severe (15-19), and severe depressive symptoms (20-24). Scores beginning at moderate severity are considered clinically significant and warrant further evaluation for depression. The PHQ-8 is widely used and has demonstrated evidence to support reliability and validity, as well as clinical utility of cut-off scores (Kroenke et al., 2001). It has also been recommended as the annual universal screening tool for depression by the American Academy of Pediatrics (Zuckerbrot, et al., 2018).

Herth Hope Index (HHI, Herth, 1992). This measure has 12 items and assesses optimism toward the future with questions such as "I have a positive outlook affects my life.", "I am able to give and receive caring/love.", and "I believe that each day has potential." These questions assess connectedness to positive expectations for the future as well as inter-connectedness with other people. Participants respond on a 4-point Likert scale ranging from "strongly disagree" to

"strongly agree." Two items are reverse scored, with the range of score being from 12 to 48, with a higher score representing more hopefulness. Herth Hope Index is the abbreviated version adapted from the Herth Hope Scale. It shows good reliability and validity properties (Herth, 1992).

Children's Hope Scale (**CHS**; **Snyder et al., 1997).** The CHS is a six-item self-report measure that is based on the premise that children are goal-directed and that their goal-directed thoughts can be understood according to agency and pathways (Snyder, 1994). Each item is scored from 0 (never) to 6 (all of the time). Scores range from 6 to 36. A higher score indicates that the children are more goal-directed. The CHS has been introduced and validated for use with children aged 7-16.

The Rosenberg Self-Esteem Scale (**RSS**, **Rosenberg, 1965).** RSS consists of 10 statements that evaluate self-worth by measuring both positive and negative feelings about the self. The items are answered using a 4-point Likert scale format ranging from strongly agree to strongly disagree. An example of the questions is "I take a positive attitude toward myself." The score ranges from 10 to 40. A high score represents high self-esteem. This is a popular scale with good reliability and validity properties (Gray-Little et al., 1997).

Coopersmith Self-Esteem Inventory (**CSI**, **Coopersmith, 2002).** The self-esteem of students is measured using the Coopersmith Self-Esteem Inventory (School Form)-Short Form (CSEI-SC Short Form). Coopersmith hypothesized multiple aspects of self-esteem as a "central feature of the definition of self-esteem…that one's overall appraisal of his/her self-esteem may not be reflective of one's self-appraisal under different role defining conditions" (as cited in Sewell, 1985, p.1). Thus, the 25-item CSEI-SC Short Form, which has been used extensively since 1981, measures children's "attitudes toward the self in social, academic, family, and

personal areas of experience" (Coopersmith, 2002, p.1). It can be completed by children as young as 8 years of age. All items are declarative statements about the self (e.g., "I often feel upset in school" and "I'm pretty sure about myself"), which are answered as either "like me" or "unlike me." Each item is scored a 1 if students respond to negative items as "unlike me" and to positive items as "like me." Items answered "unlike me" for positive items and "like me" for negative items are scored a 0. Scores are sums of item responses for five self-esteem subscales: General Self, Social Self-Peers, Home-Parents, School-Academic, and Lie Scale. The score ranges from 0 to 25. Overall score reliability was estimated at 0.86 (Johnson et al., 1983). Test-retest reliability was estimated in separate studies and ranged from 0.64 (three-year interval; Rubin, 1978) to 0.88 (five-week interval; Coopersmith, 1967).

The Toronto Empathy Questionnaire (**TEQ**; **Spreng, McKinnon, Mar, & Levine, 2009).** TEQ is a 16-item self-report scale designed to provide a unidimensional tool for measuring dispositional empathy. Items include "I remain unaffected when someone close to me is happy." and "I can tell when others are sad even when they do not say anything." Responses are given on a scale from 0 (Never) to 4 (Always). The score ranges from 0 to 64. The TEQ has high internal reliability and convergent validity (Spreng et al., 2009).

Empathy Questionnaire for Children and Adolescents (**EQ-C**; **Overgaauw et al., 2017).** The Empathy Questionnaire for Children and Adolescents is a 18-item measure with three scales: (1) affective empathy (nine items; e.g., 'When a friend is upset, I feel upset too') measuring the extent to which one feels for the emotion of another person, (2) cognitive empathy (six items; e.g., 'If a friend cries, I often understand what has happened') measuring the extent to which one understands why another person is in distress, and (3) intention to comfort (six items; e.g., 'If a friend is sad, I want to do something to make it better') measuring the extent to which one is also

inclined to actually help or support a person in need. A higher score indicates that the children have higher empathy (18 items; range 0 - 36). The EQ-C has demonstrated good internal consistency and strong convergent validity (Overgaauw et al., 2017).

The Classroom Cooperation Scale (**CCS**, **Wellborn, 1991).** The Classroom Cooperation Scale was designed as a teachers' measure of student cooperation on a 4-point Likert scale, in which they rated 10 items such as "This student works only as hard as necessary to get by" and "This student concentrates on doing his/her work in my class" (Cronbach's alpha = 0.88). The score ranges from 1 to 40 for one student. In past research, the scale has been found to have good split-half reliability (0.82) and has been significantly associated with student-reported teacher characteristics (Skinner, Wellborn, & Connell, 1990). To ascertain if cooperation increases in the more disruptive children, each teacher is asked to identify the four most disruptive students in the class and then to fill out the scale on them.

Chapter 4: Results

In this chapter, reliability and validity of the initial assessment measurement and effect of the *Agape* Love and Forgiveness Education program are discussed using students' data collected from each research site: Northern Ireland, Israel, Taiwan and using teachers' data collected from three research sites combined. The data of each measure were screened separately. A respondent's responses to the surveys at different time points were linked by a survey code. The final responses remaining in the analyses are responses that passed the following screening criteria in each measure in each of the four surveys and were linked from the four surveys.

For each measure, a participant's scale was eliminated if it has a lower than 75% completion rate of that particular measure; b. fails four or five of the five attention checks; c. has straight responding pattern in a scale with both positively-worded and negatively-worded statements.

EALI and EFI each has two additional screening criteria at initial assessment (T1). For EALI: a. in examining the 5-item pseudo-*agape* love subscale, we eliminated responses scored above the cut-off (26 for responded to five items; 20 for responded to four items; 15 for responded to three items; 10 for responded to two items; 5 for responded to one item; the response was excluded if responded to none of the five items); OR b. responses with EALI story rating of "3" (indicating no such situation/no help/no hurt/don't know). Please see Appendix C: Criteria of EALI Story Rating for details. EFI: a. examination of the 3-item pseudo-forgiveness subscale eliminate responses scored above the cut-off (12 for responded to three items; 8 for responded to two items; 4 for responded to one item; the response was excluded if responded to none of the three items); OR b. responses with EFI story rating of "3" (indicating no such situation/no hurt/don't know). Please see Appendix D: Criteria of EFI Story Rating for details.

Northern Ireland Students

Reliability and Validity of the Initial Assessment

Cronbach's alpha was assessed at the initial assessment (T1) only. Cronbach's alpha for the 24-item Enright *Agape* Love Inventory for children (EALI-C) is 0.88 and Cronbach's alpha for three subscales (EALI-6) assessing change of *agape* love across time is 0.64. EALI-C and EALI-6 both positively correlated with EFI-C (forgiveness), AL-CDL-C, AL-DS-C, AL-WS-C, hope, self-esteem, and empathy; negatively correlated with anger and depression.

Cronbach's alpha for the 30-item Enright Forgiveness Inventory for children (EFI-C) was 0.97. EFI-C positively correlated with EALI-C, AL-DS-C, AL-WS-C, hope, self-esteem and negatively correlated with anger and depression. The four scales measuring *agape* love, EALI-C, AL-CDL-C, AL-DS-C, and AL-WS-C, are positively correlated with each other. The correlations show convergent validity of the four *agape* love scales, which are similar to the correlations achieved in the US children's sample. Cronbach's alpha of internal consistency for AL-CDL-C is 0.37; for AL-DS-C is 0.78; for AL-WS-C is 0.92. Cronbach's alpha for the other measures are as follows: anger expression (0.87), anger control (0.72), depression (0.84), hope (0.81), self-esteem (0.87), and empathy (0.86).

Table 2

Correlations for Variables in the Initial Assessment (T1) of Northern Ireland Students

Measures	M	SD	*N*	1	2	3	4	5	6	7	8	9	10
1. EALI-C	104.68	17.83	250										
2. EALI-6	25.69	4.95	250	.81**									
3. EFI-C	85.64	24.99	254	.27**	.20**								
4. AL-CDL-C	5.63	2.76	293	.24**	.15*	0.08							

5. AL-DS-C	38.69	8.25	280	.31**	.23**	.22**	.22**						
6. AL-WS-C	102.59	22.12	292	.26**	.20**	0.07	.18**	.31**					
7. Anger	39.38	9.95	293	-.28**	-.18**	-.27**	-0.07	-.17**	-0.01				
8. Depression	7.13	5.77	291	-.17**	-.18**	-.23**	-0.02	-0.12	0	.49**			
9. Hope	23.58	6.43	287	.19**	.18**	.19**	-0.02	.23**	0.1	-.25**	-.34**		
10. Self-esteem	15.71	5.38	292	.20**	.18**	.21**	0.02	.15*	0.05	-.50**	-.64**	.48**	
11. Empathy	24.66	5.83	290	.37**	.24**	0.11	.23**	.31**	.31**	-0.02	-0.01	.17**	-0.01

Effect of the *Agape* Love and Forgiveness Education Program

Two classes in the experimental group and three classes in the control group were eliminated from the analysis based on teachers' self-report (not adhering to the teaching schedule) and researchers' observation (not adhering to the assessment schedule) on the intervention. Please check Appendix B: Criteria of Problem Intervention Coding for details. Only students who were able to finish four times of surveys with the responses passed screening criteria were retained in the analyses. Means and standard deviations are in Table 3. To examine the possibility of pre-test differences between the two groups, two tailed independent t-tests were conducted to compare scores on each measure between the two groups. No differences were found at the baseline (T1) for the following measures: AL-CDL-C, AL-DS-C, AL-WS-C, CHS (hope), CSI (self-esteem), EQ (empathy). The control group has a significantly higher score than the experimental group at the baseline (T1) for the following measures: EALI-C, EALI-6, EFI-C (forgiveness), anger

control. The experimental group has a significantly higher score than the control group at the baseline (T1) for anger expression and depression. Please see Table 4 and the footnote.

Table 3

Descriptive Statistics of Agape Love Measures and Psychological Compromises and Well-being of Northern Ireland Students

	Group	N	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	First follow-up to EG/ Second post-test to CG *M* (*SD*)	Second follow-up to EG/ First follow-up to CG *M* (*SD*)
EALI-C	EG	51	103.63 (16.92)	104.04 (15.02)	105.51 (16.00)	102.16 (16.57)
	CG	38	115.95 (15.29)	114.55 (16.48)	117.47 (16.47)	114.50 (15.36)
EALI-6	EG	51	25.08 (4.76)	25.67 (4.61)	25.75 (4.94)	24.88 (4.95)
	CG	38	27.89 (5.39)	28.68 (4.39)	28.89 (4.80)	28.26 (5.15)
EFI-C	EG	58	82.05 (24.61)	87.57 (24.97)	92.33 (23.34)	93.47 (23.31)
	CG	44	93.32 (23.45)	99.84 (21.55)	103.25 (18.70)	105.84 (19.07)
AL-CDL-C	EG	62	5.52 (2.70)	5.31 (3.07)	5.08 (3.00)	4.76 (3.40)
	CG	47	5.87 (2.72)	5.85 (2.96)	6.09 (2.39)	5.13 (3.25)
AL-DS-C	EG	55	39.93 (7.40)	39.55 (7.74)	41.04 (7.26)	40.00 (7.06)
	CG	42	42.19 (7.94)	43.21 (7.24)	45.45 (6.78)	44.79 (7.53)
AL-WS-C	EG	68	103.10 (21.46)	109.04 (17.63)	111.63 (13.35)	112.81 (15.61)
	CG	46	109.87 (21.10)	111.35 (18.55)	116.98 (14.95)	115.98 (20.28)
Anger Expression	EG	65	31.31 (10.05)	29.00 (9.28)	27.89 (9.62)	28.02 (11.41)
	CG	46	26.57 (6.98)	27.59 (8.42)	25.93 (9.12)	27.02 (9.98)
Anger Control	EG	65	22.46 (5.17)	23.55 (5.74)	24.38 (5.60)	23.95 (7.28)
	CG	46	24.02 (4.51)	24.28 (5.28)	25.09 (6.27)	26.28 (6.39)
Depression	EG	67	7.73 (5.90)	5.81 (5.06)	5.97 (6.93)	5.90 (6.81)
	CG	43	5.05 (4.58)	5.26 (5.14)	5.12 (5.08)	6.19 (6.34)

Hope	EG	64	23.83 (6.08)	24.02 (6.31)	24.22 (5.71)	25.61 (6.85)
	CG	46	26.02 (5.85)	25.98 (5.98)	26.11 (5.87)	26.83 (6.14)
Self-esteem	EG	66	15.59 (5.89)	16.17 (4.92)	16.67 (5.89)	16.21 (6.42)
	CG	46	17.63 (5.25)	18.15 (4.21)	17.80 (4.84)	17.98 (4.87)
Empathy	EG	64	23.48 (6.45)	24.19 (5.78)	24.13 (5.55)	24.97 (7.37)
	CG	46	25.61 (4.87)	26.33 (4.08)	26.37 (4.95)	26.52 (5.05)

Cutoff scores for the eight-item Patient Health Questionnaire depression scale (PHQ-8): 0 - 4 = no significant depressive symptoms; 5 - 9 = mild depressive symptoms; 10 - 14 = moderate; 15 - 19 = moderately severe; 20 - 24, severe. (Kroenke et al., 2001).

To test the effectiveness of Forgiveness Education intervention compared to the no treatment wait list control (teaching regular school curriculum), we first conducted a one-tailed independent t-test to compare the gain scores between the two groups from pretest to post-test using data from the remaining 11 classes (EG = 6; CG = 5) with students who completed and passed the screening criteria of all four surveys. We used a one-tailed t-test here based on our original hypotheses: we assumed that Forgiveness Education intervention would have better effects than the no treatment wait list control. The results showed that the experimental group had significantly greater decreases in anger expression ($p = 0.007$) and depression ($p = 0.009$) compared to the control group (see Table 4). In both cases, the effect sizes are small to medium.

Next, a one-tailed paired sample t-test was run within the control group compared to itself once this group had Forgiveness Education. We compared the gain scores (post-test scores minus pretest scores) when the control group had no treatment (NT, teaching regular school curriculum) with the gain scores (second post-test minus post-test scores) for the control group-turned-experimental group) once the control group became the experimental group with Forgiveness

Education intervention. This particular analysis was done with the data from the five remaining classes with students who completed and passed the screening criteria of all four surveys. The results showed that the control group neither had significant growth in any of the *agape* love measures, forgiveness, anger control, hope, self-esteem, empathy; nor had significant decrease in anger or depression after receiving Forgiveness Education compared to the teaching of the regular school curriculum with no treatment (see Table 5).

Table 4

Comparison of Gain Scores from Pretest to Posttest between the Two Groups of Northern Ireland Students

Measures	Experimental Group				Control Group				t(*df*)	*p* (one-tailed)	95% CI	Cohen's *d*	
	N	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain Scores *M* (*SD*)	*N*	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain Scores *M* (*SD*)					
EALI-C	51	103.63 (16.92)	104.04 (15.02)	0.41 (16.35)	38	115.95 (15.29)	114.55 (16.48)	-1.39 (14.24)	-0.56 (84.83)	0.290	[-Inf, 3.60]	0.12	Negligible
EALI-6	51	25.08 (4.76)	25.67 (4.61)	0.59 (5.66)	38	27.89 (5.39)	28.68 (4.39)	0.79 (5.44)	0.17 (81.52)	0.567	[-Inf, 2.18]	-0.04	Negligible
EFI-C	58	82.05 (24.61)	87.57 (24.97)	5.52 (24.31)	44	93.32 (23.45)	99.84 (21.55)	6.52 (24.06)	0.21 (93.22)	0.582	[-Inf, 9.03]	-0.04	Negligible
AL-CDL-C	62	5.52 (2.70)	5.31 (3.07)	-0.21 (3.21)	47	5.87 (2.72)	5.85 (2.96)	-0.02 (2.98)	0.32 (102.6)	0.624	[-Inf, 1.18]	-0.06	Negligible
AL-DS-C	55	39.93 (7.40)	39.55 (7.74)	-0.38 (7.88)	42	42.19 (7.94)	43.21 (7.24)	1.02 (6.85)	0.94 (93.34)	0.825	[-Inf, 3.89]	-0.19	Negligible
AL-WS-C	68	103.10 (21.46)	109.04 (17.63)	5.94 (18.63)	46	109.87 (21.10)	111.35 (18.55)	1.48 (22.49)	-1.11 (84.29)	0.135	[-Inf, 2.21]	0.22	Small
Anger Expression	65	31.31 (10.05)	29.00 (9.28)	-2.31 (7.90)	46	26.57 (6.98)	27.59 (8.42)	1.02 (6.21)	2.48 (107.74)	0.007	[1.10, Inf]	0.46	Small
Anger Control	65	22.46 (5.17)	23.55 (5.74)	1.09 (4.55)	46	24.02 (4.51)	24.28 (5.28)	0.26 (4.34)	-0.97 (99.78)	0.166	[-Inf, 0.58]	0.19	Negligible
Depression	67	7.73 (5.90)	5.81 (5.06)	-1.92 (5.31)	43	5.05 (4.58)	5.26 (5.14)	0.21 (4.02)	2.39 (104.97)	0.009	[0.65, Inf]	0.44	Small
Hope	64	23.83 (6.08)	24.02 (6.31)	0.19 (6.51)	46	26.02 (5.85)	25.98 (5.98)	-0.04 (6.18)	-0.19 (100.06)	0.425	[-Inf, 1.80]	0.04	Negligible
Self-esteem	66	15.59 (5.89)	16.17 (4.92)	0.58 (3.73)	46	17.63 (5.25)	18.15 (4.21)	0.52 (3.87)	-0.07 (94.65)	0.471	[-Inf, 1.16]	0.01	Negligible

Empathy	64	23.48 (6.45)	24.19 (5.78)	0.70 (6.01)	46	25.61 (4.87)	26.33 (4.08)	0.72 (5.22)	0.01 (104.09)	0.505	[-Inf, 1.80]	-0.00	Negligible

Table 5

Comparison of Gain Scores between NT and FE for the Control Group of Northern Ireland Students

Measure	*N*	Pretest *M* (SD)	Posttest *M* (SD)	Second Posttest *M* (SD)	NT Gain *M* (SD)	FE Gain *M* (SD)	t(*df*)	*p* (one-tailed)	95% CI	Cohen's *d*	
EALI-C	38	115.95 (15.29)	114.55 (16.48)	117.47 (16.47)	-1.39 (14.24)	2.92 (12.71)	-1.25 (37)	0.109	[-Inf, 1.50]	0.32	Small
EALI-6	38	27.89 (5.39)	28.68 (4.39)	28.89 (4.80)	0.79 (5.44)	0.21 (3.63)	0.48 (37)	0.683	[-Inf, 2.61]	-0.13	Negligible
EFI-C	44	93.32 (23.45)	99.84 (21.55)	103.25 (18.70)	6.52 (24.06)	3.41 (19.80)	0.56 (43)	0.709	[-Inf, 12.54]	-0.14	Negligible
AL-CDL-C	47	5.87 (2.72)	5.85 (2.96)	6.09 (2.39)	-0.02 (2.98)	0.23 (2.70)	-0.36 (46)	0.362	[-Inf, 0.95]	0.09	Negligible
AL-DS-C	42	42.19 (7.94)	43.21 (7.24)	45.45 (6.78)	1.02 (6.85)	2.24 (8.50)	-0.64 (41)	0.262	[-Inf, 1.96]	0.16	Negligible
AL-WS-C	46	109.87 (21.10)	111.35 (18.55)	116.98 (14.95)	1.48 (22.49)	5.63 (14.61)	-0.93 (45)	0.178	[-Inf, 3.33]	0.22	Small
Anger Expression	46	26.57 (6.98)	27.59 (8.42)	25.93 (9.12)	1.02 (6.21)	-1.65 (7.63)	1.49 (45)	0.071	[-0.33, Inf]	0.38	Small
Anger Control	46	24.02 (4.51)	24.28 (5.28)	25.09 (6.27)	0.26 (4.34)	0.80 (5.17)	-0.48 (45)	0.316	[-Inf, 1.34]	0.11	Negligible
Depression	43	5.05 (4.58)	5.26 (5.14)	5.12 (5.08)	0.21 (4.02)	-0.14 (4.46)	0.31 (42)	0.380	[-1.55, Inf]	0.08	Negligible
Hope	46	26.02 (5.85)	25.98 (5.98)	26.11 (5.87)	-0.04 (6.18)	0.13 (6.59)	-0.11 (45)	0.457	[-Inf, 2.53]	0.03	Negligible
Self-esteem	46	17.63 (5.25)	18.15 (4.21)	17.80 (4.84)	0.52 (3.87)	-0.35 (3.67)	0.90 (45)	0.814	[-Inf, 2.49]	-0.23	Small
Empathy	46	25.61 (4.87)	26.33 (4.08)	26.37 (4.95)	0.72 (5.22)	0.04 (4.90)	0.52 (45)	0.698	[-Inf, 2.84]	-0.13	Negligible

We also performed one-tailed matched pair t-tests within each of the two groups to examine the amount of change on the 12 measures after each period of treatment, to examine the intervention effectiveness as well as the maintenance or growth of the intervention effect after the treatment ended. For the first period of treatment, when comparing the pre-test scores (T1) and post-test scores (T2), the experimental group showed significant decrease in anger expression ($p = 0.011$) and depression ($p = 0.002$); as well as significant increases in forgiveness ($p = 0.045$), *agape* love (AL-WS-C, $p = 0.005$), anger control ($p = 0.029$, See Table 6). The control group showed no significant changes on any of the 12 dependent variables during the no treatment period (See Table 7). For the second period of treatment, when comparing the post-test scores (T2) and the second-post test scores (T3), the control-group-turned-experimental group (after 12-week Forgiveness Education) had significantly increased in *agape* love (AL-DS-C: $p = 0.048$ and AL-WS-C: $p = 0.006$).

Within the experimental group (see Table 6), when comparing the pre-test scores (T1) and the first follow-up scores (T3), it showed significant decreases in anger expression ($p = 0.002$) and depression ($p = 0.008$); as well as significant increases in forgiveness ($p = 0.000$), *agape* love (AL-WS-C, $p = 0.000$), anger control ($p = 0.007$) and self-esteem ($p = 0.027$). When comparing the pre-test scores (T1) and the second follow-up scores (T4), it showed significant decrease in anger expression ($p = 0.003$) and depression ($p = 0.006$); as well as significant increases in forgiveness ($p = 0.000$), *agape* love (AL-WS-C, $p = 0.001$), anger control ($p = 0.032$), and hope ($p = 0.047$).

Within the control-group-turned experimental group (see Table 7), when comparing the post-test scores (T2) and the follow-up scores (T4), it significantly increased in forgiveness ($p = 0.037$), *agape* love (AL-WS-C, $p = 0.021$), and anger control ($p = 0.002$).

We examined the maintenance or growth effect after the interventions with the variables that showed significant differences in the within-group analyses. For forgiveness, *agape* love (AL-WS-C), anger expression, anger control, and depression, no differences were found for the original experimental group from the post-test (T2) to the first follow-up (T3), the post-test (T2) to the second follow-up (T4), or the first follow-up (T3) to the second follow-up (T4) (see Table 6). This indicates the effects of Forgiveness Education intervention were maintained after a 16-week no treatment period. The results demonstrated the long-term effectiveness of Forgiveness Education. In addition, given the severity of COVID-19 in the second semester (Spring of 2022), the continued increase in forgiveness, *agape* love (AL-WS-C) and the continued decrease in depression after the treatment ended are worth noting. No differences were found for these two variables from second post-test (T3) to follow-up (T4), indicating the effects of Forgiveness Education intervention were maintained after the 4-week no treatment period for the control-group-turned-experimental group.

Table 6

Comparison of Scores within the Experimental Group of Northern Ireland Students

Measure						Intervention Effect						Maintenance Effect					
	N	Pretest *M* (*SD*)	Posttest *M* (*SD*)	1st Follow-up *M* (*SD*)	2nd Follow-up *M* (*SD*)	Pre- to post-Gain *M* (*SD*)	*p* (one-tailed)	Pre- to 1st follow-up Gain *M* (*SD*)	*p* (one-tailed)	Pre- to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)	Post to 1st follow-up Gain *M* (*SD*)	*p* (one-tailed)	Post to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)	1st follow-up to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)
EALI-C	51	103.63 (16.92)	104.04 (15.02)	105.51 (16.00)	102.16 (16.57)	0.41 (16.36)	0.429	1.88 (14.89)	0.185	1.47 (16.66)	0.734						
EALI-6	51	25.08 (4.76)	25.67 (4.61)	25.75 (4.94)	24.88 (4.95)	0.59 (5.66)	0.231	0.67 (5.70)	0.204	-0.20 (6.04)	0.591						
EFI-C	58	82.05 (24.61)	87.57 (24.97)	92.33 (23.34)	93.47 (23.31)	5.51 (24.31)	0.045	10.28 (22.51)	0.000	11.41 (21.71)	0.000	4.76 (17.76)	0.977	5.89 (16.72)	0.995	1.14 (11.91)	0.765
AL-CDL-C	62	5.52 (2.70)	5.31 (3.07)	5.08 (3.00)	4.76 (3.40)	-0.21 (3.21)	0.696	-0.44 (3.13)	0.861	-0.76 (3.51)	0.953						
AL-DS-C	55	39.93 (7.40)	39.55 (7.74)	41.04 (7.26)	40.00 (7.06)	-0.38 (7.88)	0.640	1.11 (8.47)	0.168	0.07 (7.48)	0.471						
AL-WS-C	68	103.10 (21.46)	109.04 (17.63)	111.63 (13.35)	112.81 (15.61)	5.94 (18.63)	0.005	8.52 (19.91)	0.000	9.71 (24.17)	0.001	2.59 (16.82)	0.896	3.76 (21.70)	0.921	1.18 (13.24)	0.767
Anger Expression	65	31.31 (10.05)	29.00 (9.28)	27.89 (9.62)	28.02 (11.41)	-2.31 (7.90)	0.011	-3.42 (8.96)	0.002	-3.29 (9.31)	0.003	-1.11 (6.98)	0.897	-0.98 (8.26)	0.830	0.12 (7.51)	0.448
Anger Control	65	22.46 (5.17)	23.55 (5.74)	24.38 (5.60)	23.95 (7.28)	1.09 (4.55)	0.029	1.92 (6.18)	0.007	1.49 (6.37)	0.032	0.83 (5.06)	0.905	0.40 (6.33)	0.694	-0.43 (5.56)	0.267
Depression	67	7.73 (5.90)	5.81 (5.06)	5.97 (6.93)	5.90 (6.81)	-1.92 (5.31)	0.002	-1.76 (5.79)	0.008	-1.83 (5.82)	0.006	0.16 (5.53)	0.404	0.09 (6.09)	0.452	-0.07 (4.74)	0.551
Hope	64	23.83 (6.08)	24.02 (6.31)	24.22 (5.71)	25.61 (6.85)	0.19 (6.51)	0.409	0.39 (6.34)	0.312	1.78 (8.41)	0.047						
Self-esteem	66	15.59 (5.89)	16.17 (4.92)	16.67 (5.89)	16.21 (6.42)	0.58 (3.73)	0.107	1.08 (4.47)	0.027	0.62 (4.92)	0.154					-0.45 (3.13)	0.121

Empathy	64	23.48 (6.45)	24.19 (5.78)	24.13 (5.55)	24.97 (7.37)	0.70 (6.01)	0.176	0.64 (6.84)	0.228	1.48 (8.44)	0.082

Table 7

Comparison of Scores within the Control-group-turned-experimental Group of Northern Ireland Students

								Intervention Effect			Maintenance Effect	
Measure	*N*	Pretest *M* (SD)	Posttest *M* (SD)	Second Posttest *M* (SD)	Follow-up *M* (SD)	NT Gain *M* (SD)	*p* (two-tailed)	Post- to 2nd post-) *M* (SD)	*p* (one-tailed)	Post- to follow-up) *M (SD)*	*p* (one-tailed)	2nd post- to follow-up) *M (SD)*	*p* (one-tailed)
EALI-C	38	115.95 (15.29)	114.55 (16.48)	117.47 (16.47)	114.50 (15.36)	-1.39 (14.24)	0.550	2.92 (12.71)	0.083	-0.05 (13.43)	0.510		
EALI-6	38	27.89 (5.39)	28.68 (4.39)	28.89 (4.80)	28.26 (5.15)	0.79 (5.44)	0.377	0.21 (3.63)	0.362	-0.42 (3.92)	0.744		
EFI-C	44	93.32 (23.45)	99.84 (21.55)	103.25 (18.70)	105.84 (19.07)	6.52 (24.06)	0.079	3.41 (19.80)	0.130	6.00 (21.69)	0.037		
AL-CDL-C	47	5.87 (2.72)	5.85 (2.96)	6.09 (2.39)	5.13 (3.25)	-0.02 (2.98)	0.961	0.23 (2.70)	0.277	-0.72 (3.06)	0.944		
AL-DS-C	42	42.19 (7.94)	43.21 (7.24)	45.45 (6.78)	44.79 (7.53)	1.02 (6.85)	0.338	2.24 (8.50)	0.048	1.57 (8.80)	0.127	-0.67 (7.39)	0.281
AL-WS-C	46	109.87 (21.10)	111.35 (18.55)	116.98 (14.95)	115.98 (20.28)	1.48 (22.49)	0.658	5.63 (14.61)	0.006	4.63 (15.03)	0.021	-1.00 (13.48)	0.309
Anger Expression	46	26.57 (6.98)	27.59 (8.42)	25.93 (9.12)	27.02 (9.98)	1.02 (6.21)	0.270	-1.65 (7.63)	0.074	-0.57 (8.18)	0.321		
Anger Control	46	24.02 (4.51)	24.28 (5.28)	25.09 (6.27)	26.28 (6.39)	0.26 (4.34)	0.685	0.80 (5.17)	0.149	2.00 (4.50)	0.002		
Depression	43	5.05 (4.58)	5.26 (5.14)	5.12 (5.08)	6.19 (6.34)	0.21 (4.02)	0.735	-0.14 (4.46)	0.419	0.93 (6.57)	0.821		
Hope	46	26.02 (5.85)	25.98 (5.98)	26.11 (5.87)	26.83 (6.14)	-0.04 (6.18)	0.962	0.13 (6.59)	0.447	0.85 (5.16)	0.135		
Self-esteem	46	17.63 (5.25)	18.15 (4.21)	17.80 (4.84)	17.98 (4.87)	0.52 (3.87)	0.366	-0.35 (3.67)	0.738	-0.17 (3.82)	0.621		
Empathy	46	25.61 (4.87)	26.33 (4.08)	26.37 (4.95)	26.52 (5.05)	0.72 (5.22)	0.356	0.04 (4.90)	0.476	0.20 (4.81)	0.392		

Israel Students

Reliability and Validity of the Initial Assessment

Cronbach's alpha of each measure was assessed at the initial assessment (T1) only. Cronbach's alpha for the 24-item Enright *Agape* Love Inventory for children (EALI-C) is 0.77 and Cronbach's alpha for three subscales (EALI-6) assessing change of *agape* love across time is 0.37. EALI-C and EALI-6 both positively correlated with EFI-C (forgiveness), AL-DS-C, AL-WS-C, hope, self-esteem, and empathy; negatively correlated with anger and depression.

Cronbach's alpha for the 30-item Enright Forgiveness Inventory for children (EFI-C) was 0.92. EFI-C (forgiveness) is positively correlated with EALI-C, AL-DS-C, AL-WS-C, self-esteem and negatively correlated with anger. The three scales measuring *agape* love, EALI-C, AL-DS-C, and AL-WS-C, are positively correlated with each other. The correlations show convergent validity of the three *agape* love scales, which are similar to the correlations achieved in the US children's sample. However, AL-CDL-C doesn't have positive correlations with the other three scales measuring *agape* love. Cronbach's alpha of internal consistency for AL-CDL-C is 0.44; for AL-DS-C is 0.74; for AL-WS-C is 0.91. Cronbach's alpha for the other measures are as follows: anger expression (0.87), anger control (0.80), depression (0.70), hope (0.83), self-esteem (0.71), and empathy (0.84). The first item in the Patient Health Questionnaire Screener for Depression (PHQ-8A) ("How often have you been bothered by each of the following symptoms during the past two weeks? - Feeling down, depressed, irritable, or hopeless?") was not shown properly in Qualtrics when the online questionnaire was administered at T1 to most of the Arabic-speaking classes in Israel. The seventeenth item in the Empathy Questionnaire for Children and Adolescents (EQ) was not shown properly in Qualtrics when the online questionnaire was administered at T1 and T2 to most of the Arabic-speaking classes in Israel.

Thus, these two items were eliminated from all surveys (at T1, T2, T3, T4) for all students in Israel for further analysis.

Table 8

Correlations for Variables in the Initial Assessment (T1) of Israel Students

Measures	M	SD	*N*	1	2	3	4	5	6	7	8	9	10
1. EAL-C	94.56	15.3	308										
2. EALI-6	23.39	4.53	308	.75**									
3. EFI-C	86.23	19.28	327	.33**	.18**								
4. AL-CDL-C	5.08	2.5	356	0	-0.06	-0.01							
5. AL-DS-C	36.4	8.35	351	.33**	.30**	.14*	-0.03						
6. AL-WS-C	102.51	22.92	355	.26**	.22**	.21**	-0.08	.27**					
7. Anger	40.26	10.21	358	-.23**	-.15**	-.18**	-0.03	-.19**	-.12*				
8. Depression	6.44	4.3	350	-.19**	-0.1	-0.09	0	-.11*	-.11*	.39**			
9. Hope	24.62	7	357	.23**	.26**	0.07	-.18**	.30**	.23**	-0.09	-0.04		
10. Self-esteem	14.77	4.01	352	.26**	.18**	.17**	-0.05	.23**	.17**	-.43**	-.40**	.19**	
11. Empathy	21.75	6.19	353	.23**	.24**	0.06	-.14*	.29**	.30**	0.03	0.08	.37**	0.1

Effect of the *Agape* Love and Forgiveness Education Program

One class in the experimental group and three classes in the control group were eliminated from the analysis based on teachers' self-report (not adhering to the teaching schedule) and researchers' observation (not adhering to the assessment schedule) on the intervention. Please check Appendix B: Criteria of Problem Intervention Coding for details. Only students who were able to finish four times of surveys with the responses passed screening criteria were retained in the analyses. To examine the possibility of pre-test differences between the two groups, two tailed independent t-tests were conducted to compare scores on each measure between the two groups. No differences were found at the baseline (T1) for the following measures: EALI-C, EFI-C (forgiveness), AL-DS-C, AL-WS-C, anger expression, anger control, CHS (hope), EQ (empathy). The control group had a significantly higher score than the experimental group at the baseline (T1) for depression. The experimental group has a significantly higher score than the control group at the baseline (T1) for AL-CDL-C and CSI (self-esteem). Please see Table 9 for means and standard deviations and the footnote.

Table 9

Descriptive Statistics of Agape Love Measures and Psychological Compromises and Well-being of Israel Students

	Group	N	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	First follow-up to EG/ Second post-test to CG *M* (*SD*)	Second follow-up to EG/ First follow-up to CG *M* (*SD*)
EALI-C	EG	77	95.08 (16.82)	98.18 (16.38）	98.00 (18.39)	98.61 (20.70)
	CG	54	95.85 (14.26)	94.22 (18.31)	94.31 (15.72)	94.15 (21.00)
EALI-6	EG	77	23.61 (5.09)	24.18 (4.49)	23.81 (5.25)	23.65 (5.53)
	CG	54	23.87 (3.85)	23.07 (5.14)	22.96 (4.40)	22.43 (4.41)

EFI-C	EG	83	86.20 (19.59)	88.55 (19.54)	92.43 (17.99)	88.67 (19.54)
	CG	56	84.21 (21.03)	86.91 (18.91)	87.50 (18.17)	87.50 (23.32)
AL-CDL-C	EG	101	5.33 (2.45)	5.69 (3.19)	5.46 (2.77)	5.31 (3.14)
	CG	63	4.59 (2.42)	4.17 (2.51)	4.60 (2.43)	3.81 (2.53)
AL-DS-C	EG	98	37.94 (8.30)	38.97 (8.73)	38.13 (9.03)	37.80 (8.65)
	CG	61	36.61 (8.65)	36.33 (9.03)	36.03 (8.68)	35.28 (9.23)
AL-WS-C	EG	102	105.72 (22.71)	108.01 (24.13)	112.43 (18.26)	108.55 (102)
	CG	63	102.75 (21.22)	105.03 (18.82)	102.86 (20.83)	104.22 (19.28)
Anger Expression	EG	100	29.69 (8.63)	27.78 (9.01)	28.19 (9.11)	29.63 (11.32)
	CG	60	31.60 (9.63)	31.23 (9.27)	31.87 (9.77)	30.90 (10.94)
Anger Control	EG	100	22.00 (5.96)	23.84 (6.19)	24.87 (6.00)	24.68 (6.74)
	CG	60	21.78 (5.67)	22.37 (6.49)	22.97 (4.95)	23.93 (6.04)
Depression	EG	93	5.85 (4.51)	5.43 (4.46)	5.65 (4.16)	5.68 (5.22)
	CG	63	7.21 (3.72)	6.75 (4.21)	6.21 (4.08)	5.94 (4.42)
Hope	EG	97	24.71 (6.52)	27.08 (6.36)	27.65 (6.06)	26.97 (7.19)
	CG	66	25.70 (7.29)	24.95 (7.16)	26.64 (6.73)	25.89 (7.10)
Self-esteem	EG	99	15.33 (3.93)	15.97 (4.35)	16.18 (4.50)	16.51 (4.67)
	CG	65	13.71 (4.52)	14.32 (4.96)	13.95 (4.56)	14.52 (4.92)
Empathy	EG	99	21.99 (5.95)	23.89 (6.10)	23.54 (5.97)	23.34 (7.32)
	CG	64	21.69 (5.97)	22.62 (6.33)	21.58 (6.38)	22.20 (7.09)

Cutoff scores for the eight-item Patient Health Questionnaire depression scale (PHQ-8): 0 - 4 = no significant depressive symptoms; 5 - 9 = mild depressive symptoms; 10 - 14 = moderate; 15 - 19 = moderately severe; 20 - 24, severe. (Kroenke et al., 2001).

To test the effectiveness of Forgiveness Education intervention compared to the no treatment wait list control (teaching a regular school curriculum), we first conducted a one-tailed independent t-test to compare the gain scores between the two groups from pretest to post-test using data from the remaining 18 classes (EG = 11; CG = 7) with students who completed and passed the screening criteria of all four surveys. We used a one-tailed t-test here based on our original hypotheses: we assumed that Forgiveness Education intervention would have better effects than the no treatment wait list control. The results showed that the experimental group had significantly greater increases in *agape* love (EALI-C, $p = 0.071$) and hope ($p = 0.002$) compared to the control group (see Table 10).

Next, a one-tailed paired sample t-test was run within the control group compared to itself once this group had Forgiveness Education. We compared the gain scores (post-test scores at T2 minus pretest scores at T1) when the control group had no treatment (NT, teaching a regular school curriculum) with the gain scores (second post-test at T3 minus post-test scores at T2) for the control group-turned-experimental group) once the control group became the experimental group with Forgiveness Education (FE) intervention. This particular analysis was done with the data from the seven remaining control group classes with students who completed and passed the screening criteria of all four surveys. The results showed that the control group had significant growth in hope ($p = 0.028$) after receiving Forgiveness Education compared to the teaching of a regular school curriculum with no treatment (see Table 11).

Table 10

Comparison of Gain Scores from Pretest to Posttest between the Two Groups of Israel Students

Measures	Experimental Group				Control Group				t(*df*)	*p* (one-tailed)	95% CI	Cohen's *d*	
	N	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain Scores *M* (*SD*)	*N*	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain Scores *M* (*SD*)					
EALI-C	77	95.08 (16.82)	98.18 (16.38)	3.10 (14.98)	54	95.85 (14.26)	94.22 (18.31)	-1.63 (14.17)	-1.84 (118.03)	0.034	[-Inf, -0.46]	0.32	Small
EALI-6	77	23.61 (5.09)	24.18 (4.49)	0.57 (4.81)	54	23.87 (3.85)	23.07 (5.14)	-0.80 (5.46)	-1.48 (104.75)	0.071	[-Inf, 0.16]	0.27	Small
EFI-C	83	86.20 (19.59)	88.55 (19.54)	2.35 (16.92)	56	84.21 (21.03)	86.91 (18.91)	2.70 (19.83)	0.11 (105.27)	0.543	[-Inf, 5.72]	-0.02	Negligible
AL-CDL-C	101	5.33 (2.45)	5.69 (3.19)	0.37 (3.46)	63	4.59 (2.42)	4.17 (2.51)	-0.41 (2.62)	-1.63 (155.87)	0.052	[-Inf, 0.01]	0.25	Small
AL-DS-C	98	37.94 (8.30)	38.97 (8.73)	1.03 (10.19)	61	36.61 (8.65)	36.33 (9.03)	-0.28 (10.59)	0.94 (93.34)	0.222	[-Inf, 1.51]	0.13	Negligible
AL-WS-C	102	105.72 (22.71)	108.01 (24.13)	2.29 (26.70)	63	102.75 (21.22)	105.03 (18.82)	2.29 (21.64)	-0.002 (151.31)	0.499	[-Inf, 6.28]	0.00	Negligible
Anger Expression	100	29.69 (8.63)	27.78 (9.01)	-1.91 (9.76)	60	31.60 (9.63)	31.23 (9.27)	-0.37 (6.92)	1.17 (153.53)	0.123	[-0.65, Inf]	0.18	Negligible
Anger Control	100	22.00 (5.96)	23.84 (6.19)	1.84 (6.24)	60	21.78 (5.67)	22.37 (6.49)	0.58 (6.98)	-1.15 (113.68)	0.127	[-Inf, 0.56]	0.19	Negligible
Depression	93	5.85 (4.51)	5.43 (4.46)	-0.42 (5.03)	63	7.21 (3.72)	6.75 (4.21)	-0.46 (4.69)	-0.05 (139.33)	0.521	[-1.34, Inf]	-0.01	Negligible
Hope	97	24.71 (6.52)	27.08 (6.36)	2.37 (7.33)	66	25.70 (7.29)	24.95 (7.16)	-0.74 (6.29)	-2.90 (152.47)	0.002	[-Inf, -1.34]	0.45	Small
Self-esteem	99	15.33 (3.93)	15.97 (4.35)	0.64 (4.08)	65	13.71 (4.52)	14.32 (4.96)	0.62 (4.20)	-0.03 (134.34)	0.487	[-Inf, 1.08]	0.01	Negligible

Empathy	99	21.99 (5.95)	23.89 (6.10)	1.90 (6.26)	64	21.69 (5.97)	22.62 (6.33)	0.94 (5.95)	-0.99 (139.61)	0.163	[-Inf, 0.65]	0.16	Negligible

Table 11

Comparison of Gain Scores between NT and FE for the Control Group of Israel Students

Measure	*N*	Pretest *M* (SD)	Posttest *M* (SD)	Second Posttest *M* (SD)	NT Gain *M* (SD)	FE Gain *M* (SD)	t(*df*)	*p* (one-tailed)	95% CI	Cohen's *d*	
EALI-C	54	95.85 (14.26)	94.22 (18.31)	94.31 (15.72)	-1.63 (14.17)	0.09 (14.04)	-0.52 (53)	0.304	[-Inf, 3.86]	0.12	Small
EALI-6	54	23.87 (3.85)	23.07 (5.14)	22.96 (4.40)	-0.80 (5.46)	-0.11 (4.96)	-0.55 (53)	0.293	[-Inf, 1.41]	0.13	Negligible
EFI-C	56	84.21 (21.03)	86.91 (18.91)	87.50 (18.17)	2.70 (19.83)	0.59 (21.10)	0.46 (55)	0.677	[-Inf, 9.74]	-0.10	Negligible
AL-CDL-C	63	4.59 (2.42)	4.17 (2.51)	4.60 (2.43)	-0.41 (2.62)	0.43 (2.74)	-1.44 (62)	0.078	[-Inf, 0.13]	0.31	Small
AL-DS-C	61	36.61 (8.65)	36.33 (9.03)	36.03 (8.68)	-0.28 (10.59)	-0.30 (9.06)	0.008 (60)	0.503	[-Inf, 3.61]	-0.00	Negligible
AL-WS-C	63	102.75 (21.22)	105.03 (18.82)	102.86 (20.83)	2.29 (21.64)	-2.17 (18.87)	1.11 (62)	0.864	[-Inf, 11.19]	-0.22	Small
Anger Expression	60	31.60 (9.63)	31.23 (9.27)	31.87 (9.77)	-0.37 (6.92)	0.63 (8.15)	-0.62 (59)	0.733	[-3.67, Inf]	-0.13	Negligible
Anger Control	60	21.78 (5.67)	22.37 (6.49)	22.97 (4.95)	0.58 (6.98)	0.60 (5.15)	-0.01 (59)	0.495	[-Inf, 2.15]	0.00	Negligible
Depression	63	7.21 (3.72)	6.75 (4.21)	6.21 (4.08)	-0.46 (4.69)	-0.54 (4.54)	0.08 (62)	0.470	[-1.66, Inf]	0.02	Negligible
Hope	66	25.70 (7.29)	24.95 (7.16)	26.64 (6.73)	-0.74 (6.29)	1.68 (5.95)	-1.95 (65)	0.028	[-Inf, -0.35]	0.40	Small
Self-esteem	65	13.71 (4.52)	14.32 (4.96)	13.95 (4.56)	0.62 (4.20)	-0.37 (4.01)	1.10 (64)	0.863	[-Inf, 2.47]	-0.24	Small
Empathy	64	21.69 (5.97)	22.62 (6.33)	21.58 (6.38)	0.94 (5.95)	-1.05 (6.36)	1.60 (63)	0.943	[-Inf, 4.06]	-0.32	Small

We also performed one-tailed matched pair t-tests within each of the two groups to examine the amount of change on the 12 measures after each period of treatment, to examine the intervention effectiveness as well as the maintenance or growth of the intervention effect after the treatment ended. We examined the maintenance or growth effect after the interventions with the variables that showed significant differences in the within-group analyses. For the first period of treatment, when comparing the pre-test scores (T1) and post-test scores (T2), the experimental group showed significant decrease in anger expression ($p = 0.027$) and significant increases in *agape* love (EALI-C, $p = 0.036$), anger control ($p = 0.002$), hope ($p = 0.001$) and empathy ($p = 0.002$, See Table 12). The control group showed no significant changes on any of the 12 dependent variables during the no treatment period (See Table 13). For the second period of treatment, when comparing the post-test scores (T2) and the second-post test scores (T3), the control-group-turned-experimental group (after 12-week Forgiveness Education) had significantly increased in hope ($p = 0.012$).

Within the experimental group (see Table 12), when comparing the pre-test scores (T1) and the first follow-up scores (T3), it showed significant increases in forgiveness (EFI-C, $p = 0.001$), *agape* love (AL-WS-C, $p = 0.002$), anger control ($p = 0.000$), hope ($p = 0.000$), self-esteem ($p = 0.016$), and empathy ($p = 0.010$). When comparing the pre-test scores (T1) and the second follow-up scores (T4), it showed significant increases in *agape* love (EALI-C, $p = 0.046$), anger control ($p = 0.000$), hope ($p = 0.006$), self-esteem ($p = 0.006$) and empathy ($p = 0.040$).

Within the control-group-turned experimental group (see Table 13), when comparing the post-test scores (T2) and the follow-up scores (T4), it significantly increased in anger control ($p = 0.015$).

For *agape* love (EALI-C), anger control, hope, and empathy, no differences were found for the original experimental group from the post-test (T2) to the first follow-up (T3), the post-test (T2) to the second follow-up (T4), or the first follow-up (T3) to the second follow-up (T4). For empathy, no difference was found for the original experimental group from the first follow-up (T3) to the second follow-up (T4) (see Table 12). These indicate the effects of Forgiveness Education intervention for the experimental group were maintained after the 16-week no treatment period.

The results demonstrated the long-term effectiveness of Forgiveness Education. In addition, given the severity of COVID-19 in the second semester (Spring of 2022), the continued increase in *agape* love (EALI-C) and self-esteem after the treatment ended are worth noting. For forgiveness (EFI-C) and *agape* love (AL-WS-C), significant decreases from the first follow-up (T3) to the second follow-up (T4) in the experimental group were observed. However, the scores at the second follow-up (T4) didn't drop to the baseline (T1). For anger expression, a significant increase was observed from the post-test (T2) to the second follow-up (T4) in the experimental group. However, there is no significant difference between the score at the second follow-up (T4) and the score at the baseline (T1).

For hope, no difference was found for the control-group-turned experimental group from the second post-test (T3) to the follow-up (T4), which indicate the effect of the Forgiveness Education intervention for the control-group-turned experimental group was maintained after the 4-week no treatment period.

Table 12

Comparison of Scores within the Experimental Group of Israel Students

Measure						Intervention Effect						Maintenance Effect					
	N	Pretest *M* (*SD*)	Posttest *M* (*SD*)	1st Follow-up *M* (*SD*)	2nd Follow-up *M* (*SD*)	Pre- to post-Gain *M* (*SD*)	*p* (one-tailed)	Pre- to 1st follow-up Gain *M* (*SD*)	*p* (one-tailed)	Pre- to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)	Post to 1st follow-up Gain *M* (*SD*)	*p* (one-tailed)	Post to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)	1st follow-up to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)
EALI-C	77	95.08 (16.82)	98.18 (16.38）	98.00 (18.39)	98.61 (20.70)	3.10 (14.98)	0.036	2.92 (15.97)	0.056	3.53 (18.20)	0.046	-0.18 (13.89)	0.454	0.43 (16.42)	0.590	0.61 (13.55)	0.653
EALI-6	77	23.61 (5.09)	24.18 (4.49)	23.81 (5.25)	23.65 (5.53)	0.57 (4.81)	0.150	0.19 (5.31)	0.374	0.04 (5.44)	0.475						
EFI-C	83	86.20 (19.59)	88.55 (19.54)	92.43 (17.99)	88.67 (19.54)	2.35 (16.92)	0.105	6.23 (18.50)	0.001	2.47 (20.29)	0.135					-3.75 (17.41)	0.026
AL-CDL-C	101	5.33 (2.45)	5.69 (3.19)	5.46 (2.77)	5.31 (3.14)	0.37 (3.46)	0.145	0.13 (3.37)	0.351	-0.02 (3.43)	0.523						
AL-DS-C	98	37.94 (8.30)	38.97 (8.73)	38.13 (9.03)	37.80 (8.65)	1.03 (10.19)	0.160	0.19 (10.43)	0.427	-0.14 (10.61)	0.553						
AL-WS-C	102	105.72 (22.71)	108.01 (24.13)	112.43 (18.26)	108.55 (102)	2.29 (26.70)	0.194	6.72 (22.95)	0.002	2.83 (24.88)	0.126					-3.88 (18.75)	0.020
Anger Expression	100	29.69 (8.63)	27.78 (9.01)	28.19 (9.11)	29.63 (11.32)	-1.91 (9.76)	0.027	-1.50 (9.40)	0.057	-0.06 (12.66)	0.481	0.41 (8.27)	0.311	1.85 (8.65)	0.018		
Anger Control	100	22.00 (5.96)	23.84 (6.19)	24.87 (6.00)	24.68 (6.74)	1.84 (6.24)	0.002	2.87 (7.02)	0.000	2.68 (7.55)	0.000	1.03 (5.93)	0.957	0.84 (6.92)	0.886	-0.19 (5.75)	0.371
Depression	93	5.85 (4.51)	5.43 (4.46)	5.65 (4.16)	5.68 (5.22)	-0.42 (5.03)	0.212	-0.20 (4.91)	0.345	-0.17 (6.04)	0.392						
Hope	97	24.71 (6.52)	27.08 (6.36)	27.65 (6.06)	26.97 (7.19)	2.37 (7.33)	0.001	2.94 (7.88)	0.000	2.26 (8.66)	0.006	0.57 (6.45)	0.806	-0.11 (7.20)	0.439	-0.68 (7.40)	0.184
Self-esteem	99	15.33 (3.93)	15.97 (4.35)	16.18 (4.50)	16.51 (4.67)	0.64 (4.08)	0.062	0.85 (3.87)	0.016	1.17 (4.50)	0.006					0.32 (3.30)	0.834

Empathy	99	21.99 (5.95)	23.89 (6.10)	23.54 (5.97)	23.34 (7.32)	1.90 (6.26)	0.002	1.55 (6.46)	0.010	1.35 (7.61)	0.040	-0.35 (6.12)	0.283	-0.55 (7.66)	0.240	-0.19 (6.64)	0.387

Table 13

Comparison of Scores within the Control-group-turned-experimental Group of Israel Students

								Intervention Effect			Maintenance Effect		
Measure	*N*	Pretest *M* (SD)	Posttest *M* (SD)	Second Posttest *M* (SD)	Follow-up *M* (SD)	NT Gain *M* (SD)	*p* (two-tailed)	Post- to 2nd post-) *M* (SD)	*p* (one-tailed)	Post- to follow-up) *M (SD)*	*p* (one-tailed)	2nd post-to follow-up) *M (SD)*	*p* (one-tailed)
EALI-C	54	95.85 (14.26)	94.22 (18.31)	94.31 (15.72)	94.15 (21.00)	-1.63 (14.17)	0.402	0.09 (14.04)	0.481	-0.07 (21.32)	0.510		
EALI-6	54	23.87 (3.85)	23.07 (5.14)	22.96 (4.40)	22.43 (4.41)	-0.80 (5.46)	0.289	-0.11 (4.96)	0.565	-0.65 (5.85)	0.790		
EFI-C	56	84.21 (21.03)	86.91 (18.91)	87.50 (18.17)	87.50 (23.32)	2.70 (19.83)	0.313	0.59 (21.10)	0.418	0.39 (28.14)	0.459		
AL-CDL-C	63	4.59 (2.42)	4.17 (2.51)	4.60 (2.43)	3.81 (2.53)	-0.41 (2.62)	0.215	0.43 (2.74)	0.110	-0.37 (3.29)	0.809		
AL-DS-C	61	36.61 (8.65)	36.33 (9.03)	36.03 (8.68)	35.28 (9.23)	-0.28 (10.59)	0.838	-0.30 (9.06)	0.600	-1.05 (9.67)	0.800		
AL-WS-C	63	102.75 (21.22)	105.03 (18.82)	102.86 (20.83)	104.22 (19.28)	2.29 (21.64)	0.405	-2.17 (18.87)	0.818	-0.81 (20.19)	0.624		
Anger Expression	60	31.60 (9.63)	31.23 (9.27)	31.87 (9.77)	30.90 (10.94)	-0.37 (6.92)	0.683	0.63 (8.15)	0.725	-0.33 (9.17)	0.390		
Anger Control	60	21.78 (5.67)	22.37 (6.49)	22.97 (4.95)	23.93 (6.04)	0.58 (6.98)	0.520	0.60 (5.15)	0.185	1.57 (5.43)	0.015		
Depression	63	7.21 (3.72)	6.75 (4.21)	6.21 (4.08)	5.94 (4.42)	-0.46 (4.69)	0.439	-0.54 (4.54)	0.175	-0.81 (5.02)	0.103		
Hope	66	25.70 (7.29)	24.95 (7.16)	26.64 (6.73)	25.89 (7.10)	-0.74 (6.29)	0.341	1.68 (5.95)	0.012	0.94 (7.13)	0.144	-0.74 (6.16)	0.834
Self-esteem	65	13.71 (4.52)	14.32 (4.96)	13.95 (4.56)	14.52 (4.92)	0.62 (4.20)	0.242	-0.37 (4.01)	0.770	0.20 (4.57)	0.363		
Empathy	64	21.69 (5.97)	22.62 (6.33)	21.58 (6.38)	22.20 (7.09)	0.94 (5.95)	0.212	-1.05 (6.36)	0.904	-0.42 (5.70)	0.722		

Taiwan Students

Reliability and Validity of the Initial Assessment

Cronbach's alpha of each measure was assessed at the initial assessment (T1) only. Cronbach's alpha for the 24-item Enright *Agape* Love Inventory for children (EALI-C) is 0.84 and Cronbach's alpha for three subscales (EALI-6) assessing change of *agape* love across time is 0.67. EALI-C and EALI-6 both positively correlated with EFI-C (forgiveness), AL-CDL-C, AL-DS-C, AL-WS-C, hope, self-esteem, and empathy; negatively correlated with anger and depression.

Cronbach's alpha for the 30-item Enright Forgiveness Inventory for children (EFI-C) was 0.95. EFI-C (forgiveness) is positively correlated with EALI-C, AL-DS-C, hope, self-esteem and negatively correlated with anger and depression. The four scales measuring *agape* love, EALI-C, AL-CDL-C, AL-DS-C, and AL-WS-C, are positively correlated with each other. The correlations show convergent validity of the four *agape* love scales, which are similar to the correlations achieved in the US children's sample. The *agape* love knowledge scale, AL-KS, is neither positively correlated with any of the four *agape* love scales nor negatively correlated with the anger or depression scale. Cronbach's alpha of internal consistency for AL-CDL-C is 0.64; for AL-DS-C is 0.67; for AL-WS-C is 0.93; for AL-KS-C is -0.15. We did not expect the internal consistency for AL-KS-C to be high at the pre-test because children from the general population did not have prior knowledge or understanding about *agape* love. Cronbach's alpha for the other measures are as follows: anger expression (0.92), anger control (0.85), depression (0.83), hope (0.86), self-esteem (0.82), and empathy (0.90).

Table 14

Correlations for Variables in the Initial Assessment (T1) of Taiwan Students

Measures	M	SD	*N*	1	2	3	4	5	6	7	8	9	10	11
1. EALI-C	93.98	17.98	411											
2. EALI-6	24.49	5.97	411	.88**										
3. EFI-C	75.58	20.84	414	.36**	.31**									
4. AL-CDL-C	6.18	3	430	.28**	.25**	0.05								
5. AL-DS-C	32.96	7.08	430	.33**	.32**	.22**	.13**							
6. AL-WS-C	87.59	25.48	427	.31**	.26**	.18**	.15**	.41**						
7. AL-KS-C	4.27	1.23	430	0.05	0.03	.11*	0.04	0.05	-0.02					
8. Anger	39.19	10.54	429	-.22**	-.19**	-.28**	0.07	-.17**	-.11*	-0.02				
9. Depression	4.99	4.61	429	-.14**	-.16**	-.18**	-0.01	-.10*	0.04	-0.01	.53**			
10. Hope	22.87	7.58	429	.28**	.33**	.14**	0.03	.34**	.41**	-0.09	-.24**	-.25**		
11. Self-esteem	14.99	5.04	419	.17**	.24**	.19**	-.11*	.15**	.11*	-0.05	-.45**	-.51**	.48**	
12. Empathy	20.92	7.78	429	.30**	.26**	0.09	.11*	.43**	.51**	0.02	-0.06	-0.04	.42**	.19**

Effect of the *Agape* Love and Forgiveness Education Program

None of the classes in the experimental group and none of the classes in the control group were eliminated from the analysis based on teachers' self-report (not adhering to the teaching schedule) and researchers' observation (not adhering to the assessment schedule) on the intervention. Only students who were able to finish four times of surveys with the responses passed screening criteria were retained in the analyses. To examine the possibility of pre-test differences between the two groups, two tailed independent t-tests were conducted to compare scores on each measure between the two groups. No differences were found at baseline for all 12 measures, including EALI-C, EFI-C (forgiveness), AL-CDL-C, AL-DS-C, AL-WS-C, AL-KS-C, anger expression, anger control, depression, hope, self-esteem, and empathy. Please see Table 15 for means and standard deviations and the footnote.

Table 15

Descriptive Statistics of Agape Love Measures and Psychological Compromises and Well-being of Taiwan Students

	Group	N	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	First follow-up to EG/ Second post-test to CG *M* (*SD*)	Second follow-up to EG/ First follow-up to CG *M* (*SD*)
EALI-C	EG	171	92.59 (18.09)	95.78 (18.65)	94.13 (20.24)	94.50 (17.70)
	CG	177	94.97 (18.61)	96.64 (17.99)	95.23 (16.89)	95.81 (19.91)
EALI-6	EG	171	23.93 (6.23)	25.04 (5.98)	25.02 (6.08)	25.24 (6.00)
	CG	177	24.90 (6.02)	25.58 (5.41)	25.18 (5.54)	25.37 (6.27)
EFI-C	EG	176	75.99 (20.56)	81.06 (21.51)	79.86 (22.42)	81.28 (20.93)
	CG	181	74.40 (21.86)	79.86 (20.20)	81.91 (18.93)	80.69 (20.49)

AL-CDL-C	EG	189	5.98 (2.96)	6.36 (3.34)	6.08 (3.26)	5.19 (3.36)
	CG	192	6.30 (3.12)	6.09 (3.08)	6.21 (3.17)	5.77 (3.07)
AL-DS-C	EG	187	32.68 (7.02)	34.50 (7.59)	33.75 (8.05)	34.33 (6.94)
	CG	191	33.32 (7.32)	32.84 (6.53)	34.60 (6.90)	33.86 (6.41)
AL-WS-C	EG	189	87.18 (24.46)	94.57 (25.04)	92.94 (24.94)	93.52 (25.27)
	CG	190	89.19 (26.24)	93.99 (23.24)	93.12 (24.45)	94.48 (24.70)
AL-KS-C	EG	189	4.20 (1.28)	4.70 (1.30)	4.58 (1.19)	4.63 (1.04)
	CG	192	4.28 (1.16)	4.27 (1.07)	4.65 (1.12)	4.57 (1.13)
Anger Expression	EG	189	30.93 (10.32)	31.92 (9.24)	31.17 (10.25)	31.14 (10.08)
	CG	192	29.20 (9.22)	30.03 (9.20)	30.29 (9.13)	30.28 (9.34)
Anger Control	EG	189	21.95 (6.18)	21.30 (6.14)	21.44 (6.48)	21.44 (6.10)
	CG	192	22.94 (6.34)	22.84 (5.99)	22.32 (6.25)	22.20 (6.11)
Depression	EG	189	5.23 (4.67)	4.65 (4.64)	4.02 (4.24)	3.86 (4.38)
	CG	192	4.80 (4.50)	4.94 (4.49)	4.24 (3.98)	4.44 (4.41)
Hope	EG	189	22.72 (7.64)	23.37 (7.84)	23.07 (8.45)	23.89 (8.18)
	CG	192	23.39 (7.40)	23.59 (7.25)	24.33 (7.29)	23.43 (7.10)
Self-esteem	EG	188	14.71 (5.28)	14.81 (5.19)	14.64 (5.68)	14.63 (5.76)
	CG	192	15.20 (4.93)	15.22 (4.88)	15.20 (5.19)	14.90 (5.54)
Empathy	EG	189	20.48 (7.65)	21.52 (7.57)	20.96 (8.89)	21.80 (8.28)
	CG	192	21.84 (7.80)	22.09 (7.96)	22.57 (7.79)	21.97 (8.41)

Cutoff scores for the eight-item Patient Health Questionnaire depression scale (PHQ-8): 0 - 4 = no significant depressive symptoms; 5 - 9 = mild depressive symptoms; 10 - 14 = moderate; 15 - 19 = moderately severe; 20 - 24, severe. (Kroenke et al., 2001).

To test the effectiveness of Forgiveness Education intervention compared to the no treatment wait list control (teaching regular school curriculum), we first conducted a one-tailed independent t-test to compare the gain scores between the two groups from pretest to post-test using data from the 20 classes (EG = 11; CG = 9) with students who completed and passed the screening criteria of all four surveys. We used a one-tailed t-test here based on our original hypotheses: we assumed that Forgiveness Education intervention would have better effects than the no treatment wait list control. The results showed that the experimental group had significantly greater increase in *agape* love (AL-CDL-C: $p = 0.023$ and AL-DS-C: $p = 0.001$) and in *agape* love knowledge (AL-KS-C: $p = 0.001$), compared to the control group (see Table 16).

Next, a one-tailed paired sample t-test was run within the control group compared to itself once this group had Forgiveness Education. We compared the gain scores (post-test scores minus pretest scores) when the control group had no treatment (NT, teaching a regular school curriculum) with the gain scores (second post-test minus post-test scores) for the control group-turned-experimental group) once the control group became the experimental group with Forgiveness Education (FE) intervention. This particular analysis was done with the data from the 9 control group classes with students who completed and passed the screening criteria of all four surveys. The results showed that the control group had significant growth in *agape* love (AL-DS-C, $p = 0.004$) and *agape* love knowledge (AL-KS, $p = 0.016$) after receiving Forgiveness Education compared to the teaching regular school curriculum with no treatment (see Table 17).

Table 16

Comparison of Gain Scores from Pretest to Posttest between the Two Groups of Taiwan Students

Measures	Experimental Group				Control Group				t(*df*)	*p* (one-tailed)	95% CI	Cohen's *d*	
	N	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain Scores *M* (*SD*)	*N*	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain Scores *M* (*SD*)					
EALI-C	171	92.59 (18.09)	95.78 (18.65)	3.19 (16.13)	177	94.97 (18.61)	96.64 (17.99)	1.67 (13.19)	-0.96 (328.23)	0.170	[-Inf, 1.10]	0.10	Negligible
EALI-6	171	23.93 (6.23)	25.04 (5.98)	1.11 (5.76)	177	24.90 (6.02)	25.58 (5.41)	0.68 (4.41)	-0.78 (318.35)	0.219	[-Inf, 0.48]	0.08	Negligible
EFI-C	176	75.99 (20.56)	81.06 (21.51)	5.07 (19.72)	181	74.40 (21.86)	79.86 (20.20)	5.46 (20.25)	0.19 (355)	0.574	[-Inf, 3.88]	-0.02	Negligible
AL-CDL-C	189	5.98 (2.96)	6.36 (3.34)	0.38 (3.52)	192	6.30 (3.12)	6.09 (3.08)	-0.31 (3.16)	-1.99 (373.33)	0.023	[-Inf, -0.12]	0.20	Small
AL-DS-C	187	32.68 (7.02)	34.50 (7.59)	1.82 (7.42)	191	33.32 (7.32)	32.84 (6.53)	-0.48 (6.84)	0.94 (93.34)	0.001	[-Inf, -1.09]	0.32	Small
AL-WS-C	189	87.18 (24.46)	94.57 (25.04)	7.39 (23.33)	190	89.19 (26.24)	93.99 (23.24)	4.79 (21.67)	-1.12 (374.64)	0.131	[-Inf, 1.22]	0.12	Negligible
AL-KS-C	189	4.20 (1.28)	4.70 (1.30)	0.50 (1.68)	192	4.28 (1.16)	4.27 (1.07)	-0.01 (1.46)	-3.18 (370.15)	0.001	[-Inf, -0.25]	0.33	Small
Anger Expression	189	30.93 (10.32)	31.92 (9.24)	0.99 (8.75)	192	29.20 (9.22)	30.03 (9.20)	0.83 (7.31)	-0.21 (365.47)	0.582	[-1.54, Inf]	-0.02	Negligible
Anger Control	189	21.95 (6.18)	21.30 (6.14)	-0.65 (6.59)	192	22.94 (6.34)	22.84 (5.99)	-0.10 (5.41)	0.88 (362.94)	0.812	[-Inf, 1.57]	-0.09	Negligible
Depression	189	5.23 (4.67)	4.65 (4.64)	-0.58 (4.56)	192	4.80 (4.50)	4.94 (4.49)	0.15 (4.97)	1.49 (377.14)	0.068	[-0.08, Inf]	0.15	Negligible
Hope	189	22.72 (7.64)	23.37 (7.84)	0.65 (7.69)	192	23.39 (7.40)	23.59 (7.25)	0.20 (6.96)	-0.60 (373.99)	0.276	[-Inf, 0.79]	0.06	Negligible

Self-esteem	188	14.71 (5.28)	14.81 (5.19)	0.10 (3.89)	192	15.20 (4.93)	15.22 (4.88)	0.02 (3.62)	-0.19 (374.83)	0.423	[-Inf, 0.56]	0.02	Negligible
Empathy	189	20.48 (7.65)	21.52 (7.57)	1.05 (6.76)	192	21.84 (7.80)	22.09 (7.96)	0.26 (6.56)	-1.16 (378.19)	0.123	[-Inf, 0.33]	0.12	Negligible

Table 17

Comparison of Gain Scores between NT and FE for the Control Group of Taiwan Students

Measure	*N*	Pretest *M* (SD)	Posttest *M* (SD)	Second Posttest *M* (SD)	NT Gain *M* (SD)	FE Gain *M* (SD)	t(*df*)	*p* (one-tailed)	95% CI	Cohen's *d*	
EALI-C	177	94.97 (18.61)	96.64 (17.99)	95.23 (16.89)	1.67 (13.19)	-1.41 (15.57)	1.80 (176)	0.964	[-Inf, 5.91]	-0.21	Small
EALI-6	177	24.90 (6.02)	25.58 (5.41)	25.18 (5.54)	0.68 (4.41)	-0.41 (5.34)	1.88 (176)	0.969	[-Inf, 2.04]	-0.22	Small
EFI-C	181	74.40 (21.86)	79.86 (20.20)	81.91 (18.93)	5.46 (20.25)	2.05 (17.26)	1.47 (180)	0.929	[-Inf, 7.25]	-0.18	Negligible
AL-CDL-C	192	6.30 (3.12)	6.09 (3.08)	6.21 (3.17)	-0.31 (3.16)	0.13 (3.30)	-1.09 (191)	0.138	[-Inf, 0.22]	0.13	Negligible
AL-DS-C	191	33.32 (7.32)	32.84 (6.53)	34.60 (6.90)	-0.48 (6.84)	1.76 (6.88)	-2.72 (190)	0.004	[-Inf, -0.88]	0.33	Small
AL-WS-C	190	89.19 (26.24)	93.99 (23.24)	93.12 (24.45)	4.79 (21.67)	-0.87 (19.33)	2.27 (189)	0.988	[-Inf, 9.79]	-0.28	Small
AL-KS-C	192	4.28 (1.16)	4.27 (1.07)	4.65 (1.12)	-0.01 (1.46)	0.39 (1.56)	-2.17 (191)	0.016	[-Inf, -0.09]	0.26	Small
Anger Expression	192	29.20 (9.22)	30.03 (9.20)	30.29 (9.13)	0.83 (7.31)	0.26 (7.96)	0.62 (191)	0.268	[-0.94, Inf]	0.07	Negligible
Anger Control	192	22.94 (6.34)	22.84 (5.99)	22.32 (6.25)	-0.10 (5.41)	-0.52 (5.73)	0.62 (191)	0.733	[-Inf, 1.50]	-0.07	Negligible
Depression	192	4.80 (4.50)	4.94 (4.49)	4.24 (3.98)	0.15 (4.97)	-0.70 (3.88)	1.50 (191)	0.068	[-0.09, Inf]	0.19	Negligible
Hope	192	23.39 (7.40)	23.59 (7.25)	24.33 (7.29)	0.20 (6.96)	0.73 (6.88)	-0.64 (191)	0.261	[-Inf, 0.84]	0.08	Negligible
Self-esteem	192	15.20 (4.93)	15.22 (4.88)	15.20 (5.19)	0.02 (3.62)	-0.02 (3.64)	0.09 (191)	0.537	[-Inf, 0.78]	-0.01	Negligible

Empathy	192	21.84 (7.80)	22.09 (7.96)	22.57 (7.79)	0.26 (6.56)	0.47 (6.76)	-0.27 (191)	0.396	[-Inf, 1.14]	0.03	Negligible

We also performed one-tailed matched pair t-tests within each of the two groups to examine the amount of change on the 13 measures after each period of treatment, to further examine the within-group intervention effectiveness as well as the maintenance or growth of the intervention effect after the treatment ended. We examined the maintenance or growth effect after the interventions with the variables that showed significant differences in the within-group analyses.

For the first period of treatment, when comparing the pre-test scores (T1) and post-test scores (T2), the experimental group showed significant decrease in depression ($p = 0.040$) as well as significant increases in *agape* love (EALI-C: $p = 0.005$, EALI-6: $p = 0.007$, AL-DS-C: $p = 0.000$, AL-WS-C: $p = 0.000$), *agape* love knowledge (AL-KS-C: $p = 0.000$), forgiveness ($p = 0.000$), and empathy ($p = 0.017$, See Table 18). The control group showed a significant increase in *agape* love (EALI-6: $p = 0.042$ and AL-WS-C: $p = 0.003$) and forgiveness (EFI: $p = 0.000$) during this wait/no treatment period (See Table 19). For the second period of treatment, when comparing the post-test scores (T2) and the second-post test scores (T3), the control-group-turned-experimental group (after 12-week Forgiveness Education) had significant increase in *agape* love (AL-DS-C: $p = 0.000$), *agape* love knowledge (AL-KS-C: $p = 0.000$) as well as significant decrease in depression ($p = 0.007$).

Within the experimental group (see Table 18), when comparing the pre-test scores (T1) and the first follow-up scores (T3), it showed significant increases in *agape* love (EALI-6: $p = 0.005$, AL-DS-C: $p = 0.042$, AL-WS-C: $p = 0.001$) forgiveness (EFI-C: $p = 0.007$), *agape* love knowledge (AL-KS-C: $p = 0.001$), and significant decrease in depression ($p = 0.000$). When comparing the pre-test scores (T1) and the second follow-up scores (T4), it showed significant increases in *agape* love (EALI-6: $p = 0.003$, AL-DS-C: $p = 0.003$, AL-WS-C: $p = 0.000$)

forgiveness (EFI-C: $p = 0.000$), *agape* love knowledge (AL-KS-C: $p = 0.000$), hope ($p = 0.026$), and empathy ($p = 0.011$), as well as significant decrease in depression ($p = 0.000$).

Within the control-group-turned experimental group (see Table 19), when comparing the post-test scores (T2) and the follow-up scores (T4), it had significant increases in *agape* love (AL-DS-C: $p = 0.011$) and *agape* love knowledge (AL-KS-C: $p = 0.003$) as well as a significant decrease in depression ($p = 0.042$).

For *agape* love (EALI-6, AL-DS-C, AL-WS-C), forgiveness, *agape* love knowledge (AL-KS-C) depression, and empathy, no differences were found for the original experimental group from the post-test (T2) to the first follow-up (T3), the post-test (T2) to the second follow-up (T4), or the first follow-up (T3) to the second follow-up (T4) (see Table 18). These indicate the effects of Forgiveness Education intervention for the experimental group were maintained after the 16-week no treatment period. The results demonstrated the long-term effectiveness of Forgiveness Education intervention. A significant decrease in *agape* love (AL-CDL-C) from the first follow-up (T3) to the second follow-up (T4) in the experimental group was observed. Upon a further check, the score at the second follow-up (T4) is significantly lower compared to the score at the baseline (T1), indicating the washout of the intervention effect for this variable.

No differences were found for *agape* love knowledge (AL-KS-C) and depression from second post-test (T3) to the follow-up (T4), indicating the effects of Forgiveness Education intervention were maintained after 4-week no treatment period for the control-group-turned-experimental group. For *agape* love (AL-DS-C), there was a significant decrease between the second post-test to the follow-up. However, the score at the second follow-up (T4) didn’t drop to the baseline (T1).

Table 18

Comparison of Scores within the Experimental Group of Taiwan Students

Measure						Intervention Effect						Maintenance Effect					
	N	Pretest *M* (*SD*)	Posttest *M* (*SD*)	1st Follow-up *M* (*SD*)	2nd Follow-up *M* (*SD*)	Pre- to post-Gain *M* (*SD*)	*p* (one-tailed)	Pre- to 1st follow-up Gain *M* (*SD*)	*p* (one-tailed)	Pre- to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)	Post to 1st follow-up Gain *M* (*SD*)	*p* (one-tailed)	Post to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)	1st follow-up to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)
EALI-C	171	92.59 (18.09)	95.78 (18.65)	94.13 (20.24)	94.50 (17.70)	3.19 (16.13)	0.005	1.53 (16.60)	0.114	1.91 (17.07)	0.072	-1.65 (14.50)	0.069	-1.27 (15.79)	0.146	0.37 (13.38)	0.643
EALI-6	171	23.93 (6.23)	25.04 (5.98)	25.02 (6.08)	25.24 (6.00)	1.11 (5.76)	0.007	1.09 (5.40)	0.005	1.31 (6.03)	0.003	-0.02 (4.69)	0.481	0.20 (5.68)	0.681	0.22 (4.55)	0.738
EFI-C	176	75.99 (20.56)	81.06 (21.51)	79.86 (22.42)	81.28 (20.93)	5.07 (19.72)	0.000	3.87 (20.61)	0.007	5.28 (19.35)	0.000	-1.20 (17.35)	0.180	0.22 (17.35)	0.566	1.41 (15.67)	0.884
AL-CDL-C	189	5.98 (2.96)	6.36 (3.34)	6.08 (3.26)	5.19 (3.36)	0.38 (3.52)	0.072	0.10 (3.34)	0.348	-0.79 (3.31)	0.999						
AL-DS-C	187	32.68 (7.02)	34.50 (7.59)	33.75 (8.05)	34.33 (6.94)	1.82 (7.42)	0.000	1.06 (8.40)	0.042	1.65 (8.20)	0.003	-0.75 (7.38)	0.082	-0.17 (7.04)	0.370	0.58 (6.80)	0.879
AL-WS-C	189	87.18 (24.46)	94.57 (25.04)	92.94 (24.94)	93.52 (25.27)	7.39 (23.33)	0.000	5.76 (25.32)	0.001	6.34 (24.75)	0.000	-1.63 (19.53)	0.126	-1.05 (20.80)	0.245	0.59 (18.85)	0.666
AL-KS-C	189	4.20 (1.28)	4.70 (1.30)	4.58 (1.19)	4.63 (1.04)	0.50 (1.68)	0.000	0.38 (1.63)	0.001	0.43 (1.62)	0.000	-0.12 (1.42)	0.120	-0.07 (1.47)	0.260	0.05 (1.31)	0.709
Anger Expression	189	30.93 (10.32)	31.92 (9.24)	31.17 (10.25)	31.14 (10.08)	0.99 (8.75)	0.940	0.25 (9.67)	0.638	0.21 (10.85)	0.606						
Anger Control	189	21.95 (6.18)	21.30 (6.14)	21.44 (6.48)	21.44 (6.10)	-0.65 (6.59)	0.912	-0.50 (7.22)	0.830	-0.51 (7.46)	0.825						
Depression	189	5.23 (4.67)	4.65 (4.64)	4.02 (4.24)	3.86 (4.38)	-0.58 (4.56)	0.040	-1.21 (4.57)	0.000	-1.37 (5.28)	0.000	-0.63 (3.77)	0.989	-0.78 (4.46)	0.992	-0.15 (4.07)	0.698
Hope	189	22.72 (7.64)	23.37 (7.84)	23.07 (8.45)	23.89 (8.18)	0.65 (7.69)	0.123	0.35 (8.71)	0.288	1.17 (8.27)	0.026						

Self-esteem	188	14.71 (5.28)	14.81 (5.19)	14.64 (5.68)	14.63 (5.76)	0.10 (3.89)	0.368	-0.07 (4.67)	0.586	-0.08 (4.91)	0.588						
Empathy	189	20.48 (7.65)	21.52 (7.57)	20.96 (8.89)	21.80 (8.28)	1.05 (6.76)	0.017	0.48 (7.95)	0.203	1.32 (7.88)	0.011	-0.57 (7.26)	0.143	0.28 (7.18)	0.701	0.84 (6.22)	0.968

Table 19

Comparison of Scores within the Control-group-turned-experimental Group of Taiwan Students

Measure	*N*	Pretest *M* (SD)	Posttest *M* (SD)	Second Posttest *M* (SD)	Follow-up *M* (SD)	NT Gain *M* (SD)	*p* (two-tailed)	Intervention Effect				Maintenance Effect	
								Post- to 2nd post-) *M* (SD)	*p* (one-tailed)	Post- to follow-up) *M (SD)*	*p* (one-tailed)	2nd post-to follow-up) *M (SD)*	*p* (one-tailed)
EALI-C	177	94.97 (18.61)	96.64 (17.99)	95.23 (16.89)	95.81 (19.91)	1.67 (13.19)	0.093	-1.41 (15.57)	0.885	-0.84 (16.12)	0.755		
EALI-6	177	24.90 (6.02)	25.58 (5.41)	25.18 (5.54)	25.37 (6.27)	0.68 (4.41)	0.042	-0.41 (5.34)	0.844	-0.21 (5.28)	0.701		
EFI-C	181	74.40 (21.86)	79.86 (20.20)	81.91 (18.93)	80.69 (20.49)	5.46 (20.25)	0.000	2.05 (17.26)	0.056	0.82 (18.84)	0.279		
AL-CDL-C	192	6.30 (3.12)	6.09 (3.08)	6.21 (3.17)	5.77 (3.07)	-0.31 (3.16)	0.179	0.13 (3.30)	0.300	-0.32 (3.36)	0.904		
AL-DS-C	191	33.32 (7.32)	32.84 (6.53)	34.60 (6.90)	33.86 (6.41)	-0.48 (6.84)	0.332	1.76 (6.88)	0.000	1.02 (6.16)	0.011	-0.74 (5.94)	0.043
AL-WS-C	190	89.19 (26.24)	93.99 (23.24)	93.12 (24.45)	94.48 (24.70)	4.79 (21.67)	0.003	-0.87 (19.33)	0.733	0.49 (19.43)	0.363		
AL-KS-C	192	4.28 (1.16)	4.27 (1.07)	4.65 (1.12)	4.57 (1.13)	-0.01 (1.46)	0.921	0.39 (1.56)	0.000	0.31 (1.51)	0.003	-0.08 (1.29)	0.202
Anger Expression	192	29.20 (9.22)	30.03 (9.20)	30.29 (9.13)	30.28 (9.34)	0.83 (7.31)	0.121	0.26 (7.96)	0.675	0.26 (8.15)	0.668		
Anger Control	192	22.94 (6.34)	22.84 (5.99)	22.32 (6.25)	22.20 (6.11)	-0.10 (5.41)	0.790	-0.52 (5.73)	0.893	-0.64 (5.93)	0.932		
Depression	192	4.80 (4.50)	4.94 (4.49)	4.24 (3.98)	4.44 (4.41)	0.15 (4.97)	0.685	-0.70 (3.88)	0.007	-0.51 (4.04)	0.042	0.19 (3.42)	0.218
Hope	192	23.39 (7.40)	23.59 (7.25)	24.33 (7.29)	23.43 (7.10)	0.20 (6.96)	0.686	0.73 (6.88)	0.070	-0.16 (7.11)	0.623		

Self-esteem	192	15.20 (4.93)	15.22 (4.88)	15.20 (5.19)	14.90 (5.54)	0.02 (3.62)	0.937	-0.02 (3.64)	0.532	-0.32 (3.71)	0.886
Empathy	192	21.84 (7.80)	22.09 (7.96)	22.57 (7.79)	21.97 (8.41)	0.26 (6.56)	0.590	0.47 (6.76)	0.166	-0.12 (7.20)	0.591

All Teachers Combined

Reliability and Validity of the Initial Assessment

Cronbach's alpha of each measure was assessed at the initial assessment (T1) only. Cronbach's alpha for the 36-item Enright *Agape* Love Inventory for adults (EALI-36) is 0.92 and Cronbach's alpha for three subscales (EALI-18) assessing change of *agape* love across time is 0.85. EALI-36 and EALI-18 were both positively correlated with EFI-30 (forgiveness) and AL-WS, but showed neither significant positive correlation with AL-CDL, AL-DS, hope, self-esteem, or empathy nor significant negative correlation with anger or depression. Cronbach's alpha for the 30-item Enright Forgiveness Inventory for adults (EFI-30) was 0.98. EFI-30 (forgiveness) is positively correlated with EALI-36, EALI-18 and AL-WS, but showed neither significant positive correlation with AL-CDL, AL-DS, hope, self-esteem, or empathy nor significant negative correlation with anger or depression. The four scales measuring *agape* love, only EALI-36 and AL-WS are positively correlated with each other. The correlations show convergent validity of these two *agape* love scales, which are similar to the correlations achieved in the US adults' sample. AL-CDL and AL-DS didn't show significant positive correlations with either EALI-36 or AL-WS. AL-KS, the *agape* love knowledge scale, didn't show significant positive correlation with either one of the four scales measuring *agape* love. Cronbach's alpha of internal consistency for AL-CDL is 0.69; for AL-DS is 0.73; for AL-WS is 0.91; for AL-KS (measured within Taiwanese teachers only) is -0.53. We did not expect the internal consistency for AL-KS to be high at the pre-test because adults from the general population did not have prior knowledge or understanding about *agape* love. Cronbach's alpha for the other measures are as follows: anger (0.72), depression (0.80), hope (0.84), self-esteem (0.85), empathy (0.66) and class cooperation (0.86).

Table 20

Correlations for Variables in the Initial Assessment (T1) of All Teachers

Measures	M	SD	*N*	1	2	3	4	5	6	7	8	9	10	11	12
1. EALI-36	159.31	23.91	52												
2. EALI-18	76.5	13.14	52	.93**											
3. EFI-30	113.6	38.78	55	.33*	.30*										
4. AL-CDL	9.84	4.19	56	0.12	0.16	0.01									
5. AL-DS	38.44	5.6	55	0.11	-0.01	.35*	-0.09								
6. AL-WS	89.14	14.42	56	.48**	.40**	.37**	-0.04	0.25							
7. Anger	6.96	4.42	56	-0.26	-0.2	-0.19	-0.08	-0.21	-.30*						
8. Depression	3.61	3.02	56	-0.19	-0.18	-0.09	0.01	-0.22	-0.01	.48**					
9. Hope	40.29	4.34	56	0.23	0.19	0.11	-0.19	.36**	.39**	-.28*	-.38**				
10. Self-esteem	31.66	4.24	56	0.13	-0.01	0.17	-0.08	.31*	.38**	-.28*	-.37**	.70**			
11. Empathy	49.39	4.83	56	0.17	0.18	0.18	0	.44**	.50**	-0.26	-0.05	.31*	0.19		
12. Class Cooperation	91.51	17.27	55	0.01	-0.06	0.1	-0.15	0.13	.41**	-.27*	0.04	0.1	0.03	.42**	
13. AL-KS	8.63	1.12	19	-0.16	-0.18	0.01	-0.23	.52*	0.07	0.1	0.06	0.05	0.06	0.19	0.03

Effect of the *Agape* Love and Forgiveness Education Program

None of the teachers' responses in the experimental group and none of the teachers' responses in the control group was eliminated from the analysis based on teachers' self-report (not adhering to the teaching schedule) or researchers' observation (not adhering to the assessment schedule) on the intervention. Teachers who were able to finish four times of surveys with the responses passed screening criteria were retained in the analyses. Thus, the substitute teachers' responses were excluded from the analyses. To examine the possibility of pre-test differences between the two groups, two tailed independent t-tests were conducted to compare scores on each measure between the two groups. No differences were found at baseline for all 13 measures, including EALI-36, EALI-18, EFI-30 (forgiveness), AL-CDL, AL-DS, AL-WS, AL-KS, anger (CAS), depression (PHQ-8), hope (HHI), self-esteem (RSS), empathy (EQ), and class cooperation (CCS). Please see Table 21 for means and standard deviations and the footnote.

Table 21

Descriptive Statistics of Agape Love Measures and Psychological Compromises and Well-being of All Teachers

	Group	N	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	First follow-up to EG/ Second post-test to CG *M* (*SD*)	Second follow-up to EG/ First follow-up to CG *M* (*SD*)
EALI-36	EG	26	159.92 (20.79)	169.92 (17.59)	168.73 (18.36)	167.88 (21.69)
	CG	20	156.20 (28.37)	157.55 (24.02)	169.60 (19.61)	172.15 (15.95)
EALI-18	EG	26	77.15 (10.86)	83.77 (9.21)	82.23 (9.81)	82.27 (10.24)
	CG	20	74.35 (15.79)	75.35 (12.63)	82.90 (10.84)	84.10 (8.27)
EFI-30	EG	27	119.56 (38.10)	143.67 (33.57)	141.30 (35.70)	141.22 (34.47)
	CG	22	107.41 (40.40)	109.45 (36.88)	126.91 (31.84)	130.41 (34.02)

AL-CDL	EG	29	11.14 (3.88)	11.45 (3.62)	11.55 (3.42)	11.72 (3.65)
	CG	22	8.91 (4.13)	9.14 (4.28)	10.36 (3.74)	10.55 (3.70)
AL-DS	EG	29	38.45 (6.21)	40.21 (5.46)	40.24 (5.30)	40.28 (6.82)
	CG	21	37.71 (4.85)	38.38 (5.30)	40.29 (5.52)	40.33 (5.57)
AL-WS	EG	29	88.93 (12.54)	93.41 (11.96)	96.10 (12.37)	94.86 (14.07)
	CG	22	87.68 (17.51)	87.23 (12.26)	93.55 (14.34)	91.09 (12.47)
AL-KS	EG	10	8.40 (0.97)	9.40 (1.07)	9.50 (0.85)	9.10 (0.99)
	CG	8	8.88 (1.36)	8.88 (0.99)	9.88 (0.83)	9.63 (1.51)
Anger	EG	29	6.76 (4.22)	4.38 (3.91)	4.21 (2.68)	4.86 (5.27)
	CG	22	7.23 (4.67)	5.68 (4.44)	4.82 (3.85)	4.91 (4.32)
Depression	EG	28	3.68 (3.04)	2.29 (2.80)	2.36 (2.31)	3.32 (3.56)
	CG	22	3.64 (2.98)	3.36 (2.44)	3.09 (2.65)	2.82 (2.30)
Hope	EG	28	40.79 (4.52)	42.79 (4.84)	42.18 (4.70)	41.57 (4.21)
	CG	22	39.68 (4.12)	39.27 (4.09)	41.36 (4.05)	41.91 (4.19)
Self-esteem	EG	29	32.17 (4.52)	33.90 (3.70)	32.66 (5.77)	33.62 (4.72)
	CG	22	31.32 (4.24)	30.82 (4.64)	32.09 (3.94)	31.05 (5.60)
Empathy	EG	28	49.93 (5.03)	50.68 (5.54)	51.75 (5.08)	48.96 (6.11)
	CG	22	48.27 (4.03)	49.00 (6.04)	49.64 (5.28)	49.36 (5.69)
Class Cooperation	EG	27	87.15 (16.46)	91.37 (17.11)	97.70 (15.46)	96.89 (12.64)
	CG	20	92.95 (13.84)	92.10 (15.91)	97.85 (17.52)	100.00 (22.62)

Cutoff scores for the eight-item Patient Health Questionnaire depression scale (PHQ-8): 0 - 4 = no significant depressive symptoms; 5 - 9 = mild depressive symptoms; 10 - 14 = moderate; 15 - 19 = moderately severe; 20 - 24, severe. (Kroenke et al., 2001).

To test the effectiveness of Forgiveness Education intervention compared to the no treatment wait list control (teaching regular school curriculum), we first conducted a one-tailed independent t-test to compare the gain scores between the two groups from pretest to post-test using data from the 46 teachers from three sites (EG = 26; CG = 20) who completed and passed the screening criteria of all four surveys. We used a one-tailed t-test here based on our original hypotheses: we assumed that those who taught the Forgiveness Education curriculum would have better effects than the no treatment wait list control. The results showed that the experimental group had significantly greater increase in *agape* love (EALI-18: $p = 0.027$ and AL-WS: $p = 0.033$), forgiveness (EFI-30: $p = 0.004$), hope ($p = 0.010$), and self-esteem ($p = 0.005$), compared to the control group (see Table 22).

Next, a one-tailed paired sample t-test was run within the control group compared to itself once this group taught the Forgiveness Education curriculum. We compared the gain scores (post-test scores minus pretest scores) when the control group had no treatment (NT, teaching regular school curriculum) with the gain scores (second post-test minus post-test scores) for the control group-turned-experimental group) once the control group became the experimental group with Forgiveness Education (FE) intervention. This particular analysis was done with the data from the 20 teachers in the control group who completed and passed the screening criteria of all four surveys. The results showed that the control group had significant growth in *agape* love (EALI-36: $p = 0.035$, EALI-18: $p = 0.015$) and hope ($p = 0.041$) after receiving Forgiveness Education compared to the teaching regular school curriculum with no treatment (see Table 23). The variables that had very close to the level of significant growth ($p = 0.050$) are forgiveness ($p = 0.052$), *agape* love (AL-WS, $p = 0.052$), and class cooperation ($p = 0.051$).

Table 22

Comparison of Gain Scores from Pretest to Posttest between the Two Groups of All Teachers

Measures	Experimental Group				Control Group				t(*df*)	*p* (one-tailed)	95% CI	Cohen's *d*	
	N	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain Scores *M* (*SD*)	*N*	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain Scores *M* (*SD*)					
EALI-36	26	159.92 (20.79)	169.92 (17.59)	10.00 (18.45)	20	156.20 (28.37)	157.55 (24.02)	1.35 (19.46)	-1.53 (39.87)	0.067	[-Inf, -0.88]	0.46	Small
EALI-18	26	77.15 (10.86)	83.77 (9.21)	6.62 (9.40)	20	74.35 (15.79)	75.35 (12.63)	1.00 (9.61)	-1.98 (40.57)	0.027	[-Inf, -0.85]	0.59	Medium
EFI-30	27	119.56 (38.10)	143.67 (33.57)	24.11 (30.99)	22	107.41 (40.40)	109.45 (36.88)	2.05 (24.60)	-2.78 (46.98)	0.004	[-Inf, -8.74]	0.78	Medium
AL-CDL	29	11.14 (3.88)	11.45 (3.62)	0.31 (4.26)	22	8.91 (4.13)	9.14 (4.28)	0.23 (3.48)	-0.08 (48.70)	0.470	[-Inf, 1.73]	0.02	Negligible
AL-DS	29	38.45 (6.21)	40.21 (5.46)	1.76 (6.20)	21	37.71 (4.85)	38.38 (5.30)	0.67 (5.72)	-0.64 (45.14)	0.262	[-Inf, 1.76]	0.18	Negligible
AL-WS	29	88.93 (12.54)	93.41 (11.96)	4.48 (8.14)	22	87.68 (17.51)	87.23 (12.26)	-0.45 (10.02)	-1.89 (39.80)	0.033	[-Inf, -0.53]	0.55	Medium
AL-KS	10	8.40 (0.97)	9.40 (1.07)	1.00 (1.33)	8	8.88 (1.36)	8.88 (0.99)	0.00 (1.20)	-1.68 (15.74)	0.057	[-Inf, 0.04]	0.78	Medium
Anger	29	6.76 (4.22)	4.38 (3.91)	-2.38 (4.17)	22	7.23 (4.67)	5.68 (4.44)	-1.54 (3.73)	0.75 (47.62)	0.228	[-1.03, Inf]	0.21	Small
Depression	28	3.68 (3.04)	2.29 (2.80)	-1.39 (3.35)	22	3.64 (2.98)	3.36 (2.44)	-0.27 (1.98)	1.47 (44.94)	0.074	[-0.15, Inf]	0.40	Small
Hope	28	40.79 (4.52)	42.79 (4.84)	2.00 (3.53)	22	39.68 (4.12)	39.27 (4.09)	-0.41 (3.46)	-2.42 (45.64)	0.010	[-Inf, -0.74]	0.69	Medium
Self-esteem	29	32.17 (4.52)	33.90 (3.70)	1.72 (3.37)	22	31.32 (4.24)	30.82 (4.64)	-0.50 (2.52)	-2.70 (49.00)	0.005	[-Inf, -0.84]	0.73	Medium

Empathy	28	49.93 (5.03)	50.68 (5.54)	0.75 (3.88)	22	48.27 (4.03)	49.00 (6.04)	0.72 (4.89)	-0.02 (39.43)	0.493	[-Inf, 2.13]	0.01	Negligible
Class Cooperation	27	87.15 (16.46)	91.37 (17.11)	4.22 (15.24)	20	92.95 (13.84)	92.10 (15.91)	-0.85 (11.75)	-1.29 (44.90)	0.102	[-Inf, 1.54]	0.37	Small

Table 23

Comparison of Gain Scores between NT and FE for the Control Group of All Teachers

Measure	*N*	Pretest *M* (SD)	Posttest *M* (SD)	Second Posttest *M* (SD)	NT Gain *M* (SD)	FE Gain *M* (SD)	t(*df*)	*p* (one-tailed)	95% CI	Cohen's *d*	
EALI-36	20	156.20 (28.37)	157.55 (24.02)	169.60 (19.61)	1.35 (19.46)	12.05 (12.98)	-1.91 (19)	0.035	[-Inf, -1.02]	0.65	Medium
EALI-18	20	74.35 (15.79)	75.35 (12.63)	82.90 (10.84)	1.00 (9.61)	7.55 (6.15)	-2.34 (19)	0.015	[-Inf, -1.70]	0.81	Large
EFI-30	22	107.41 (40.40)	109.45 (36.88)	126.91 (31.84)	2.05 (24.60)	17.45 (24.93)	-1.70 (21)	0.052	[-Inf, 0.17]	0.62	Medium
AL-CDL	22	8.91 (4.13)	9.14 (4.28)	10.36 (3.74)	0.23 (3.48)	1.23 (2.89)	-0.82 (21)	0.211	[-Inf, 1.10]	0.31	Small
AL-DS	21	37.71 (4.85)	38.38 (5.30)	40.29 (5.52)	0.67 (5.72)	1.90 (6.99)	-0.50 (20)	0.312	[-Inf, 3.06]	0.19	Negligible
AL-WS	22	87.68 (17.51)	87.23 (12.26)	93.55 (14.34)	-0.45 (10.02)	6.23 (11.66)	-1.69 (21)	0.052	[-Inf, 0.10]	0.62	Medium
AL-KS	8	8.88 (1.36)	8.88 (0.99)	9.88 (0.83)	0.00 (1.20)	1.00 (1.07)	-1.53 (7)	0.085	[-Inf, 0.24]	0.88	Large
Anger	22	7.23 (4.67)	5.68 (4.44)	4.82 (3.85)	-1.54 (3.73)	-0.86 (4.13)	-0.48 (21)	0.681	[-3.14, Inf]	-0.17	Negligible
Depression	22	3.64 (2.98)	3.36 (2.44)	3.09 (2.65)	-0.27 (1.98)	-0.27 (2.53)	0.00 (21)	0.500	[-1.38, Inf]	0.00	Negligible
Hope	22	39.68 (4.12)	39.27 (4.09)	41.36 (4.05)	-0.41 (3.46)	2.09 (3.62)	-1.83 (21)	0.041	[-Inf, -0.15]	0.71	Medium
Self-esteem	22	31.32 (4.24)	30.82 (4.64)	32.09 (3.94)	-0.50 (2.52)	1.27 (4.26)	-1.44 (21)	0.082	[-Inf, 0.34]	0.51	Medium
Empathy	22	48.27 (4.03)	49.00 (6.04)	49.64 (5.28)	0.72 (4.89)	0.64 (4.88)	0.05 (21)	0.519	[-Inf, 3.35]	-0.02	Negligible

Class Cooperation	20	92.95 (13.84)	92.10 (15.91)	97.85 (17.52)	-0.85 (11.75)	5.75 (12.02)	-1.72 (19)	0.051	[-Inf, 0.02]	0.56	Medium

We also performed one-tailed matched pair t-tests within each of the two groups to examine the amount of change on the 13 measures after each period of treatment, to further examine the within-group intervention effectiveness as well as the maintenance or growth of the intervention effect after the treatment ended. We examined the maintenance or growth effect after the interventions with the variables that showed significant differences in the within-group analyses.

For the first period of treatment, when comparing the pre-test scores (T1) and post-test scores (T2), the experimental group showed significant decrease in depression and anger, as well as significant increases in *agape* love (EALI-36: $p = 0.005$, EALI-18: $p = 0.001$, AL-WS: $p = 0.003$), *agape* love knowledge (AL-KS: $p = 0.021$), forgiveness (EFI-30: $p = 0.000$), hope ($p = 0.003$), and self-esteem ($p = 0.005$, See Table 24). The control group showed no significant changes on any of the 13 dependent variables during the no treatment period (See Table 25). For the second period of treatment, when comparing the post-test scores (T2) and the second-post test scores (T3), the control-group-turned-experimental group (after the 12-week Forgiveness Education) had significant increases in *agape* love (EALI-36: $p = 0.000$, EALI-18: $p = 0.000$, AL-CDL: $p = 0.030$, AL-WS: $p = 0.009$), *agape* love knowledge (AL-KS: $p = 0.017$), forgiveness ($p = 0.002$), hope ($p = 0.007$), and class cooperation ($p = 0.023$).

Within the experimental group (see Table 24), when comparing the pre-test scores (T1) and the first follow-up scores (T3), it showed significant increases in *agape* love (EALI-36: $p = 0.010$, EALI-18: $p = 0.015$, AL-WS: $p = 0.000$), forgiveness (EFI-30, $p = 0.000$), *agape* love knowledge (AL-KS: $p = 0.006$), hope ($p = 0.025$), empathy ($p = 0.008$), class cooperation ($p = 0.006$), as well as significant decreases in anger ($p = 0.001$) and depression ($p = 0.029$). When comparing the pre-test scores (T1) and the second follow-up scores (T4), it showed significant increases in *agape* love (EALI-18: $p = 0.013$, AL-WS: $p = 0.000$), forgiveness (EFI-30: $p =$

0.000), *agape* love knowledge (AL-KS: $p = 0.012$), self-esteem ($p = 0.016$), and class cooperation ($p = 0.002$).

Within the control-group-turned experimental group (see Table 25), when comparing the post-test scores (T2) and the follow-up scores (T4), it had significant increase in *agape* love (EALI-36: $p = 0.000$, EALI-18: $p = 0.000$, AL-CDL: $p = 0.006$; AL-WS: $p = 0.035$), forgiveness (EFI-30: $p = 0.001$), hope ($p = 0.003$), and class cooperation ($p = 0.006$).

For *agape* love (EALI-36, EALI-18, AL-WS), forgiveness, *agape* love knowledge (AL-KS) anger, hope, self-esteem, class cooperation, no differences were found for the original experimental group from the post-test (T2) to the first follow-up (T3), the post-test (T2) to the second follow-up (T4), or the first follow-up (T3) to the second follow-up (T4) (see Table 24). These indicate the effects of Forgiveness Education intervention for the experimental group were maintained after the 16-week no treatment period. The results demonstrated the long-term effectiveness of the Forgiveness Education intervention. A significant decrease in empathy from first the follow-up (T3) to the second follow-up (T4) in the experimental group was observed. Upon a further check, the score at the second follow-up (T4) is not significantly lower compared to the score at the baseline (T1), so there is no washout of the intervention effect. A significant increase in depression from first follow-up (T3) to second follow-up (T4) in the experimental group was observed, however, the score at the second follow-up (T4) was still lower compared to the baseline (T1).

No differences were found for *agape* love (EALI-36, EALI-18, AL-CDL), forgiveness, *agape* love knowledge (AL-KS), hope, and class cooperation from second post-test (T3) to the follow-up (T4), indicating the effects of the Forgiveness Education intervention were maintained after

the 4-week no treatment period for the control-group-turned-experimental group. For *agape* love (AL-WS), there was a significant decrease between the second post-test to the follow-up. However, the score at the second follow-up (T4) didn't go down to the baseline (T1).

Table 24

Comparison of Scores within the Experimental Group of All Teachers

Measure						Intervention Effect						Maintenance Effect					
	N	Pretest *M* (*SD*)	Posttest *M* (*SD*)	1st Follow-up *M* (*SD*)	2nd Follow-up *M* (*SD*)	Pre- to post-Gain *M* (*SD*)	*p* (one-tailed)	Pre- to 1st follow-up Gain *M* (*SD*)	*p* (one-tailed)	Pre- to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)	Post to 1st follow-up Gain *M* (*SD*)	*p* (one-tailed)	Post to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)	1st follow-up to 2nd follow-up Gain *M* (*SD*)	*p* (one-tailed)
EALI-36	26	159.92 (20.79)	169.92 (17.59)	168.73 (18.36)	167.88 (21.69)	10.00 (18.45)	0.005	8.81 (18.21)	0.010	7.96 (24.20)	0.053	-1.19 (10.97)	0.292	-2.04 (13.83)	0.230	-0.85 (14.10)	0.381
EALI-18	26	77.15 (10.86)	83.77 (9.21)	82.23 (9.81)	82.27 (10.24)	6.62 (9.40)	0.001	5.08 (11.30)	0.015	5.12 (11.14)	0.013	-1.54 (6.60)	0.123	-1.50 (5.28)	0.080	0.04 (6.80)	0.511
EFI-30	27	119.56 (38.10)	143.67 (33.57)	141.30 (35.70)	141.22 (34.47)	24.11 (30.99)	0.000	21.74 (26.02)	0.000	21.67 (24.27)	0.000	-2.37 (16.79)	0.235	-2.44 (15.72)	0.213	-0.07 (12.84)	0.488
AL-CDL	29	11.14 (3.88)	11.45 (3.62)	11.55 (3.42)	11.72 (3.65)	0.31 (4.26)	0.349	0.41 (4.57)	0.315	0.59 (4.07)	0.222						
AL-DS	29	38.45 (6.21)	40.21 (5.46)	40.24 (5.30)	40.28 (6.82)	1.76 (6.20)	0.069	1.79 (5.77)	0.052	1.83 (6.43)	0.068						
AL-WS	29	88.93 (12.54)	93.41 (11.96)	96.10 (12.37)	94.86 (14.07)	4.48 (8.14)	0.003	7.17 (9.20)	0.000	5.93 (7.91)	0.000	2.69 (9.02)	0.940	1.45 (7.51)	0.846	-1.24 (7.40)	0.187
AL-KS	10	8.40 (0.97)	9.40 (1.07)	9.50 (0.85)	9.10 (0.99)	1.00 (1.33)	0.021	1.10 (1.10)	0.006	0.70 (0.82)	0.012	0.10 (0.74)	0.661	-0.30 (0.95)	0.172	-0.40 (0.70)	0.052
Anger	29	6.76 (4.22)	4.38 (3.91)	4.21 (2.68)	4.86 (5.27)	-2.38 (4.17)	0.002	-2.55 (4.10)	0.001	-1.90 (6.08)	0.052	-0.17 (3.83)	0.595	0.48 (4.99)	0.303	0.66 (5.46)	0.262
Depression	28	3.68 (3.04)	2.29 (2.80)	2.36 (2.31)	3.32 (3.56)	-1.39 (3.35)	0.018	-1.32 (3.52)	0.029	-0.36 (3.93)	0.317	0.07 (3.18)	0.453	1.04 (2.57)	0.021	0.96 (3.58)	0.083
Hope	28	40.79 (4.52)	42.79 (4.84)	42.18 (4.70)	41.57 (4.21)	2.00 (3.53)	0.003	1.39 (3.60)	0.025	0.79 (3.27)	0.107	-0.61 (3.85)	0.206	-1.21 (4.11)	0.065	-0.61 (2.30)	0.086
Self-esteem	29	32.17 (4.52)	33.90 (3.70)	32.66 (5.77)	33.62 (4.72)	1.72 (3.37)	0.005	0.48 (5.91)	0.332	1.45 (3.47)	0.016	-1.24 (5.93)	0.135	-0.28 (3.01)	0.313	0.97 (5.84)	0.810

Empathy	28	49.93 (5.03)	50.68 (5.54)	51.75 (5.08)	48.96 (6.11)	0.75 (3.88)	0.158	1.82 (3.71)	0.008	-0.96 (5.36)	0.825	1.07 (3.33)	0.950	-1.71 (3.61)	0.009	-2.79 (4.09)	0.001
Class Cooperation	27	87.15 (16.46)	91.37 (17.11)	97.70 (15.46)	96.89 (12.64)	4.22 (15.24)	0.081	10.56 (20.03)	0.006	9.74 (15.51)	0.002	6.33 (18.50)	0.957	5.52 (17.68)	0.942	-0.81 (15.77)	0.395

Table 25

Comparison of Scores within the Control-group-turned-experimental Group of All Teachers

									Intervention Effect			Maintenance Effect	
Measure	*N*	Pretest *M* (SD)	Posttest *M* (SD)	Second Posttest *M* (SD)	Follow-up *M* (SD)	NT Gain *M* (SD)	*p* (two-tailed)	Post- to 2nd post-) *M* (SD)	*p* (one-tailed)	Post- to follow-up) *M (SD)*	*p* (one-tailed)	2nd post- to follow-up) *M (SD)*	*p* (one-tailed)
EALI-36	20	156.20 (28.37)	157.55 (24.02)	169.60 (19.61)	172.15 (15.95)	1.35 (19.46)	0.760	12.05 (12.98)	0.000	14.60 (16.05)	0.000	2.55 (11.71)	0.829
EALI-18	20	74.35 (15.79)	75.35 (12.63)	82.90 (10.84)	84.10 (8.27)	1.00 (9.61)	0.647	7.55 (6.15)	0.000	8.75 (8.55)	0.000	1.20 (7.37)	0.762
EFI-30	22	107.41 (40.40)	109.45 (36.88)	126.91 (31.84)	130.41 (34.02)	2.05 (24.60)	0.701	17.45 (24.93)	0.002	20.95 (23.89)	0.001	3.50 (20.90)	0.780
AL-CDL	22	8.91 (4.13)	9.14 (4.28)	10.36 (3.74)	10.55 (3.70)	0.23 (3.48)	0.762	1.23 (2.89)	0.030	1.41 (2.40)	0.006	0.18 (3.67)	0.591
AL-DS	21	37.71 (4.85)	38.38 (5.30)	40.29 (5.52)	40.33 (5.57)	0.67 (5.72)	0.599	1.90 (6.99)	0.113	1.95 (5.39)	0.056		
AL-WS	22	87.68 (17.51)	87.23 (12.26)	93.55 (14.34)	91.09 (12.47)	-0.45 (10.02)	0.834	6.23 (11.66)	0.009	3.86 (9.50)	0.035	-2.45 (6.39)	0.043
AL-KS	8	8.88 (1.36)	8.88 (0.99)	9.88 (0.83)	9.63 (1.51)	0.00 (1.20)	1.00	1.00 (1.07)	0.017	0.75 (2.12)	0.143	-0.25 (1.28)	0.299
Anger	22	7.23 (4.67)	5.68 (4.44)	4.82 (3.85)	4.91 (4.32)	-1.54 (3.73)	0.065	-0.86 (4.13)	0.169	-0.77 (4.43)	0.211		
Depression	22	3.64 (2.98)	3.36 (2.44)	3.09 (2.65)	2.82 (2.30)	-0.27 (1.98)	0.525	-0.27 (2.53)	0.309	-0.55 (2.70)	0.177		
Hope	22	39.68 (4.12)	39.27 (4.09)	41.36 (4.05)	41.91 (4.19)	-0.41 (3.46)	0.585	2.09 (3.62)	0.007	2.63 (4.11)	0.003	0.55 (2.56)	0.836
Self-esteem	22	31.32 (4.24)	30.82 (4.64)	32.09 (3.94)	31.05 (5.60)	-0.50 (2.52)	0.363	1.27 (4.26)	0.088	0.23 (6.43)	0.435		

Empathy	22	48.27 (4.03)	49.00 (6.04)	49.64 (5.28)	49.36 (5.69)	0.72 (4.89)	0.493	0.64 (4.88)	0.274	0.36 (5.07)	0.370		
Class Cooperation	20	92.95 (13.84)	92.10 (15.91)	97.85 (17.52)	100.00 (22.62)	-0.85 (11.75)	0.750	5.75 (12.02)	0.023	7.90 (12.76)	0.006	2.15 (10.04)	0.825

Chapter 5: Discussion

General Discussion of the Results

Hypotheses Focused on the Students

Six out of the eight hypotheses focused on the students were tested in each of the research site: Northern Ireland, Israel, and Taiwan. Hypothesis VII (the effectiveness of Forgiveness Education on class cooperation) was tested by teachers' report on the most disruptive students in classes from teachers of three sites combined. Hypothesis VIII (the effectiveness of Forgiveness Education on academic grades) was not tested in this study because the scores obtained from the schools were not directly comparable among schools. In other words, in each research site, the schools that participated in our study did not use the same assessment for math and literacy in the academic year of 2021-22.

With the students in Northern Ireland, of the six hypotheses proposed for effectiveness of Forgiveness Education intervention, only Hypothesis III (anger) and Hypothesis IV (depression) were statistically supported for the original experimental group. Because none of the eight hypotheses was statistically supported for the control-group-turned-experimental-group relative to the control group, we didn't carry out a further analysis of comparing between the gain achieved from the pre-test (T1) to the post-test (T2) in the first Forgiveness Education intervention for the original experimental group and the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention for the students in Northern Ireland.

With the students in Israel, of the six hypotheses proposed for effectiveness of Forgiveness Education intervention, Hypothesis I (*agape* love measured by EALI-C) and Hypothesis V (hope)

were statistically supported for the original experimental group; Hypothesis V (hope) was statistically supported for the control-group-turned-experimental-group relative to the control group. Because Hypothesis V was statistically supported for both the original experimental group and the control-group-turned-experimental-group relative to the control group, we carried out a further analysis of comparing between the gain of hope achieved from the pre-test (T1) to the post-test (T2) in the first Forgiveness Education intervention for the original experimental group and the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention for the students in Israel using a two-tailed independent t-test. The result showed that for hope, the gain achieved from the pre-test (T1) to the post-test (T2) in the first Forgiveness Education intervention for the original experimental group is statistically similar to the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention for the students in Israel.

With the students in Taiwan, of the six hypotheses proposed for effectiveness of Forgiveness Education intervention, Hypothesis I (*agape* love measured by AL-CDL-C, AL-DS-C; and *agape* love knowledge measured by AL-KS-C) was statistically supported for the original experimental group; Hypothesis I (*agape* love measured by AL-DS-C and *agape* love knowledge measured by AL-KS-C) was statistically supported for the control-group-turned-experimental-group relative to the control group. Because Hypothesis I (*agape* love measured by AL-DS-C and *agape* love knowledge measured by AL-KS-C) was statistically supported for both the original experimental group and the control-group-turned-experimental-group relative to the control group, we carried out two further analyses: a. compare the gain of *agape* love (AL-DS-C) achieved from the pre-test (T1) to the post-test (T2) in the first Forgiveness Education intervention for the original

experimental group and the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention for the students in Israel using a two-tailed independent t-test; b. compare the gain of *agape* love knowledge (AL-KS-C) achieved from the pre-test (T1) to the post-test (T2) in the first Forgiveness Education intervention for the original experimental group and the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention for the students in Israel using a two-tailed independent t-test. For the first comparison, the result showed that for *agape* love (AL-DS-C), the gain achieved from the pre-test (T1) to the post-test (T2) in the first Forgiveness Education intervention for the original experimental group is statistically similar to the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention. For the second comparison, the result showed that for *agape* love knowledge (AL-KS-C), the gain achieved from the pre-test (T1) to the post-test (T2) in the first Forgiveness Education intervention for the original experimental group is statistically similar to the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention for the students in Taiwan.

Hypotheses Focused on the Teachers

With the teachers from three sites combined, Hypothesis I (*agape* love measured by EALI-18 and AL-WS), Hypothesis II (forgiveness), Hypothesis V (hope), and Hypothesis VI (self-esteem) were statistically supported for the original experimental group; Hypothesis I (*agape* love measured EALI-36 and EALI-18) and Hypothesis V (hope) were statistically supported for the control-group-turned-experimental-group relative to the control group. Hypothesis I (*agape* love

measured by EALI-18) and Hypothesis V (hope) were statistically supported for both the original experimental group and the control-group-turned-experimental-group relative to the control group. In addition, Hypothesis I (*agape* love measured by AL-WS) and Hypothesis II (forgiveness) were almost supported (very close to significant level, $p = 0.052$ for AL-WS and forgiveness) for both groups. As a result, we carried out four further analyses to compare the gain score achieved from the pre-test (T1) to post-test (T2) in the first Forgiveness Education intervention for the original experimental group and the gain score achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention for the teachers using a two-tailed independent t-test with the following variables: EALI-18, AL-WS, forgiveness, hope. The result showed that gain achieved from the pre-test (T1) to the post-test (T2) in the first Forgiveness Education intervention for the original experimental group is statistically similar to the gain achieved from the first post-test (T2) to the second post-test (T3) of the control-group-turned-experimental group in the second Forgiveness Education intervention for all four variables (EALI-18, AL-WS, forgiveness, hope).

Hypotheses Related to the Maintenance of Effects

Hypothesis I (maintenance of effectiveness within the original experimental group after the treatment) was tested with variables that showed significance in within-group changes from the pre-test (T1) to the post-test (T2) (12-week Forgiveness Education intervention with the original experimental group) in students from each site and in teachers. Hypothesis II (maintenance of effectiveness within the control-group-turned-experimental group after the treatment) was tested with variables that showed significance in within-group changes from the post-test (T2) to second post-test (T3) (12-week Forgiveness Education intervention with the control-group-

turned-experimental group) in students from each site and in teachers. Hypothesis III (comparing the maintenance of effectiveness between the original experimental group at the second follow-up test (T4) and the control-group-turned-experimental group at the follow-up test (T4)) was tested with variables that showed maintenance or growth after treatment in both groups in students from each site and in teachers. Hypothesis IV (comparing the maintenance of effectiveness between the original experimental group at the first follow-up test (T3) and the control-group-turned-experimental group at the follow-up test (T4) was tested with variables that showed maintenance or growth after treatment in both groups in students from each site and in teachers.

For the students in Northern Ireland, Hypothesis I was met for all the variables that showed significant differences within the original experimental group from the pre-test (T1) to the post-test (T2), including forgiveness, AL-WS-C, anger expression, anger control, depression. Among them, the scores of forgiveness, *agape* love (AL-WS-C) kept increasing and the score of depression kept decreasing across the four time points (T1 to T4), which indicates not only maintenance but also growth of the intervention effectiveness 16-week after the treatment ended. Given the severity of COVID-19, especially in the second semester (Spring of 2022), the continued increase in forgiveness, *agape* love (AL-WS) and the continued decrease in depression after the treatment ended are worth noting. Hypothesis II was met for both variables that showed significant differences within the control-group-turned-experimental group from the first post-test (T2) to the second post-test (T3): AL-DS-C & AL-WS-C. Only AL-WS-C had maintenance of effectiveness after treatment ended for both the original experimental group and the control-group-turned-experimental group. To test Hypothesis III and IV, we carried out two analyses: a. comparing the score of the original experimental group at the 16-week second follow-up test (T4)

and the control-group-turned-experimental group at the 4-week follow-up test (T4); b. comparing the score of the original experimental group at the first 12-week follow-up test (T3) and the control-group-turned-experimental group at the 4-week follow-up test (T4) using two-tailed independent t-test. Hypotheses III and IV were both met as the scores of AL-WS-C at the first follow-up test (T3) and at the second follow-up test (T4) of the original experimental group are statistically similar to the score of the control-group-turned-experimental group at their follow-up test (T4).

For the students in Israel, Hypothesis I was met for the following variables that showed significant differences within the original experimental group from the pre-test (T1) to the post-test (T2): EALI-C, anger control, hope, and empathy. Among them, the scores of *agape* love (EALI-C) kept increasing across the four time points (T1 to T4), which indicates not only maintenance but even growth of the intervention effectiveness 16-week after the treatment ended. Hypothesis II was met for hope, which is the only variable that showed significant differences within the control-group-turned-experimental group from the first post-test (T2) to the second post-test (T3). Only hope had maintenance of effectiveness after treatment ended for both the original experimental group and the control-group-turned-experimental group. To test Hypothesis III and IV, we carried out two analyses: a. comparing the score of the original experimental group at the 16-week second follow-up test (T4) and the control-group-turned-experimental group at the 4-week follow-up test (T4); b. comparing the score of the original experimental group at the 12-week first follow-up test (T3) and the control-group-turned-experimental group at the 4-week follow-up test (T4) using two-tailed independent t-test. Hypotheses III and IV were both met as the scores of hope at the first follow-up test (T3) and at

the second follow-up test (T4) of the original experimental group are statistically similar to the score of the control-group-turned-experimental group at their follow-up test (T4).

For the students in Taiwan, Hypothesis I was met for the following variables that showed significant differences within the original experimental group from the pre-test (T1) to the post-test (T2): EALI-C, EALI-6, forgiveness, AL-DS-C, AL-WS-C, AL-KS-C, depression, and empathy. Hypothesis II was met for the following variables that showed significant differences within the control-group-turned-experimental group from the first post-test (T2) to the second post-test (T3): AL-DS-C, AL-KS-C, and depression. AL-KS-C, depression, had maintenance of effectiveness after treatment ended for both the original experimental group and the control-group-turned-experimental group. To test hypothesis III and IV, we carried out two analyses for each variable: a. comparing the score of the original experimental group at the 16-week second follow-up test (T4) and the control-group-turned-experimental group at the 4-week follow-up test (T4); b. comparing the score of the original experimental group at the 12-week first follow-up test (T3) and the control-group-turned-experimental group at the 4-week follow-up test (T4) using two-tailed independent t-test. For the two variables, AL-KS-C, and PHQ-8A, Hypotheses III and IV were both met as the scores achieved by the original experimental group at the first follow-up test (T3) and at the second follow-up test (T4) are statistically similar to the scores achieved by the control-group-turned-experimental group at their follow-up test (T4).

For the teachers of three sites combined, Hypothesis I was met for the following variables that showed significant differences within the original experimental group from the pre-test (T1) to the post-test (T2): EALI-36, EALI-18, forgiveness, AL-WS, AL-KS, anger, depression, hope, self-esteem, and class cooperation. Hypothesis II was met for the following variables that showed significant differences within the control-group-turned-experimental group from the first

post-test (T2) to the second post-test (T3): EALI-36, EALI-18, forgiveness, AL-CDL, AL-KS, hope, and class cooperation. For Hypothesis III and IV, EALI-36, EALI-18, forgiveness, AL-KS, hope, and class cooperation had maintenance of effectiveness after treatment ended for both the original experimental group and the control-group-turned-experimental group. We carried out two analyses for each variable: a. comparing the score of the original experimental group at the 16-week second follow-up test (T4) and the control-group-turned-experimental group at the 4-week follow-up test (T4); b. comparing the score of the original experimental group at the 12-week first follow-up test (T3) and the control-group-turned-experimental group at the 4-week follow-up test (T4) using two-tailed independent t-test. For the six variables, EALI-18, EALI-36, forgiveness, AL-KS, hope, class cooperation, Hypotheses III and IV were both met as the scores achieved by the original experimental group at the first follow-up test (T3) and at the second follow-up test (T4) are statistically similar to the scores achieved by the control-group-turned-experimental group at their follow-up test (T4).

The significant growth of *agape* love was observed in the between-group comparison for the students in Israel and Taiwan, as well as in the teachers of three sites combined. The significant growth of *agape* love was observed within the original experimental group for the students of all three sites as well as in teachers.

Among them, EALI-36 (or EALI-C), which measures the degree of expressing *agape* love towards a particular person in a particular situation, had significant growth favoring the original experimental group in the between-group comparison for the students in Israel ($p = 0.034$, see Table 10). EALI-18, three sub-scales of EALI-36 measuring the change of *agape* love towards a particular person in a particular situation, had significant growth favoring the original experimental group in the between group comparison for teachers ($p = 0.027$, see Table 22). AL-

CDL (or AL-CDL-C), measuring the understanding of *agape* love as the highest form of love, had significant growth favoring the original experimental group in the between-group comparison for the students in Taiwan ($p = 0.023$, see Table 16). AL-DS (or AL-DS-C), measuring the tendency of expressing *agape* love to different groups of other people when the needs of the others are in conflict with the needs of the self, had significant growth favoring the original experimental group in the between group comparison in the students in Taiwan ($p = 0.001$, see Table 16). AL-WS (or AL-WS-C), as measuring the willingness of engaging in *agape* love by hypothetical examples, had significant growth favoring the original experimental group in the between group comparison in teachers ($p = 0.033$, see Table 22). AL-KS (or AL-KS-C), which measures the knowledge of *agape* love and was only administered to the students and to the teachers only in Taiwan, had a significant growth favoring the original experimental group in the between group comparison for the students in Taiwan ($p = 0.001$, see Table 16) and a close to significant level of growth in teachers in Taiwan ($p = 0.057$, see Table 22).

It is interesting to observe that children (ages 10 to 12) had more significant growth in the cognitive understanding of *agape* love (measured by AL-CDL and AL-KS) while adults didn't. The comparison of the baseline scores of children and adults shows that the teachers had higher scores of AL-CDL and AL-KS at the pre-test (T1) than the students. One possible explanation is that the teachers received training and learned the concepts of *agape* love and forgiveness in the curriculum, while the students didn't begin to learn about *agape* love and forgiveness until the teachers began teaching the curriculum. After having Forgiveness Education, teachers had significant growth in changing the current feelings, view of the meaning, self-identity as the *agape* servant when they recalled the real-life incident of helping someone that involves physical or psychological pain (EALI-18, the three subscales measuring change in EALI) as well as

expressing *agape* love in the hypothetical situations (AL-WS). The students in Israel had significant growth in expressing *agape* love in a real-life incident, as they recalled helping another person caused discomfort in their mind or in their body (EALI-24). The students in Taiwan started to develop more of a tendency, at least in their minds, to take care of others' needs instead of their own, as is evidenced in the significant growth of AL-DS-C. A plausible explanation of the results is: children (ages 10 to 12) are capable of developing cognitive understanding of *agape* love after deliberate engagement in Forgiveness Education, while they might not be mature enough to keep expressing *agape* love as an identity, they started to develop the willingness or tendency of practicing *agape* love in actions as their minds had changed. This could further be supported by the within-group changes after the treatment: the original experimental group of Taiwan students grew significantly in AL-CDL-C (understanding *agape* love as the highest form of love); both original experimental group and the control-group-turned-experimental group of Taiwan students as well as the control-group-turned-experimental group of Northern Ireland students grew significantly in AL-DS-C (tending others' needs first).

The within-group comparisons showed some interesting results: while students from none of the three sites had significant growth in AL-WS-C (expressing *agape* love in hypothetical situations) in between-group analyses, the original experimental group in three sites all showed significant growth (Northern Ireland: $p = 0.005$ at the post-test, Table 6; Israel: $p = 0.002$ at the first follow-up, Table 12; Taiwan: $p = 0.000$ at the post-test, Table 18). It was because the control group also had some growth in AL-WS-C during the no treatment (teaching regular school curriculum) period, which was not found in teachers' results. It is possible that children tend to choose more desirable moral actions when they imagine hypothetical situations and this tendency grew naturally with their age. This is also observed in the students in Northern Ireland: the

original experimental group grew significantly in AL-WS-C after the Forgiveness Education intervention and continued to grow in the two follow-ups (after 12-week: $p = 0.000$; after 16-weesk: $p = 0.001$, see Table 6) while the control-group-turned-experimental group had small, non-significant growth of AL-WS-C after the no treatment period (two-tailed t-test: $p = 0.658$) then grew significantly after the treatment ($p = 0.006$, see Table 7) and the effectiveness was maintained at the 4-week follow-up.

Although the significant growth of forgiveness (measured by EFI-C) was not observed in the between-group comparisons of students in the three sites, the significant growth of forgiveness after the treatment was found in the within-group comparisons of the original experimental group in all three sites (Northern Ireland: $p = 0.045$ at post-test, Table 6; Israel: $p = 0.001$ at the first follow-up, Table 12; Taiwan: $p = 0.000$ at post-test, Table 18), as well as in the within-group comparisons of the control-group-turned-experimental group in Northern Ireland ($p = 0.037$, , see Table 7) and in Taiwan ($p = 0.056$, close to the significant level, see Table 19). Similar to AL-WS-C, the non-significance of forgiveness in between-group comparisons in all three sites was due to the control group's growth of forgiveness during the wait period (teaching a regular curriculum). A plausible explanation to the results is: upon checking the stories of EFI-C (the first five questions in EFI-C, asking the respondent to recall an incident of being hurt by another person), most of the stories reported by children of ages 10 to 12 are not a deep injustice or a serious abuse, but daily, transient moments of experiencing injustice. In comparison, the stories of EFI-30 reported by the teachers generally involved a deeper injustice – a hurtful event that was remembered for years. A study with college students found that forgiveness can grow naturally without treatment (Wang Xu, Kim, Olmstead, Enright, 2021), which might also be true with elementary students, especially with those who have not experienced any deep injustice in

life. To draw a more accurate conclusion, further analyses of the Forgiveness Education effectiveness on forgiveness should be conducted with some sub-samples, such as students who reported more severe injustice or deeper hurtful feelings or students who had lower forgiveness and higher anger at the baseline.

For teachers, the original experimental group had significant growth of forgiveness ($p = 0.004$, see Table 22) in the between-group comparison and the control-group-turned-experimental group had very close to a significant level of growth of forgiveness ($p = 0.052$, see Table 23) after teaching the Forgiveness Education. The within-group analyses showed that both groups of teachers had significant growth of forgiveness after teaching the Forgiveness Education ($p = 0.000$ for the original experimental group, Table 24; $p = 0.002$ for the control-group-turned-experimental group, Table 25), and the effectiveness maintained for the original experimental group at the 16-week follow-up and kept growing for the control-group-turned-experimental group at the 4-week follow-up. The results are congruent with the studies of forgiveness interventions with adults (see meta-analyses of forgiveness interventions in Lundahl et al., 2008; Wade et al., 2014; Akhtar & Barlow, 2018; as well as a cross-over designed forgiveness intervention study done in a correctional institution by Yu et al., 2021).

Rapp et al., (2022) point out in the meta-analysis that the Forgiveness Education interventions have a significant effect on anger reduction. The between-group comparison of the students in Northern Ireland echoes the finding: the original experimental group had a significant decrease in anger expression after having the Forgiveness Education ($p = 0.007$, see Table 4), in comparison to the control group. The within-group analysis for the original experimental group of the students in Israel showed a significant decrease in anger expression after the treatment ($p = 0.027$, see Table 12) at the post-test (T2) and the effectiveness maintained at the 12-week follow-

up (T3) and the 16-week follow-up (T4). The students in the original experimental groups in Northern Ireland ($p = 0.029$, see Table 6) and in Israel ($p = 0.002$, see Table 12) both showed significant increases in anger control at the post-test (T2). For the students in the control-group-turned-experimental group, their anger control increased after the treatment, but the increase did not have a significant difference at the second post-test (T3). However, the anger control kept growing and showed a significant difference at the follow-up (T4, 4-week after the treatment ended), in comparison to the first post-test (T2), for students in Northern Ireland ($p = 0.002$, see Table 7) and in Israel ($p = 0.015$, see Table 13). It is important to see that the Forgiveness Education intervention had an impact on reducing anger expression and increasing anger control in students because persistent anger can lead to increased violence and aggression (Hawes et al., 2016; Aseltine et al., 2000), academic difficulty (Loveland et al., 2007; Strauss et al., 1987; Wiesner & Windle, 2004), and depressive symptoms (Carey et al., 1991; Taysi et al., 2015) in children and adolescents. The average score of the students achieved at the pre-test in three sites is 26.57 to 31.60 in anger expression (score ranges from 16 to 64) and 21.78 to 24.02 in anger control (score ranges from 9 to 36). The baseline scores are not considered as high in anger in each site, which reflect the nature of the samples of a large number of regular elementary school children. A further analysis could investigate the effectiveness of Forgiveness Education on anger with a sub-sample of the ‘angriest’ students (high baseline of anger expression and low baseline of anger control) in each site.

For the teachers, there was no significant decrease of anger observed in the between-group comparison. The original experimental group of teachers showed a significant decrease in anger ($p = 0.002$) at the post-test (T2) and the effectiveness was maintained at the 12-week follow-up (T3) and the 16-week follow-up (T4) in the within-group analysis. Even though the average of

the anger score of teachers was not considered as high (M = 6.76, SD = 4.22) in the anger score ranges from 0 to 60, teaching Forgiveness Education could further reduce the degree of anger within the original experimental group (M = 4.38, SD = 3.91). The control-group-turned-experimental group also showed lower anger score after teaching Forgiveness Education, but the change was not significant (from M = 5.68, SD = 4.44 to M = 4.82, SD = 3.85).

It is interesting to find that the Forgiveness Education intervention in this study had an impact on reducing depressive symptoms in students and in teachers. In Northern Ireland, the between-group analysis showed the significant reduction in depressive symptoms favoring the original experimental group (p = 0.00, see Table 4). The within-group analysis of the original experimental group of the students in Northern Ireland showed a significant decrease in depressive symptoms after the treatment (p = 0.002, see Table 6) at the post-test (T2) and the effectiveness maintained at the 12-week follow-up (T3) and the 16-week follow-up (T4). The within-group analysis for the original experimental group of the students in Taiwan showed a significant decrease in depressive symptoms after the treatment (p = 0.040, see Table 18) at the post-test (T2) and the effectiveness was maintained at the 12-week follow-up (T3) and the 16-week follow-up (T4). The within-group analysis for the control-group-turned-experimental group of the students in Taiwan showed significant decrease of depressive symptoms (p = 0.007, see Table 19) at the second post-test (T3) and effectiveness maintained at the follow-up (T4). The within-group analysis for both groups of students in Israel showed a reduction of depressive symptoms after their respective treatments, but the changes were not considered to be statistically significant. The findings among the students in Northern Ireland were encouraging because the original experimental group had a high depression score at the baseline (T1, M = 7.73, SD = 5.90), which is considered as having mild depressive symptoms according to the

guidelines of PHQ-8 (Kroenke et al., 2001). After having Forgiveness Education intervention, the depression score of the original experimental group dropped two points at the post-test (T2, $M = 5.81$, $SD = 5.06$). Whereas the control-group-turned-experimental group had some but not statistically significant reduction in depressive symptoms after the treatment partly because the depression score was not high at the baseline (T2, $M = 5.05$, $SD = 4.58$), compared to the original experimental group. Similar encouraging findings were observed with the students in Taiwan as well: the original experimental group's depression score (T1, $M = 5.23$, $SD = 4.67$), considered as having mild depressive symptoms according to Kroenke et al., (2001), went down to having no significant depressive symptoms at the post-test (T2, $M = 4.65$, $SD = 4.64$) and kept decreasing at the 12-week follow-up (T3, $M = 4.02$, $SD = 4.24$) and at the 16-week follow-up (T4, $M = 3.86$, $SD = 4.38$). Although the control-group-turned-experimental group of students in Taiwan had a baseline depression score (T1, $M = 4.80$, $SD = 4.50$), indicating no significant depressive symptoms, it had a significant reduction ($p = 0.007$) after the treatment at the second post-test (T3, $M = 4.24$, $SD = 3.98$), in comparison with the first post-test (T2, $M = 4.94$, $SD = 4.49$) and the effectiveness was maintained at the 4-week follow-up (T4, $M = 4.44$, $SD = 4.41$). In the meta-analysis study with Forgiveness Education interventions (Rapp et al., 2022), depression did not have a significant effect as one of the outcomes within a relatively small number of studies ($N = 7$). This current study shows the effectiveness of the Forgiveness Education intervention on reducing the depressive symptoms with general school children (ages 10 to 12) who didn't have clinical depressive symptoms prior to the intervention, which could potentially contribute to meta-analytical findings of the Forgiveness Education interventions effects on depression. In addition, population-based studies have found a cut point of ≥ 10, and therefore this can be used for defining current depression (Kroenke et al., 2009; Gómez-Gómez

et al, 2022). In the future, a sub-sample of students with PHQ-8A baseline score ≥ 10 (indicating moderate to severe depressive symptoms) should be generated from each site (or three sites combined) to further investigate the Forgiveness Education interventions' effects on depression.

For teachers, both the original experimental group (T1, $M = 3.68$, $SD = 3.04$) and the control-group-turned-experimental group (T1, $M = 3.64$, $SD = 2.98$) had low baseline depression scores, indicating the teachers in the sample in general had good mental health. Interestingly, both groups had a reduction in depressive symptoms after teaching Forgiveness Education. The original experimental group showed a significant reduction ($p = 0.018$, see Table 24) in depressive symptoms right after teaching the curriculum at the post-test (T2, $M = 2.29$, $SD = 2.80$) and the effectiveness maintained at the 12-week follow-up (T3, $M = 2.36$, $SD = 2.31$). The depression score of the control-group-turned-experimental group, although not showing statistically significant reduction in depressive symptoms after the educators taught Forgiveness Education (T3, $M = 3.09$, $SD = 2.65$), kept decreasing at the 4-week follow-up (T4, $M = 2.82$, $SD = 2.30$). This result echoes the finding from the previous meta-analytical studies of forgiveness interventions with adult population (Wade et al., 2014; Akhtar & Barlow, 2018) that forgiveness interventions, although not targeting mental health symptoms directly, resulted in reductions in depressive symptoms.

Hope, as one mental health construct that has been frequently examined with forgiveness interventions in pervious empirical and meta-analytical studies, showed significant gains favoring the original experimental group in the between-group comparison ($p = 0.002$, see Table 10) as well as significance favoring the control-turned-experimental-group (gain from T2 to T3) relative to the control group (gain from T1 to T2) ($p = 0.028$, see Table 11) in students in Israel. The significant within-group gain of hope was also observed in both groups of students in Israel

with the effectiveness maintained at each follow-up as well as in the original experimental group of students in Northern Ireland at the second follow-up (T4). This finding resonates with the meta-analysis of Forgiveness Education interventions (Rapp et al., 2022), in which hope was very close to a statistically significant level ($p = 0.06$, 95% CI [-0.09, 3.49]) with five studies that measured hope as an outcome of Forgiveness Education interventions. Hope is also found to be a statistically significant outcome of teaching Forgiveness Education for teachers in the between-group analysis ($p = 0.010$, see Table 22), comparison of control group and control-group-turned experimental group ($p = 0.041$, see Table 23), as well as the within-group analysis of both groups with effectiveness maintained at each follow-up. In the meta-analytical study, Wade et al, (2014) investigated hope with a subset of six studies, it was found that hope showed a statistically significant improvement in the forgiveness treatment condition vs. in the no-treatment condition, which is consistent with what we found in the current study.

Self-esteem, as another mental health construct that has been frequently examined in forgiveness interventions, did not show any statistically significant improvement in either the between-group analysis or the within-group analysis in students of all three sites. The result is not unexpected because Rapp et al. (2022) also found it not to be a strong outcome with a small effect size (g = 0.02, $p = 0.92$, 95% CI [-0.39, 0.43]) in the meta-analysis of four forgiveness education intervention studies measuring self-esteem. The scores of self-esteem are similar (from $M = 13.71$, $SD = 4.52$ to $M = 17.63$, $SD = 5.25$) in both experimental and control groups at baseline in the three sites. Although there is no specific cutoff point for low self-esteem on the Self-esteem Inventories Manual (Coopersmith, 2002), the baseline scores in the children's samples in the current study are comparable to the other studies using the Coopersmith Self-Esteem Inventory (School Form)-Short Form (Watkins, 1982; Francis, 1998; Hill et al., 2011)

with general children's populations. A further analysis using a sub-sample of students with low self-esteem at the baseline is worth doing.

Conversely, self-esteem showed a significant improvement in the between-group analysis of teachers ($p = 0.005$, see Table 22) as well as a significant change in the within-group analysis with the original experimental group after teaching the Forgiveness Education and the effectiveness was maintained at the 12-week and the 16-week follow-ups. The control-group-turned-experimental group of teachers also had an improvement in self-esteem after teaching the curriculum, but the change was not statistically significant. Lundahl et al. (2008)'s meta-analysis had a similar finding of self-esteem being a statistically significant outcome of investigating nine studies of forgiveness interventions.

Empathy is described as a positive emotion developed in the work phase of the process model of forgiveness (Enright & Fitzgibbons, 2015). There was no significant gain in empathy observed in the between-group analysis of students and teachers. However, the significant gains in empathy were found in the within-group analysis of the experimental groups of students in Israel and in Taiwan and the effectiveness was maintained at the 12-week and the 16-week follow-ups. Rapp et al. (2022) reported that empathy was a statistically significant outcome of Forgiveness Education interventions with a small to medium effect ($g = 0.32$, $p = 0.04$, 95% CI [0.008, 0.64]). In current study, the students in the control group in three sites all showed some improvement in empathy during the no treatment/teaching regular curriculum period, thus leading to the non-significant results in the between-group comparison. Similarly in teachers, there is no significant gain in empathy in the between-group analysis because the control group teachers grew in empathy during the no treatment/teaching regular curriculum period. The

within-group analysis showed that the experimental group had significant improvement in empathy after teaching Forgiveness Education, however, the effectiveness didn't maintain at the 16-week's second follow-up, with a washout effect.

Class cooperation is a behavior measure, reported by each class teacher who taught the Forgiveness Education, on four students who were regarded as the most disruptive in the class. Although class cooperation did not show significance in the between-group analysis, a significant improvement was observed in the within-group analysis of both the original experimental group and control-group-turned-experimental group after the respective treatment and kept improving at the 12-week follow-up of the original experiment group and the 4-week follow-up of the control-group-turned-experimental group.

The findings from this study have some cultural implications. Taiwan being a collectivist culture traditionally encourages the suppression of anger, which might explain why there was neither significant reduction of anger expression nor significant growth of anger control observed as the effectiveness of Forgiveness Education for students in Taiwan: the items in the Anger Expression Scale for Children (AESC) might not be the most appropriately worded to accurately measure the levels of anger expression or anger control with the children in Taiwan. On the contrary, the anger expression showed significant reduction and anger control showed significant improvement in students in Northern Ireland in both the between-group analysis and the within-group analysis while only in the within-group analysis in Israel. Interestingly, Northern Ireland is an individualistic culture while Israel is a blend of individualist and collectivistic cultures. Children brought up in different cultures are under certain cultural expectations for expressing and controlling anger. This study adopted the same anger measure (AESC) for children across the three research sites. One example of the item is "I feel like breaking things."

It might be more difficult for children in collectivist culture to imagine such expressive feelings of anger (breaking things). In other words, children in a collectivist culture might not choose to break things even though they feel very angry because it is not approved by their culture. Although using the same objective measures is parallel for cross-cultural comparison, it might not be the most appropriate strategy to capture the actual changes in anger expression or anger control within each culture. Another observation is that Taiwan is the only site having significant reductions in depressive symptoms in the within-group analysis for both groups. It is possible that being in a collectivistic culture, children in Taiwan tend to suppress their emotions more than children in individualistic cultures (Huwaë & Schaafsma, 2016), which is related to depressive symptoms shown in previous research (Gross & John, 2003; Flynn et al., 2010). While Forgiveness Education encourages students to express their feelings, especially those uncomfortable feelings, to themselves (through journaling or visualization exercises) or to teachers and peers, in small groups or in class, it could help students who are brought up in collectivistic culture to express their feelings, which could possibly lead to reducing depressive symptoms.

Limitations of the Study

One major limitation of the study is that the two Forgiveness Education interventions and four times of data collection were done during the COVID-19 pandemic, which caused an unprecedented challenge for data collection as well as disruption to the teaching plan. For example, data collection had to be postponed in schools that experienced suspension because of detection of COVID cases. The originally planned 12-week Forgiveness Education program had to be accomplished longer than 12 weeks because of school suspension and other disruptions caused by the pandemic. The sickness absence of students was more common and lengthier

during pandemic. In the Spring of 2022, the situations in schools were worsened by the Omicron variant: more school suspensions and absences were reported by the local research team at each site. Occasionally some lessons needed to be taught online rather than in person, as was reported by the researchers in Israel. In Israel, due to the school suspensions, some survey administrations at T3 and T4 were shifted from in-person to synchronously online as one researcher and the whole class of students met online (e.g., through a Zoom meeting) and worked on the survey at the same time. In Taiwan, the original plan of paper-and-pencil surveys had to be switched to online Qualtrics surveys for many schools during T3 and T4. Because of these disruptions, our research group had to create a coding manual of problem situations to exclude classes that had serious issues of not adhering to the original teaching plan from the analysis, based on teachers' self-report and researchers' observation (See Appendix B). However, we are still unsure of how the COVID-19 pandemic impacted the intervention and data collection in general, including the situation that the control-group-turned-experimental group had fewer statistically significant outcomes observed than the experimental group. The control-group-turned-experimental group had their Forgiveness Intervention during the most challenging COVID period, the spring semester of 2022.

We should also be cautious when making a comparison between the experimental group vs. the control group and that same control group vs. itself as the control-group-turned-experimental group. This is the case because, for the control-group-turned-experimental group, it is within the same group, comparing the gain between the no treatment period and the Forgiveness Education intervention at a different period in time. For the original experimental group vs. the control group, this comparison of gains is between two groups during the same period: one with treatment and the other without.

Despite the challenges caused by pandemic, another limitation of the study is switching teachers: three classes in Israel and one class in Northern Ireland had substitute teachers when the original teachers couldn't commit to teaching the curriculum due to various reasons. The cases are all in the control-group-turned-experimental group: three substitute teachers took over at the beginning of the curriculum and one substitute teacher stepped in after the original teacher taught a few lessons and had sick leave. Switching teachers may not only cause problems to the research study (e.g., the substitute teachers' responses were not included in the analysis), but also have negative impact on the intervention (e.g., students might feel lost because of the disconnection with the original teacher).

One limitation is technical difficulties experienced during data collection. In Northern Ireland and Israel, all survey data were collected using online electronic questionnaires through Qualtrics. In most cases, the whole class of students used the school computers in a computer room at a designated time to answer the online questionnaires. However, a few classes in Israel had to use their own smartphones to answer the questionnaires because of the unavailability of the school computers, either due to the malfunction of internet connection or, on one occasion, the school computer room was broken into by a thief, who stole all the computers before the data collection. In addition, because of the unfamiliarity with the Qualtrics online questionnaire system, the researchers in Northern Ireland accidentally administered the wrong survey link to two classes at the post-test (T2). The two classes were excluded from further analysis. In Israel, the researchers accidentally used a survey link that could not store unsubmitted responses permanently at the pre-test (T1) to Hebrew-speaking students. Because of this, quite a few students' responses were not included in the analysis because we were unable to retrieve the pre-test responses (T1) and link them to their later survey responses (T2, T3, T4).

Another limitation is the problem with using the objective survey measures across three sites. Although the scales used in this study are all validated with demonstrated reliability and validity, not all of them were validated cross culturally. As was discussed previously, the respondents in collectivist culture may have cultural barrier in understanding the meaning of the scale items that essentially were developed by researchers in individualistic culture, which might not be culturally relevant to respondents brought up in collectivistic cultures. Thus, the responses might not accurately reflect the actual psychological state of the respondents. In addition, the cultural barrier could be further impeded by translation. For example, some words translated from English might not be as commonly understood in their own language.

Further, researchers cannot control confounding variables in schools. For example, we do not know what kind of curriculum or program was taught during the control group's no treatment/teaching regular school curriculum period. Other types of moral virtue-based curriculum or social emotional learning program could be taught during the time with no treatment of Forgiveness Education. No other site but Taiwan's student control group experienced significant within-group growth in forgiveness (EFI-C), *agape* love (AL-WS-C), and hope during the no treatment period. It is possible that students had another moral virtue-based curriculum as it is part of the regular school curriculum in Taiwan. Another possibility is that the teachers mentioned the concepts forgiveness and *agape* love in their teaching or daily conversations with the students prior to the post-test (T2) as the teachers started to learn about the Forgiveness Education curriculum before the students did.

Directions for Future Research

Replication studies are needed to demonstrate the validity and reliability of Forgiveness Education interventions in different cultures. Replication studies should be conducted in more

places, especially countries with a collectivist culture, to evaluate the impact of Forgiveness Education interventions on the negative psychological constructs such as anger expression, anger control, anxiety, and depression as well as the positive psychological constructs such as hope, empathy, and self-esteem. Replication studies should be conducted with different ages and grades to expand our understanding of what impact Forgiveness Education interventions may bring from a developmental perspective.

For future studies investigating the effectiveness of Forgiveness Education, more behavioral measures and measures of peer relationships should be included. In the qualitative analysis of the interview responses from the students, our research group discovered that many students, who had growth in *agape* love and forgiveness, mentioned that after the Forgiveness Education program, they had better relationships with classmates and made more friends. If measures of peer relationships could be included in the surveys, we would have been able to observe if such changes are generally applied to the majority of students in the quantitative analysis.

We recommend future studies to include longer times for the follow-up tests. In our current study, more maintenance or even growth of the variables were observed at the 12-week and the 16-week follow-ups for the original experimental group than the washout of the intervention effects. There is no washout of intervention effects of any variable observed at the 4-week follow-up of the control-group-turned-experimental group. Rapp et al. (2022) also reported in the meta-analysis that the effect for forgiveness and anger maintained at the follow-ups ranged from 4-week to 1 year. Thus, a 1-year or even longer follow-up is suggested to be implemented in future studies.

We received feedback from teachers, who taught the Forgiveness Education curriculum in our current study, with regard to the cultural relevance of the stories used in the curriculum. A few teachers in Israel suggested that more stories from their own culture should be used in the curriculum so students would feel more culturally engaged. Research has found out that the important factor that may influence comprehension of stories is culture (Bock, 2006). The stories used in the Forgiveness Education curriculum were mostly written by authors from Western and individualistic cultural backgrounds. Future studies could be conducted to investigate the impact of cultural relevance of the curriculum on the efficacy of the Forgiveness Education curriculum.

Conclusion

This study demonstrates that through educating children about forgiveness, which is a conscious and deliberate expression of *agape* love, it fostered at least to some degree the development of the moral virtues of *agape* love and forgiveness in children and their teachers, in the cognitive and socio-emotional developmental domains. The improvement in psychological well-being and in class cooperation were observed, again at least to some degree, as the results of the Forgiveness Education intervention.

References

Akhtar, S. & Barlow, J. (2018). Forgiveness therapy for the promotion of mental well-being: A systematic review and meta-analysis. *Trauma, Violence, & Abuse*, *19*, 107-122.

Al-Mabuk, R., Enright, R. D., & Cardis, P. (1995). Forgiveness education with parentally love-deprived college students. *Journal of Moral Education*, *24*, 427-444.

American Psychiatric Association (1994). *American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders (4th Edition)*, American Psychiatric Association, Washington, DC.

Aseltine Jr, R. H., Gore, S., & Gordon, J. (2000). Life stress, anger and anxiety, and delinquency: An empirical test of general strain theory. *Journal of health and social behavior*, 256-275.

Baskin, T. W., & Enright, R. D. (2004). Intervention studies on forgiveness: A meta-analysis. *Journal of Counseling and Development*, *82*, 79-90.

Bock, T. (2006). A consideration of culture in moral theme comprehension: Comparing native and European American students. *Journal of Moral Education*, *35*, 71-87. https://doi.org/10.1080/03057240500495310

Carey, T. C., Finch, A. J., & Carey, M. P. (1991). Relation between differential emotions and depression in emotionally disturbed children and adolescents. *Journal of Consulting and Clinical Psychology*, *59*, 594-597. https://doi.org/10.1037/0022-006X.59.4.594

Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco: W. H. Freeman and Company.

Coopersmith, S. (2002). *Coopersmith Self-Esteem Inventories: Manual*. Palo Alto, CA: Mindgarden.

Corson K., Gerrity M. S., Dobscha S. K. (2004). Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. *The American Journal of Managed Care*, *10*, 839-845

Egan, L. A., & Todorov, N. (2009). Forgiveness as a coping strategy to allow school students to deal with the effects of being bullied: Theoretical and empirical discussion. *Journal of Social and Clinical Psychology*, *28*, 198-222.

Enright, R. D., Santos, M. J., & Al-Mabuk, R. (1989). The adolescent as forgiver. *Journal of Adolescence*, *12*, 99-110.

Enright, R. D., Freedman, S., & Rique, J. (1998). The psychology of interpersonal forgiveness. In R. D. Enright, & J. North (Eds.). *Exploring forgiveness* (pp. 4-62). The University of Wisconsin Press.

Enright, R. D. (2000). *Enright Forgiveness Inventory for Children*. Madison, WI: Department of Educational Psychology, University of Wisconsin-Madison.

Enright, R. D., & Fitzgibbons, R. P. (2015). *Forgiveness therapy: An empirical guide for resolving anger and restoring hope*. Washington, DC: American Psychological Association.

Enright, R. D., & Rique, J. (2004). *The Enright forgiveness inventory sampler set manual, instrument, and scoring guide*. Menlo Park, CA: Mind Garden.

Enright, R. D., Knutson, J., Holter, A. C., Baskin, T., & Knutson, C. (2007). Waging peace through forgiveness in Belfast, Northern Ireland II: Educational programs for mental health improvement of children. *Journal of Research in Education*, *Fall*, 63-78.

Enright, R., & Knutson, J. (2010). The journey toward forgiveness and love: a guided curriculum for children ages 10-12 [Program of studies]. https://internationalforgiveness.com/product-category/curriculum-guides/

Enright, R. D. (2012a). The anti-bullying forgiveness program: Reducing the fury within those who bully. Retrieved from www.internationalforgiveness.com

Enright, R. D. (2012b). *The forgiving life*: *A pathway to overcoming resentment and creating a Legacy of Love*. Washington, DC, US: American Psychological Association.

Enright, R. D., Rhody, M., Litts, B., & Klatt. J. S. (2014). Piloting forgiveness education in a divided community: Comparing electronic pen-pal and journaling activities across two groups of youth. *Journal of Moral Education*, *43*, 1-17.

Enright, R.D. (Eds.). (2020). Manuals for Assessing *Agape* Love: Adult and Adolescent Versions. Unpublished manual, June 30, 2020.

Enright R. D., Rique J., Lustosa R., Song M. J., Komoski M. C., Batool I., Bolt D. M., Sung H. J., Huang S. T., Park Y. H., Leer-Salvesen P. E., Andrade T., Naeem A., Viray J., Costuna E. (2021). Validating the Enright Forgiveness Inventory – 30 (EFI-30): International studies. *European Journal of Psychological Assessment*. Advance online publication. https://doi.org/10.1027/1015-5759/a000649

Enright, R. D., Wang Xu, J., Rapp, H., Evans, M., Song, J. Y. (2022a) The Philosophy and Social Science of *Agape* Love. *Journal of Theoretical and Philosophical Psychology*. https://doi.org/10.1037/teo0000202

Enright, R. D. (2022b, July 19-20). Defining our Terms: What are *Agape* Love and Forgiveness? [Conference presentation]. International Educational Conference on *Agape* Love and Forgiveness, Madison, WI, United States.

Exline, J. J., Worthington, E. L., Jr., Hill, P., & McCullough, M. E. (2003). Forgiveness and justice: A research agenda for social and personality psychology. *Personality and Social Psychology Review*, *7*, 337-348. https://doi.org/10.1207/S15327957PSPR0704_06

Flyn, J. J., Hollenstein, T., Mackey, A. (2010). The effect of suppressing and not accepting emotions on depressive symptoms: Is suppression different for men and women? *Personality and Individual Differences*, *49*, 582-586. https://doi.org/10.1016/j.paid.2010.05.022.

Francis, L. J., & Gibbs, D. (1996). Prayer and self-esteem among 8- to 11-year-olds in the United Kingdom. *Journal of Social Psychology*, *136*, 791-793.

Franklin, Benjamin. (August 23, 1750). [Letter from Benjamin Franklin to Samuel Johnson, 1750]. Retrieved from https://founders.archives.gov/documents/Franklin/01-04-02-0009

Freedman, S. R., and Enright, R. D. (1996). Forgiveness as an intervention goal with incest survivors. *Journal of Consulting and Clinical Psychology*, *64*, 983-992.

Freedman, S. R., & Zarifkar, T. (2015). The psychology of interpersonal forgiveness and guidelines for forgiveness therapy: What therapists need to know to help their clients forgive. *Spirituality in Clinical Practice*, *3*, 45-58.

Gambaro, M. E., Enright, R. D., Baskin, T. A., & Klatt, J. (2008). Can school-based forgiveness counseling improve conduct and academic achievement in academically at-risk adolescents? *Journal of Research in Education*, *18*, 16-27.

Gómez-Gómez, I., Benítez, I., Bellón, J., Moreno-Peral, P., Oliván-Blázquez, B., Clavería, A., … Motrico, E. (2022). Utility of PHQ-2, PHQ-8 and PHQ-9 for detecting major depression in primary health care: A validation study in Spain. *Psychological Medicine*, *52*, 1-11.

Gray-Little, B., Williams, V. S. L., & Hancock, T. D. (1997). An item response theory analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *23*, 443-451.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*, 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Hawes, S. W., Perlman, S. B., Byrd, A. L., Raine, A., Loeber, R., & Pardini, D. A. (2016). Chronic anger as a precursor to adult antisocial personality features: The moderating influence of cognitive control. *Journal of Abnormal Psychology*, *125*, 64-74.

Hebl, J. H., & Enright, R. D. (1993). Forgiveness as a psychotherapeutic goal with elderly females. *Psychotherapy*, *30*, 658-667.

Hendrick, C., & Hendrick, S. (1986). A theory and method of love. *Journal of Personality and Social Psychology*, *50*, 392-402. https://doi.org/10.1037/0022-3514.50.2.392

Herth, K. (1992). Abbreviated instrument to measure hope: development and psychometric evaluation. *Journal of Advanced Nursing*, *17*, 1251-1259.

Hills, P. R., Francis, L. J., & Jennings, P. (2011). The School Short-Form Coopersmith self-esteem inventory: Revised and Improved. Canadian Journal of School Psychology, 26(1), 62-71. https://doi.org/10.1177/0829573510397127

Ho, M. Y., & Fung, H. H. (2011). A dynamic model of forgiveness: A cross-cultural perspective. *Review of General Psychology*, *15*, 77-84.

Holmgren, M. (1993). Forgiveness and the intrinsic value of persons. *American Philosophical Quarterly*, *30*, 341-352. https://www.jstor.org/stable/20014472

Holter, A. C., Magnuson, C., Knutson, C., Knutson-Enright, J., & Enright, R. D. (2008). The forgiving child: The impact of forgiveness education on excessive anger for elementary-aged children in Milwaukee's central city. *Journal of Research in Education*, *18*, 82-93.

Hui, K. P. E., & Chau, T. S. (2009). The impact of a forgiveness intervention with Hong Kong Chinese children hurt in interpersonal relationships. *British Journal of Guidance & Counseling*, *37*, 141-156.

Johnson, Brian W., Doris L. Redfield, Richard L. Miller, & Robert E. Simpson. (1983). The Coopersmith self-esteem inventory: A construct validation study. *Educational and Psychological Measurement*, *43*, 907-913.

Johnson, J. G., Harris, E. S., Spitzer, R. L., & Williams, J. B. (2002). The patient health questionnaire for adolescents. Journal of Adolescent Health, 30, 196–204.

Jones, S. M., & Kahn, J. (2018). The evidence base for how learning happens: A consensus on

social, emotional, and academic development. *American Educator*, *41*, 16.

Kroenke, K., Spitzer, R. L., Williams, J. B. W. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606-613.

Kroenke K., Spitzer R.L. (2002). The PHQ-9: a new depression and diagnostic severity measure *Psychiatric Annals*, *32*, 509-521

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. Journal of *Affective Disorders*, *114*, 1-3, 163-173. https://doi.org/10.1016/j.jad.2008.06.026.

Lickona, T. (1996) Eleven principles of effective character education, *Journal of Moral Education*, *25*, 93-100.

Lin, W. N., Enright, R., & Klatt, J. (2011). Forgiveness as character education for children and adolescents. *Journal of Moral Education*, *40*, 237-253.

Loveland, J. M., Lounsbury, J. W., Welsh, D., & Buboltz, W. C. (2007). The validity of physical aggression in predicting adolescent academic performance. *British Journal of Educational Psychology*, *77*, 167–176. https://doi.org/10.1348/000709905X79563

Lundahl, W. B., Taylor, J. M., Stevenson, R., & Daniel, K. R. (2008). Process-based forgiveness interventions: a meta-analytic review. *Research on Social Work Practice*, *18*, 465.

North, J. (1987). Wrongdoing and forgiveness. *Philosophy*, *62*, 499-508. https://doi.org/10.1017/S003181910003905X

Oman, D., & Meyer, K. N. (2013). Research on unlimited love: First-wave findings and future directions. Retrieved from http://unlimitedloveinstitute.org/wpcontent/uploads/2015/02/RESEARCH-ON-UNLIMITED-LOVE-OMAN- REPORT.pdf

Overgaauw, S., Rieffe, C., Broekhof, E., Crone, E. A., & Güroğlu, B. (2017). Assessing empathy across childhood and adolescence: Validation of the empathy questionnaire for children and adolescents (EmQue-CA). *Frontiers in Psychology*, *8*, 870. doi:10.3389/fpsyg.2017.00870

Post, S. G., Underwood, L. G., Schloss, J. P., & Hurlbut, W. P. (2002). *Altruism and altruistic love.* Oxford University Press.

Park, J. H., Enright, R. D., Essex, M. J., Zahn-Waxler, C., & Klatt, J. S. (2013). Forgiveness intervention for south Korean female adolescent aggressive victims. *Journal of Applied Developmental Psychology*, *34*, 268-276.

Rahman, A., Iftikhar, R., Kim, J. J., & Enright, R. D. (2018). Pilot study: Evaluating the effectiveness of forgiveness therapy with abused early adolescent females in Pakistan. *Spirituality in Clinical Practice*, *5*, 75-87. https://doi.org/10.1037/scp00 00160

Rapp, H., Wang Xu, J., & Enright, R. D. (2022). A meta-analysis of forgiveness education interventions' effects on forgiveness and anger in children and adolescents. *Child Development*. Online publication. https://doi.org/10.1111/cdev.13771

Reed, G. L., & Enright, R. D. (2006). The effects of forgiveness therapy on depression, anxiety, and posttraumatic stress for women after spousal emotional abuse. *Journal of Consulting and Clinical Psychology*, *74*, 920-929.

Reyes, V. A., & Hicklin, T. A. (2005). Anger in the combat zone. *Military Medicine*, *170*, 483-487. doi:10.7205/ MILMED.170.6.483

Rubin, R. A. (1978). Stability of self-esteem ratings and their relation to academic achievement: A longitudinal study. *Psychology in the Schools*, *15*, 430-433.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Sewell, T. E. (1985). *Review of the Coopersmith self-esteem inventory*. In Mitchell, J. (ed.), The Ninth Mental Measurements Yearbook. University of Nebraska Press, Lincoln.

Shechtman, Z., Wade, N., & Khoury, A. (2009). Effectiveness of a forgiveness program for Arab Israeli adolescents in Israel: An empirical trial. *Peace and Conflict: Journal of Peace Psychology*, *15*, 415-438.

Simon, Y. R. (1986). *The definition of moral virtue*. Fordham University Press.

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology*, *82*, 22-32.

Snell, W. E., Gum, S., Shuck, R. L., Mosley, J. A., & Kite, T. L. (1995). The clinical anger scale: Preliminary reliability and validity. *Journal of Clinical Psychology*, *51*, 215-226.

Snyder, C. R. (1994). *The psychology of hope: You can get there from here*. New York: Free Press.

Snyder, C. R., Hoza, B., Pelham, W.E., Rapoff, M., Ware, L., Danovsky, M., *et al*. (1997). The development and validation of the Children's Hope Scale. *Journal of Pediatric Psychology*, *22*, 399-421.

Spitzer, R. L., Kroenke, K., Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study: Primary Care Evaluation of Mental Disorders: Patient Health Questionnaire. *Journal of the American Medical Association*. *282*:1737-1744.

Spielberger, C.D. (1983). *State-trait anxiety inventory for adults*. Palo Alto, CA: Mind Garden.

Sprecher, S., & Fehr, B. (2005). Compassionate love for close others and humanity. *Journal of Social and Personal Relationships*, *22*, 629-651.

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to

multiple empathy measures. *Journal of Personality Assessment*, *91*, 62-71. doi:10.1080/00223890802484381

Steele, R. G., Legerski, Nelson, T. D., & Phipps, S. (2009). The Anger Expression Scale for Children: Initial validation among healthy children and children with cancer. *Journal of Pediatric Psychology*, *34*, 51-62, 2009. doi:10.1093/jpepsy/jsn054

Strauss, C. C., Frame, C. L., & Forehand, R. (1987). Psychosocial impairment associated with anxiety in children. *Journal of Clinical Child Psychology*, *16*, 235-239.

Subkoviak, M. J., Enright, R. D., Wu, C. R., Gassin, E. A., Freedman, S., Olson, L. M., *et al*. (1995). Measuring interpersonal forgiveness in late adolescence and middle adulthood. *Journal of Adolescence*, *18*, 641-655.

Taysi, E., & Vural, D. (2016). Forgiveness education for fourth grade students in Turkey. *Child Indicators Research*, *9*, 1095-1115.

Vassilopoulos, S. P., Koutsoura, A., Brouzos, A., & Tamami, D. (2020). Promoting forgiveness in Greek preadolescents: a universal, school-based group intervention. *British Journal of Guidance & Counselling*, *48*, 670-684. https://doi.org/10.1080/03069 885.2018.1527285

Wang Xu, J., Kim, J. K., Olmstead, N., & Enright, R. D. (2021). The development of forgiveness and other-focused love. *Journal of Psychology and Theology*, *50*. https://doi.org/10.1177/00916471211034514

Wade, N. G., Hoyt, W. T., Kidwell, J. E. M., & Worthington, Jr., E. L. (2014). Efficacy of psychotherapeutic interventions to promote forgiveness: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *82*, 154-170.

Watkins, D. (1982). Sex-role perceptions of Filipino adolescents. *International Journal of Psychology*, *17*, 359-368.

Wellborn, J. G. (1991). *Engaged and disaffected action: The conceptualization and measurement of motivation in the academic domain*. Rochester, NY: University of Rochester.

Wiesner, M. & Windle, M. (2004). Assessing covariates of adolescent delinquency trajectories: A latent growth mixture modeling approach. *Journal of Youth and Adolescence*, *33*, 431-442.

Wivestad, S. (2008) The educational challenges of *agape* and phronesis. Journal of Philosophy of Education, 42, 307-324.

Worthington, E. L., Jr. (2020). Understanding forgiveness of other people: Definitions, theories, and processes. In E. V. Worthington, Jr. (Ed.), *Handbook of forgiveness* (2nd ed., pp. 11-21). Routledge.

Yu, L., Gambaro, M., Song, J. Y., Teslik, M., Song, M., Komoski, M. C., . . .Enright, R. D. (2021). Forgiveness therapy in a maximum-security correctional institution: A randomized clinical trial. *Clinical Psychology & Psychotherapy*. Advance online publication. https://doi.org/10.1002/cpp.2583

Zuckerbrot, R. A., Cheung, A., Jensen, P. S. (2018). Guidelines for adolescent depression in primary care (GLAD-PC): part I. practice preparation, identification, assessment, and initial management. *Pediatrics*, *141*.

Appendix A: Fidelity Checklist

Check each item that occurs during the lesson.

1. ___ The instructor reviews the previous lesson by making 1-3 statements about what was covered. (Not applicable to Lesson 1). This often will include a review of what forgiveness is.
2. ___ The instructor introduces the day's lesson using the script provided in the manual.
3. ___ The instructor conducts the discussion for the lesson using at least three questions provided in the manual.
4. ___ The instructor concludes the lesson by summarizing 1-2 main ideas for that lesson.

Number of items completed: _____/4

Percent of task items completed: _____%

Evaluating Teacher Questionnaires

Based on the teacher questionnaire, did the teacher perceive that the students were actively participating? (yes/no)

What percent of the "did students learn" items were indicated/checked by the teacher? _____%

Appendix B: Criteria of Problem Intervention Coding

31 - abnormally long time (one SD above the group mean) between pre- and post- surveys of the intervention

21- teacher reported more than 4 weeks of having 2 or more lessons taught in the same week

11 - teacher reported 3 or 4 weeks of having 2 or more lessons taught in the same week

1 - teacher reported 1 or 2 weeks of having 2 or more lessons taught in the same week

0 - no intervention issue observed

*If the class has more than one issue, the add-up score will be the code.

*Only if the add-up score (code) is >=31, the class will be eliminated from analysis.

Appendix C: Criteria of Enright *Agape* Love Inventory (EALI) Story Rating

0 - acceptable story

1 - No story: if EALI_Q5 is blank or filled with random characters

2 - I don't want to say/share/secret/lots of things

3 - I haven't had such situation/no help/no hurt/don't know

4 - ambiguous theme

Appendix D: Criteria of Enright Forgiveness Inventory (EFI) Story Rating

0 - acceptable story

1 - No story: if EFI_Q6 is blank or filled with random characters

2 - I don’t want to say/share/secret/lots of things

3 - I haven’t had such situation/no help/no hurt/don’t know

4 - ambiguous theme