Study of effectiveness of prevention programs

Mihić, Josipa

University of Zagreb

FACULTY
OF EDUCATION AND REHABILITATION SCIENCES

Josipa Mihić

STUDY OF EFFECTIVENESS OF PREVENTION PROGRAMS

DOCTORAL THESIS

Zagreb, 2013

Sveučilište u Zagrebu

EDUKACIJSKO-REHABILITACIJSKI FAKULTET

Josipa Mihić

PROCJENA UČINKOVITOSTI PREVENTIVNIH PROGRAMA

DOKTORSKI RAD

Zagreb, 2013

University of Zagreb

FACULTY

OF EDUCATION AND REHABILITATION SCIENCES

Josipa Mihić

STUDY OF EFFECTIVENESS OF PREVENTION PROGRAMS

DOCTORAL THESIS

Supervisor:

Clemens Hosman

Zagreb, 2013

Sveučilište u Zagrebu

EDUKACIJSKO-REHABILITACIJSKI FAKULTET

Josipa Mihić

PROCJENA UČINKOVITOSTI PREVENTIVNIH PROGRAMA

DOKTORSKI RAD

Mentor:

Clemens Hosman

Zagreb, 2013

Na temelju članka 46. Statuta Edukacijsko-rehabilitacijskog fakulteta Sveučilišta u Zagrebu, Fakultetsko vijeće Edukacijsko-rehabilitacijskog fakulteta na 4. redovitoj sjednici održanoj dana 28. siječnja 2013. godine donijelo je odluku o imenovanju Povjerenstva za obranu doktorske disertacije u sastavu:

1. **Prof. dr. sc. Josipa Bašić**, redovita profesorica Edukacijsko-rehabilitacijskog fakulteta Sveučilišta u Zagrebu, predsjednica
2. **Prof. dr. sc. Gordana Pavleković**, redovita profesorica na Školi narodnog zdravlja Andrija Štampar, Medicinski fakultet Sveučilišta u Zagrebu, članica
3. **Prof. dr.sc. Damir Ljubotina**, izvanredni profesor Filozofskog fakulteta Sveučilišta u Zagrebu, član

Imenovano povjerenstvo pozitivno je ocijenilo doktorsku disertaciju Josipe Mihić pod naslovom „Study of Effectiveness of Prevention Programs“.

Disertacija je uspješno obranjena pred navedenim povjerenstvom 11. ožujka 2013. godine na Edukacijsko-rehabilitacijskom fakultetu Sveučilišta u Zagrebu.

CURRICULUM VITAE OF THE DOCTORAL STUDY SUPERVISOR

Clemens Hosman is emeritus Professor of Mental Health Promotion and Prevention of Mental Disorders at the Maastricht University (Health Promotion) and Radboud University Nijmegen (Clinical Psychology) in The Netherlands and former director of their common Prevention Research Centre. He is also guest professor of prevention at the University of Zagreb within the PhD program on Prevention and Prevention Science in the field of mental health. His research in collaboration with many Dutch mental health and public health centres is targeted at prevention of depression, prevention of transmission of mental disorders from parents to children, development of preventive community interventions, principles of program development, effectiveness studies, effect management, and innovation of prevention approaches and policies. He is offering (post)graduate prevention seminars and training at both Dutch universities and in other European countries. He is involved in prevention and mental health promotion since 1969 and played a crucial role in the development of this field in the Netherlands and Europe. He has occupied many international functions, e.g. chair of the European WHO Task Force on MH Promotion and Prevention, chair of the Clifford Beers Foundation, board member of the international Society for Prevention Research, and chair/member of the Global Consortium for Prevention/Promotion in Mental Health. He was co-project leader of several European prevention projects (e.g. IMHPA, DataPrev) and serves as prevention consultant to NGO's and professional and governmental agencies in many countries. He is co-author of the PREFFI 2.0, an effect management instrument that aims to enhance the implementation of scientific knowledge on effectiveness in prevention policy and practice. He was principle editor of the WHO Report on Evidence-based Prevention of Mental Disorders (Hosman, Jane-Llopis & Saxena, 2004).

A selection of supervisor's publications:

Kersten-Alvarez, L.E., Hosman, C.M.H., Riksen-Walraven, M., van Doesum, K.T.M. and Hoefnagels, C. (2010). Which Preventive Interventions Effectively Enhance Depressed Mothers' Sensitivity? A Meta-analysis. *Infant Mental Health Journal, 32, 3, 362-376.*

Kraag, G., Van Breukelen, G.J.P, Kok, G. and Hosman, C.M.H. (2009). 'Learn Young, Learn Fair', a Stress Management Program for Fifth and Sixth Graders: Longitudinal Results from an Experimental Study. *Journal of Child Psychology and Psychiatry, 50, 9,* 1185-1195.

Van Doesum, K., Riksen-Walraven, M., Hosman, C.M.H. and Hoefnagels, C. (2008). A Randomized Controlled Trial of an Early Intervention Aimed at Preventing Relationship Problems in Depressed Mothers and their Infants. *Child Development,79, 3, 547-561.*

Patel, V., Araya, R., Chatterjee, S., Chisholm, D., Cohen, A., De Silva, M., Hosman, C.M.H., McGuire, H., Rojas, G. and van Ommeren, M. (2007). Treatment and Prevention of Mental Disorders in Low-income and Middle-income Countries. *Lancet, 15, 370 (9591),* 991-1005.

Molleman, G.R.M., Peters, L.W.H., Hosman, C.M.H., Kok, G. and Oosterveld, P. (2005). Project Quality Rating by Experts and Practitioners with Preffi 2.0 as a Quality Assessment Instrument. *Health Education Research, 21, 2,* 219-229.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to many people and institutions for the support they have given me throughout the course of my Ph.D. Special thanks go to my colleague and friend, Miranda Novak. I will always be thankful to the Universe for having her as an invaluable companion in going through this stage of my life. I have shared with her many wonderful moments during the last couple of years. I want to thank her for being next to me and supporting me during all the challenging moments. You have encouraged my soul to grow.

I am very much grateful to my supervisor, Professor Clemens Hosman. I would like to thank him for his patience, wisdom and encouragement. Working with him has truly been an honour and an inspiration. The expertise, knowledge, rationality and optimism he has generously shared with me over the last four years has been invaluable. Most of all, I want to thank him for "running the marathon" right next to me and encouraging me during the whole process.

Special thanks go to Gerard Molleman, PhD, for his willingness to share his knowledge and expertise and for showing his interest in international collaboration established within this doctoral project.

I would also like to thank Professor Josipa Bašić. With her compassion in developing prevention science in Croatia and through establishing a doctoral programme in this field, she has provided me and my colleagues the unique opportunity to learn from renowned experts and in that way to grow professionally. Her work has resulted in many valuable projects and initiatives and I am thankful that I had a chance to be a part of them.

The project would not have been possible without the support of the Department of Health and Social Services of the County of Istria, and Head of the Department, Sonja Grozić-Živolić and her colleagues. Their readiness and sensibility for investment into the development of evidence-based mental health promotion and prevention practice was crucial for conducting this project. Also, very special thanks go to all the participants of this study. I'm glad I met so many enthusiastic individuals dedicated to their work within a civil sector. It was inspiring to collaborate with them.

I would also like to express my gratitude to Celene Domitrovich for supporting me and generously sharing her knowledge during the process of this study development. I'm also grateful to Martina Lotar and Professor Lajos Szirovicza, for their invaluable comments, advice and support during the process of data analysis. Special thanks go to Tena Vukasović

for her generous willingness to share her knowledge and experience in conducting the meta-analysis.

I would also like to thank Martina and Valentina for awakening my interest in prevention science and practice at the time when I was a graduate student, and motivating me to develop professionally in this field. I am thankful that you have believed in me.

I wish to express gratitude and love to all of my friends for their support and help. My deepest gratitude goes to Jernej, for his endless understanding, patience, support and encouragement when it was most required. I am particularly grateful to you for reminding me about what is truly important in life and for sharing your peace with me. You have helped me in every way imaginable… I would also like to thank my friend Vesna for being a truly fantastic friend and for providing amazing support. Miss r., thank you for the laughter, endless talks and for believing in my "digging-the-way-to-China" projects. My friend Mirna was also tremendously supportive. Thank you for sharing with me all the wonderful moments in my life and for all the fun, but also for keeping me sane in times of crises. I am also very much thankful to Ivana, Una, Iva, Kiki and Kristina for showing their support in so many nice ways.

Finally, I would like to express my gratitude and love to my family, my parents Vlatka and Stjepan and to my brother Ivan. *Hvala vam što ste uvijek davali sve od sebe da mi omogućite uvjete kako bih se razvijala osobno i profesionalno. Zahvalna sam na svemu što sam od vas naučila i dobila.*

SUMMARY

There is an increasing focus on the effectiveness of mental health promotion and prevention interventions worldwide. Many studies have shown that mental health promotion and prevention interventions can be effective in reducing mental, emotional and behavioural problems (Botvin et al., 2006; Hawkins et al., 2005; Hosman et al., 2004; Tobler and Stratton, 1997). This doctoral study was conducted with the intention to provide deeper understanding of mental health promotion and prevention programs' effectiveness in the Croatian context. A study was a part of project "Preffi – Quality Assurance in the County of Istria", initiated by the University of Zagreb, Faculty of Education and Rehabilitation Sciences and the Department of Health and Social Services in the County of Istria. The main assumption of the project was that incorporation of evidence-based principles is crucial in improving the quality and effectiveness of mental health promotion and prevention practice. The study included 24 mental health promotion and prevention programs proposed by NGOs and other institutions from Istria and financed by the Department. All programs were assessed with the quality assessment instrument Preffi 2.0. The programs were divided into two groups, experimental and control. The program's managers and deliverers from the experimental group were involved in the Training for Prevention intervention. In order to assess whether the Training had an impact on programs' effectiveness, an outcome evaluation of programs from both conditions was conducted.

The results of the study on metric characteristic of Preffi 2.0 have shown that the instrument has a high content validity and that it is reliable in assessing the quality of program proposals. Concerning the predictive validity of Preffi 2.0, the results of a correlation analysis have shown that there is no significant correlation between total scores on Preffi 2.0 and the effect sizes of programs. However, it was found that there is a moderate, positive linear relationship between two Preffi clusters and effects of programs.

A method of meta-analysis has shown that programs whose managers and deliverers were involved in the Training did not achieve significantly higher scores on effectiveness compared to those programs which were not involved in the Training. However, the analysis has shown that the Training for prevention influence on the programs' effect sizes is marginally significant at the level of 10% (p=.11). Program managers and deliverers involved in the Training achieved significantly higher scores on three Preffi clusters after the Training, in comparison with programs which were not involved in the Training.

Assessment of written program proposals with the Preffi 2.0 instrument has offered an insight into each program's quality level and into the average quality level of 24 programs from the County of Istria. The results suggest that the general quality level of assessed program proposals from the County of Istria is rather low. Existing strengths and weaknesses of assessed programs are detected and described.

Based on the study results, recommendations for Preffi 2.0 and Training for Prevention improvement and future studies are given. Perspectives on Preffi 2.0 and Training for Prevention in improving prevention capacities in Croatia are also elaborated in this doctoral dissertation.

Key words: *mental health promotion, prevention, effect predictors, effectiveness, quality assessment, quality assurance*

SAŽETAK

Znanstvenici diljem svijeta sve su više zainteresirani za temu učinkovitosti programa promocije mentalnog zdravlja i prevencije mentalnih, emocionalnih i ponašajnih problema. Brojna istraživanja ukazuje na to da kvalitetno osmišljene preventivne intervencije mogu značajno utjecati na smanjenje problema i poremećaja u ponašanju kao što su nasilničko i delinkventno ponašanje (Botvin i sur., 2006), pijenje alkohola i korištenje droga (Tobler i Stratton, 1997) ili rizično seksualno ponašanje (Kirby i sur. 1994). Provedene meta-analize su pokazale kako postoje značajne razlike između preventivnih programa s obzirom na učinkovitost te su uz pomoć takvih istraživanja utvrđeni čimbenici koji utječu na učinkovitost programa nazvani prediktorima, odnosno moderatorima učinkovitosti (Stice i sur., 2007; Hosman, 2005; Jane Llopis, Hosman, Jenkins i sur., 2003; Brown i sur., 2000; Hosman i Engels, 1999; Kok i sur., 1997). Istraživanja su ukazala kako je za unaprjeđenje kvalitete programa promocije mentalnog zdravlja i prevencije, u njih nužno integrirati znanja o principima učinkovitih intervencija.

U posljednje je vrijeme i u Hrvatskoj sve izraženija svijest znanstvenika i praktičara o važnosti i učinkovitosti znanstveno utemeljenih intervencija tog tipa. Istraživanje koje će biti prikazano u ovoj disertaciji provedeno je unutar projekta "Preffi - osiguranje kvalitete u Istarskoj županiji" nastalog u suradnji znanstvenika sa Sveučilišta u Zagrebu, Edukacijsko-rehabilitacijskog fakulteta te stručnjaka iz Upravnog odjela za zdravstvo i socijalnu zaštitu Istarske županije. Upravni odjel svake godine u okviru natječaja za Programe javnih potreba dodjeljuje financijska sredstva nevladinim organizacijama za programe prevencije poremećaja u ponašanju djece i mladih te liječenja ovisnosti. Za ostvarivanje financijske potpore projekti trebaju zadovoljiti kriterije razvijene od strane Upravnog odjela. Tijekom dugogodišnje suradnje Fakulteta i Upravnog odjela primijećena je potreba za povećanjem kvalitete preventivnih intervencija u Istarskoj županiji te je upravo s toga pokrenut navedeni projekt. Glavni pretpostavka projekta bila je da je za unaprjeđenje programa promocije mentalnog zdravlja i prevencije nužno integrirati znanja o indikatorima i principima učinkovitih intervencija u postojeće i nove programe te u kriterije za odabir i financiranje programa.

U istraživanje su bila uključena 24 programa promocije mentalnog zdravlja i prevencije mentalnih, emocionalnih i ponašajnih problema financirana od strane Upravnog odjela za zdravstvo i socijalnu skrb Istarske županije ili nekih drugih institucija u Istarskoj županiji u 2011. godini. Tijekom prosinca 2010. i siječnja 2011. godine tri nezavisna procjenjivača su procijenila kvalitetu 24 pisana prijedloga programa uz pomoć Preffi 2.0 instrumenta

(Molleman, 2005). Riječ je o nizozemskom instrumentu koji se sastoji od 121 čestice koje reflektiraju čimbenike koji se smatraju neophodnima za kvalitetu i učinkovitost programa. Čestice čine 39 indikatora kvalitete, učinkovitosti programa koji su grupirani u 8 Preffi klastera; (1) Kontekstualni uvjeti i provedivost intervencije, (2) Analiza problema, (3) Determinante problema, ponašanja i okruženja, (4) Ciljna skupina, (5) Ciljevi intervencije, (6) Razvoj intervencije, (7) Implementacija i (8) Evaluacija. Nakon što su projektne prijave programa procijenjene uz pomoć Preffi 2.0 instrumenta, programi su raspoređeni u eksperimentalnu i kontrolnu skupinu te je svaka skupina sadržavala 12 programa. Pri svrstavanju programa u skupine korištena je metoda izjednačenih parova s obzirom na: vrstu programa, problem na koji je program usmjeren, broj korisnika u programu, broj godina tijekom kojih je program već bio financiran te duljinu trajanja provedbe programa. Važan čimbenik izjednačavanja skupina bili su i rezultati procjene programa Preffi 2.0 instrumentom.

Voditelji organizacija i provoditelji programa iz eksperimentalne skupine su tijekom veljače i ožujka 2011. godine bili uključeni u Trening za prevenciju, intervenciju osmišljenu za potrebe ove doktorske studije. Trening se sastoji od 35 sati edukacije o principima učinkovitosti programa promocije mentalnog zdravlja i prevencije mentalnih, emocionalnih i ponašajnih problema. Podijeljen je u pet tematskih cjelina: (1) Znanstveno-utemeljena promocija mentalnog zdravlja i prevencija, (2) Logički model i kvaliteta, (3) Implementacija, (4) Evaluacija te (5) Zagovaranje. Kako bi se utvrdilo da li je Trening utjecao na učinkovitost programa onih organizacija koje su sudjelovale u Treningu, provedena je evaluacija učinkovitosti programa iz obje skupine. Korišten je pre-test i post-test istraživački dizajn te je za svaki program pripremljena upitnička baterija, ovisno o ciljevima programa. Programi su implementirani tijekom 2011. godine. Isto tako, kako bi se ispitalo da li je Trening utjecao na kvalitetu pisanih prijedloga programa, sudionici iz obje skupine su zamoljeni da napišu nove programske prijedloge koji su zatim procijenjeni Preffi 2.0 instrumentom tijekom prosinca 2011. i siječnja 2012. godine.

Prvi istraživački zadatak ovog doktorskog rada bio je ispitati metrijske karakteristike Preffi 2.0 instrumenta - sadržajnu valjanost, pouzdanost i prediktivnu valjanost. Sadržajnu valjanost je procjenjivalo 10 stručnjaka iz područja promocije mentalnog zdravlja i prevencije. Procijenjeno je da je Preffi 2.0 sadržajno valjan instrument, odnosno da njegove čestice dobro reflektiraju teoriju vezanu uz prediktore učinkovitosti i značajne su za procjenu prisutnosti i kvalitete prediktora učinkovitosti u nekom programu. Isto tako, utvrđeno je kako među svim Preffi 2.0 klasterima postoje visoke i pozitivne korelacije što je i očekivano s

obzirom da svi Preffi klasteri reflektiraju generalan koncept kvalitete programa. U odnosu na pouzdanost, utvrđeno je da je tijekom obje primjene Preffi instrumenta na pisanim prijedlozima programa, pouzdanost instrumenta bila visoka. Međutim, treba naglasiti kako su pri tom pogreške mjerenja bile više od dopuštenih vrijednosti. Rezultat ukazuje da se adekvatna preciznost mjerenja Preffi instrumentom za sada ne može postići uz pomoć samo tri procjenjivača. Istraživanje je također pokazalo da je najveći izvor varijance za ukupan Preffi 2.0 rezultat u obje procjene upravo prijedlog programa. Što se tiče prediktivne valjanosti Preffi instrumenta, rezultati parcijalne korelacijske su pokazali kako ne postoji značajna korelacija između ukupnog Preffi 2.0 rezultata te veličine efekta programa. Međutim, analiza korelacije između rezultata na pojedinim Preffi klasterima i učinaka programa je pokazala da postoji umjerena, pozitivna korelacija između trećeg Preffi klastera „Determinante problema, ponašanja i okruženja“ i petog klastera „Ciljevi intervencije“ te veličina efekata programa.

Drugi istraživački zadatak bio je ispitati utjecaj Treninga za prevenciju na učinkovitost i kvalitetu programa. Kako bi se ispitala razlika u veličinama efekata programa koje su provodili provoditelji koji su sudjelovali u Treningu za prevenciju i onih koji nisu sudjelovali u Treningu, provedena je metoda meta-analize. Rezultati su pokazali kako programi uključeni u Trening postižu veće veličine efekata od programa koji nisu bili uključeni u Trening, ali da ta razlika nije statistički značajna. Provedena je i analiza moderatora veličine efekata koja je pokazala kako je Trening za prevenciju marginalno značajan na razini od 10% ($p = 0,11$). Što se tiče utjecaja Treninga za prevenciju na kvalitetu programa, analiza varijance za ponovljena mjerenja je pokazala da nema razlike između eksperimentalne i kontrolne skupine na ukupnom Preffi 2.0 rezultatu. Međutim, rezultati su pokazali da su programi koji su bili uključeni u Trening postigli značajno bolje rezultate na tri Preffi klastera u odnosu na programe čiji predlagatelji i provoditelji nisu sudjelovali u Treningu. Navedeno se odnosi na klastere: (3) Determinante problema, ponašanja i okruženja, (5) Ciljevi intervencije te (8) Evaluacija.

Treći istraživački zadatak bio je identificirati kvalitetu, odnosno snage i slabosti programa promocije mentalnog zdravlja i prevencije iz Istarske županije koji su bili uključeni u istraživanje. Uvid u to je pružila prva procjena prijedloga programa Preffi 2.0 instrumentom u prosincu 2010. i siječnju 2011. godine. Prosječni rezultat za 24 procijenjena programa na Preffi 2.0 instrumentu je iznosio 5.68 (mogući raspon Preffi 2.0 rezultata je od vrijednosti 3.33 do 10). Ova prosječna razina kvalitete svih procijenjenih programa je ispod teoretske sredine rezultata koja iznosi 6.65. Rezultat upućuje na to da je opća razina kvalitete

procijenjenih prijedloga programa prilično niska. Samo 4 od 24 procijenjena programa je dostiglo Preffi 2.0 rezultat viši od teoretske sredine. Analiza rezultata na pojedinim Preffi 2.0 klasterima je pokazala kako je kvaliteta sedam od osam Preffi 2.0 klastera niža od teoretske sredine. Samo je klaster „Ciljevi intervencije“ postigao prosječni rezultat veći da teoretske sredine.

S obzirom na sve dobivene rezultate osmišljene su preporuke za unaprjeđenje Preffi 2.0 instrumenta. Jedan od zaključaka je da bi prije procjene projektnih prijedloga s instrumentom, procjenjivači za to trebali biti posebno pripremljeni i educirani. Također je utvrđeno da bi pojedine Preffi čestice trebale biti jasnije formulirane te da bi uz sam instrument bili korisno pripremiti rječnik pojmova iz područja prevencijske znanosti koji se pojavljuju u instrumentu. Isto tako je zaključeno da se procjena prijedloga programa Preffi 2.0 instrumentom ne bi trebala temeljiti samo na pisanim prijedlozima već da bi bilo korisno provesti i intervjue s predlagateljima programa kako bi se osigurao uvid u sve karakteristike programa potrebne za procjenu Preffi instrumentom. Također se ističe kako bi se u slučajevima kada nekoliko procjenjivača procjenjuje isti prijedlog programa uz pomoć Preffi 2.0 instrumenta, konačni Preffi 2.0 rezultati trebali definirati kroz diskusiju i konsenzus svih procjenjivača. Predlaže se i priprema digitalne verzije Preffi 2.0 instrumenta koja bi olakšala njegovu primjenu.

Što se tiče preporuke budućih istraživanja vezanih uz Preffi 2.0, predlaže se istraživanje iskustva praktičara u korištenju instrumenta te njihove percepcije korisnosti i razumljivosti instrumenta. Također bi bilo korisno u jednom od budućih istraživanja provesti faktorsku analizu instrumenta. Predlaže se i ponovna studija pouzdanosti instrumenta na većem uzorku prijedloga programa kao i ponovljena studija njegove prediktivne valjanosti.

U odnosu na Trening za prevenciju i s obzirom na dobivene rezultate istraživanja, utvrđeno je da bi tijekom Treninga posebna pozornost trebala biti dana temama poput kontekstualni uvjeti i provedivost intervencije, analiza problema, determinante problema, ponašanja i okruženja, ciljna skupina, ciljevi intervencije, razvoj intervencije te implementacija. Također je utvrđeno kako bi bilo korisno pripremiti skriptu sa sadržajem Treninga te kako bi tijekom Treninga sudionici trebali biti što više uključeni u praktične aktivnosti. Individualne konzultacije sa sudionicima Treninga o kvaliteti njihovih programa bi također doprinijele učinku Treninga na kvalitetu i učinkovitost programa. Potvrda o uspješnom sudjelovanju u Treningu bi se trebala osigurati svim sudionicima koji su sudjelovali u cijelom Treningu i koji su na kraju treninga unaprijedili prijedloge svojih programa. Vjeruje se kako bi se na taj način sudionike Treninga motiviralo na aktivnu uključenost i usvajanje znanja. Primijećena je i potreba da se razviju dva modula Treninga.

Jedan modul bi bio usmjeren na usvajanje osnovnih znanja iz područja promocije mentalnog zdravlja i prevencije dok bi napredni modul bio namijenjen onima koji već imaju iskustvo s razvijanjem i provođenjem programa u navedenom području. Također se ističe kako bi sadržaj Treninga trebalo kontinuirano usklađivati i nadograđivati sukladno novim znanstvenim spoznajama iz područja učinkovitosti programa promocije mentalnog zdravlja i prevencije.

Zaključno je istaknuto kako Preffi 2.0 instrument i Trening za prevenciju predstavljaju značajan potencijal za unaprjeđenje programa promocije mentalnog zdravlja i prevencije u Hrvatskoj. Također je naglašeno kako je to unaprjeđenje dugotrajan proces koji zahtjeva suradnju znanstvenika, praktičara i donositelja odluka na lokalnim razinama kao i na nacionalnoj razini. Tek višerazinskim i sustavnim ulaganjem u razvoj znanstveno-utemeljenih programa promocije mentalnog zdravlja i prevencije je moguće očekivati značajne, pozitivne ishode i na javnozdravstvenoj razini.

Ključne riječi: *promocija mentalnog zdravlja, prevencija, prediktori učinkovitosti, učinkovitost, procjena kvalitete, osiguravanje kvalitete*

CONTENTS

1. INTRODUCTION

To introduce the main theme of this dissertation - program effectiveness - the field of mental health promotion and prevention science will briefly be described. Since mental health promotion and prevention science in Croatia is still developing, an overview of the state of the art of mental health promotion and prevention science in Croatia will also be presented. The main theme of the dissertation is elaborated through different sections of the introductory chapter. That includes relevant findings about effect predictors in mental health promotion and prevention, and a description of an evaluation process of program effectiveness. The accent of the second part of the introductory chapter is on two concepts closely related to mental health promotion and prevention effectiveness – concepts of quality assessment and quality assurance. In the last part of the chapter, a project within which this doctoral study was conducted will be presented, including the Training for Prevention intervention developed within the project.

1.1. Mental Health Promotion and Prevention in Croatia and Worldwide

Positive mental health is crucial in today's society because it has a power to stimulate growth and development and contribute to prosperity, solidarity, social justice and increased quality of life across the world. During the last decade, promotion of mental health and prevention have been given high priority on the agenda of the World Health Organization and in health policies of many Western European countries, the United States, Canada and Australia (Herman, Saxena and Moodie, 2005; Hosman, Jane-Llopis, and Saxena, 2004). The term *mental health promotion and prevention* in this dissertation refers to a wide range of initiatives focused on promotion of mental health and prevention of mental, emotional and behavioural problems (MEB) among individuals and groups.

Mental, emotional and behavioural problems among young people are associated with substantially increased morbidity and reduced health-related quality of life. These problems are connected with psychological suffering as well as increased risks of physical illnesses (Vreeland, 2007). They represent an enormous burden during childhood (Glied and Cuellar, 2003), and are also correlated with significantly increased risks to health and reduced productivity in adulthood (Kessler et al., 2005). Mental disorders lead to lost productivity and functioning not only for the children, but also for the parents and caregivers of the children (Tolan and Dodge, 2005). An individual's mental health condition may affect his or her peers

also; in particular, substance abuse and suicidal behaviour (Gould et al., 2003, according to O'Connell, Boat and Warner, 2009; Gaviria and Raphael, 2001) are thought to spread among peers via a contagion effect. Untreated mental illness may have intergenerational effects also. Having a depressed mother, or having two parents with poor mental health, is associated with mental, emotional and behavioural problems in children (Kahn et al., 2004, according to O'Connell, Boat and Warner, 2009).

According to Global Burden of Disease Study (Murray and Lopez, 1996), the burden of mental, emotional and behavioural problems has been seriously underestimated. While mental health problems are responsible for little more than 1% of deaths, they account for almost 11% of the disease burden worldwide. Predictions suggest that by 2020 the disease burden of mental health conditions in the world may increase to almost 15%.

In 2010, mental health problems and disorders were the second leading cause of all hospitalizations in Croatia (http://www.hzjz.hr/epidemiologija/kron_mas/dusevne.htm) with 13.3% of all cases. Epidemiological data shows that in 2010, 18.8% of all hospitalizations are due to mental disorders and were caused by alcohol drinking problems, 15.7% were caused by schizophrenia disorders, 13.2% of hospitalizations occurred because of depressive problems and 10.2% were caused by post-traumatic stress disorder.

Figures from ESPAD research conducted in Croatia in 2007 on smoking, drinking, and drug consumption (Kuzman, Pejnović Franelić, Pavić Šimetin and Pejak, 2008) indicate a sharp increase in the frequency of alcohol consumption amongst minors in Croatia. Drinking six and more times over the past few months more than doubled for boys (13-39%), and quadrupled amongst girls (4-16%) in the period from 1995 to 2007. A similar trend was noticed with excessive consumption (five or more drinks in one evening), recorded with nearly every other girl and more than half of all boys. The same research showed that the number of minors experimenting with drugs in the monitored period (1995-2007) has peaked, stabilizing at 20% of boys and 15% of girls. The results of the last ESPAD survey conducted in 2011 (Hibell et al., 2011) indicate that cigarette smoking in the past 30 days is more common in Croatia than in other European countries, and alcohol use is also higher in terms of past-30-days use and the amount consumed on the most recent drinking day. In addition, the proportion of students reporting that they had engaged in heavy episodic drinking during the past 30 days is also above average. Lifetime use of inhalants, which was reported by 28%, is considerably above average. The authors of the study stress that this is three times higher than the average for all countries, and of the eight variables studied, it is the one that differs the most from ESPAD average. According to the Croatian National Institute of Public Health

(http://www.hzjz.hr/), the number of drug addicts in the Republic of Croatia more than quintupled over the past ten years (1995, 1,340 addicts were treated, against 7,427 in 2007). It is of great concern, however, that nearly 35% of the total population of addicts consists of children and youth in reach of the education system, specifically, up to 24 years of age.

Figures from the Ministry of the Interior of the Republic of Croatia (2011) over the past few years show an increase in the number of crimes committed by minors aged 14 years and younger. In 2004, the UNICEF Office in Croatia (Pregrad, 2007) conducted an in-depth research into the phenomenon of violent behaviour covering 84 elementary and 9 high schools in Croatia. An average 10.4% of children were continuously exposed to violence from their peers (every week); while 22.3% had been exposed to violent behaviour one to two times over the past few months. Children were most often abused in a period of weeks (16%), although 3% of boys and 4% of girls claimed that it had gone on for a number of years.

The increasing burden of mental, emotional and behavioural problems and poor mental health for individuals, families, society and the economy calls for action to promote mental health and prevent behavioural problems (Jané-Llopis, Katschnig, McDaid and Wahlbeck, 2010). Strong advocacy during the last decade has led to increased funding for practice and research in these domains, and to nation-wide strategic support (e.g. policy, databases, training and etc.) in many countries. Numerous interventions aimed at reducing or delaying the onset of a wide range of mentally and behaviourally related problems have been developed (Brown, Berndt, Brinales, Zong and Bhagwat, 2000). There has been an expansion in the development of theoretically sound interventions that have been tested in controlled efficacy trials, which has led to a science of mental health promotion and prevention (Mrazek and Haggerty, 2010; O'Connell, Boat and Warner, 2009; Kellam and Langevin, 2003).

1.1.1. Field of Mental Health Promotion and Prevention Science

From 1970s, health and mental health care became more and more promotion and prevention focused, and societies were developing towards promotion and prevention oriented societies. In the early nineties of the last century, Mrazek and Haggerty (1994) have developed a worldly recognized and accepted spectrum of interventions for mental, emotional and behavioural problems and disorders (Figure 1). The spectrum included three main approaches to those problems: prevention, treatment and maintenance. The model shows three levels of prevention interventions – universal, selective, and indicated - which are followed by

treatment interventions that include identification of individuals with mental, emotional and behavioural problems, and provision of a standard treatment to these individuals. According to the authors, the last stage of the intervention spectrum is the one focused on long-term treatment and after care interventions.

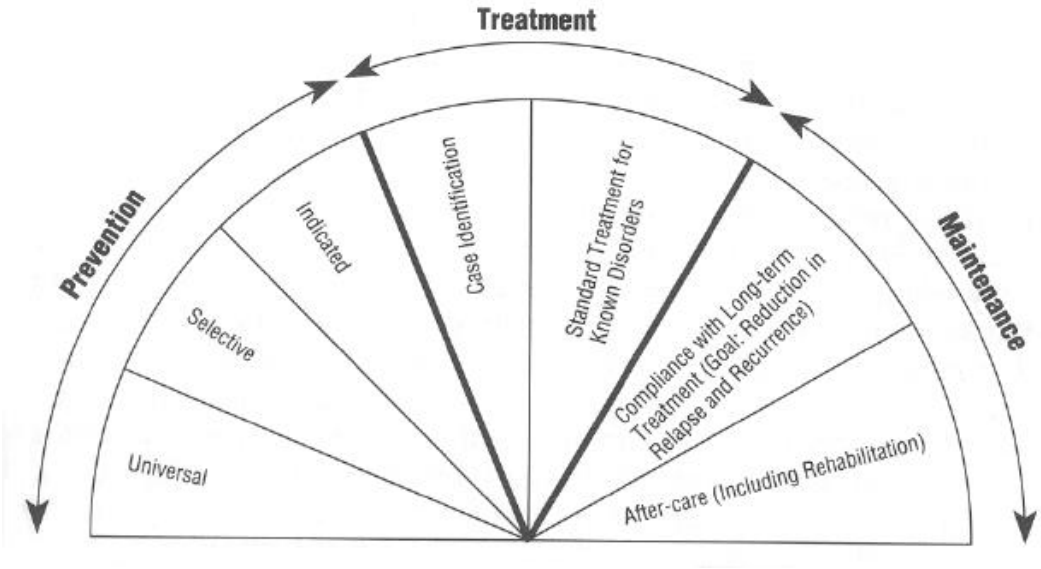


Figure 1.

Spectrum of Interventions for Mental, Emotional and Behavioural Problems and Disorders (Mrazek and Haggerty, 1994)

As noticeable from the model proposed by Mrazek and Haggert (1994), promotion interventions were not recognized as a part of this model. As scientists became more and more aware of the importance of mental health promotion and prevention activities, the model was upgraded (O'Connell, Boat and Warner, 2009; Barry, 2001) in a way that mental health promotion interventions were also incorporated into the spectrum (Figure 2).

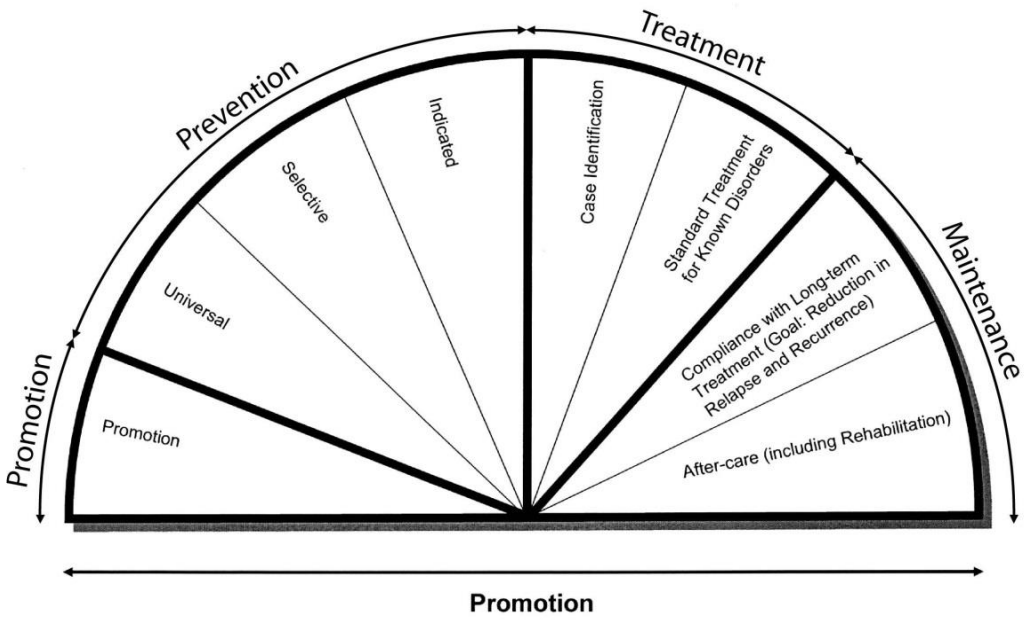


Figure 2.

Expanded Spectrum of Interventions for Mental, Emotional and Behavioural Problems and Disorders (O'Connell, Boat and Warner, 2009; Barry, 2001)

The expanded model starts with mental health promotion interventions, continues with prevention interventions, followed by treatment and maintenance approach. This direction of interventions continuum implies the level of risk of an individual or a group. It is important to note that the model presents an idealized conceptualization. In reality, the boundaries between various sectors of the model are blurred. In practice, it may be difficult to classify an intervention as pure promotion, prevention or early intervention as many interventions combine elements of all of these.

There is increasing evidence that promotion of positive aspects of mental health is an important approach to reducing mental, emotional, behavioural and related problems and disorders (Catalano, Berglund, Ryan, Lonczak and Hawkins 2004; Catalano, Hawkins, Berglund, Pollard and Arthur, 2002).

Mental health promotion implies the creation of individual, social and environmental conditions that empower and enable optimal health and development (Jané-Llopis et al., 2010). Such initiatives involve individuals in the process of achieving positive mental health and enhancing quality of life. In general, mental health promotion is any action

taken to maximize mental health and wellbeing among populations and individuals (Commonwealth Department of Health and Aged Care, 2000).

It focuses on improving environments (social, physical and economic) that affect mental health and enhancing the coping capacity of communities as well as individuals (Wood and Wise, 1997). These environmental determinants are broadly based in all aspects of life and, as a consequence, the gains from mental health promotion activities generalize to improvements in physical health as well as productivity in the school, home and workplace. It is important to recognize that mental health promotion is 'a process': a process aimed at giving power, knowledge, skills and necessary resources to individuals, families, the community and whole populations (European Commission, 1999). *Mental health promotion interventions* aim to enhance individuals' ability to achieve developmentally appropriate tasks (competence) and a positive sense of self-esteem, mastery, well-being, and social inclusion, and strengthen their ability to cope with adversity (SAMHSA, 2007a; WHO, 2004; Hosman and Jane-Llopis, 1999, according to O'Connell, Boat and Warner, 2009).

Prevention may generally be defined as the process focused on decreasing the incidence and prevalence of mental, emotional and behavioural disorders in children and youth (Prevention Term Glossary, according to Bašić, 2009). Mrazek and Haggerty (1994, p. 23.) define prevention as "interventions that occur before the initial onset of a disorder to prevent the development of disorder". The prevention of mental, emotional and behavioural problems relies on reducing the risk factors for problem development, as well as enhancing the protective factors that promote mental health.

The level of risk of an individual to develop a mental, emotional or behavioural problem or disorder can be determined by the exposure and vulnerability to risk factors and the presence and strength of protective factors associated with the development of such problems or disorders. Effective prevention requires an understanding of the risk and protective factors for mental health, identification of the groups and individuals who can potentially benefit from interventions, and the development, dissemination and implementation of effective interventions (Commonwealth Department of Health and Aged Care, 2000).

Risk factors are "characteristics, variables, or hazards that, if present for a given individual, make it more likely that this individual, rather than someone selected at random from the general population, will develop a disorder" (Mrazek and Haggerty,

1994, p. 127). They exist before the onset of a mental, emotional or behavioural problem or disorder, and may be time-limited or continue over time. Risk factors can derive from the individual, the family, the community, institutions or the general environment and wider society. They can play a causal role or be a marker for a problem.

Like risk factors, protective factors derive from all domains of life, from the individual, family, community and wider environment (Mrazek and Haggerty, 1994). Some protective factors are internal, such as a person's temperament and level of intelligence, while others are external, related to social, economic and environmental supports.

Protective factors enable individuals to maintain their emotional and social wellbeing and cope with life experiences and adversity (Commonwealth Department of Health and Aged Care, 2000).

The relationship between risk and protective factors is complex. It is not simply the presence of risk and protective factors, but their interaction and the accumulation of factors over time that affects the development of mental, emotional and behavioural problems and disorders. As it is presented in figures 1 and 2, prevention interventions are divided into universal, selective and indicated prevention interventions.

Universal preventive interventions are targeted to the general public or a whole population that has not been identified on the basis of individual risk. Universal prevention includes strategies based on evidence that it is likely to provide some benefit to all (reduce the probability of disorder), which clearly outweighs the costs and risks of negative consequences (Mrazek and Haggerty, 1994).

Examples of that kind of interventions are school-based programs offered to all children to teach social and emotional skills or to avoid substance abuse or programs offered to all parents of sixth graders to provide them with skills to communicate to their children about resisting substance use.

Selective preventive interventions are targeted to individuals or a population subgroup whose risk of developing mental, emotional or behavioural problems is significantly higher than average (Mrazek and Haggerty, 1994). The risk may be imminent or it may be a lifetime risk. Risk groups may be identified on the basis of biological,

psychological, or social risk factors that are known to be associated with the onset of a mental, emotional, or behavioural problem.

Examples of selective prevention intervention are programs offered to children exposed to risk factors, such as parental divorce, parental mental illness, death of a close relative, or abuse, to reduce risk for adverse mental, emotional, and behavioural outcomes.

Indicated preventive interventions are targeted to high-risk individuals who are identified as having minimal but detectable signs or symptoms foreshadowing mental, emotional, or behavioural disorder, or biological markers indicating predisposition for such a disorder, but who do not meet diagnostic levels at the current time (Mrazek and Haggerty, 1994). Indicated interventions might be reasonable even if intervention costs are high and even if the intervention entails some risk.

Indicated preventive interventions are, for example, interventions for children with early problems of aggression or elevated symptoms of depression or anxiety.

It is very important to explain that for the past two decades various prevention researchers have argued for a synthesis of promotion and prevention approaches (Greenberg et al., 2003; Catalano et al., 2002; Commonwealth Department of Health and Aged Care, 2000; Hosman, 1997). The reason is that in practice there is already considerable overlap between prevention and promotion. As it was already mentioned, sometimes it is difficult to classify an intervention as pure promotion or prevention, as many interventions combine elements of both approaches. Meta-analytic and qualitative reviews of preventive intervention studies demonstrate that many psychosocial prevention programs involve the promotion of child competencies or healthy functioning of families, schools, or communities (Greenberg, Domitrovich and Bumbarger, 2000; Durlak and Wells, 1998). For example, reviews of mental health promotion programs for children and young people cite many programs that have been demonstrated both to reduce problems and to increase positive aspects of development (Catalano et al., 2004). Catalano and colleagues (2004) concluded that several youth development programs that were effective in building positive development in such areas as social, emotional, and cognitive competence as well as self-determination and efficacy were also effective in reducing a range of problem behaviours, such as alcohol and drug use, violence, and aggression.

In 1997, Hosman presented a model that links promotion and prevention as functionally related fields, with positive mental health seen as a cluster of individual competencies that is conditional to a wide range of positive outcomes, including the prevention of mental disorders (Hosman, 1997). The author stresses that adopting a more inclusive approach may also be less stigmatizing for young people and their families, and increase participation in relevant programs, as the focus shifts from avoiding the possibility of disorder toward helping young people realize their potential. Based on these arguments, mental health promotion and prevention of mental, emotional and behavioural problems are more and more recognized as a unique scientific field.

As it is shown in Figure 3, mental health promotion and prevention science represent an interdisciplinary discipline that uses methods and knowledge of several established disciplines – human development, psychopathology, education, epidemiology, public health, and psychology. Since it is a rather new scientific field, it is possible that in the future mental health promotion and prevention scientists will also use methods and knowledge of some other disciplines.

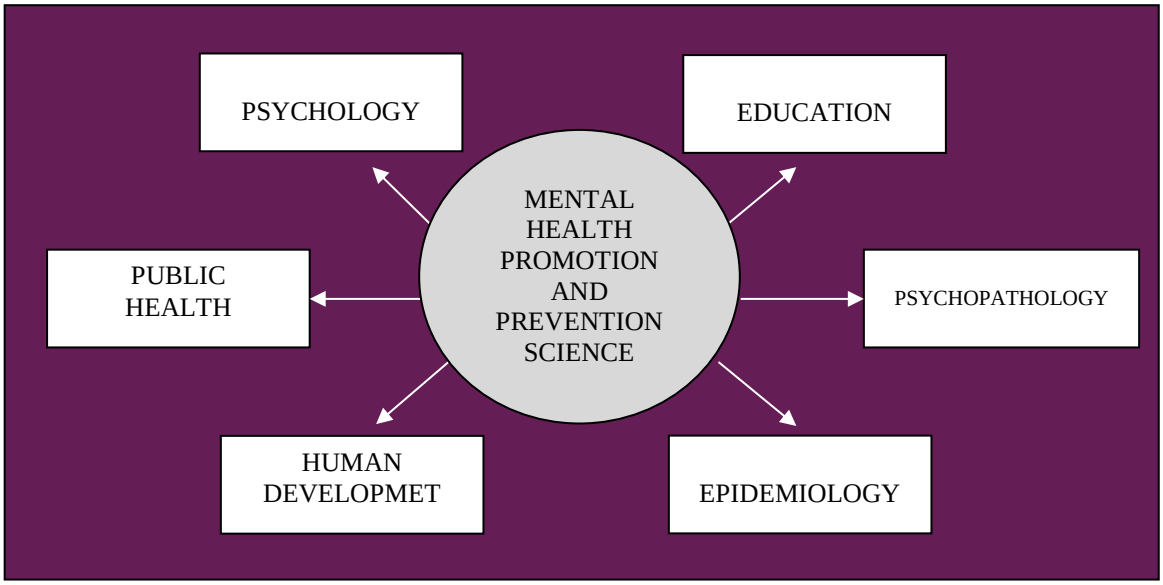


Figure 3.

Interdisciplinarity of Mental Health Promotion and Prevention Science

The focus of mental health promotion and prevention studies has changed over the years (Figure 4). Mrazek and Haggerty (1994) stated that studies in this field follow a trajectory of research phases usual in biomedical and social sciences.

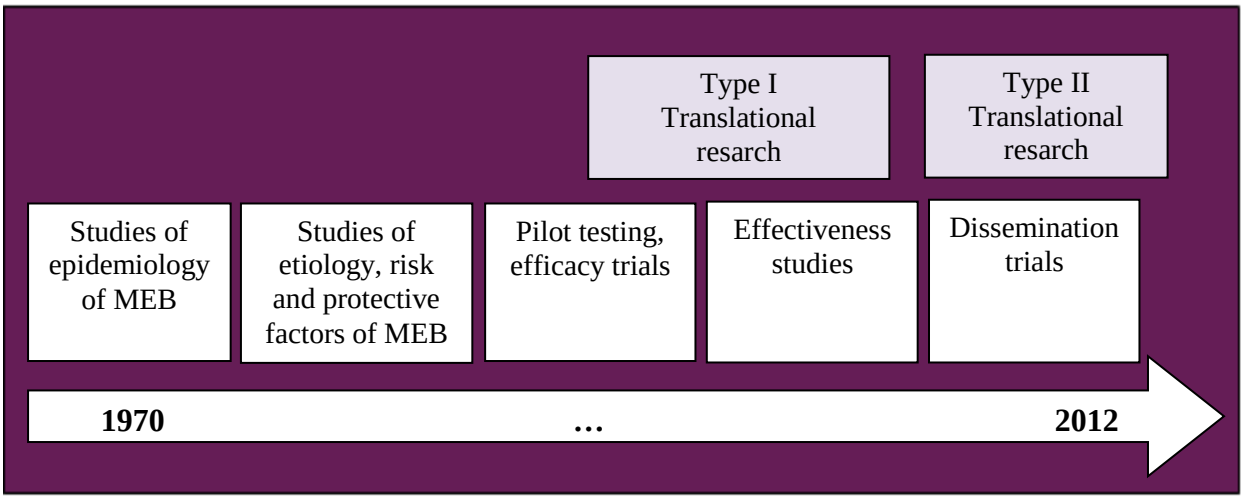


Figure 4.

Phases of Research Studies in the Field of Mental Health Promotion and Prevention

The first research studies in this field were mostly focused on the epidemiology of phenomena, assessing the prevalence and incidence of mental, emotional and behavioural problems. Later on, mental health promotion and prevention scientists conducted studies aimed to detect etiological, risk and protective factors related with specific mental, emotional or behavioural problems. Gathered knowledge was subsequently used for the development of specific mental health promotion and prevention interventions. Since 1990s, mental health promotion and prevention researchers have increasingly focused on conducting controlled outcome studies and assessing effects of mental health promotion and prevention interventions.

It is significant to mention the presence of translational studies in the field of mental health promotion and prevention. There are two types of translational research – Type I and Type II. *Type I* translational research applies discoveries generated through basic science research to the development and preliminary testing of mental health promotion and prevention interventions (Spoth et al., 2008). Currently, there is a strong movement towards conducting *Type II* translational research in this scientific field. Spoth and colleagues (2008) describe this kind of research as that "which examines a broad range of factors necessary for successful adoption, implementation, and sustainability of evidence-based interventions across diverse populations, through the application of naturalistic methods and experimental trials". Type II translational research contributes to the process of translating mental health promotion and prevention interventions into an effective practice. This doctoral study is an example of Type II translational research, however studying the determinants of effectiveness

can also be considered as a contribution to fundamental knowledge about processes of planned change.

1.1.2. Development and State of the Art of Mental Health Promotion and Prevention in Croatia

The origin of promotion of mental health and prevention within Croatia began in the early nineties of the last century. Before that period, scientists in Croatia studied criminology, juvenile delinquency and behavioural disorders of children and youth mostly from the aspect of treatment of mentioned phenomena. The shift away from treatment towards a prevention approach began with the efforts of scientists from the University of Zagreb's Faculty of Education and Rehabilitation Sciences and the project Integral Method (Bašić, Koller Trbović, Žižak, 1993). This project was developed for parents of kindergarten children and pre-school teachers, and designed to establish a foundation of universal prevention and positive development for children and youth in Croatia. The Integral Method project was one of the first initiatives to bring together a diverse group of researchers, experts and practitioners to focus on the social and emotional development of Croatian children.

Reviewing the history of prevention in Croatia, it is also important to recognize the activities of the National Council for Children as well as the Government Commission for Prevention of Behavioural Disorders of Children and Youth in developing mental health promotion and prevention practice in Croatia. These two boards gathered influential members of government and researchers across academic disciplines, who researched children and youth. The Government Commission for Prevention of Behavioural Disorders of Children and Youth was constituted in 1997 (with professor Josipa Bašić as the first president) and was composed of representatives of different ministries, the State Attorney's Office, judiciary practice, the Institute for Family, Motherhood and Youth, as well as scientists researching family, children and youth.

Prevention science and mental health promotion in Croatia were also strongly influenced by the public health sector, which has a long and prosperous tradition in this geographical area. Many mental health promotion and prevention activities in local communities were conducted within the Croatian Healthy Cities Network (http://www.zdravi-gradovi.com.hr/), which was established in 1993.

Also, systematic education of future experts in the field of prevention of behavioural problems and promotion of mental health at academic level is considered as significant for

further development of this field in Croatia. The Department of Behavioural Disorders of University of Zagreb, Faculty of Education and Rehabilitation Sciences, provides continuous education of social pedagogues in the field of prevention of behavioural problems. Prevention experts at the Faculty are delivering prevention courses at the undergraduate and graduate levels, preparing future social pedagogues for implementation of science-based prevention practice. In 2005, prevention scientists from the Faculty led by Josipa Bašić, regular professor of prevention at the Department of Behavioural Disorders, have established the Centre for Prevention Research (http://www.erf.unizg.hr/CPI/CentarZaPrevencijskaIstrazivanja.html). The goal of the Centre is to develop and improve mental health promotion and prevention science and practice in Croatia. Researchers from the Centre are focused on conducting research studies in the field of mental health promotion and prevention, implementation of evidence-based programs at the local and national levels, and collaboration with other research centres in Croatia and from abroad.

Based on more than a decade of investments in the field of mental health promotion and prevention, in 2007/2008 the University of Zagreb's Faculty of Education and Rehabilitation Sciences initiated an international doctoral program "Prevention science: prevention of mental and behavioural disorders and promotion of mental health". The doctoral program was initiated by professor Josipa Bašić. This comprehensive doctoral program targeting prevention and promotion in mental health can be considered as rather unique in Europe, as so far no equivalent doctoral program exists on the Europe. In the program, prevention experts from Croatia and from abroad are empowering future prevention scientists in Croatia with the most recent knowledge and skills in this field.

Over the last decade, Croatian prevention scientists established *collaboration with scientists and centres for prevention science worldwide*, especially with the Prevention Research Centre of Penn State University (USA) and the Prevention Research Centre of the Radboud University Nijmegen and Maastricht University (the Netherlands). They also started a collaborative relationship with prevention scientists from Scuola Universitaria Professionale della Svizzera Italiana (Switzerland), Verwey-Jonker Institute (the Netherlands), and other relevant institutions from abroad. Collaboration with foreign mental health promotion and prevention experts enables an exchange of knowledge and expertise and it encourages international projects that have a significant influence on mental health promotion and prevention science in Croatia.

It is important to emphasize that several world renowned prevention *"*model programs*"* have been adopted, implemented and researched in Croatia until this very moment. Model

programs are well-implemented, well-evaluated programs whose developers are willing to disseminate them and to provide training and technical assistance to practitioners who wish to adapt them. All model programs implemented in Croatia are American, which is a result of several years of collaboration with scientists from the Society for Prevention Research (http://www.preventionresearch.org/). One of the model programs, the Communities that Care model has been implemented in the Croatian County of Istria since 2002 (Bašić, Ferić Šlehan, Kranželić Tavra, 2007a and 2007b; Bašić, Grozić-Živolić, 2010). The Northland project was implemented in the city of Split during 2002 within the context of the international Healthy Cities network, the Life Skills Training in the city of Rijeka from 2005 and the PATHS model program on socio-emotional learning in the County of Istria, Zagreb and Rijeka since 2008 (Bašić, Grozić-Živolić, 2010).

Although a review of the history of prevention in Croatia suggests that there are a lot of initiatives going on in such a small country, it is still evident that science-based prevention practice in Croatia is still in its roots and is facing a lot of bottlenecks (Bašić, 2009). There are several general dimensions of existing problems in this field:

- lack of a science-based approach to mental health promotion and prevention,
- lack of coordination between institutions, stakeholders and activities concerned with mental health promotion and prevention, and
- lack of consistent implementation of existing law regulations and policies concerning the well-being of children, youth and families.

Bašić (2009) has emphasized a strong need for using scientific knowledge and a systematic approach in organizing, developing, implementing and evaluating prevention interventions and initiatives in Croatia. In general, there is a lack of evidence-based programs widespread across local communities in Croatia. Local and national authorities often do not demand any evidence of quality assurance or evidence of program effectiveness. Croatian mental health promotion and prevention programs are sometimes run by local practitioners non-trained in prevention and not familiar with a science-based approach. Prevention programs are rarely theory-based and their outcomes are often not evaluated (Bašić, 2009, Bašić, Mihić and Novak, 2010). Another problem is a lack of attunement of the interventions to the specific needs of the targeted populations.

Also, coordination between institutions, stakeholders and activities concerned with mental health promotion and prevention is not strong enough. Croatia doesn't have an active

"umbrella" institution that takes care of policies and interventions of promotion and prevention (Bašić, 2009; Bašić, Mihić, Novak, 2010). Related to this, there exists no national database of evidence-based programs as in several other countries (e.g. US, the Netherlands, Norway). Deficiency in national coordination of prevention in Croatia has resulted with prevention initiatives of some national departments such as the Ministry of Science, Education and Sports, the Ministry of Health and Social Welfare or initiatives of local authorities. Various mental health promotion and prevention interventions get remarkable financial support from local and state agencies, but they are treated as single and incidental actions, rather than a part of a more comprehensive strategy.

The problem of coordination is closely connected with non-consistent implementation of regulations and policies concerning the well-being of children, youth and families. Even though high quality regulations and policies exist, such as the National strategy for prevention of behavioural problems of children and youth for 2009-2012 (http://www.mobms.hr/media/17218/strategija%20pup%20izvje%C5%A1ce.pdf), they are not implemented very effectively and they have a narrow reach. One of the reasons for this condition is surely connected with a lack of an infrastructure for mental health promotion and prevention. Other reasons could be that there is a need for investment in knowledge and skills of the Strategy carriers, the division of tasks between the various institutions in this field is not transparent, there is a lack of supervision and that the consequences for not achieving desirable goals are not considered.

If we consider all of the mentioned problems, there are some possible approaches which could improve the state of this field in Croatia. It is evident that there is a *strong need for investing in the knowledge of developers and deliverers of the interventions* focused on mental health promotion and prevention. Intervention developers and deliverers should be much more aware of the advantages of science-based practice and of continuously being trained to incorporate science-based principles into their practice. A systematic investment into knowledge of mental health promotion and prevention intervention developers and providers is the first step in improving the effectiveness of this field. At the same time, there is a *strong need for promotion of evaluation and implementation of evaluation studies* in Croatia. Assessment of interventions' impact and effectiveness will lead to detection and dissemination of the best practices on the one side, and improvement of current interventions on the other.

1.2.Effectiveness of Mental Health Promotion and Prevention Interventions

There is an increasing focus on the effectiveness of mental health promotion and prevention interventions worldwide.

> *Effectiveness* is concerned with the intervention's ability to actually affect the causal pathway of some phenomenon (McQueen, 2007).

Growing evidence suggests that with high-quality interventions, a wide variety of behavioural and mental health problems can be reduced, including violence and delinquency (Botvin, Grifin and Nichols, 2006), tobacco and alcohol use (Tobler and Stratton, 1997), risk sexual behaviour (Kirby et al. 1994) and other emotional problems (Hawkins, Kosterman, Catalano, Hill and Abbott, 2005; Hosman et al., 2004). Randomized control studies provided evidence that many interventions are effective in promoting mental health and preventing behavioural problems. Effect sizes for these interventions average between .20 and .30, but can range from .10 for individual programs to .60 for preventive programs which combine several interventions (Stice and Shaw, 2004; Jane-Llopis, Hosman, Jenkins and Anderson, 2003; Brown et al., 2000; Tobler and Stratton, 1997). Stice and colleagues (2009) even report on the effect size of 0.68 for interventions focused on the prevention of depression symptoms. Cuijpers and colleagues (2008) report on the basis of their meta-analysis that prevention of depression interventions show an incidence rate ratio of 0.78, which is equal to a prevention of 22% of new cases of depression.

Wilson and Lipsey (2007) stress that school-based violence prevention programs have effects that would lead to a 25% to 33% reduction in the base rate of aggressive problems in an average school. For instance, the Good Behaviour Game prevention program reduced disruptive and aggressive behaviour and reduced the likelihood that persistently highly aggressive boys would receive a diagnosis of antisocial personality disorder as a young adult (Petras et al., 2008). The same program also significantly reduced the risk of illicit drug abuse or dependence disorder at age 19-21 (Kellam et al., 2008). Horowitz and Garber (2006) have found that interventions to prevent depression can both effectively reduce the number of new cases of depression in adolescents and reduce depressive symptomatology among children and youth. Life Skills Training program significantly reduced drug and polydrug (tobacco, alcohol, and marijuana) use three years after the program (Botvin et al., 2000). Regarding the prevention of risky sexual behaviour, the Seattle Social Development Project, a combined parent and teacher training intervention, demonstrated fewer sexual partners, greater condom use, and (among the girls) fewer pregnancies and births by age 21 (Hawkins et al., 2005).

While some interventions are highly effective, others are only moderately effective or have little or no impact leaving much room for intervention improvement (Hosman, Llopis and Saxena, 2004). For that reason, evidence on intervention effectiveness is crucial. Mittelmark and colleagues (2007) emphasize that studies of effectiveness in the field of mental health promotion and prevention are needed for a number of reasons:

- to enable the development of the effective interventions,
- to identify the best practices,
- to demonstrate to the decision makers that mental health promotion and prevention works and offers an effective public strategy,
- to support decisions in policy development and funding allocation,
- to show the wider community the benefits of promotion and prevention actions and
- to advocate for further investments in development of interventions of mental health promotion and prevention of mental and behavioural problems.

Stern (2005, according to Shaw, Green and Melvin, 2007) distinguishes the following five purposes of the interventions' evaluation, providing a view of how evaluation can have an impact on political decisions for planning, learning, developing, and termination of a program:

(1) Planning/efficiency – ensuring that there is a justification for a policy/program and that resources are efficiently used,
(2) Accountability - demonstrating how far a program has achieved its objectives and how well it has used its resources,
(3) Implementation - improving the performance of programs and the quality of how they are delivered and managed,
(4) Knowledge production - increasing our understanding of what works in what circumstances and how different measures and interventions can be made more effective, and
(5) Institutional and community strengthening – improving and developing capacity among program participants and their networks and institutions.

1.2.1. Evaluation of Effectiveness of Mental Health Promotion and Prevention Interventions

"Evaluation", or at least its root word "value", finds its origin in the Old French *value* and *valoir* and the Latin *valére*, which had the sense of "to be worth (something)" and "to work out the value of something" (Shaw et al., 2007). A key way in which definitions of evaluation differ is in terms of the components they include. Some definitions of evaluation focus on the general function evaluation serves, while some other specify the purpose that evaluation has.

> The *evaluation process* involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of the evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation (Schriven, 1991, according to Shaw et al., 2007).

> *Program evaluation* is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgements about the program, improve program effectiveness, and/or inform decisions about future programming (Patton, 1997, according to Shaw et al., 2007).

The systematic evaluation of social programs first became a commonplace in education and public health. From 1960s, social scientists from all over the world showed an interest in assessing the effectiveness of delinquency prevention programs, psychotherapeutic treatment, educational activities and numerous other initiatives (Suchman, 1967, Freeman et al., 1980, Levin et al., 1981, according to Rossi, Lipsey and Freeman, 2004). In the early 1970s, evaluation research emerged as a distinct speciality field in social science (Riecken and Boruch, 1974; Bernstein and Freeman, 1975, according to Rossi et al., 2004) within which different evaluation theories were developed.

> Alkin (2004) describes *evaluation theories* as a set of rules, prescriptions, prohibitions, and guiding frameworks that specify what a good or proper evaluation is, and how evaluation should be done. They are thus theories of evaluation practice that address questions like how to understand the nature of what we evaluate, how to assign value to programs and their performance, how to construct knowledge, and how to use the knowledge generated by evaluation (Shadish, 1998).

The concept of evaluation entails, on the one hand, a description of the performance of the entity being evaluated and, on the other, some standards or criteria for judging that performance. Evaluation process can be focused on comprehensive systems, initiatives, policies or systems on national or local level. But also, increasingly usual are evaluations of some specific interventions or programs.

> *Program evaluation* in general is the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions (Rossi et al., 2004).

Rossi and colleagues (2004) explain that program evaluation generally involves assessing one or more of five domains: (1) the need for the program, (2) the program's design, (3) its implementation and service delivery, (4) its impact or outcomes, and (5) its efficiency**.** The same authors stress that there is a certain hierarchy in the process of conducting these five evaluation domains presented in Figure 5.

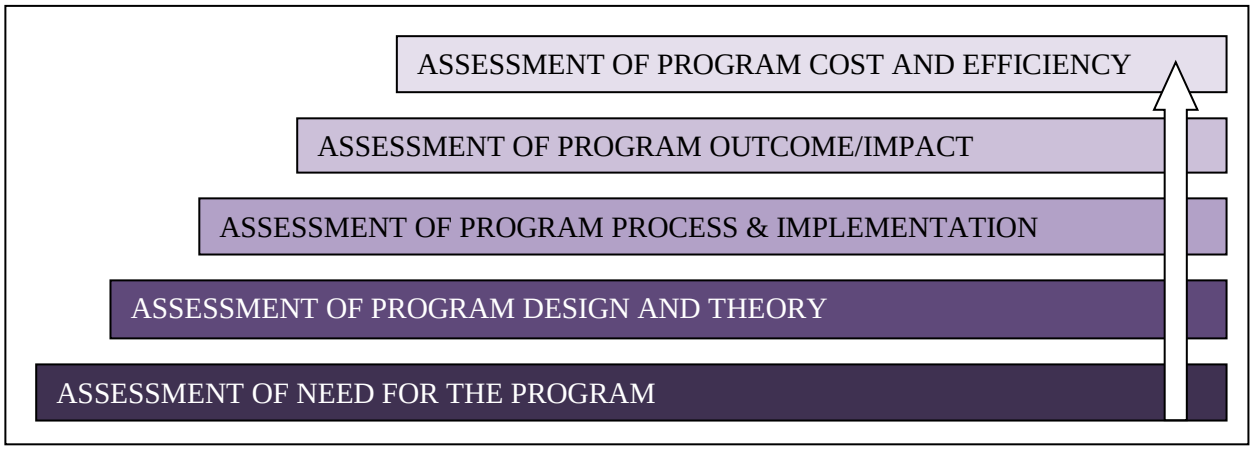


Figure 5.

The Evaluation Hierarchy (adapted according to Rossi et al., 2004)

Rossi and colleagues (2004) emphasize that the foundational level of the evaluation hierarchy relates to the assessment of the need for the program. Only if we know that the social need is properly understood, that the program theory for addressing it is reasonable, and that the corresponding program activities and services are well implemented, then it may be meaningful to assess program outcomes. The same authors also describe that depending on

the stage of program development, different evaluation questions can be raised that are connected with specific evaluation functions (Table 1.1.).

Table 1.1.

Stages of Program Development and Related Evaluation Functions

(adapted according to Rossi et al., 2004)

STAGE OF PROGRAM DEVELOPMENT	QUESTION TO BE ASKED	EVALUATION FUNCTION
1. Assessment of social problems and needs	To what extent are the community needs and standards met?	Needs assessment; Problem description
2. Determination of goals	What must be done to meet those needs and standards?	Needs assessment; Service needs
3. Design of program alternatives	What services could be used to produce the desired changes?	Assessment of program logic or theory
4. Selection of alternative	Which of the possible program approaches is best?	Feasibility study; Formative evaluation
5. Program implementation	How should the program be put into operation?	Implementation assessment
6. Program operation	Is the program operating as planned?	Process evaluation; Program monitoring
7. Program outcomes	Is the program having the desired outcomes?	Outcome evaluation
8. Program efficiency	Are program's effects attained at a reasonable cost?	Cost-benefit analysis; Cost-effectiveness analysis

As it is visible from the table, impact of a program is measured through the process of impact evaluation. In addition to understanding the meaning of impact evaluation, it is very important to understand the meaning of outcome evaluation. These two terms are sometimes unreasonably used as synonyms, though they represent different concepts of evaluation. According to the Glossary of Evaluation Terms (http://pdf.usaid.gov/pdf_docs/PNADO820.pdf), *outcomes* are the more immediate and

tangible results of program activities that can be observed, monitored, measured, and evaluated in the short to midterm of a project and can be intended or unintended. An outcome indicator attempts to capture whether or not actual project results have brought tangible benefits to the targeted beneficiary groups. An *impact* is a high level result or effect that is caused by a project or program which can be intended or unintended and positive or negative. Impact indicators seek to demonstrate how the project has affected the big picture issues, problems, or challenges that the intervention was designed to ameliorate. Impacts are the broader changes that occur within the community, organization, society, or environment as a result of program outcomes (Rubin, 2004).

The ultimate goal of all social programs, including mental health promotion and prevention programs, is to affect a problem or social condition in beneficial ways. Rossi and colleagues (2004) explain that an *outcome* is the state of the target population or the social conditions that a program is expected to have changed. Various program stakeholders have their own understanding of what the program is supposed to accomplish and, correspondingly, what outcomes they expect it to affect. A program's intended outcomes are ordinarily identified in the program's impact theory. The same authors also stress that the most direct sources of information about these expected outcomes usually are the stated objectives, goals, and mission of the program. For the evaluator's purposes, an outcome description must indicate the pertinent characteristics, behaviour, or conditions that the program is expected to change. Rossi and colleagues (2004) described the difference between proximal and ultimate outcomes of a program. *Proximal outcomes* are those that the program services are expected to affect most directly and immediately. These proximal outcomes are usually attitudes, knowledge, awareness, skills, motivation, behavioural intentions, and other conditions that are susceptible to relatively direct influence by program's processes and services. Proximal outcomes are rarely ultimate outcomes the program intends to generate, but accomplishment of proximal outcomes will consequently lead to the realization of *ultimate outcomes* that are usually defined as some changes on a more public and broader level. Sensitive and valid measurement of those outcomes is technically challenging but essential for assessing a program's success. The challenge for evaluators, then, is to assess not only the outcomes that actually obtain but also the degree to which any change in outcomes is attributable to the program itself.

It is important to explain that usually two types of research contribute to determining the outcomes and impact of interventions: efficacy and effectiveness studies (Commonwealth Department of Health and Aged Care, 2000). Aveline (1997) describes *efficacy studies*, as

usually randomized controlled trials which are undertaken under experimental or 'controlled' conditions to develop and refine strategies. They provide important, but limited, information regarding the outcomes of interventions under ideal circumstances. They do not, however, yield information related to all the outcomes of interest. *Effectiveness studies* test the 'real world' impact of interventions that have been shown to be efficacious, and implemented under normal conditions in the 'daily' routine of practice. These studies are imperative to determine the generalizability of controlled studies in the real world, because interventions conducted under highly controlled conditions may not translate well into the less controlled environment of normal practice, which represents the real world.

Rossi and colleagues (2004) illustrated and explained relations between crucial elements of the programs' effect concept which can be helpful in understanding a logic of programs' efficacy and effectiveness studies (Figure 6).

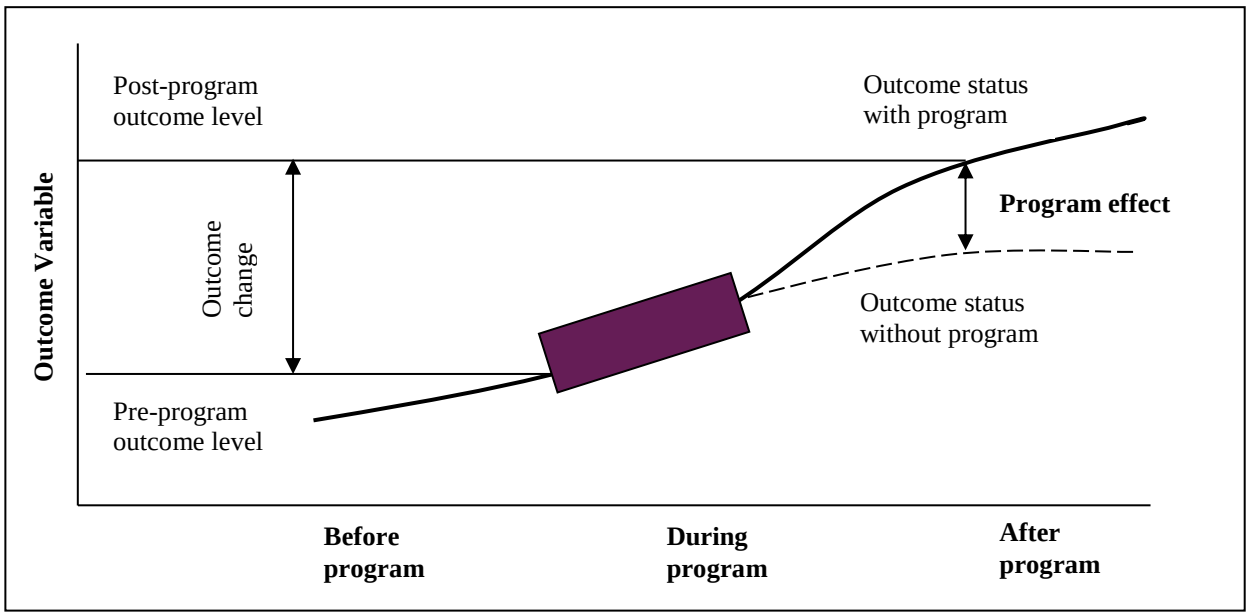


Figure 6.

Program Effect (adapted according to Rossi et al., 2004)

Authors explain that the *outcome level* is the status of an outcome at some point in time. *Outcome change* is the difference between outcome levels at different points in time. *Program effect* is that portion of an outcome change that can be attributed uniquely to a program as opposed to the influence of some other factor. According to the same authors, assessing the degree to which a program produces these effects is a core function of the outcome evaluation.

In conclusion, it is important to mention that in practice, program evaluation presents many challenges to the evaluator. As Weiss (1972, according to Rossi et al., 2004) once observed, social programs, which also include mental health promotion and prevention programs, are inherently inhospitable environments for research purposes. The challenges to the evaluator are to match the research procedures with the evaluation questions and circumstances as well as possible and, whatever procedures are used, to apply them at the highest possible standard feasible to those questions and circumstances. Cronbach (1982) also emphasized that the evaluation in social sciences should meet high standards of scientific research and at the same time be dedicated to serving the information needs of program decision makers. He noticed that evaluation should be dedicated to providing the maximally useful information that the political circumstances, program constraints, and available resources allow. The role of the evaluation is to provide answers to questions about the program that will be useful and will actually be used. This form is fundamental for evaluation – its purpose is to inform action.

1.3. Quality Assessment and Quality Assurance in Mental Health Promotion and Prevention

As the pool of evidence-based interventions has grown, and political and fiscal support for using these types of strategies has increased, so have the challenges facing mental health promotion and prevention science, practice, and policy. Even in large developed countries, there is variation in the quality of promotion and prevention practices (Jané-Llopis and Anderson, 2005).

Concepts of quality assessment and quality assurance are closely related to the concept of effectiveness. Kahan and Goodstart (1999) promote the interesting idea that the tradition of quality assessment and assurance popular within the industry and health care can be very successfully transferred to the field of mental health promotion and prevention also. The authors believe that through implementation of quality assessment and quality assurance principles in this field, effectiveness of mental health promotion and prevention programs could be increased.

Quality is a term which is often used in a variety of ways even though it is not precisely defined. Mullen and colleagues (1992, according to Kok, 1997) elaborate the term "quality" very broadly. They emphasize that the quality of the intervention is the only and the most important determinant of the intervention's effectiveness. According to these authors, quality

of the intervention can be measured as the level of application of five principles during an intervention's development process:

(1) Level of intervention's relevance - tailoring of the program to knowledge, beliefs, circumstances, and prior experience of the learner, as assessed by pretesting or other means,
(2) Individualization – the provision of opportunities for learners to have personal questions answered or instructions paced according to their individual progress,
(3) Feedback – information given to the learner regarding the extent to which learning is being accomplished,
(4) Reinforcement - any component of the intervention that is designed to reward the behaviour (other than feedback) after the behaviour has been enacted (e.g. social support) and
(5) Facilitation - the provision of means for the learner to take action and/or to reduce barriers to action.

A more precise and generally accepted explanation of the term quality is the one in which quality represents the level to which key "effect predictors" are incorporated into an intervention. Nation and colleagues (2003) stress that mental health promotion and prevention science has a sufficient knowledge base of the characteristics of effective prevention programming. This knowledge is needed for quality assessment and quality assurance of the intervention.

It is important to explain that processes of quality assessment and quality assurance are actually parallel processes. Assessment of the quality level of effect predictors within some specific intervention is the base for further improvement and assurance of interventions' quality. Measurement of the presence and quality level of "effect predictors" in the intervention is the *quality assessment* process. Ader and colleagues (2001) stress that the concept of *quality assurance* is a broader concept and that it encompasses methods for describing, measuring, evaluating and, when needed, taking measures aimed at the improvement of what, in a broad sense, is described as intervention's quality.

Speller, Evans and Head (1997) emphasize that in mental health promotion and prevention science, despite improvements in outcome research, not enough attention is given to quality assurance in order to maximize program effectiveness. The authors believe that this could be solved through the application of a Systematic Planning Approach (Bartholomew et

al., 2001; Kok et al., 1997) and effectiveness guidelines (Molleman, Peters, Hosman, and Kok, 2005a; Nation et al., 2003; Hosman and Engels, 1999) while developing and implementing mental health promotion and prevention interventions. In general, it means that interventions' developers and deliverers should be trained in understanding the science-based principles and theory of the intervention, and should be aware of the importance of incorporation of effect predictors during the intervention development and implementation process.

Hosman (2008) also stresses that available scientific knowledge from earlier successful and unsuccessful trials to prevent behavioural disorders and to promote mental health, offers a base for designing and implementing effective prevention programs and effective local, national and international prevention policies. He explains that this knowledge could become part of quality assurance strategies if it is used to guide the development of interventions, to assess the extent to which on-going interventions adhere to quality standards, and to identify targets for quality improvement of interventions that are already being implemented. It can also be used by funders to select future projects to be financed.

Hosman and Engels (1999), just like Molleman (2005), emphasize that there are various strategies which can be conducted to improve the effectiveness and quality of mental health promotion and prevention practices. The authors notice that one commonly used strategy is to develop and test prevention model programs, to make them widely available and to enhance their dissemination and implementation, supported by the instructions from program developers. The second approach is to focus on the research of determinants of the efficacy and effectiveness of programs or interventions and to translate such knowledge into generally applicable principles and guidelines for effect management in mental health promotion and prevention. Subsequently, local mental health promotion and prevention professionals need to be stimulated and educated on a large scale to apply such principles and guidelines to their mental health promotion and prevention practice. The authors stress that a combination of these strategies is required to achieve a significant improvement of the total quality and effectiveness of promotion and prevention practices at the local or national level.

It is reasonable to assume that systematic investment in the knowledge of professionals about mental health promotion and their skills regarding effect predictors could be an effective approach in improving mental health promotion and prevention quality.

1.3.1. Effect predictors in Mental Health Promotion and Prevention

Outcome and impact research provides the knowledge necessary to identify predictors of efficacy and effectiveness in mental health promotion and prevention programs. Nation and colleagues (2003) have found, just as Dryfoos (1998), that there is substantial overlap in the principles of effective programs across mental health promotion and prevention domains that allow us to identify general principles of the effectiveness. Determinants of an intervention's impact or effect are referred to as "effect predictors" or "effect moderators" (Hosman and Engels, 1999; Raphael, 1999; Hosman, 1994). Table 1.2. represents effect predictors detected within different research studies in the field of mental health promotion and prevention.

Table 1.2.

An Overview of Effect Predictors Detected by Different Authors

Study Author/s and Year	Effect Predictors
Stice, Gau, Presnell and Shaw (2007)	- *fit between the program and the population it targets* - *characteristics of the intervention itself (e.g., duration, methods, socio-cultural relevance)*
Bartholomew et al. (2001)	- *systematic application of available theoretical and empirical knowledge during processes of intervention development and implementation*
Kok et al. (1997); Tobler and Stratton (1997); Brown et al. (2000); Jane-Llopis and Barry (2005)	- *clear goals and objectives* - *theoretical basis of the program*
Jane-Llopis and Barry (2005)	- *theoretical basis of the program, clear goals and objectives, high quality evaluation and research methods, infrastructural support from management, program fidelity and transferability to different countries and cultures* - *high quality implementation, training and supervision of program providers, high participation in program sessions*

Durlak et al. (2011) **Dryfoos (1990)**	*- quality of implementation process* *- provision of intense individualized attention, multilevel interventions, early identification of a problem, training based on the skills development, engagement of peers and parent in the intervention*
Nation and colleagues (2003)	*- comprehensiveness, various teaching methods, sufficient dosage, theoretical basis, opportunities for positive relationships, appropriate timing, socio-cultural relevance, outcomes evaluation, well-trained staff*
Ader and colleagues (2001)	*- quality of the program's structure - goals, target group, design, responsibility, resources, and organization* *- quality of the program's process – network, commitment, exposure, participation* *- quality of the program's outcomes – behavioural changes, environmental changes, epidemiological changes and maintenance*
Tobler and Stratton (1997)	*- quality of research design (evaluation)*
Nation et al. (2003); Jane-Llopis, Hosman, Jenkins et al. (2003)	*- variety of intervention methods* *- appropriate timing*

Stice and colleagues (2007) stress that crucial effect predictors in mental health promotion and prevention are *characteristics of the intervention itself* (e.g., duration, methods, socio-cultural relevance) and *the fit between the program and the population it targets*. Bartholomew and colleagues (2001) noticed that the potential effect of the intervention could be much higher when in the processes of intervention development and implementation available theoretical and empirical knowledge is systematically applied. Programs that have *clear goals and objectives*, and that are *theory-based* both in terms of targeted risk and protective factors and the mechanisms of change used in the intervention program have a more positive impact (Jane-Llopis and Barry, 2005; Brown et al., 2000; Kok et al., 1997; Tobler and Stratton, 1997).

High *quality of implementation* is found to be a core effect predictor, associated with positive intervention outcomes (Durlak et al., 2011). Jane-Llopis and Barry (2005) stress that high quality implementation, including *training and supervision of program providers* and *high participation in the program sessions* predicted higher program effectiveness. Those authors gave a systematic review of the crucial factors identified in determining program success – *theoretical basis of the program, clear goals and objectives, program provider training and support, evaluation* and *high quality research methods, infrastructural support from management, program fidelity* and *transferability* to different countries and cultures. Also, comprehensive programs that utilize a *variety of methods* and that are *delivered at the appropriate time* are more successful (Nation et al., 2003; Jane-Llopis, Hosman, Jenkins et al., 2003). Tobler and Stratton (1997) also found that programs rating higher in the quality of research design were significantly more effective than programs that rated lower in *quality of the program evaluation*.

Ader and colleagues (2001) have detected 14 quality indicators that have proved to be necessary and important in mental health promotion and prevention, and need to be clarified during interventions' development. They include:

(1) *Indicators that refer to the program's structure* - goals, target group, design, responsibility, resources, and organization,
(2) *Indicators that refer to the program's process* – network, commitment, exposure, participation, and
(3) *Indicators of the program's outcomes* – behavioural changes, environmental changes, epidemiological changes and maintenance.

Dryfoos (1990) reviewed over 100 prevention programs related to substance abuse, teen pregnancy, school dropout and juvenile delinquency. Her review yielded several key characteristics associated with successful programs such as *provision of intense individualized attention, multilevel interventions, early identification of a problem, training based on skills development*, and *engagement of peers and parents in the intervention*.

Nation and colleagues (2003) have identified 9 characteristics that were consistently associated with effective prevention programs across 4 areas – substance abuse, risky sexual behaviour, school failure and juvenile delinquency, and violence. According to their findings, effective programs were: (1) *comprehensive,* (2) *included varied teaching methods,* (3) *provided sufficient dosage,* (4) *were theory-driven,* (5) *provided opportunities for positive*

relationships, (6) *were appropriately timed*, (7) *were socio-culturally relevant*, (8) *included outcomes evaluation,* and (9) *involved well-trained staff*. Those nine characteristics could be related to the four broad areas of prevention programming:

(1) Program characteristics,
(2) Matching programs to target population,
(3) Implementation quality, and
(4) Evaluation of the interventions.

General principles gleaned from effective interventions may help mental health promotion and prevention practitioners to select, modify or create more effective programs.

1.3.2. Preffi 2.0 - Quality Assurance and Quality Assessment Instrument

Examples of the translation of "effect predictor" findings into tools for quality assessment are found in the United States and some Western European countries (Aro, Van den Broucke and Räty, 2005). For example, the RE-AIM framework developed in the United States (Glasgow, Vogt, and Boles, 1999) and the Preffi 2.0 instrument developed in the Netherlands (Molleman et al., 2005a, 2005b) are designed to promote systematic application of evidence-based principles that are associated with higher quality of the interventions and better outcomes.

There were several reasons why Preffi 2.0 was chosen among other instruments to be analysed in this doctoral study. As it will be presented in later chapters of the dissertation, Preffi 2.0 is a very comprehensive quality assessment instrument and it compares very favourably with all other quality assessment tools in the world (Vermeulen et al., 2005). The instrument was intensely examined (Molleman, Peters, Hommels and Ploeg, 2003; Molleman 2005a, 2005b). Also, it is important to stress that the expertise of Preffi 2.0 authors was available to Croatian researchers. Preffi 2.0 is interesting on the international level too and it was adopted by scientists in Hungary, Spain, France, and Norway.

Since the early 1990s, the Dutch mental health promotion specialists have been trying to assess what determines the effectiveness of health promotion and prevention programs and how interventions can be designed in such a way as to maximize their effectiveness. The main assumption of the Dutch mental health promotion and prevention scientists was that the knowledge about effect predictors translated into practical guidelines that will be used by prevention practitioners in developing and implementing prevention programs could

systematically increase the effectiveness of the intervention. With that intention the PREvention EFFect-management Instrument (Preffi 1.0) was developed (Molleman, 2005).

Preffi 1.0 instrument was designed for professionals in the field of mental health promotion and prevention in the form of guidelines that can be used in developing and improving their interventions to maximize their effectiveness. The main constructs of the Preffi 1.0 instrument are effectiveness and effect management as a tool to increase the likelihood of achieving the greatest possible effectiveness in mental health promotion and prevention programs. The Preffi 1.0 instrument is based on international research findings about program's aspects affecting effectiveness and quality. The instrument contains four dimensions that contribute to the effectiveness: (1) program development, (2) the program itself, (3) implementation, and (4) evaluation. It focuses on the theoretical steps that have to be taken and choices that have to be made during the design, implementation and evaluation of a project or program. These are principles that can be applied to all interventions and can considerably increase their effectiveness.

The purpose of the Preffi 1.0 instrument was to achieve systematic improvements in quality, particularly in the effectiveness of mental health promotion and prevention programs in practice. Its intention was to stimulate the practitioner "to learn how to learn" and to allow an assessment and systematic evaluation of the current prevention and health promotion practice. Currently most mental health promotion and prevention practices in The Netherlands use this instrument on a regular basis. After researching the application and implementation process of the Preffi 1.0 (Molleman, 2005), it was found that there was a great need for a new and updated Preffi version which would incorporate the latest research findings and the experiences of practitioners who had used Preffi 1.0. The instrument needed improvements of its content, norms, format, and its role.

In the early 2000s, the Preffi 1.0 instrument was thoroughly revised into a new version; now more generally called the Health Promotion Effect Management Instrument, the Preffi 2.0 (Molleman et al., 2003). The Preffi 2.0 instrument was developed by a joint project group including experts from the Netherlands Institute for Health Promotion and Disease Prevention and Radboud University Nijmegen (Molleman et al., 2005a, 2005b). Preffi 2.0 invites users to assess projects against various criteria, indicate points to be improved, prioritize them and achieve improvements. From this perspective, the Preffi 2.0 instrument is a quality assessment and quality assurance instrument that allows users to assess whether health promotion and prevention programs have been designed in such a way as to maximize the chances of being effective.

The Preffi 2.0 instrument consists of 39 quality criteria – effect predictors, variables that are demonstrably related to the program's intended output. Preffi 2.0 items reflect research findings on effect predictors, as well as insights into such predictors derived from critical discussions with practitioners. A number of criteria were used in selecting effect predictors that were incorporated in the instrument – relevance, scientific evidence, generalizability, modifiability, and measurability. These quality criteria are distributed within eight clusters: (1) contextual conditions and feasibility, (2) problem analysis, (3) determinants of behaviour and environment, (4) target group, (5) objectives, (6) intervention development, (7) implementation and (8) evaluation.

The authors of Preffi 2.0 (Molleman et al., 2005a, 2005b) emphasize that the instrument promotes a systematic and evidence based approach, which is expected to lead to high quality programs and better outcomes. A Model of Preffi 2.0 is presented in Figure 7.

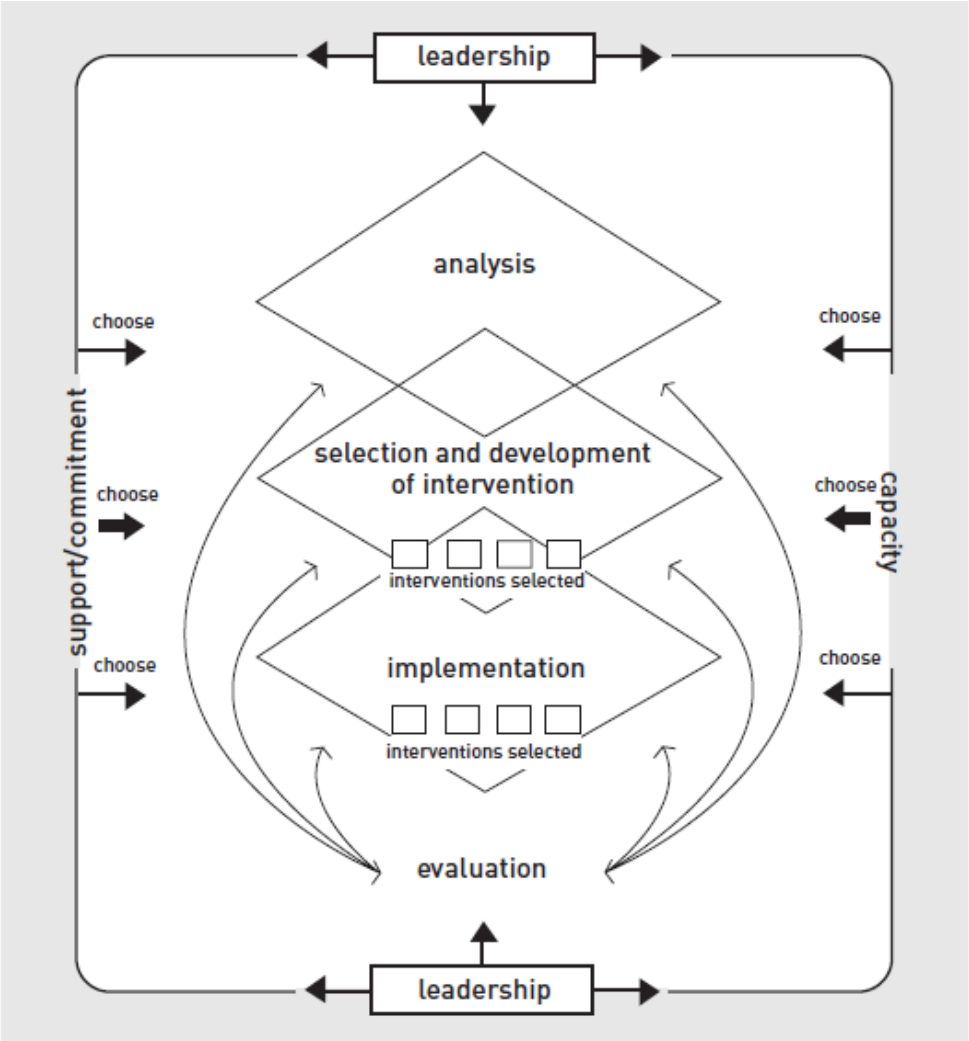


Figure 7.

Preffi 2.0 Model (Molleman, 2005)

Preffi 2.0 clusters follow the logical steps of a Planning model (Bartholomew et al., 2006) and were arranged in the order to follow that model:

- Analysing the problem, including its nature and scale, as well as its determinants,
- Making successive decisions about objectives, target groups and suitable intervention types,
- Paying special attention to the inclusion of effective elements, as derived mostly from social learning theory,
- Pre-testing and implementing the interventions,
- Evaluating in terms of both process and effect, and

- Contextual conditions and project management, including the personal characteristics of the project manager.

The model illustrates the dynamics of all concepts necessary for the development of a program. From the model it is evident that there are key elements for the development of a program presented in the center of the model - the problem analysis and analysis of a need for the program, the process of developing and planning the program, program implementation and evaluation. On the sides are those elements that represent contextual conditions for the development of a program - the quality of leadership in the organization that conceived the project, the capacity to implement the program, support organizations and deliverers of the program. Also, it is clear from the model that the different phases of program development are interrelated and interdependent. For example, the design of a program evaluation should be synchronized with the objectives of the program, but it also depends on some contextual conditions, such as the financial resources necessary for evaluation.

Studies published on Preffi 1.0 and Preffi 2.0 are those in which Preffi 1.0 was provided to mental health promotion and prevention practitioners as a quality enhancement guideline during the intervention development phase (Molleman, 2005a), and in which Preffi 2.0 was used and validated as a quality assessment instrument (Molleman et al., 2005b). In the latter study, some metric characteristics of the Preffi 2.0 instrument (e.g., content validity and reliability) were examined. The Preffi 2.0 reliability study showed that two assessors are needed for a sufficiently reliable and accurate assessment of a prevention program or project as a whole, while three assessors are needed for a reliable and accurate assessment of each of the individual Preffi 2.0 clusters.

1.3.3. Training for Prevention – Intervention for Quality Assurance in Mental Health Promotion and Prevention

A Training for Prevention was developed within the project "Preffi – Quality Assurance in the County of Istria" initiated by the University of Zagreb, Faculty of Education and Rehabilitation Sciences and the Department of Health and Social Services in the County of Istria. The intervention program for practitioners was designed by Miranda Novak and Josipa Mihić, doctoral students and researchers from the University of Zagreb, Faculty of Education and Rehabilitation Sciences, members of the project team. The project started in 2011 and was completed in 2012.

Since 2008, the authors of the Training for Prevention were members of the Department of Health and Social Services' committee for projects' appraisal in the County of Istria. The role of the committee was to evaluate the quality of written program proposals and to assess whether proposed programs involved a sufficient level of effect predictors. Evaluated projects were those focused on mental health promotion and prevention of mental, emotional and behavioural problems, and proposed by NGOs active in the County of Istria. Within the evaluated programs, a range of overall weaknesses and gaps were identified. During several years of that experience, the authors of the Training for Prevention realized that most of the programs have similar difficulties in transferring science-based principles into practice. It was evident that the authors and deliverers of the evaluated programs came from different professional backgrounds and had poor knowledge of mental health promotion and prevention. The weakest elements of written program proposals were problem analysis, target group, theory behind their programs, relations between programs' goals, activities and expected outcomes as well as evaluation design.

Introduction and Theoretical Background

Regarding detected problems, the Training for Prevention was developed as an intervention aimed at program managers, program developers and deliverers who were included as intervention's participants. The theoretical concept of the Training for Prevention is based on:

- Recent knowledge and research on effect predictors (Durlak et al., 2011; Cuijper et al., 2008; Stice et al., 2007; Jane-Llopis and Barry, 2005; Brown et al., 2000; Cuijper, 2002; Hosman and Engels, 1999; Raphael, 1999; Tobler and Stratton, 1997; Kok et al., 1997; Hosman, 1994)
- Preffi 2.0, its model and scientific base (Molleman, 2005; Molleman et al., 2005a; 2005b; Molleman et al., 2003)
- Intervention mapping approach (Kok, Peters and Ruiter, 2011; Bartholomew, Parcel, Kok and Gotlieb, 2006),
- Theory of planned behaviour (Montano and Kasptzyk, 2000; Itzak, 1991; Ajzen, 1991; Huchting, Lac and LaBrie, 2008),
- Transtheoretical model (Prochaska, Redding and Evers, 2002) and
- Logic modeling (Wyatt Knowlton and Philips, 2009; Rogers, 2005).

Theories were chosen based on the current state of the art of the field of mental health promotion and prevention in Croatia and regarding the goals of the Training for Prevention. The common intention of all mentioned theories is to understand and explain the process of behavioural change. Achieving behavioural change is in the focus of Training for Prevention, but also in the focus of most mental health promotion and prevention projects of NGO leaders and providers involved in the Training for Prevention. Within the Training, mentioned theoretical concepts were used on three levels:

1. for the transfer of knowledge about processes of change to the Training's participants,
2. for the incorporation of effect indicators in participants' programs, and
3. for the development of Training participants' skills needed for initiating the process of change in their target groups.

To achieve an impact on all three levels, authors of the Training incorporated featured theoretical backgrounds in the Training's content. In general, the intervention Training for Prevention was aimed at *transferring knowledge about effect predictors* to interventions' developers, managers and deliverers in order to increase the effectiveness and quality of their programs. For a successful transfer of knowledge and research on effect predictors to Training participants, Training authors applied the *Intervention mapping approach* which follows exact steps wherein effect predictors are incorporated. Intervention mapping follows six steps: (1) need assessment, (2) program objectives, (3) methods and application, (4) program development, (5) planning for program implementation, and (6) planning for evaluation. Those steps are needed for theory-based and research-based development of interventions (Bartholomew et al., 2006). The same stages are also contained in Preffi 2.0. The approach requires that people who are developing the program identify their change objectives and specify methods proven effective for behaviour change. It describes iterative paths from problem identification to problem solving. By basing such decisions on previous evidence and documenting the way in which intervention materials are designed, interventions can communicate clearly about the intervention content and principles, which facilitates subsequent intervention improvement.

The importance of intention is more especially elaborated in the *Theory of the planned behaviour* (Ajzen, 1985, 1991). Theory of the planned behaviour provides a useful model for identifying intervention targets because it proposes a number of potentially modifiable determinants of behaviour. It is a model of rational decision making which underlines that

behaviour is determined by intention and perceived behavioural control. Intention is determined by three independent cognitions: attitude, subjective norms i.e. perceived social pressure from important others and perceived behavioural control.

Another approach to intentional health behavioural change is the *Transtheoretical Model* (Prochaska, Redding and Evers, 2002). This staged model provides a framework for understanding and facilitating the process of health behaviour change. It incorporates four related concepts considered central to behaviour change: stages of change, self- efficacy, decisional balance and process of change. The Transtheoretical Model postulates that behaviour change is accomplished through a series of stages, rather than a single or sudden event. These five stages of change are (1) pre-contemplation, (2) contemplation, (3) action, (4) preparation, and (5) maintenance. The first three stages describe the development of intention to take action, whereas the last two stages describe the process of fully actualizing the intent to change. Progression through the stages is linked to differences in self-efficacy, decisional balance and process of change.

Also, the approach of *Logic Modelling* is useful in the process of program design, planning, implementation, and evaluation. Rogers (2005) describes a logic model as a representation of how an activity (such as a project, a program, or a policy) is intended to produce particular results. It offers a way to visually describe and share an understanding of relationships among elements necessary to operate a program or change effort. Wyatt Knowlton and Philips (2009) stress that logic modelling represents the use of program theory in program design and evaluation. Program logic models vary in detail but offer additional information that assists design, planning, managing, and monitoring/evaluation. Program models support a display that can be tested for feasibility.

It is important to stress that theoretical backgrounds described and used as a base for the Training can be applicable and useful for program development in different professional fields, not only the mental health promotion and prevention field. The process of change follows principles which are common to diverse areas, and empirical understanding of that process is helpful for achieving desirable outcomes.

Model of Training for Prevention

The model of the Training for Prevention shown in Figure 8 was developed on the principles of previously described theoretical concepts and findings. The model presents relations between crucial effect predictors and their impact on one another, which in synergy

result in overall program effectiveness. It is expected that incorporation of principles of science-based mental health promotion and prevention can result in better understanding of the theory and logic model of a program, improvement of the quality of written project proposals, and finally lead to higher implementation quality and better behavioural and mental outcomes.

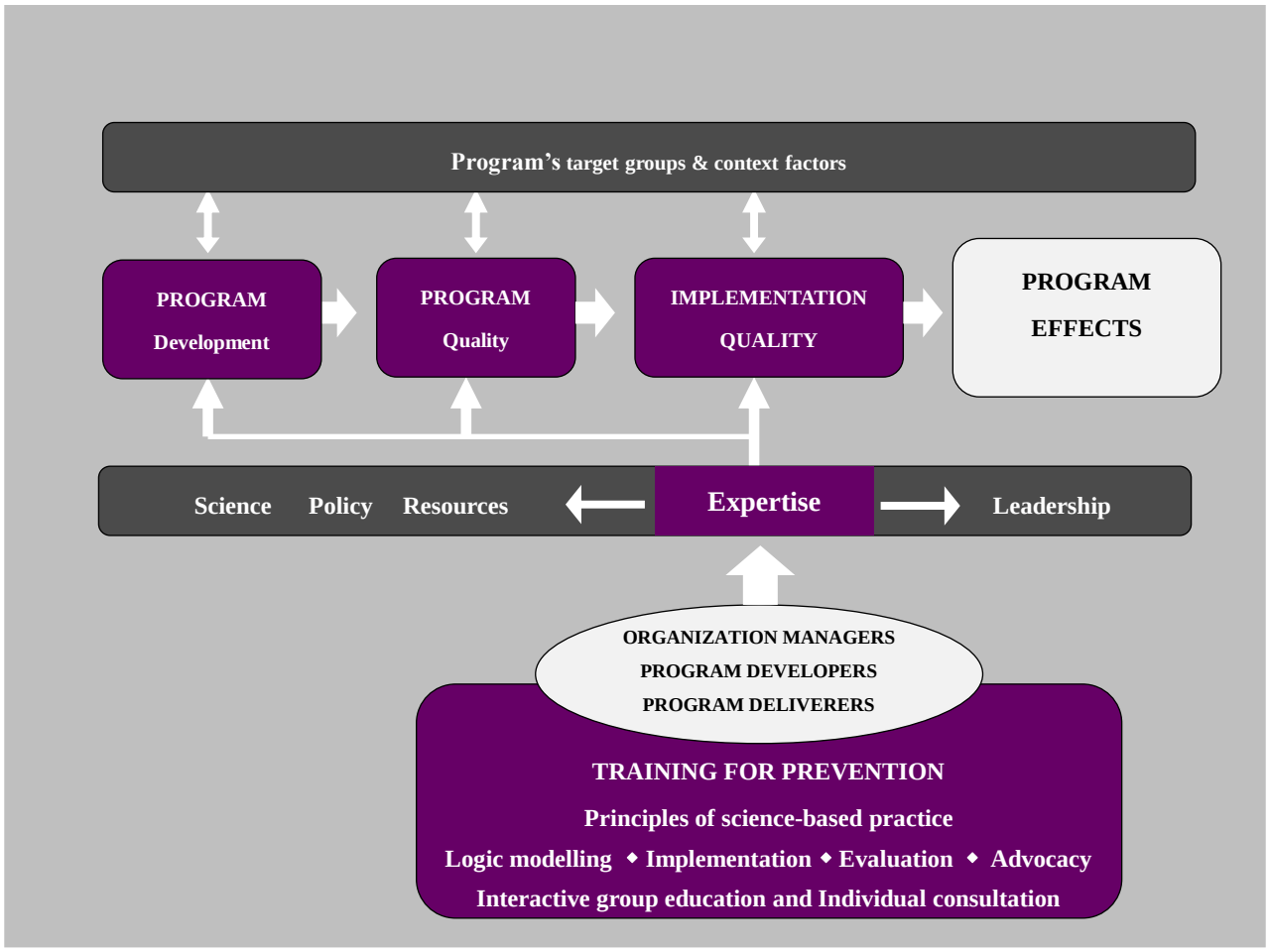


Figure 8.
Model of Training for Prevention
(developed by Novak and Mihić, 2010)

The diagram of the model shows that the Training for Prevention is aimed at enhancing the expertise of organization's managers, program developers and deliverers. As presented in the model, the expertise is not the only condition needed for comprehensive program development, program quality, implementation quality and, finally, program effects.

Although science, policy, resources and leadership are also important conditions for effective prevention and promotion practice, within the project "Preffi – Quality Assurance in the County of Istria" the enhancement of expertise was chosen as the most suitable and most easily reachable strategy of change. Improvement of the level of expertise of managers, program authors and deliverers can also reinforce policy change, develop other resources and encourage research.

1.4. Research Project »Preffi – Quality Assurance in the County of Istria«

To develop mental health promotion and prevention in a country with not very well developed national infrastructure and governance in that field, the research team from the Faculty of Education and Rehabilitation Sciences concentrated their efforts in a geographical area which would offer the best perspective on success. Since 2002, a team from the Faculty started with the process of developing a national "laboratory" for research, policy making, implementation and quality assessment in the County of Istria. It is one of the most developed and relatively more prosperous counties in Croatia. Members of the research team from the Faculty realized that conditions for investment and development were more favourable in that county and that the outcomes of learning experience can serve in future initiatives of developing mental health promotion and prevention on a nation-wide scale.

In 2002, the Faculty was supported by the Department of Health and Social Care of Istria County in running a project "Communities That Care: Development of a Model for Behavioural Disorders Prevention" (Bašić et al., 2007a and 2007b; Bašić, Grozić-Živolić, 2010), which was followed by the project "Communities that care – development, implementation and evaluation of the community model of prevention of behavioural disorders" (Bašić et al., 2007a and 2007b; Bašić, Grozić-Živolić, 2010). These two projects focused on prevention of behavioural disorders and promotion of mental health in children and youth in Istria County. Aims of the research team during these projects were to promote the principles of science-based mental health promotion and prevention practice and to enhance the collaboration of science and practice in reaching positive outcomes. During projects, important steps were taken to improve the quality of mental health promotion and prevention practice in Istria. These include: (1) assessment of readiness for mental health promotion and prevention, (2) needs assessment, (3) setting of mental health promotion and prevention priorities, (4) systematic identification of resources, (5) implementation of mental health promotion and prevention programs according to the defined needs, and (6) evaluation

of those programs and whole project. Given the many years of systematic investment in mental health promotion and prevention in Istria County, it is reasonable to assume that Istria has the appropriate base for further productive mental health promotion and prevention investments.

Programs of mental health promotion and prevention in Croatia are in most cases initiated and implemented by members of the civil sector and nongovernmental organizations. For that reason, the Department of Health and Social Care of Istria County strives to systematically develop mental health promotion and prevention through annual financing of programs of nongovernmental organizations provided by local practitioners. To get the financial support, projects must meet various criteria developed by the Department. In 2002, the Department started to develop a systematic procedure for allocating its funds, and financed 11 mental health promotion and prevention projects. Initially, the criteria for financing were that proposals should provide a solution to a particular problem and that the proposed services had to be broadly offered within the community (Bašić, Ferić Šlehan, Kranželić Tavra, 2007a; internal materials from Department of Health and Social Care, County of Istria, 2010). In order to develop criteria for program selection, the Department began including other criteria to the selection process as advised by researchers from the Faculty of Education and Rehabilitation Sciences. These included (internal materials from Department of Health and Social Care, County of Istria, 2010):

(1) Clear and specific program goals,
(2) Firm organizational structure for program implementation,
(3) Partnership with other organizations in the community, and
(4) Involvement of volunteers.

In 2004, 27 projects were financed and this expanded to 32 projects in 2006. The financed projects were focused on the prevention of behavioural disorders and problems, prevention and treatment of drug abuse, counselling programs, parent education programs and programs promoting partnership between kindergartens, schools, families and local communities. According to the 2009 report of the Department of Health and Social Care of Istria County, more than nine million kunas were invested in the mental health promotion and prevention programs in Istria since 2002. Even though financing criteria have been adapted during the years, there was a strong need for further improvement of the criteria. As the demands for quality, accountability and sustainability grew, the existing approach to funding

was seen as insufficient. The research team from the Faculty of Education and Rehabilitation Sciences and the Department of Health and Social Care of Istria county noticed there is a:

- Need for developing a quality assessment tool for mental health promotion and prevention programs,
- Need for improvement of mental health promotion and prevention programs' quality, and
- Need for evaluation of mental health promotion and prevention program effectiveness in Istria county.

Based on the detected needs, in 2010 the Department of Health and Social Care of Istria County decided to continue its collaboration with the Faculty of Education and Rehabilitation Sciences through the project »Preffi – Quality Assurance in the County of Istria« (project team: professor Josipa Bašić, PhD, Miranda Novak, M.A., Josipa Mihić, M.A.). It is important to stress that this project was a continuation of "Communities That Care" project implementation in Istria. The general aim of the project was to decrease mental and behavioural problems of children and youth in Istria County through the incorporation of evidence-based principals into the mental health promotion and prevention practice.

In order to achieve this long-term goal, the aims of the project were:

1. To enhance the knowledge and capacities of NGO's leaders, program managers and deliverers, financed by Istria County, about the principles of science-based practice by providing them with "Training for Prevention";
2. To improve the quality of written proposals of mental health promotion and prevention programs proposed by NGOs in Istria;
3. To improve the outcomes of mental health promotion and prevention programs financed by the Department of Health and Social Care of Istria County, and
4. To create science-based criteria for financing mental health promotion and prevention programs in Istria County through incorporation of effect predictors in financing criteria.

The main assumption of the project was that the incorporation of evidence-based principles is crucial in improving the quality and effectiveness of mental health promotion and prevention practice in the County of Istria. For that reason, the project team intended to

encourage and prepare the Department of Health and Social Care to be oriented towards implementation of an evidence-based policy in their County. Evidence-based policy has been defined as an approach that "helps people make well-informed decisions about policies, programs and projects by putting the best available evidence form research at the hearth of policy development and implementation" (Davies, 1999, according to Shaw et al., 2007). This approach stands in contrast to opinion-based policy, which relies heavily on either the selective use of evidence or on the untested views of individuals or groups, often inspired by ideological standpoints, prejudices or speculative conjecture. Gray (1997, according to Shaw et al., 2007) has suggested that there is a new dynamic to decision making in mental health promotion and other areas of public policy, whereby the speculation of opinion-based policy is being replaced by a more rigorous approach that gathers, critically appraises, and uses high-quality research evidence to inform policy-making and professional practice.

The project included 24 mental health promotion and prevention programs proposed by NGOs and other institutions from Istria financed during 2011 by the Department of Health and Social Care, County of Istria. Programs were divided in two groups, experimental and control, tied by matched pairs and assessed with the Preffi 2.0 instrument, using a pretest-posttest evaluation design. After the first assessment of programs with the Preffi 2.0 instrument, the experimental group, i.e. program leaders, was involved in Training for Prevention intervention.

1.4.1. Logic Model of the Project

Figure 9 represents a logic model of the project »Preffi – Quality Assurance in the County of Istria«. As is evident from the model and as it was already elaborated in the previous sections, Istria County had many conditions affordable for successful implementation of this project. The most important one was that Istria County stakeholders showed a high level of readiness for developing science-based mental health promotion and prevention practice. Furthermore, the history of collaboration between Istria County and University of Zagreb Faculty of Education and Rehabilitation Sciences contributed to the initiation of this project. Two doctoral students of the doctoral program »Prevention Science« at the mentioned Faculty developed the idea for this project during a course by professor Clemens Hosman, Phd – Theories and principles of change and effect management. Support and mentorship of the researchers from the Netherlands – professor Clemens Hosman, PhD and Gerard Molleman, PhD provided significant capacity needed for initiating this project.

Available scientific knowledge on effect predictors and effectiveness of mental health promotion and prevention formed the knowledge base for the development of 'Training for Prevention' by the two doctoral students. It is important to note that a crucial condition for implementing this project was the formal position and power of Istria County to recruit leaders of NGOs, designers and providers of funded programs for participation in the Training for Prevention. The Department of Health and Social Care of Istria has a decision-making position. That position created conditions for the implementation of projects' activities, but will also serve in changing the mental health promotion and prevention policy in Istria County in the future.

Based on all the inputs and conditions for project realization, the Training for Prevention was developed and delivered to the NGO leaders, program developers and deliverers, and to one member of the Istria County Department of Health and Social Care. It resulted in more than 50 participants of the Training. The assumption that the Training for Prevention will result in its targeted outcomes is based on the theories and science-based approaches behind the Training for prevention intervention.

One of the expected short-term outcomes is enhanced awareness of the Training for Prevention participants about evidence-based mental health promotion and prevention practice. Also, it is expected that the Training for Prevention will improve the knowledge of Training participants about evidence-based mental health promotion and prevention practice, improve their skills of quality program development and skills of writing quality program proposals. Other assumptions are that the application of Preffi 2.0 in assessing the programs' quality will lead to the incorporation of effect predictors and evidence-based principles into the criteria for financing programs proposed by NGOs. It is assumed that this will result in increased quality of written program proposals.

As is evident from the figure, all short-term outcomes are expected to be related with a range of medium-term outcomes. The assumption is that the medium-term outcomes of this project are a general improvement of the quality of NGO programs and that the programs which will be financed by the Istria County in the future will be in accordance with specific needs of the community. Also, it is expected that the several years of applying evidence-based criteria for program funding will result in increased effectiveness of NGO's mental health promotion and prevention programs. It is believed that this will enable sustainable implementation of effective programs in Istria County. Changes and improvements of criteria for financing the programs will also promote partnerships between different NGOs in implementing more comprehensive mental health promotion and prevention programs.

In a long-term perspective, the final vision of this project is that sustainable application of evidence-based criteria for financing programs and sustainable investment into the knowledge and skills of program designers and deliverers will have a significant impact on the level of public health and public mental health. As is shown in the figure, the final long-term expected outcome of this project is a decrease in mental, emotional, and behavioural problems and improvement of quality of life in Istria County.

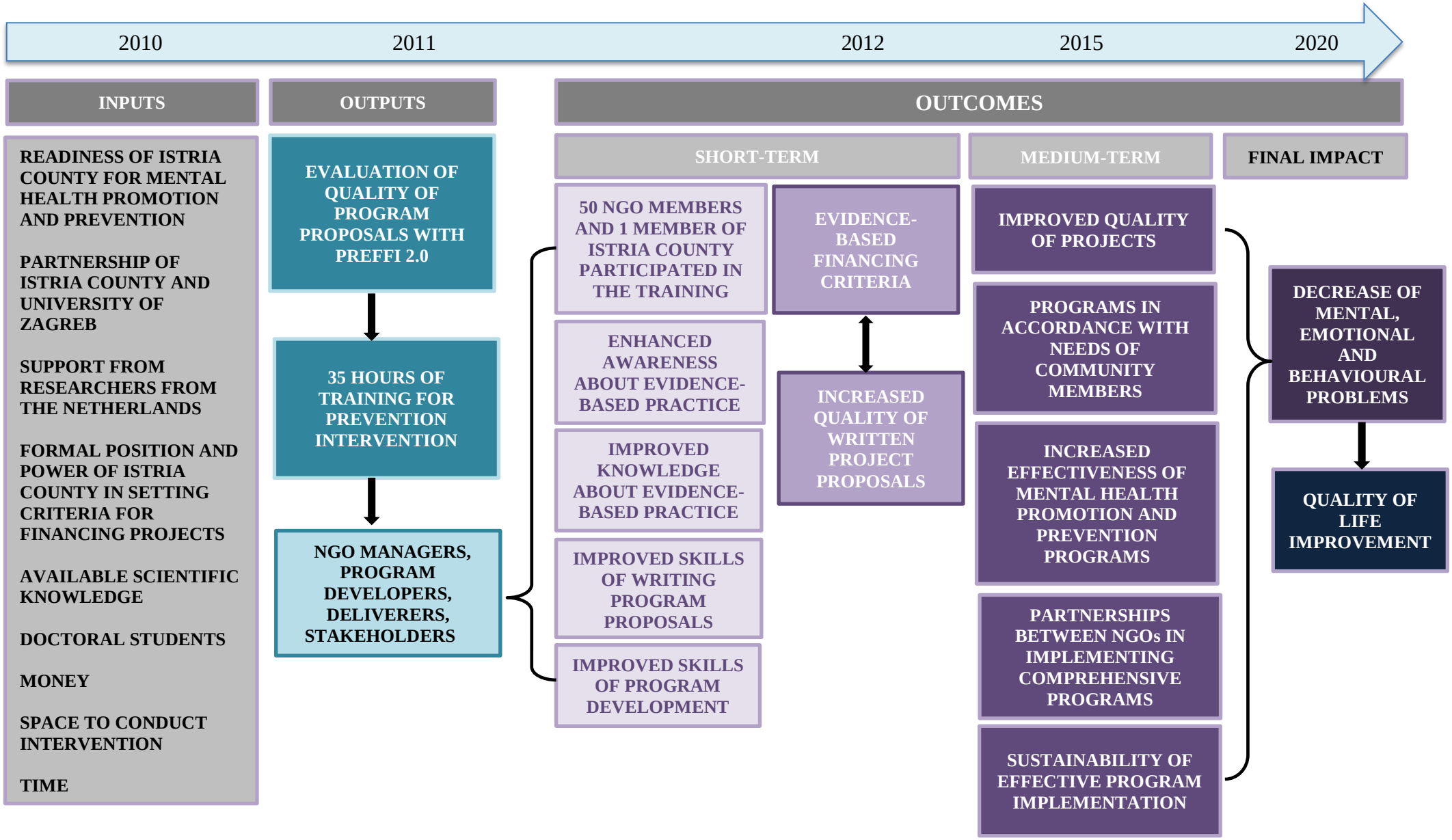


Figure 9.

Logic Model of the Project »Preffi – Quality Assurance in the County of Istria«

1.4.2. Doctoral Research Studies Within the Project

During the project »Preffi – Quality Assurance in the County of Istria« two doctoral research studies were designed and conducted simultaneously:

- *Study of effectiveness of prevention programs* (doctoral student: Josipa Mihić, University of Zagreb, Faculty of Education and Rehabilitation Sciences, mentor: Clemens Hosman, PhD, University of Nijmegen and Maastricht, the Netherlands) and
- *An empirical study on implementation quality in prevention programs* (doctoral student: Miranda Novak, University of Zagreb, Faculty of Education and Rehabilitation Sciences, mentors: Clemens Hosman, PhD, Radboud University of Nijmegen and Maastricht University, the Netherlands and Celene Domitrovich, PhD, Pennstate University, USA).

Study of effectiveness of prevention programs is the doctoral study which will be presented in this dissertation. One of the aims of this study is to adapt the Preffi 2.0 quality assessment instrument and to assess some of its metric characteristics. Other aims of the study are to measure the Training for Prevention impact on the effectiveness of programs of mental health promotion and prevention in achieving desired outcomes and to measure its impact on the quality of written program proposals. The assumption of the study is that training of the NGO leaders, program managers and deliverers about effect predictors can improve the effectiveness of their programs and the quality of written program proposals. Also, the predictive validity of the Preffi 2.0 instrument in predicting the effectiveness of programs is assessed within this study. It is expected that the results of this study will provide an insight into the quality of NGO programs of mental health promotion and prevention involved into a study. Study results will offer suggestions for creating a science-based mental health promotion and prevention practice in Istria County.

The main objective of the doctoral study *An empirical study on implementation quality in prevention programs* is to monitor the quality of implementation processes in 24 mental health promotion and prevention programs. This study attempts to answer the question whether the Training for Prevention which will be delivered to an experimental group has an impact on improving the quality of implementation of 24 mental health promotion and prevention programs. In order to answer this question, three new measures of implementation process were constructed, relying on the literature and trends in mental health promotion and

prevention science. These include a measure for monitoring the quality of implementation of programs from the perspective of an NGO leader, a measure which monitors the quality of implementation of programs from the perspective of program developers and providers, and a measure that tracks the quality of implementation of preventive program from the perspective of program users.

2. AIMS AND RESEARCH PROBLEMS OF THE STUDY

Many studies have shown that mental health promotion and prevention interventions can be effective in reducing mental, emotional and behavioural problems (Botvin et al., 2006; Hawkins et al., 2005; Hosman et al., 2004; Tobler and Stratton, 1997; Kirby et al., 1994). Still, there are many challenges facing mental health promotion and prevention science, practice, and policy effectiveness. As discussed in the introduction, scientists are more and more interested in detecting the general principles of effectiveness, the so called "effect predictors" or "effect moderators" (Hosman and Engels, 1999; Raphael, 1999; Hosman, 1994). Previously conducted studies have found different effect predictors of programs aimed at preventing specific mental, emotional and behavioural problems (Durlak et al., 2011; Stice et al., 2007; Jane-Llopis and Barry, 2005; Bartholomew et al., 2001; Ader et al., 2001; Brown et al., 2000), but there are few published studies focused on quality assessment and quality assurance in the field of mental health promotion and prevention in general. These two concepts are closely connected with program effectiveness, so a deeper understanding of those concepts can offer a significant contribution towards more effective mental health promotion and prevention science and practice.

Within this doctoral study, both the concept of quality assessment and the concept of quality assurance in mental health promotion and prevention are considered and examined.

The study has two aims. The first one is focused on the analysis of the quality assessment process through the application of the Dutch quality assessment instrument Preffi 2.0 on written proposals of mental health promotion and prevention programs. The second aim is related to the concept of quality assurance. The aim is to assess if investment in the knowledge and skills of mental health promotion and prevention program managers and deliverers about the principles of effectiveness can improve the level of quality of their written program proposals and improve the outcomes of the programs they develop and deliver.

For this doctoral study the following research tasks and hypotheses were defined:

FIRST RESEARCH TASK

To assess the metric characteristics of Preffi 2.0, i.e. its content validity, reliability and predictive validity.

Related to this task, the following hypotheses will be tested:

Hypothesis 1.1. The items of the Preffi 2.0 instrument are theoretically connected with specific quality indicators and are essential for assessing those indicators. The Content Validity Ratio of the whole Preffi 2.0 is 0.70 or higher.

Hypothesis 1.2. Preffi 2.0 is a reliable instrument with at least a medium or high value of concordance between three assessors (G= 0.70 or higher).

Hypothesis 1.3. Programs that accomplish higher total scores on Preffi 2.0 achieve more effective outcomes than programs that accomplish lower total scores on Preffi 2.0.

Regarding the predictive validity of Preffi 2.0, it was also assumed that the programs which accomplish higher scores on particular Preffi 2.0 clusters achieve more effective outcomes than programs that accomplish lower results on these clusters. Since it is an explorative research task, no directive hypothesis was defined.

SECOND RESEARCH TASK

To examine the impact of the Training for Prevention on the effectiveness and quality of mental health promotion and prevention programs.

Hypothesis 2.1. Programs whose managers and deliverers were involved in the Training for Prevention achieve significantly higher scores on effectiveness, i.e. higher effect sizes, than programs whose managers and deliverers were not involved in the Training.

Hypothesis 2.2. There is a difference between the experimental and control group on the total Preffi 2.0 score, i.e. programs involved in the Training for Prevention achieve significantly higher total scores on Preffi 2.0 after the Training compared to the programs that were not involved in the Training.

Hypothesis 2.3. There is a difference between experimental and control groups on specific Preffi 2.0 cluster scores in a way that programs involved in the Training for Prevention achieve significantly higher scores on individual Preffi 2.0 clusters after the Training compared to the programs that were not involved in the Training.

THIRD RESEARCH TASK

To identify strengths and weaknesses of the programs of mental health promotion and prevention from the County of Istria that were involved in the study.

Application of Preffi 2.0 will provide information about the quality of the assessed programs, i.e. their strengths and weaknesses. Regarding the explorative characteristic of the described research task, no directive hypothesis will be defined.

3. METHODS

Within this doctoral research, different but interrelated studies were conducted in order to examine and test research tasks and hypotheses, as defined in Chapter 2. All the studies were conducted within the project "Preffi – Quality Assurance in Istria County" conducted by researchers from the University of Zagreb, Faculty of Education and Rehabilitation Sciences in cooperation with the Department of Health and Social Services, the County of Istria (see the Section 1.4., p. 37).

Figure 10 represents the timeline of different studies conducted within this research. All conducted studies could be grouped into three main studies:

1. Study on Metric Characteristics of the Preffi 2.0 Instrument;
2. Study on the Quality of Mental Health Promotion and Prevention Programs in Istria, and
3. Study on the Impact of the Training for Prevention on Mental Health Promotion and Prevention Programs' Effectiveness and Quality.

It is important to stress that studies involved different samples and applied measures which will be explained in detail in the following sections of this chapter.

The aim of the first study was to test metric characteristics of Preffi 2.0, which refers to the first research task of this doctoral study. Within this study, reliability, content and predictive validity of Preffi 2.0 were tested. The reliability analysis of Preffi 2.0 was based on the assessment of 24 program proposals with Preffi 2.0 at two time points and analysed by using the generalizability theory. The predictive validity of Preffi 2.0 was examined by using the results from the second and third study, which is explained in Section 3.4.

The second study was conducted with the aim to assess the quality and to identify the strengths and weaknesses of the Istrian programs of mental health promotion and prevention. This study is based on the first quality assessment of 24 written proposals of programs of mental health promotion and prevention from Istria with Preffi 2.0.

The third study, i.e. the study on the impact of the Training for Prevention on mental health promotion and prevention program's effectiveness and quality, had two aims. The first aim was to assess the influence of the Training for Prevention on the effectiveness of mental health promotion and prevention programs developed and delivered by the Training participants. For that reason, the evaluation of all programs' outcomes was conducted by

applications of different measures on different program participants. A method of meta-analysis was used in order to analyse the results of this part of the study. The second aim of the same study was to examine if the Training for Prevention had an impact on the level of quality of written program proposals of mental health promotion and prevention programs developed by the Training participants 10 months after the Training. For that purpose a method of repeated measures analysis of variance was applied in the analysis of differences between 24 program proposals written before and after the Training for Prevention.

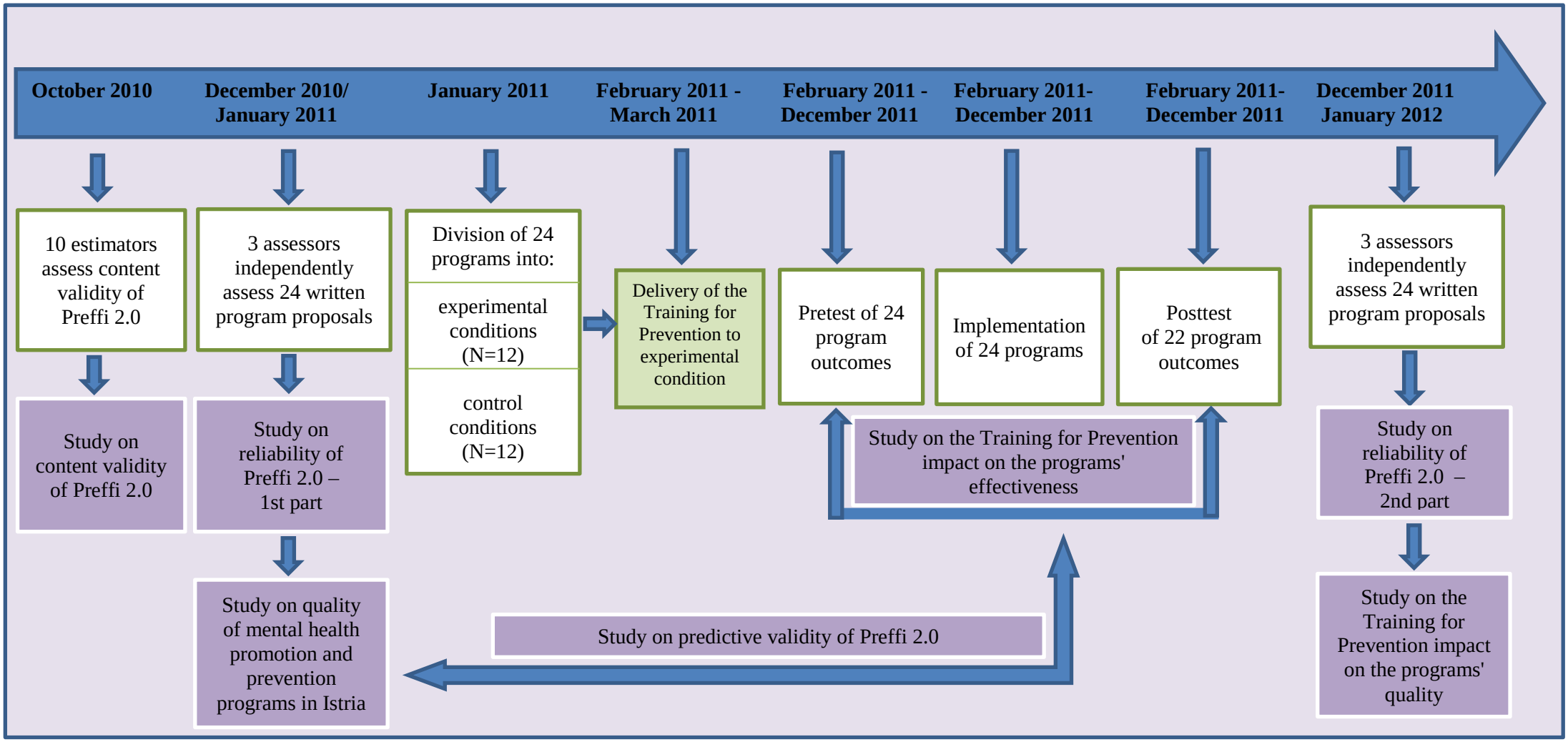


Figure 10.

Timeline and Interrelations of Research Studies Conducted Within Doctoral Study

3.1. Study on Metric Characteristics of the Preffi 2.0 Instrument

3.1.1. Study on Content Validity of the Preffi 2.0 Instrument

Participants

Content validity of Preffi 2.0 was estimated by 10 mental health promotion and prevention experts from Croatia. On average, estimators had more than 10 years of experience in the mental health promotion and prevention field. Most of the estimators were graduated social pedagogues (N=9), while one was a graduated psychologist. Four of them had a PhD in the field of prevention of mental, emotional and behavioural problems while others were doctoral students in the program "Prevention Science: Prevention of Mental and Behavioural Disorders and Promotion of Mental Health" (University of Zagreb, Faculty of Education and Rehabilitation Sciences). Estimators were from 30 to 50 years old. Seven of them were females and three were males.

Procedure

In September 2010, Preffi 2.0 was translated from English to Croatian and then translated back to English. After back translation, the English version of Preffi 2.0 was submitted to the authors of the instrument in the Netherlands to get their approval of the Preffi items translation. This procedure was chosen to ensure the preservation of the content of Preffi 2.0 during its translation into Croatian. After the successful translation of the instrument, the content validity of Preffi 2.0 was assessed in October 2010 by a group of Croatian mental health promotion and prevention experts. Participants (N=10) were introduced to the theoretical background of Preffi 2.0 through a five-page document. After having read the chapter about the theory base of Preffi 2.0, the experts were asked to fill in the Questionnaire on the content validity of Preffi 2.0. This questionnaire was sent to the participants electronically and they had to fill it in individually.

Measures

The questionnaire on the content validity of Preffi 2.0 was developed by the author of this doctoral research study and it was based on the items of the Preffi 2.0 instrument (Appendix A, p. 188). Preffi 2.0 (Molleman et al., 2005a, 2005b) consists of 121 items which are representing 39 quality criteria – effect predictors, i.e. variables that in research literature

are found to be demonstrably related to intended program output. These 39 quality criteria are distributed across eight clusters:

- The first Preffi 2.0 cluster includes 14 items and reflects the "Contextual conditions and feasibility" of the intervention being considered. It describes the quality of support and commitment of internal and external partners, capacities for the project, leadership by the project manager, including expertise and characteristics of the manager.
- The second cluster includes 13 items and reflects the "Problem analysis". This cluster presents the quality level of the analysis of nature, severity and scale of the problem, the distribution of the problem, and problem perception by stakeholders.
- The third cluster reflects the "Determinants of behaviour and environment" and consists of 13 items. It refers to the quality level of the program's theoretical model, description of contributions of determinants to the problem, amenability of factors to change and the quality of how determinants are prioritized and selected.
- The fourth cluster includes 7 items and reflects the "Target group" of the intervention. It describes a quality level of how general and demographic characteristics of the target group, motivation and opportunities of the target group to change and accessibility of the target group are analysed and described.
- The fifth cluster concerns "Objectives" and includes 12 items. It is assessing if project's objectives are fitting in with the problem analysis, if they are specific, specified in time and measureable, but also if they are acceptable to the main stakeholders and feasible. It also describes if objectives are considered achievable given the available resources, contextual conditions and intended period of time.
- The sixth cluster, "Intervention development", is the most comprehensive one and consists of 33 items. It reflects the rationale of the intervention strategy, previous experience with the intervention, duration, intensity and timing of the intervention, its fit to the target group and the culture, participation of the target group in the planning process and usage of effective techniques. It also shows the feasibility in existing practice, characteristics of implementability of the intervention and coherence of the included interventions/activities.
- The seventh Preffi 2.0 cluster, "Implementation", has 14 items. It reflects the model of implementation, the fit of implementation interventions to intervention deliverers ('intermediaries'), appropriateness of the supplier for intermediating intervention

deliverers, monitoring and generating feedback, and incorporation of the intervention in an existing organizational structure.

- The last Preffi 2.0 cluster, "Evaluation", consists of 16 items. This cluster evaluates the degree of clarity and agreement on the principles of evaluation between different stakeholders and the quality of process and effect evaluation. Effect evaluation refers to the assessment of planned changes after program delivery, and if it is plausible that changes were caused by the intervention. The same cluster also assesses the quality level of the feedback on evaluation findings to the relevant stakeholders in a community.

In order to assess the content validity of Preffi 2.0, 10 estimators were asked to estimate, according to their understanding of the Preffi 2.0 theoretical background, the level in which each of the 121 items of the Preffi 2.0 instrument is in accordance with the specific quality indicators and essential for assessing these indicators. The theoretical background of Preffi 2.0 was explained to the estimators at the beginning of the Questionnaire. For each Preffi 2.0 item estimators marked their scores on the five-point Likert scale in which 1 stood for "Completely not in accordance with the theory and not essential", 2 for "Not in accordance with the theory and not essential", 3 for "Partially in accordance with the theory and essential", 4 for "Mostly in accordance with the theory and essential" and 5 stood for "Completely in accordance with the theory and essential". At the end of the Questionnaire, estimators could write additional comments and elaborate their opinion about the accordance of the instrument items with the theoretical background of the Preffi 2.0 instrument. Besides assessing the content validity of each Preffi 2.0 item, the content validity scores were computed at the level of each Preffi 2.0 cluster and at the level of the whole instrument.

3.1.2. Study on Reliability of the Preffi 2.0 Instrument

Participants

The reliability of Preffi 2.0 was assessed by three experts in the field of mental health promotion and prevention. One of the assessors had a PhD in prevention science and more than 40 years of expertise in mental health promotion and prevention. The other two assessors were attenders of the doctoral programme "Prevention Science: Prevention of Mental and Behavioural Disorders and Promotion of Mental Health" (University of Zagreb, Faculty of

Education and Rehabilitation Sciences) with 5 years of experience in the field of mental health promotion and prevention. It is important to stress that one of the program's assessors was also an author of this doctoral study.

Programs which were assessed with Preffi 2.0 were selected among 2011 applicants for financial support from the County of Istria's Department of Health and Social Care initiative entitled "Prevention of Behavioural Disorders and Prevention of Substance Abuse." All 24 projects included in this research represent different mental health promotion and prevention activities run by various organizations in the County of Istria. The 24 selected programs come from the whole County of Istria; most of them have been implemented over several years, and financed by the Department of Health and Social Services for several years. Their organizations, organization managers, experts and practitioners are the County's most active stakeholders in the field of mental health promotion and prevention.

The 24 selected programs are briefly described in Table 3.1. In order to ensure the anonymity of organizations whose programs were involved in this research, codes and alternative program names given by the author of this doctoral study are presented instead of actual program names. For the same reason, names of organizations whose programs were assessed in this research remain confidential. The Table includes the programs' coded names, content and methods, the prevention level of the program and duration. As shown in the Table, programs differ in terms of length, participants, number of sessions and techniques used in program delivery.

Table 3.1.

Description of Mental Health Promotion and Prevention Programs from Istria Involved Into the Study of Programs' Quality

CODE OF THE PROGRAM	PROGRAM CONTENT/ LEVEL OF PREVENTION	PARTICIPANTS	PROGRAM DURATION
(1) MH promotion through the theatre	Assertiveness training using theatre techniques. *Universal prevention.*	Children and adolescents, age 7-14	46 sessions
(2) Mentor program	Mentor program promoting positive adult and child relationships. *Selective prevention.*	Children from 7 to 15 years old	Once per week during several months.
(3) Parenting program I.	Parent training program for kids from kindergarten. *Universal prevention.*	Parents	3 sessions
(4) Media literacy	Program for prevention of cyber-bullying and promotion of responsible behaviour on the Internet. *Universal prevention.*	Elementary school children, age 9-11	4 sessions
(5) Training for group leaders	Program for academic support for children with learning difficulties. *Selective prevention.*	University students age 20-23	5 sessions
(6) Substance abuse prevention for parents	Substance abuse prevention program for parents of high-school children. *Universal prevention.*	Parents of high-school children	1 lecture
(7) Substance abuse prevention for teachers	Substance abuse prevention program for high-school teachers. *Universal prevention.*	High-school teachers	1 lecture

(8) Parenting program II.	Parent training program. *Universal prevention.*	Parents	10 sessions
(9) Parenting program III.	Parent training for Roma parents. *Selective prevention.*	Parents of preschool children.	7 sessions
(10) Parenting program IV.	Parent training program for elementary and high school children. *Universal prevention.*	Parents	7 sessions
(11) Substance abuse prevention in the community	Substance abuse prevention program for adolescents. *Universal prevention.*	Adolescents, age 15-17	6 sessions
(12) Creative free time program I.	Structured free time health promotion program using creative techniques. *Universal prevention.*	Children, age 7-14	10 sessions
(13) Free time for children in foster care	Structured free time health promotion program for children in foster care. *Selective prevention.*	Children and adolescents in foster-care	Meetings each Friday from February 2011 until September 2011.
(14) Parenting program V.	Parent training program, mixed age of children. *Universal prevention.*	Parents	10 sessions
(15) Peer-violence prevention program	Peer-violence prevention program. *Universal prevention.*	Children in 4th and 5th grade of elementary school	4 sessions
(16) Self-confidence training	Health promotion program for self-confidence training. *Universal prevention.*	Elementary school children, age 10-11	4 sessions

(17) Substance abuse prevention	Substance abuse prevention program. *Universal prevention.*	Adolescents, age 15-16	1 lecture
(18) Parenting program VI.	Parent training program. *Universal prevention.*	Parents	8 sessions
(19) Underage drinking prevention	Prevention of alcohol consumption in youth. *Universal prevention.*	Adolescents, age 12-15	6 sessions
(20) MH promotion through volunteerism	Positive development promotion program. *Universal prevention*	School children, age 11-12	30 sessions
(21) MH promotion through dance	Program of health promotion aimed at life skills training. *Universal prevention.*	School children, age 12-13	12 sessions
(22) Creative free time program II.	Structured free time health promotion program using art techniques. *Universal prevention.*	School children, age 7 -15	12 sessions
(23) Parenting program VII.	Parent training program for parents of mixed age children. *Universal prevention.*	Parents	8 sessions
(24) Parenting program VIII.	Parent training program for parents of kindergarten kids *Universal prevention.*	Parents	6 sessions

Procedure

For the purpose of this study, 24 written proposals of mental health promotion and prevention programs from the County of Istria were assessed with the Preffi 2.0 instrument. In November 2010, researchers from the University of Zagreb, Faculty of Education and Rehabilitation Sciences have in collaboration with stakeholders from the Istrian Department of Health and Social Care started the procedure of selecting a representative sample of mental health promotion and prevention programs. Written program proposals, which were assessed with Preffi 2.0, were chosen from the cohort of 2011 applicants for financial support from the Department initiative entitled "Prevention of Behavioural Disorders and Prevention of Substance Abuse" and from the pool of local organizations conducting interventions in the field of mental health promotion and prevention in Istria. The procedure of selecting programs for this study began with an assessment done by the Department's commission for financing programs. A commission assessed all submitted applications and decided which organizations and programs will get financial support from the Department. The Department's application form is standardized and contains 13 sections to be filled in by applicants. It includes questions about the organization's previous experience, a description of the outcomes targeted by the intervention, the community needs assessment, goals and targeted results of the project, description of participants and activities, evaluation of efficiency, planned staff, partners and volunteers as well as the planned budget. The researchers from the Faculty of Education and Rehabilitation Sciences supplemented the form with a structured questionnaire about organizational issues and internal communication.

After the Department's commission has selected a total of 30 programs to be financed by the Department, researchers from the Faculty chose from that group 24 programs focusing on mental health promotion and prevention of mental, emotional and behavioural disorders. Six of the programs that got financial support from the Department were excluded from this doctoral study since they were only treatment oriented. The final sample of programs included in this research consisted of 24 programs described in Table 3.1.

In the course of December 2010 and January 2011, three assessors read the 24 written program proposals and independently assessed each program with the Preffi 2.0 instrument. A year after the first assessment of the 24 program proposals, program developers and authors were asked to write new proposals of the same programs. The aim of this request was to examine whether the Training for Prevention had an impact on the programs' quality (study will be explained in Section 3.3.). In December 2011 and January 2012, new program

proposals were again independently assessed by three assessors. Both applications of Preffi 2.0 on the written program proposals were used to analyse the reliability of the instrument.

Measure

The proposals of all the selected 24 mental health promotion and prevention programs, both the ones written in 2010 and the ones from 2011, were assessed with Preffi 2.0 (Appendix B, p. 195). Each of the eight Preffi clusters (see Section 3.1.1., p. 53) contains a different number of subclusters (39 all together) that consist of a different number of items. Each of the 121 Preffi 2.0 items can be scored with mark 1 for weak or non-assessable, 2 for moderate and 3 for strong. Box 1 presents how the scores on Preffi 2.0 subclusters are operationalized.

5.2. Objectives are specific, time-limited and measurable

Operationalization:

1. Do objectives specify factors that need to be changed? (Suggestion: This question has been analysed in 5.1)
2. Has for the objectives a target group been specified in which these objectives need to be achieved?
3. Do objectives specify the desired magnitude of effects that wants to be achieved (e.g.: 10% decrease)?
4. Do objectives specify the time period in which they need to be realised?

Norms:

- Weak: questions 1 and/or 2=no
- Moderate: question 1=yes, question 2=yes, question 3=no, question 4=no
- Strong: question 1=yes, question 2=yes and questions 3 and/or 4=yes

Box 1.

Example of an Operationalization and Norm for one Preffi Subcluster

The final score for each Preffi 2.0 cluster is calculated as the sum of the ratings per that cluster's subclusters divided by the maximum possible score for cluster, and multiplied with 10. The total Preffi rating for the whole program is calculated as an average score of all the cluster scores. Following that procedure, total program scores calculated with Preffi 2.0 could range from 3.33 to 10 just as the scores for each of the Preffi clusters.

In order to assess the reliability of Preffi 2.0, concordance between the three assessors in assessing the written program proposals was calculated, both for the 2010 and the 2011

proposals. The concordance was observed on the level of the total Preffi score for each of the programs, on the level of scores for 8 Preffi clusters and on the level of Preffi 2.0 subclusters.

3.2. Study on Quality of Mental Health Promotion and Prevention Programs in Istria

Participants

The study on the quality of mental health promotion and prevention programs in Istria was conducted on 24 programs which represent key mental health promotion and prevention activities run by various organizations in the County of Istria (see Table 3.1., p. 56). Most of the mental health promotion and prevention activities in that County are initiated by NGOs and almost all of these organizations are applying for the financial support of the County of Istria's Department of Health and Social Care. Also, most of them have been financed by the Department regularly for the past 5 years. Their sustainability guarantees that those programs reflect the current state of the art of mental health promotion and prevention programs in the County of Istria.

Procedure

The selection process of the 24 programs assessed in this study is described in Section 3.1.2. (p. 59). During December 2010 and January 2011 all 24 program proposals were assessed by three independent assessors with the Preffi 2.0 instrument in order to describe the quality of the selected programs in the County of Istria.

Measures

Written program proposals were assessed with the Preffi 2.0 instrument. The number of items, content and composition of Preffi 2.0, and the scoring procedure for the total and cluster scores were already described in Section 3.1.2. (p. 60).

After individual appraisal of the 24 written program proposals, the 3 assessors have agreed upon the general ratings for each of the eight Preffi clusters. This method of discussing the scores between assessors was proposed by the instrument's authors (Molleman et al., 2005b) because it can assure a deeper understanding of program proposals and more precise scores. After the agreement on scores on the level of eight clusters for each program, total scores for each program were calculated as an average of eight clusters' scores. For all 24 programs, results could be shown by score per each Preffi cluster and as a total Preffi score

for the whole program. Finally, average scores for the whole group of 24 programs on the level of clusters and total Preffi scores were calculated.

3.3. Study on the Impact of the Training for Prevention on Mental Health Promotion and Prevention Programs' Effectiveness and Quality

This study represents the main part of this doctoral research. Two parallel and interrelated studies were conducted within it:

1. A study on the impact of the Training for Prevention on programs' effectiveness and
2. A study on the impact of the Training for Prevention on the quality of mental health promotion and prevention programs.

The main aim of the first study was to assess whether program managers and deliverers involved in the Training for Prevention achieve more effective mental, emotional and behavioural outcomes with their programs than those who were not involved in the Training.

The aim of the second study was to examine if there is a difference in the level of quality between the experimental and comparison group of programs on Preffi 2.0 scores, i.e. to test whether program managers and deliverers involved in the Training for Prevention develop and write programs which achieve higher results on the Preffi 2.0 scores compared to those who were not involved in the Training.

3.3.1. Participants and Matching

In order to examine the impact of the Training for Prevention on the effectiveness of mental health promotion and prevention programs and its impact on program quality, two groups of participants were involved in the research.

Participants of the study on the impact of the Training for Prevention on the quality of mental health promotion and prevention programs were managers and deliverers of 24 community-based mental health promotion and prevention programs in Istria (Table 3.1., p. 56). Those 24 programs were divided in an experimental and control group by means of the equal pairs matching method, each containing 12 programs. Pairs of programs were created based on the following criteria: whether participants were children, teenagers or adults, the type of program, the locality from which the program was coming, the duration of the

intervention, the number of participants and the total Preffi 2.0 result. Researchers intended to have both groups as similar as possible, both containing similar types of programs and similar levels of program quality measured with Preffi. Since most organization managers and program deliverers from the sample are acquainted with each other and sometimes even collaborate because the County of Istria is rather small, researchers have tried to prevent the experimental and control groups from overlapping locally in order to reduce the possibility of contamination, i.e. the risk of communication between experimental and control groups about the content of the Training for Prevention. For example, when one organization or local community had several programs included in this study, all of those programs had to be in the same conditions, experimental or control. Division of the programs in experimental or control conditions is shown in Table 3.2.

Table 3.2.

Division of Programs in Control and Experimental Conditions and Their Scores on Preffi 2.0

PAIRS	EXPERIMENTAL CONDITION	CONTROL CONDITION
1.	*(1) MH promotion through the theatre* **Preffi 2.0 score: 5.33**	*(21) MH promotion through dance* **Preffi 2.0 score: 6.90**
2.	*(2) Mentor program* **Preffi 2.0 score: 6,22**	*(13) Free time for children in foster care* **Preffi 2.0 score: 6.29**
3.	(3) *Parenting program I* **Preffi 2.0 score: 4.17**	*(14) Parenting program V* **Preffi 2.0 score: 4.19**
4.	*(4) Media literacy* **Preffi 2.0 score: 7.10**	*(19) Underage drinking prevention* **Preffi 2.0 score: 5.84**
5.	(5) *Training for group leaders* **Preffi 2.0 score: 7.40**	*(20) MH promotion through volunteerism* **Preffi 2.0 score: 8.56**
6.	*(6) Substance abuse prevention for parents* **Preffi 2.0 score: 5.01**	*(17) Substance abuse prevention* **Preffi 2.0 score: 5.09**
7.	*(7) Substance abuse prevention for teachers* **Preffi 2.0 score: 5.18**	*(11) Substance abuse prevention in the community* **Preffi 2.0 score: 5.41**
8.	*(8) Parenting program II* **Preffi 2.0 score: 5.78**	*(18) Parenting program VI* **Preffi 2.0 score: 5.55**
9.	*(16) Self-confidence training* **Preffi 2.0 score: 4.53**	*(15) Peer-violence prevention program* **Preffi 2.0 score: 4.69**
10.	*(22) Creative free time program II* **Preffi 2.0 score: 6.62**	*(12) Creative free time program I* **Preffi 2.0 score: 4.74**
11.	*(23) Parenting program VII* **Preffi 2.0 score: 5.50**	*(10) Parenting program IV* **Preffi 2.0 score: 6.16**
12.	*(24) Parenting program VIII* **Preffi 2.0 score: 6.00**	*(9) Parenting program III* **Preffi 2.0 score: 4.21**
	AVERAGE PREFFI 2.0 SCORE IN EXPERIMENTAL CONDITIONS: 5.74	**AVERAGE PREFFI 2.0 SCORE IN CONTROL CONDITIONS: 5.64**

Managers and deliverers of programs in experimental conditions were included in the Training for Prevention. From the 12 programs in experimental conditions, 21 person attended the Training for Prevention, i.e. 5 organization managers and 16 program deliverers.

Participants of the study on the impact of Training for Prevention on the programs' effectiveness were participants of all assessed programs. Since the programs had different target groups, participants of programs differed in gender, age and level of risks. The number of participants involved and assessed in each program is shown in Appendix F (p. 236).

3.3.2. Procedure

After selection and assessment of 24 programs with Preffi 2.0, programs were divided into experimental and control conditions by the equal pairs matching method. In November 2010, the Department of Health and Social Services and researchers from the Faculty of Education and Rehabilitation Sciences organized two meetings - one with the program managers and deliverers from the experimental group of programs and another with program managers and deliverers from control conditions. During the meetings, program managers and deliverers were explained the purpose and methods of the study program managers in order to get their approval for inclusion into the study. Also, the Department made a formal agreement with each organization, stating that they will receive financing for 2011 on the condition that they continue collaborating with research staff regularly. Each of the 24 included organizations signed the agreement with the Department of Health and Social Services and researchers from the Faculty. Participants from experimental conditions were asked to respect discretion rules and secrecy regarding the content of the Training for Prevention intervention. Organization managers and program deliverers had to sign a confidentiality agreement, which was attached to the financing contract. Also, they were asked to commit that at least one member of organization and one program deliverer will be present on all sessions of the Training for Prevention intervention.

The research team explained to the participants from the control condition that they would receive the Training for Prevention intervention after the whole study and measurement has finished. Therefore, Training for Prevention was delivered to control group participants in April and May 2012.

As presented in Figure 11, during December 2010 and January 2011 all programs were scored with the Preffi 2.0 instrument to get a quantitative appraisal of program quality.

Managers and deliverers of the 12 programs in the experimental group attended the Training for Prevention during February and March 2011. The 32-hour training sessions by the Training for Prevention were delivered mostly in March 2011, the exact dates being 25th February 2011, 3rd March 2011, 11th March 2011 and both 17th and 18th March 2011. It was decided that the timing of the Training for Prevention intervention would be scheduled in the first trimester of 2011 because programs differed in their starting date and length.

	PROGRAM QUALITY ASSESSMENT – PREFFI PRETEST *December 2010 and January 2011*	TRAINING FOR PREVENTION *February and March 2011*	ASSESSMENT OF PROGRAM OUTCOMES - BASELINE *February - December 2011*	24 PROGRAMS IMPLEMENTATION *February - December 2011*	ASSESSMENT OF PROGRAM OUTCOMES - POSTTEST *February - December 2011*	PROGRAM QUALITY ASSESSMENT – PREFFI POSTTEST *December 2011 and January 2012*
EXPERIMENTAL GROUP *(12 programs)*	**O1 (A)**	**X**	**O1 (B)**	**X**	**O2 (B)**	**O2 (A)**
CONTROL GROUP *(12 programs)*	**O1 (A)**		**O1 (B)**	**X**	**O2 (B)**	**O2 (A)**

Figure 11.

Research Design of the Study on the Impact of Training for Prevention on Mental Health Promotion and Prevention Programs' Quality and Effectiveness

It is important to stress that programs were implemented in different periods of time between February and December 2011. In order to assess the effectiveness of individual programs, the author of this study prepared an outcome evaluation instrument for each of the 24 programs. Prepared measures were administered to organization's managers and program deliverers who organized group data collection from program participants in two time points – before and after program implementation. Research was anonymous for program participants. Dates of program outcomes assessments are presented in Appendix C (p. 218). Unfortunately, in two programs measures were not administrated to program participants: (2) Mentor program and (3) Parenting program I. These two programs were excluded from analysis of the Training for Prevention impact on programs' outcomes.

In September 2011, managers and deliverers of all programs were asked to write new proposals for programs they have implemented in the period between February and December 2011. In December 2011 and January 2012, the new program proposals were again assessed with Preffi 2.0 by three assessors in order to examine if the Training for Prevention had an impact on the quality of written program proposals.

3.3.3. Measures

Measures which were used in assessing program outcomes are presented in the table in Appendix D (p. 221). The Table presents the constructs which were measured in each program and contains a brief description of applied measures. As it is shown in the table, used measures focused on measuring short-term changes in specific behaviours, attitudes and skills of participants, depending on the objectives of each individual program. Some measures were already validated by different authors while others were developed by the author of this doctoral study. Alpha values of measures are presented in Appendix F (p. 236). It is important to stress that some additional measures were used in assessing program outcomes but they were excluded from further analysis because of very low reliability coefficients. However, it is still noticeable from Appendix F that Alpha values of some measures included in the analysis were rather low which can be due to the small number of participants in some programs. These measures were not excluded from analysis because they were found to be reliable in studies conducted by other authors. The research participants' results on each measure were calculated in accordance with procedures defined by the author of the measure.

For the purpose of assessing the impact of the Training for Prevention on the quality of written program proposals, the Preffi 2.0 was applied to the 24 new program proposals. The results on Preffi 2.0 were calculated in the same way as it was described in Section 3.2.

3.3.4. Content of the Training for Prevention

Training for Prevention consists of 32 hours of direct interactive group education and 3 hours of additional, individual consultations. Group education is based on lectures, group activities, exercises, case studies and is accompanied with continuous feedback from Training deliverers. Activities within group education follow a precise structure of five main topics:

1. Science-based mental health promotion and prevention practice (4 hours),
2. Logic modelling and quality (8 hours),
3. Implementation (8 hours),
4. Evaluation (8 hours), and
5. Advocacy (4 hours).

In this doctoral project, all themes were delivered within one month, leaving time between five group sessions during which participants could integrate the knowledge, work on assigned tasks and practice skills. Parallel with the group work and activities, Training deliverers had individual consultations with program leaders, authors and deliverers during which their learning process was discussed. Individual consultation included reflection on the tasks fulfilled during group training with feedback on the level of achieved quality. Special attention was given to the transfer of knowledge gained during the training into specifics of their program demands. Each topic covered several sub-areas, which are elaborated below.

1. **Science-based mental health promotion and prevention practice**

   Regarding the differences between participants' professional background, experience and level of education about mental health promotion and prevention, at the beginning of the Training for Prevention, participants were introduced to the recent concepts of prevention science and practice. This topic included lectures on theoretical models of mental health promotion, prevention, intervention continuum, risk and protective factors and mental health promotion and prevention effectiveness and cost-effectiveness.

2. Logic modelling and quality

The topic of Logic modelling and quality focused on the transfer of knowledge about all the phases and processes needed to develop a comprehensive and precise logic model of a program. During this phase, participants continuously worked on the development of a logic model for their own program. Firstly, the target group was informed about the importance of an elaborate problem analysis in program development and taught how to define problems they want to address with their program activities. This was followed by the topic of need assessment during which the connection between problem analysis and need assessment was emphasized. Participants were educated about the methods of need assessment, how to use available research, resources and data and conduct need assessment for their programs. In this phase, the difference between detected problems and existing needs was emphasised and explained on several examples. According to the results of need assessment, participants were taught how to precisely describe the target group which they want to include in their intervention. After gaining that knowledge, participants were instructed how to define specific and quality program goals, based on the conducted problem analysis, need assessment and analysis of available resources. Having defined goals, a description of short and long-term outcomes was given as a projection of how goals are to be achieved.

All described processes in this phase of the Training for Prevention focused on a better understanding of the theory behind the program. Participants were educated about the principles of person-centred theories of change inherent to each program. Participants analysed the causal assumptions behind their programs, were directed to connect their activities with the existing theoretical models and possibilities how to detect and overcome potential barriers in the process of project development and delivery. The importance of this part of the Training for Prevention was to raise participants' awareness of the role of all described elements in overall program quality.

3. Implementation

Discussing the topic of Implementation, participants were educated that implementation quality is crucial for programs effectiveness and quality. This part of the Training contained an overview of effect predictors related to the implementation process. For each type of prevention and promotion program, specific knowledge gathered from up to date research was transferred.

This phase was started with emphasizing the crucial role of professional capacities of program deliverers, which include their professional education, level of training and experience in delivering similar programs. Also, a possibility of in-service training in their organization was recommended as a method to enhance their professional capacities. This part of the Training described moderators of implementation quality, which included deliverer's motivation for conducting the program as well as beliefs and expectations of programs effectiveness. The Training also paid special attention to the development of group management skills, which increase the engagement of participants and their motivation to change. Further, it was stressed that providing constructive and continuous feedback from deliverer to the target group has a positive effect on outcomes. In addition, an added value to implementation quality was emphasized in providing organizational support to program deliverer through organizing supervisions and program monitoring, regular organizational meetings, involvement of an organizational manager in program implementation process and assuring administrative conditions. Program activities have to follow developmental trajectories of the target group in order to address crucial developmental demands relevant for individual change. Participants were educated how to tailor their activities according to the characteristics and needs of the target group, about what the optimal number of program participants is, the appropriate dosage of activities and which techniques are the most innovative and efficient for specific programs to achieve expected outcomes. Regarding outcomes enhancement, Training participants were directed to encourage their target group to practice skills and generalize content learned during the program to other social environments. It was discussed that standardization of program content and delivery model contributes to implementation quality as well. Finally, participants were encouraged to follow a fixed schedule of topics and activities in their programs, and to develop structured written materials and program manuals.

4. Evaluation

The introduction to the topic of evaluation emphasized the need for continuous and comprehensive assessment of program outcomes and quality of program implementation. The evaluation process, which includes defining evaluation concepts, theoretical overview of qualitative and quantitative indicators of program effectiveness, research methods of data collection and data sources, was thoroughly

described. Using a logic model as a starting point, participants were taught which steps they have to follow during the evaluation process. It was explained that concepts, which they have to measure, are defined within the logic model and program objectives that serve as indicators of desirable change. Based on program objectives, clear evaluation questions need to be formulated in a way to be measurable. Experimental and quasi-experimental designs were also presented. Regardless of the design used in evaluation research, measures which are planned to be used have to be standardized, reliable and in accordance with theoretical concepts behind their programs. Participants were informed about different types of evaluation (process and outcome evaluation, implementation quality research and costs analysis) and about how to develop an evaluation plan of their program. Regarding the stakeholders who might/should be involved in evaluation planning, the Training stressed the advantages of participative evaluation which stimulates the integration of science-based and practice-based principles, and the collaboration between program author, researcher, target groups and other stakeholders. Concerning the level of independence of the researcher, evaluation can be internal or external. Training deliverers encouraged participants to plan external evaluation and engage research experts in order to assure objective conclusions about program effectiveness. Training participants learned that adequate evaluation research enables them to gain an insight into successes and failures as well as to get an overview of gaps in the planning process and required improvements. Information gathered during the evaluation process is significant for program development and changes of program content and implementation, all of which can contribute to program sustainability in the community.

5. Advocacy

The final phase of the Training for Prevention focused on the role of advocacy for setting the conditions for success and quality of programs. Participants learned that through the process of advocacy they can assure sufficient resources for program development and implementation. Adequate funding, community support, networking and partnership are benefits of quality advocacy and assure program sustainability. All the above has an influence on increasing the visibility of the program and organization, affecting the motivation of target groups to participate in the program, and giving credibility to program deliverers/authors in the process of policy development. This part of the Training explained characteristics of quality advocacy,

steps of the advocacy process, starting with a clear definition of advocacy goals. The process of advocacy is a continuous process and its activities start before the program and last during program implementation and after the program is finished. Participants were trained how to detect stakeholders and decision makers they want to address, how to adjust the message and their interests to the interests of key people, to find a common language and to use key moments for lobbying (good timing). Special attention was given to methods of advocacy, especially to the usage of media for communicating the message and how to make more impact on the decision makers. Apart from lobbying, participants were taught how to recognize available funding resources.

3.4. Study on Predictive Validity of the Preffi 2.0 instrument

The study was conducted to examine whether higher quality of a written program proposal is related to program effectiveness, i.e. whether programs that accomplish higher total scores on Preffi 2.0 are more effective than programs with lower total scores on Preffi. This study was based on the assessments of the two previously described studies conducted within this doctoral research. Preffi scores of all 24 programs were estimated during the first application of Preffi 2.0 by assessors in December 2010 and January 2011 within the Study of quality of mental health promotion and prevention programs in Istria (see Section 3.1.2.). Effectiveness of all 24 programs was measured within the Study on the impact of the Training for Prevention on program effectiveness (see Section 3.3.).

3.5. Ethical principles of the research

The Ethical Committee of the Faculty of Education and Rehabilitation Sciences has confirmed that this doctoral study respects and that it is in accordance with ethical principles of research. During this study, several actions were taken in order to ensure compliance with research ethics. The research procedure was described and presented to the Department of Health and Social Services of the County of Istria after which the Department and the Faculty signed an agreement for conducting the project. Also, the procedure of the research was explained to managers of 24 organizations involved into the project. After they were informed about the project, each of the 24 included organizations signed an agreement with the Department of Health and Social Services and researchers from the Faculty. The purpose of

the agreement was to ensure their participation in all project activities and their commitment to follow ethical standards while conducting some of the project activities.

In order to assess the effectiveness of all mental health promotion and prevention programs in two measurements (pre-test and post-test), program managers asked participants of their programs for consent for their participation in the research. Since the participants of some of the assessed programs were children, their parents were asked to sign the consent for their children to participate in the research. Also, with the intention to ensure anonymity of organizations whose program effectiveness was assessed in this research, the author of this doctoral study used codes and alternative program names instead of their actual names.

As already mentioned, after the whole study and measurement were completed, the research team provided the Training for Prevention to organizations from control conditions. That was done in order to ensure that organizations from both groups, experimental and control, are provided with the same intervention before submitting applications to the Department for funding programs.

4. RESULTS AND DISCUSSION

4.1. Content Validity and Reliability of the Preffi 2.0 Instrument

Results of the Content Validity Study

The first research task of this doctoral study was to assess metric characteristics of Preffi 2.0 – content validity, reliability and predictive validity. Content validity refers to the extent to which a measure represents all facets of a given social construct (Lawshe, 1975). The main aim of this study was to assess if Preffi 2.0 items fit to the targeted concept. Ten mental health promotion and prevention experts from Croatia have therefore estimated the level to which each Preffi 2.0 item is theoretically connected with a specific quality indicator and essential for assessing that indicator, as discussed and specified in Section 1.3.2. The study was conducted to test the following hypothesis of the first research task:

> *Hypothesis 1.1.* The items of the Preffi 2.0 instrument are theoretically connected with specific quality indicators and are essential for assessing those indicators. The Content Validity Ratio of the whole Preffi 2.0 is 0.70 or higher.

Estimators marked their scores for each Preffi 2.0 item on a five-point Likert scale in which 1 stood for "Completely not in accordance with the theory and not essential", 2 for "Not in accordance with the theory and not essential", 3 for "Partially in accordance with the theory and essential", 4 for "Mostly in accordance with the theory and essential" and 5 stood for "Completely in accordance with the theory and essential". One widely used method of measuring content validity – **method of calculating Content Validity Ratio** - was developed by Lawshe (1975). It is essentially a method for assessing agreement among raters or judges regarding how essential a particular item is. Lawshe proposed that each subject matter expert rater (SMEs) on the judging panel respond to the following question for each item: "Is the skill or knowledge measured by this item 'essential,' 'useful, but not essential,' or 'not necessary' to the performance of the construct?" According to Lawshe, if more than half the panelists indicate that an item is essential, that item has at least some content validity. Greater levels of content validity exist as larger numbers of panelists agree that a particular item is essential.

The following equation for calculating the CVR has been used in this study (Lawshe, 1975):

$$CVR = (n_e - N/2)/(N/2)$$

where CVR =content validity ratio, n_e =number of SME panelists indicating "essential", N =total number of SME panelists.

While calculating the Content Validity Ratio (CVR) for each Preffi 2.0 item, assessors' scores were coded in a way that scores 4 and 5 were combined into one category "In accordance with the theory and essential". Table 4.1. presents the mean CVRs of eight Preffi 2.0 clusters and the mean CVR for the whole Preffi 2.0.

Table 4.1.

Mean Content Validity Ratios of the Eight Preffi Clusters and the Whole Preffi 2.0

PREFFI CLUSTERS	CVR
FIRST PREFFI CLUSTER "Contextual conditions and feasibility"	0.97
SECOND PREFFI CLUSTER "Problem analysis"	0.87
THIRD PREFFI CLUSTER "Determinants of behaviour and environment"	0.98
FOURTH PREFFI CLUSTER "Target group"	0.94
FIFTH PREFFI CLUSTER "Objectives"	0.97
SIXTH PREFFI CLUSTER "Intervention development"	0.85
SEVENTH PREFFI CLUSTER "Implementation"	0.83
EIGHTH PREFFI CLUSTER "Evaluation"	0.93
TOTAL PREFFI 2.0	0.90

A Content Validity Ratio can measure between -1.0 and 1.0 (Lawshe, 1975). The closer to 1.0 the CVR is, the more essential the item is considered to be in assessing the construct. A positive CVR value indicates that at least half of the subject matter expert raters rated the item as essential. The mean CVR across items may be used as an indicator of overall instrument content validity. As can be seen in Table 4.1., mean CVR of all 121 Preffi 2.0 items (total Preffi 2.0) is 0.90 which implicates that the instrument has a very high level of content

validity. This result also indicates that Preffi 2.0 items are considered very essential and in accordance with theory concepts on which the instrument is founded. If we analyse the mean CVR of eight Preffi clusters individually, results from Table 4.1. are showing that all eight Preffi clusters have a CVR between 0.83 and 0.98 which is generally a very high level of content validity. The clusters with the *highest CVR values* are the third Preffi cluster "Determinants of behaviour and environment" (CVR=0.98), the first Preffi cluster "Contextual conditions and feasibility" (CVR=0.97) and the fifth Preffi cluster "Objectives" (CVR=0.97). Two Preffi clusters with the *lowest but still rather high CVR values* are the seventh Preffi cluster "Implementation" (CVR=0.83) and the sixth Preffi cluster "Intervention development" (CVR=0.85).

In order to provide a deeper understanding of the content validity of Preffi 2.0, CVRs of each of the 121 Preffi 2.0 items were calculated and presented in Appendix E (p. 229). The number of assessors who have estimated each item are also shown in the table. Preffi items with CVRs lower than 0.60 are marked light blue, items whose CVRs are between 0.60 and 0.79 are marked purple while items with CVRs of 0.80 or higher are white.

If we look at the results presented in Appendix E, it is noticeable that most of the Preffi 2.0 items (75 of 121 items) achieved a CVR of 1.0, the highest possible score. It shows that all assessors have estimated that these items are connected with a specific quality indicator and that they are essential for assessing that indicator. Also, a high CVR value of 0.80 was achieved for 31 Preffi items. The relatively moderate CVR values, those between 0.60 and 0.79 were achieved for 12 Preffi 2.0 items. Only three Preffi items have a CVR lower than 0.60. That is one item of the sixth Preffi cluster – "Intervention development" and two items of the seventh Preffi cluster – "Implementation". The sixth Preffi cluster's item "If the intervention is to be implemented by intermediate groups: has the timing of the interventions been adapted to the intermediate groups?" has the lowest CVR of all assessed items, a CVR of 0.40. Items "Have specific objectives been set for each stage of the process of diffusion and use and for each intermediate target group or target group segment?" and "Do the implementation interventions fit in with the objectives that have been set for each stage of diffusion and use and for each intermediate target group or target group segment?" of the seventh Preffi cluster have a CVR of 0.55. It is important to stress that even though these three items have the lowest CVR values of all assessed items, at least half of the assessors have considered them essential and in accordance with the theory concept of Preffi 2.0.

In order to provide a deeper understanding of Preffi 2.0 content validity, a correlation analysis between individual Preffi 2.0 clusters and the total score on Preffi 2.0 was also conducted.

Table 4.2.

Correlations between Scores on Individual Preffi 2.0 Clusters and Total Score on Preffi

		CORRELATIONS								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	FIRST CLUSTER "Contextual conditions and feasibility"									
2	SECOND CLUSTER "Problem analysis"	.56**								
3	THIRD CLUSTER "Determinants of behaviour and environment"	.58**	.67**							
4	FORTH CLUSTER "Target group"	.46*	.56**	.55**						
5	FIFTH CLUSTER "Objectives"	.67**	.48*	.73**	.43*					
6	SIXTH CLUSTER "Intervention development"	.85**	.57**	.78**	.52*	.86**				
7	SEVENTH CLUSTER "Implementation"	.78**	.30	.50*	.23	.50*	.73**			
8	EIGHT CLUSTER "Evaluation"	.61**	.45*	.62**	.53*	.55**	.71**	.49*		
9	TOTAL PREFFI 2.0	.87**	.73**	.85**	.67**	.80**	.94**	.72**	.78**	

* $p<.05$; ** $p<.01$

For the purpose of this analysis, Preffi scores assessed during the first measurement of 24 programs with Preffi 2.0 were used. As it was expected, all Preffi 2.0 clusters are strongly and positively correlated with the total Preffi score. The highest correlation between total Preffi 2.0 score and individual cluster is the one with the sixth cluster "Intervention development" (r=.94, p<.01) and the first cluster "Contextual conditions and feasibility" (r=.87, p<.01) while the total score has the lowest, but still moderate and positive linear relationship with the "Target group" cluster (r=.67, p<.01). What is also noticeable from Table 4.2. is that there are strong and positive linear interrelationships between different Preffi 2.0 clusters.

The *first Preffi cluster "Contextual conditions and feasibility"* is strongly and positively correlated with the sixth "Intervention development" (r=.85, p<.01) and the seventh cluster "Implementation" (r=.78, p<.01). It has a strong, positive linear correlations with all other clusters except with the fourth cluster "Target group" (r=.46, p<.01).

The *second cluster "Problem analysis"* has the strongest, positive linear relationship with the third cluster "Determinants of behaviour and environment"(r=.67, p<.01) and it is significantly correlated with all clusters except with the seventh cluster "Implementation" (r=.30, p=.18).

The *third cluster "Determinants of behaviour and environment"* is positively correlated with all clusters. It has the strongest correlation with the sixth cluster "Intervention development" (r=.78, p<.01) while it has the weakest correlation with the seventh cluster "Implementation" (r=.50, p<.05).

Fourth Preffi 2.0 cluster "Target group" is correlated with all clusters except with the seventh cluster "Implementation" (r=.23, p=.31). It has the strongest positive linear relationship with the second cluster "Problem analysis" (r=.56, p<.01).

The *fifth cluster "Objectives"* is significantly correlated with all Preffi 2.0 clusters. It has the strongest correlation with the sixth cluster "Intervention development" (r=.86, p<.01) and the weakest correlation, but still moderately strong with the forth cluster "Target group" (r=43, p<.05).

The *sixth cluster "Intervention development"* is also significantly correlated with all other Preffi 2.0 clusters. It has very strong, positive correlation with the fifth cluster "Objectives" (r=.86, p<.01), first "Contextual conditions and feasibility" (r=.85, p<.01), and with the third cluster "Determinants of behaviour and environment" (r=.78, p<.01).

The *seventh cluster "Implementation"* is significantly and strongly correlated with all clusters except with the fourth "Target group" (r=.23, p=.31) and second cluster "Problem

analysis" (r=.30, p=.18). It has the strongest correlation with the first cluster "Contextual conditions and feasibility" (r=.78, p<.01).

The *eight Preffi 2.0 cluster, "Evaluation"*, is significantly correlated with all other Preffi 2.0 clusters. It has the lowest but still significant correlation with the second cluster (r=.45, p<.05) and with the forth cluster "Target group" (r=.53, p<.05). The eight cluster is strongly correlated with the sixth cluster "Intervention development" (r=.71, p<.01).

Discussion of the Results of Content Validity Study

According to the results of this study, all Preffi 2.0 clusters have a CVR between 0.83 and 0.98 which is generally a very high level of content validity ratio while the CVR of the total Preffi 2.0 is 0.90. The clusters with the *highest CVR values* are the third cluster "Determinants of behaviour and environment" (CVR=0.98), the first cluster "Contextual conditions and feasibility" (CVR=0.97) and the fifth cluster "Objectives" (CVR=0.97).

The third Preffi cluster reflects the quality level of the program's theoretical model, description of contributions of determinants to the problem, amenability of factors to change and the quality of how determinants are prioritized and selected. The first cluster describes the quality of support and commitment of internal and external partners, capacities for the project, leadership by the project manager including expertise and characteristics of the manager. The fifth cluster is assessing if program objectives are fitting in with problem analysis, if they are specific, specified in time and measureable, but also if they are acceptable to the main stakeholders and feasible. It also describes whether objectives are considered achievable given the available resources, contextual conditions and intended period of time. These three broader concepts are often mentioned as predictors of programs effectiveness (Jane-Llopis and Barry, 2005; Nation et al., 2003; Bartholomew et al., 2001; Ader et al., 2001; Brown et al., 2000; Kok et al., 1997). It is reasonable to expect that prevention experts who were involved in this study considered this clusters' items theoretically connected with a specific quality indicator and essential for assessing that indicator.

Two Preffi clusters with the *lowest but still rather high CVR values* are the seventh Preffi cluster "Implementation" (CVR=0.83) and the sixth Preffi cluster "Intervention development" (CVR=0.85). The seventh cluster reflects the model of implementation, fitness of implementation of interventions to the intervention deliverers, appropriateness of the supplier for intermediating intervention deliverers, monitoring and generating feedback and incorporation of the intervention in an existing organizational structure. The sixth Preffi

cluster describes the rationale of the intervention strategy, previous experience with intervention, duration, intensity and timing of the intervention, fitting in the target group and in the culture, participation of the target group and usage of effective techniques. It also shows the feasibility in existing practice, characteristics of implementability of the intervention and coherence of interventions/activities. As stated in the introduction of this dissertation, awareness of the importance of implementation quality for achieving desirable outcomes of the program is growing, but it is still not high enough. Recently, more and more studies in the field of mental health promotion and prevention science have focused on the quality of the process of implementation and how it affects program effectiveness (Domitrovich et al., 2010; Durlak, 1998; Durlak and DuPre, 2008; Dusenbury et al. 2005; Greenberg, Domitrovich, Graczyk and Zins, 2001). It can be expected that with increased awareness of the importance of implementation process quality, mental health promotion and prevention experts will in the future consider these Preffi items as more essential in assessing a program quality. These results could also be explained with a possibility that the terminology regarding the implementation process is rather new and should be more clarified to mental health promotion and prevention experts from Croatia while using Preffi 2.0. Findings gained in this study are very important for the process of Preffi 2.0 adaptation in Croatia.

If we analyse CVRs of individual Preffi 2.0 items, it is interesting to notice that almost all items achieve CVR values higher than 0.60. The relatively moderate CVR values, those between 0.60 and 0.79 were achieved for 12 Preffi 2.0 items. Only three Preffi items have a CVR lower than 0.60 which still implicates that at least half of the assessors have considered them essential and in accordance with the theory concept of Preffi 2.0. In the case of Preffi 2.0 items with the lowest CVR values, it is very much possible that the assessors could not understand the meaning of the term "intermediate target group". These three items should be formulated more precisely and clearly in future versions of Preffi 2.0 if it will be offered to Croatian professionals.

The results of the correlation study shows that all individual Preffi 2.0 clusters are strongly and positively correlated with the total Preffi 2.0 score. As explained in the introduction to this dissertation, Preffi 2.0 consists of 39 quality criteria – effect predictors distributed within eight Preffi 2.0 clusters which all reflect a broader concept of program quality. Because of that, high correlations between different Preffi 2.0 clusters and between clusters and total score on Preffi 2.0 are understandable and expected.

All presented results show that **Hypothesis 1.1 is confirmed** in this study. The items of the Preffi 2.0 instrument are theoretically connected with specific quality indicators and are essential for assessing those indicators. The Content Validity Ratio of the whole Preffi 2.0 is 0.90 which is higher than 0.70.

Results of Reliability Study

As described in Section 3.1.2. (p. 54), the written proposals of mental health promotion and prevention programs from the County of Istria were assessed with Preffi 2.0 at two time points. During December 2010 and January 2011, three assessors read the 24 written program proposals and independently assessed each program with the Preffi 2.0 instrument. A year after the first assessment of the 24 program proposals, program developers and authors were asked to write new proposals for the programs assessed in the first assessment. During December 2011 and January 2012, 21 new program proposals were again independently assessed by three assessors. Results from both applications of Preffi 2.0 on written program proposals were used in analysing Preffi 2.0 reliability. This study was conducted to test the following hypothesis of the first research task:

> *Hypothesis 1.2*. Preffi 2.0 is a reliable instrument with at least a medium or high value of concordance between three assessors (G= 0.70 or higher).

In order to assess the reliability of Preffi 2.0, concordance between the three assessors in assessing the written program proposals in two time points was calculated. The reliability of Preffi 2.0 was analysed by applying the **generalizability theory** and calculating the generalizability coefficient (G) and the standard error of measurement – SEM (Brennan, 2001). The variance components of the facets and their interactions are used to estimate the true score variance (Var(T)) and the error variance (Var(E)). The true score variance consists of the variance in the discrimination facets and their interactions with each other. The error score variance consists of the variance in the instrumentation facets, their interactions with each other and their interactions with the discrimination facets.

The G coefficient or reliability coefficient is then calculated as follows:

$$G = \frac{\text{var}(T)}{Var(T) + Var(E)}$$

The Standard Error of Measurement is the square root of error variance:

$$SEM = \sqrt{Var(E)}$$

Cronbach's Alpha could not be used as a reliability estimate as both raters and items may contribute to the measurement error. While Cronbach's Alpha is only applicable in situations where there is only one source of measurement error, the generalizability theory accommodates complex measurement designs with more sources of error. In an ideal situation, the differences in scores can be attributed to differences between the objects assessed, in this case the program, but they may also be caused by various error sources, like different views of the assessors or the interaction between programs and assessors. This theory, an extension of the classical test theory, addresses the question how accurately the criterion scores can be generalized across factors contributing to the scores. It checks the extent to which differences in measurements (i.e., the scores assigned per criterion, per cluster or to the entire program) can be attributed to true score components (in this case the characteristics of the programs' scores) or to potential error sources, in this case the assessors and the measurement instrument at criterion and cluster level (Shavelson and Webb, 1991; Webb, Shavelson and Haertel, 2006).

The G coefficient indicates the ratio between the true and total variance while SEM reflects the accuracy level (Molleman, 2005). Accuracy in statistics is defined as closeness of agreement between a measured value and a true value (ISO, 1993; Taylor, 1997). G is very similar to Cronbach's Alpha and is interpreted the same way. The conventional minimum reliability threshold for reliability coefficients like G or Cronbach's Alpha is 0.70 (Nunnally and Berstein, 1994). However, according to DeVellis (1991), the minimum reliability threshold for reliability coefficients such as G is the one between 0.65 and 0.70. There is no generally accepted maximum value for SEM. Molleman and colleagues (2005b) stressed that a reasonable requirement would be that the confidence interval of the score remains within the

rounding zone for the score category, that is, half a point above or below the category. In other words, the confidence interval around a score of 2 does not include scores <1.5 or >2.5. To achieve this level of precision, SEM needs to be <0.26.

For the purpose of this study, G coefficients and SEM values were computed on the different levels of aggregation: for each of the eight Preffi 2.0 clusters, for subclusters and for the whole Preffi 2.0.

Table 4.3. shows the reliability (G) and accuracy (SEM) of the programs assessments by the three assessors during the first measurement with Preffi 2.0. It also shows the relative size of the variance estimates for each source of variance – program, assessors and interaction effect between program and assessors for each cluster, subcluster and for Preffi 2.0 as a whole.

Table 4.3.

Reliability and Accuracy Values per Preffi 2.0 Clusters, Subclasters and Preffi 2.0 as a Whole - First Assessment

PREFFI CLUSTERS AND SUBCLUSTERS	G	SEM	% of total variance		
			PROGRAM	ASSESSORS	PROGRAM x ASSESSORS
FIRST PREFFI CLUSTER "Contextual conditions and feasibility"	**.70**	**.77**	**33.9**	**22.6**	**43.5**
1.1 Support/commitment	.50	.36	17.3	31.5	51.2
1.2 Capacity	.50	.35	25.2	0.5	74.3
1.3a Expertise and characteristics of the project Manager	.68	.32	36	12.2	51.8
1.3b Focal points for the leadership	.60	.35	30.1	9.6	60.3
SECOND PREFFI CLUSTER "Problem analysis"	**.60**	**.80**	**32,1**	**3.8**	**64.2**
2.1 Nature, severity, scale of the problem	.65	.28	31.9	16.4	51.7
2.2 Distribution of the problem	.46	.37	20.6	6.1	73.3
2.3 Problem perception by stakeholders	.36	.28	15.3	3.9	80.8
THIRD PREFFI CLUSTER "Determinants of behaviour and environment"	**.67**	**.67**	**37.4**	**7.5**	**55.1**
3.1 Theoretical model	.51	.35	24.9	4.2	70.9
3.2 Contributions of determinants	.46	.27	19.8	9.5	70.7
3.3 Amenability of determinants to change	.70	.27	44	0.0	56
3.4 Priorities and selection	.50	.32	20.8	15.7	63.5
FOURTH PREFFI CLUSTER "Target group"	**.62**	**.62**	**25.7**	**27.4**	**46.9**
4.1 General and demographic characteristics	.58	.30	21.7	31.1	47.2
4.2 Motivation and possibilities of the target group(s)	.77	.15	53.3	0.0	46.7
4.3 Accessibility of the target group(s)	.47	.31	17.5	22.5	60
FIFTH PREFFI CLUSTER	**.53**	**.76**	**17.7**	**34.6**	**47.7**

“Objectives”					
5.1 Objectives fit in with the analysis	.69	.31	35.5	17.3	47.1
5.2 Objectives: specific, specified in time, measurable	.34	.29	12.8	14	73.3
5.3 Objectives are acceptable	.00	.30	0.0	41.6	58.4
5.4 Objectives are feasible	.37	.37	11.5	30.6	57.9
SIXTH PREFFI CLUSTER “Intervention development”	**.78**	**.55**	**38.6**	**28.5**	**32.9**
6.1a Fitting of strategies/methods with target, target groups	.58	.27	24.6	22.8	52.6
6.1b Previous experiences with the intervention(s)	.52	.32	22.2	17.4	60.4
6.2a Duration and intensity of the intervention	.33	.31	13.3	6.6	80.2
6.2b Timing of the intervention	.39	.34	12.4	28.7	58.8
6.3a Participation of the target group	.52	.32	24	8.7	67.2
6.3b Fitting to the 'culture'	.42	.30	15.9	17.8	66.2
6.4 Effective techniques (recommended)	.66	.31	33.2	15.5	51.3
6.5a Fitting to the intermediary target groups	.49	.42	18.6	23.8	57.6
6.5b Characteristics of implementability of intervention(s)	.00	.32	0.0	30.8	69.2
6.6 Coherence of interventions/activities	.73	.31	44.4	6.7	48.8
SEVENTH PREFFI CLUSTER ”Implementation”	**.63**	**.61**	**23.5**	**34.7**	**41.8**
7.1a Model of implementation: top-down, bottom-Up	.21	.28	6.6	18.8	74.5
7.1b Fitting the implementation of the intervention to the intermediaries	.11	.27	3.5	14.7	81.8
7.1c Appropriateness of supplier for the intermediaries	.36	.38	9.6	38.1	52.2
7.2 Monitoring and generating feedback	.76	.29	49.5	4.5	46
7.3 Incorporation in the existing structure	.11	.37	2.2	44	53.8
EIGHT PREFFI CLUSTER “Evaluation”	**.86**	**.42**	**55.8**	**16.5**	**27.7**

8.1 Clarity and agreement on principles of the Evaluation	.60	.26	30.4	7.8	61.7
8.2 Process evaluation	.61	.25	27.3	21.5	51.2
8.3a Has a change been measured?	.71	.24	38.5	14.9	46.5
8.3b Was the change caused by the intervention?	.59	.09	32.4	1.5	66.2
8.4 Feedback to the stakeholders	.72	.32	43	6.5	50.5
TOTAL PREFFI SCORE	**.79**	**.44**	**39.2**	**30.5**	**30.3**

Table 4.3. shows that the reliability coefficient (G) in the first assessment with Preffi 2.0 for eight clusters ranges from G=.53 to G=.86. The cluster with the *highest reliability coefficient* from the first measurement is the eighth Preffi 2.0 cluster, "Evaluation", (G=.86, SEM=.42) which has a very high level of reliability. The sixth cluster - "Intervention development" - also has a high level of reliability (G=.78, SEM=.55). The first cluster "Contextual conditions and feasibility" is also found as highly reliable (G=.70, SEM=.77). The third Preffi 2.0 cluster, "Determinants of behaviour and environment", has a reliability level of G=.67 (SEM=.67).

Clusters with *reliability levels lower than .65* are the seventh cluster "Implementation" (G=.63, SEM=.61), the fourth cluster "Target group" (G=.62, SEM=.62), the second Preffi 2.0 cluster "Problem analysis" (G=.60, SEM=.80) and the fifth cluster "Objectives" (G=.53, SEM=.76).

Even though some Preffi 2.0 clusters had a low reliability level in the first assessment of programs, the reliability of the whole Preffi 2.0 was found to be very high (G=.79, SEM=.44). Preffi 2.0 was assumed to yield accurate conclusions if the standard error or measurement (SEM) was below .26. Calculation of SEM values across the three assessors for each program showed that they were higher than .26 on all Preffi 2.0 clusters and on Preffi 2.0 as a whole.

If we analyse the data on the *source of variance*, the program as a source of variance is an estimate of the true score variance which ideally should be as high as possible. The program showed to be the largest source of variance for three of eight clusters – the eighth cluster "Evaluation" (55.8%), the sixth cluster "Intervention development" (38.6%) and the first cluster "Contextual conditions and feasibility" (33.9%). The variance attributable to differences in views between the assessors should be as low as possible. It is the lowest source of variance for five clusters – the second "Problem analysis" (3.8%), first "Contextual conditions and feasibility" (22.6%), third "Determinants of behaviour and environment" (7.5%), sixth "Intervention development" (28.5%) and the eighth cluster "Evaluation" (16.5%). For six Preffi 2.0 clusters, most of the variance resulted from the interaction between programs and assessors – the first cluster "Contextual conditions and feasibility" (43.5%), the second cluster "Problem analysis" (64.2%), the third cluster "Determinants of behaviour and environment" (55.1%), the fourth cluster "Target group" (46.9%), the fifth cluster "Objectives" (47.7%) and the seventh cluster "Implementation" (47.8%). If we look at the sources of variance for the whole Preffi 2.0, it is noticeable that the program is the largest source of variance for the total score of Preffi 2.0 (39.2%), followed by variance attributable

to differences in views between the assessors (30.5%) and the combination of programs and assessors (30.3%).

Table 4.4. represents the reliability and accuracy of the assessments by the three assessors gathered during the second measurement with Preffi 2.0. It also shows the relative size of the variance estimates for each source of variance - program, assessors and interaction effect between program and assessors for each cluster, subclusters and for Preffi 2.0 as a whole.

Table 4.4.

Reliability and Accuracy Values per Preffi 2.0 Clusters, Subclasters and Preffi 2.0 as a Whole - Second Assessment

PREFFI CLUSTERS AND SUBCLASTERS	G	SEM	% of total variance		
			PROGRAM	ASSESSORS	PROGRAM x ASSESSORS
FIRST PREFFI CLUSTER "Contextual conditions and feasibility"	**.68**	**.73**	**32.1**	**23.7**	**44.3**
1.1 Support/commitment	.50	.27	14.5	41.7	43.8
1.2 Capacity	.50	.31	22.7	10.1	67.2
1.3a Expertise and characteristics of the project manager	.46	.28	18.8	15.4	65.8
1.3b Focal points for the leadership	.68	.34	39.6	4.9	55.5
SECOND PREFFI CLUSTER "Problem analysis"	**.70**	**.79**	**37.6**	**15**	**47.4**
2.1 Nature, severity, scale of the problem	.60	.34	25.7	24.1	50.2
2.2 Distribution of the problem	.73	.31	43.2	9.7	47.1
2.3 Problem perception by stakeholders	.48	.37	22.1	6.3	71.6
THIRD PREFFI CLUSTER "Determinants of behaviour and environment"	**.77**	**.75**	**49.1**	**8**	**42.9**
3.1 Theoretical model	.68	.37	41.3	0.2	58.5
3.2 Contributions of determinants	.74	.30	46.5	4.2	49.3
3.3 Amenability of determinants to change	.73	.31	39.9	15.4	44.7
3.4 Priorities and selection	.14	.35	4.6	9.2	86.3
FOURTH PREFFI CLUSTER "Target group"	**.59**	**.66**	**28.8**	**10.1**	**61.1**
4.1 General and demographic characteristics	.10	.34	3	16	81
4.2 Motivation and possibilities of the target group(s)	.53	.28	25.4	7	67.5
4.3 Accessibility of the target group(s)	.54	.26	24.2	14.2	61.6
FIFTH PREFFI CLUSTER	**.61**	**.80**	**32.3**	**5.6**	**62.2**

"Objectives"					
5.1 Objectives fit in with the analysis	.55	.35	29.1	0.0	70.9
5.2 Objectives: specific, specified in time, measurable	.60	.31	31.4	5	63.6
5.3 Objectives are acceptable	.41	.28	17.1	9.8	73.2
5.4 Objectives are feasible	.61	.30	34.1	0.4	65.5
SIXTH PREFFI CLUSTER "Intervention development"	**.90**	**.47**	**59.8**	**19.6**	**20.6**
6.1a Fit of strategies/methods to target, target groups	.64	.29	37.5	0.0	62.5
6.1b Previous experiences with the intervention(s)	.60	.28	33	1.3	65.6
6.2a Duration and intensity of the intervention	.83	.27	57.5	7.3	35.3
6.2b Timing of the intervention	.80	.25	39.9	29.4	30.7
6.3a Participation of the target group	.71	.31	30.8	31.4	37.7
6.3b Fitting to the 'culture'	.46	.30	8.9	59.7	31.4
6.4 Effective techniques (recommended)	.68	.37	41.1	0.0	58.9
6.5a Fitting to the intermediary target groups	.56	.35	29.1	3.7	62.7
6.5b Characteristics of the implementability of intervention(s)	.70	.29	35.2	20.6	44.2
6.6 Coherence of interventions/activities	.71	.32	44.9	0.0	55.1
SEVENTH PREFFI CLUSTER "Implementation"	**.68**	**.59**	**29.6**	**29**	**41.4**
7.1a Model of the implementation: top-down, bottom-up	.67	.24	18.8	53.4	27.7
7.1b Fitting the implementation of the intervention to the intermediaries	.24	.31	4.4	53.4	42.2
7.1c Appropriateness of supplier for the Intermediaries	.00	.39	0.0	0.0	100
7.2 Monitoring and generating feedback	.39	.44	17.5	0.0	82.5
7.3 Incorporation in the existing structure	.57	.30	29.1	5.6	65.3
EIGHT PREFFI CLUSTER "Evaluation"	**.83**	**.59**	**48.8**	**21**	**30.3**

8.1 Clarity and agreement on principles of the evaluation	.68	.31	39.3	5.6	55.1
8.2 Process evaluation	.70	.31	31.1	28.9	40
8.3a Has a change been measured?	.82	.26	60	0.0	40
8.3b Was the change caused by the intervention?	.64	.22	36.8	0.0	63.2
8.4 Feedback to the stakeholders	.62	.32	20.4	43	36.7
TOTAL PREFFI SCORE	**.85**	**.43**	**51.1**	**21.4**	**27.5**

Results presented in Table 4.4. show that the reliability coefficient (G) in the second assessment with Preffi 2.0 for eight clusters range from G=.59 to G=.90. Six of eight clusters have a higher reliability coefficient than the minimum acceptable level of reliability .65. Clusters with the *reliability coefficient lower than the minimum acceptable level* are the fourth cluster, "Target group" (G=.59, SEM=.66), which has the lowest reliability level of all clusters and the fifth cluster, "Objectives" (G=.61, SEM=.80). The first cluster, "Contextual conditions and feasibility" (G=.68, SEM=.73), and the seventh cluster, "Implementation" (G=.68, SEM=.59), also have a lower, but still acceptable reliability level.

The cluster with the *highest reliability coefficient* in the second measurement is the sixth cluster, "Intervention development" (G=.90, SEM=.47). It is important to stress that the reliability of the whole Preffi 2.0 in the second assessment was found to be very high (G=.85, SEM=.43).

Just as in the first assessment with Preffi 2.0, it is noticeable that in the second assessment calculation of SEM values across the three assessors for each program showed that they were higher than .26 on all Preffi 2.0 clusters and on Preffi 2.0 as a whole.

If we analyse the *source of variance* in the second assessment with Preffi 2.0, the program showed to be the largest source of variance for three of eight clusters – the third cluster "Determinants of behaviour and environment" (49.1%), the sixth "Intervention development" (59.8%) and the eighth cluster "Evaluation" (48.8%). It is important to notice that the variance attributable to differences in views between the assessors is the lowest source of variance for all Preffi 2.0 clusters. It ranges from 5.6% to 29%. For five Preffi 2.0 clusters, most of the variance resulted from the interaction between programs and assessors – the first cluster "Contextual conditions and feasibility" (44.3%), the second cluster "Problem analysis" (47.4%), the fourth cluster "Target group" (61.1%), the fifth cluster "Objectives" (62.2%) and the seventh cluster "Implementation" (41.4%).

If we look at the sources of variance for the whole Preffi 2.0 in the second assessment, it is noticeable that the program showed to be the largest source of variance for the total score of Preffi 2.0 (51.1%), it is followed by the variance attributable to the combination of a program and assessors (27.5%) while the variance attributable to the differences in views between the assessors is the lowest (21.4%) source of variance.

Discussion of Results of Preffi 2.0 Reliability Study

Results from this study show that the Preffi 2.0 as a whole was found to be highly reliable in assessing the quality of programs during two assessments conducted within this study (G1=.79, SEM1=.44 and G2=.85, SEM2=.43). According to presented results, **Hypothesis 1.2 is confirmed** - Preffi 2.0 as a whole is a reliable instrument with at least a medium or high value of concordance between three assessors (G= 0.70 or higher).

It is noticeable that the whole Preffi 2.0 was slightly more reliable in the second assessment compared to the first application of the instrument. From tables 4.3. and 4.4. can be seen that the reliability coefficients (G) for eight clusters in the first assessment with Preffi 2.0 ranged from G=.53 to G=.86 while their range in the second assessment was higher, from G=.59 to G=.90. It is possible that higher values of the reliability coefficient in the second assessment are caused by the assessors' experience in applying Preffi 2.0 in the first assessment. In the first assessment with Preffi 2.0, four clusters had reliability levels lower than the minimally acceptable level of .65 - "Implementation" (G=.63, SEM=.61). "Target group" (G=.62, SEM=.62), "Problem analysis" (G=.60, SEM=.80) and "Objectives" (G=.53, SEM=.76). In the second assessment, two clusters had reliability coefficients lower than the minimum acceptable level - "Target group" (G=.59, SEM=.66) and "Objectives" (G=.61, SEM=.80). An extremely low reliability level was found for the first subcluster of the "Target group" cluster - General and demographic characteristics (G=.10). According to this result, future studies on Preffi 2.0 should involve a factor analysis of the instrument; special attention should be given to these two clusters in order to increase their reliability.

Lower reliability levels of some Preffi 2.0 clusters and subcluster might be explained by the fact that the reliability of a measure depends on the number of items. Many items of some Preffi 2.0 clusters and subclusters are rather low, which implies more threats to reliability.

Molleman and colleagues (2005b) explain that certain improvements can be conducted to increase Preffi's reliability. Firstly, according to these authors assessors should discuss their assessments of programs in a consensus meeting. Secondly, a consultation with the program leader should be included in the assessment procedure in order to supply information that was lacking in the project description. This is particularly true for aspects relating the program management, support and commitment, and the question whether the right people have been approached or involved, which is difficult to assess purely on the basis of program descriptions. Finally, authors stress that adequate scoring of some Preffi criteria requires an expert's opinion. This aspect cannot be adequately addressed in a list of general guidelines

like Preffi; it means that the general insights produced by Preffi will have to be specified for certain themes. Such studies can then indicate relevant theories for the specific domain, suitable interventions and specific success and failure factors. In the future, expert systems on the Internet, linked to a digital version of Preffi, could provide this.

Zhu and Han (2011) emphasize that in improving the measurement reliability it is important to train the scorers in using the instrument. This point is particularly important when scoring is very subjective. The scoring of the composition, for example, should not be assigned to anyone who has not learned to score compositions accurately before. They also stress that patterns of scoring should be analysed. Individuals whose scoring deviates markedly and inconsistently from the norm should be taught how to do it correctly. In addition, a detailed scoring key should be provided. The same authors also suggest that assessed subjects be identified by number instead of names. Scorers inevitably have expectations of candidates whom they know. Facts have shown that even when the candidates are unknown to the scorers, the name on a script will make a significant difference to the scores given. The identification of the candidates by number will reduce such effects. It is also important for the assessors to get items with a clear meaning. The best way to get unambiguous items is to draft and check the items, and then subject them to the critical scrutiny of colleagues, who should try as hard as they can to find alternative interpretations to the ones intended. In this way most of the problems can be identified before the test is administered.

Regarding the high SEM values found in both assessments with Preffi 2.0, it is important to stress that Molleman and his colleagues (2005b) have found that the accuracy of Preffi 2.0 is not yet sufficient to allow a project to be assessed by only one assessor. In one of their studies on Preffi 2.0 they have found that the number of assessors needed for accuracy levels lower than .26 ranges from 4 to 12 for the various Preffi 2.0 criteria, with an average of 6.36 assessors per criterion. Other study they have conducted have shown that 3 assessors are sufficient for accurate assessment with Preffi on a cluster level. Accuracy of assessment with Preffi 2.0 could be improved by the provision of training to assessors and by more exact scoring instructions. Also, it seems that some Preffi 2.0 items need to be more unambiguous.

Concerning the source of variance for the whole Preffi 2.0 score, it is noticeable that the program is the largest source of variance for the total score of Preffi 2.0 in both assessments. Variance attributable to the differences in views between the assessors was higher in the first

assessment (30.5%) compared to the second (21.4%). It is possible that this variance could be decreased if the assessors received training in applying the Preffi 2.0 instrument.

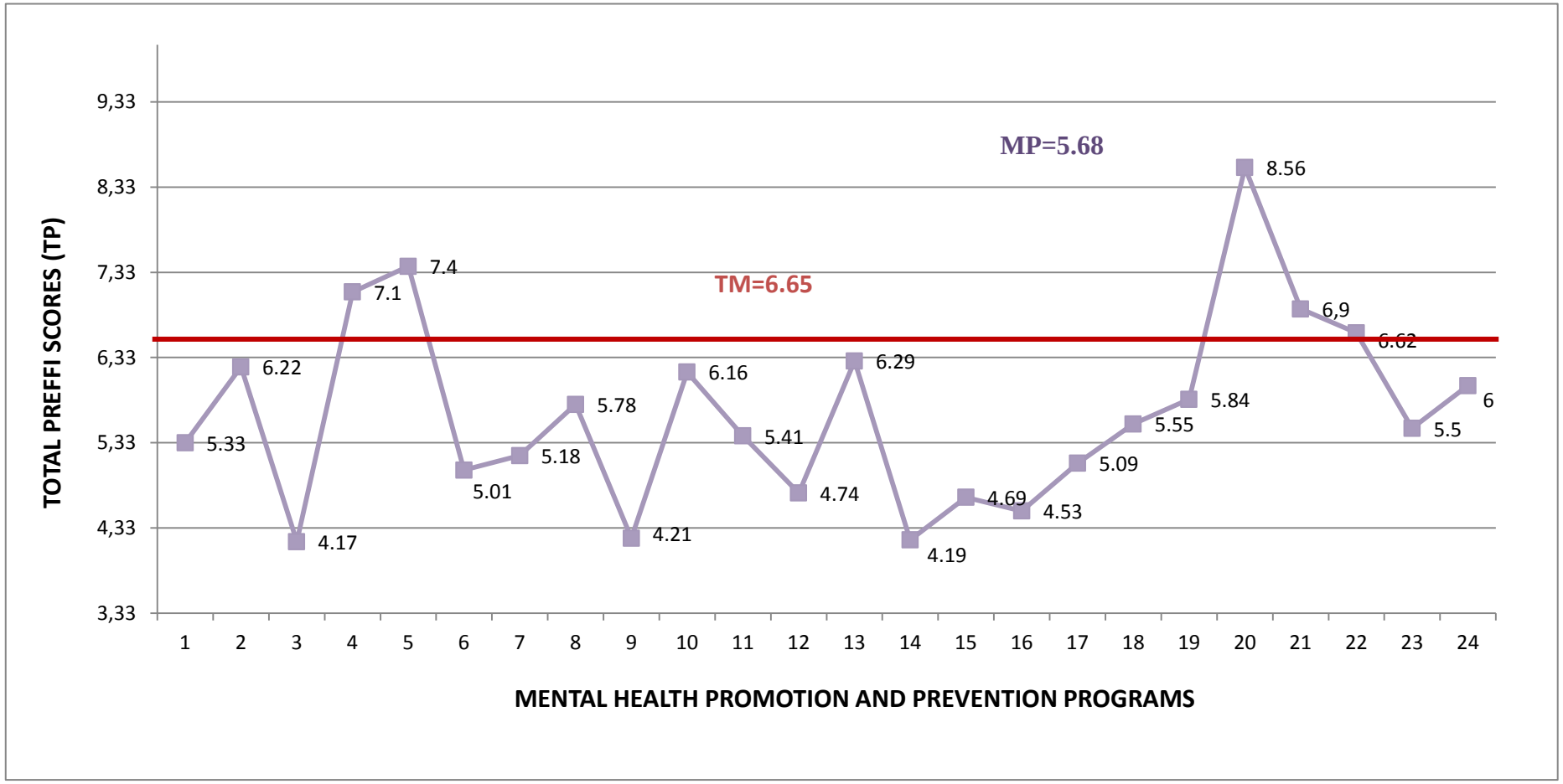
4.2. Quality of Mental Health Promotion and Prevention Programs in Istria

Results of the Study

The third research task of this doctoral study was to identify strengths and weaknesses of the mental health promotion and prevention programs from the County of Istria. It was expected that the application of the Preffi 2.0 instrument on 24 mental health promotion and prevention programs would provide information about their quality in general and identify strengths and weaknesses of the assessed programs.

As described in Section 3.2. (p. 61), during December 2010 and January 2011, three experts from the field of mental health promotion and prevention assessed a representative sample of programs from the County of Istria with the Preffi 2.0 Instrument. Programs were selected from the 2011 applicants for financial support from the County of Istria's Department of Health and Social Care initiative entitled "Prevention of Behavioural Disorders and Prevention of Substance Abuse". All 24 programs which were involved in this study are described in Table 3.1. (p. 56).

Figure 12 presents the **total Preffi scores** of 24 mental health promotion and prevention programs from the County of Istria assessed during December 2010 and January 2011. The total Preffi scores (TP) represent the general level of a program's quality and range from 3.33 to 10. As it is evident from Figure 12, total Preffi scores for the 24 assessed programs vary between the minimum of 4.17 and the maximum of 8.56 (SD=1.10). The program with the highest total Preffi score in this sample of programs is program (20) "Mental health promotion volunteerism". The program with the lowest total Preffi score is program (3) "Parenting program I".



TM= theoretical mean of possible total Preffi scores' range; MP= average mean of total Preffi scores for these 24 programs

Figure 12.

Total Preffi Scores of 24 Mental Health Promotion and Prevention Programs from the County of Istria

Within this study Preffi 2.0 has been used in Croatia for the first time and no quality norms have been established yet, neither in Croatia nor the Netherlands where it was developed. For this reason the theoretical mean (TM=6.65) of the possible **total Preffi scores range** was used as a quality reference point. Taking this reference point, only 4 of the 24 assessed programs score above the criterion value (i.e., programs 20, 5, 4 and 21). These programs are "MH promotion through volunteerism" (TP=8.56), "Training for group leaders" (TP=7.4), "Media literacy" (TP=7.1) and the program "MH promotion through dance" (TP=6.9). Programs which have the lowest total Preffi score are programs 3, 14 and 9. These programs are "Parenting program I" (TP=4.17) which is followed by the "Parenting program V" (TP=4.19) and "Parenting programme III" (TP=4.21).

The mean total Preffi score (MP) across all 24 Istrian programs is 5.68. This overall quality level of all assessed written program proposals is almost one standard deviation (-.97 SD) below the chosen quality criterion, the theoretical mean (TM) of 6.65.

Besides calculating the total Preffi scores, it is possible to gain an insight in the **mean scores of each of the eight Preffi clusters**. Mean Preffi clusters' scores provide information on which of the program's concepts reflected in the eight Preffi clusters are of the highest or the lowest quality in a certain group of assessed programs. Table 4.5. represents average scores for each Preffi cluster based on individual cluster scores of all 24 programs. This table shows the differences in quality level between eight Preffi clusters.

Table 4.5.

Mean Scores on the Eight Preffi Clusters

for 24 Mental Health Promotion and Prevention Programs from the County of Istria

PREFFI CLUSTERS	M	SD	MIN.	MAX.
FIRST PREFFI CLUSTER "Contextual conditions and feasibility"	6.63	1.56	4.17	10.00
SECOND PREFFI CLUSTER "Problem analysis"	4.76	1.45	3.30	7.78
THIRD PREFFI CLUSTER "Determinants of behaviour and environment"	4.86	1.47	3.33	8.33
FOURTH PREFFI CLUSTER "Target group"	6.12	1.39	3.30	8.89
FIFTH PREFFI CLUSTER "Objectives"	7.46	1.06	5.83	9.17
SIXTH PREFFI CLUSTER "Intervention development"	5.66	1.36	3.40	8.48
SEVENTH PREFFI CLUSTER - "Implementation"	5.42	1.39	3.33	8.00
EIGHT PREFFI CLUSTER "Evaluation"	4.58	1.32	3.33	8.67

Scores on each of the Preffi 2.0 clusters could range from 3.33 to 10. From scores presented in Table 4.5 is visible that only the fifth Preffi cluster representing *"Objectives"* (C5=7.46, SD=1.06) has a mean score higher than the quality theoretical mean (TM=6.65). It is also noticeable that the first Preffi cluster score which reflects the *"Contextual conditions and feasibility"* is very close to the theoretical mean (C1=6.63, SD=1.56). The remaining six clusters achieve scores below the theoretical mean. The lowest mean cluster score was achieved on the cluster *"Evaluation"* (C8=4.76, SD=1.45). Other Preffi clusters which also show very low quality levels in this sample of assessed programs are the second cluster *"Problem analysis"* (C2=4.81, SD=1.49) and the third cluster *"Determinants of problems, behaviour or environment"* (C3=4.86, SD=1.47).

Besides assessing each of the eight clusters' general quality, the Preffi allows a more specific understanding of each of the clusters' quality. Each Preffi cluster consists of multiple subscales which represent specific effect predictors connected with the cluster. Because of that, it is possible to evaluate the **quality of particular subscales of the eight Preffi clusters** in each program individually and across the total group of programs as it is the case in this doctoral study. For instance, within the 'Objectives' cluster one of the subscales specifically evaluates the estimated acceptability of program objectives to main stakeholders. Subscales' scores could range from 1 to 3. Since there are no quality norms established yet, the theoretical mean of 1.5 was used as a minimal quality norm. Subscales with mean scores below the theoretical mean are considered as program elements which have a low quality and represent weak points of the assessed programs. At the same time, subscales with mean score above 2 were considered as program elements of high quality representing strengths of the assessed programs from the County of Istria. Figure 13 shows that all subscales of the first Preffi cluster "Contextual conditions and feasibility" achieved mean scores higher than the theoretical mean (TM=1.5).

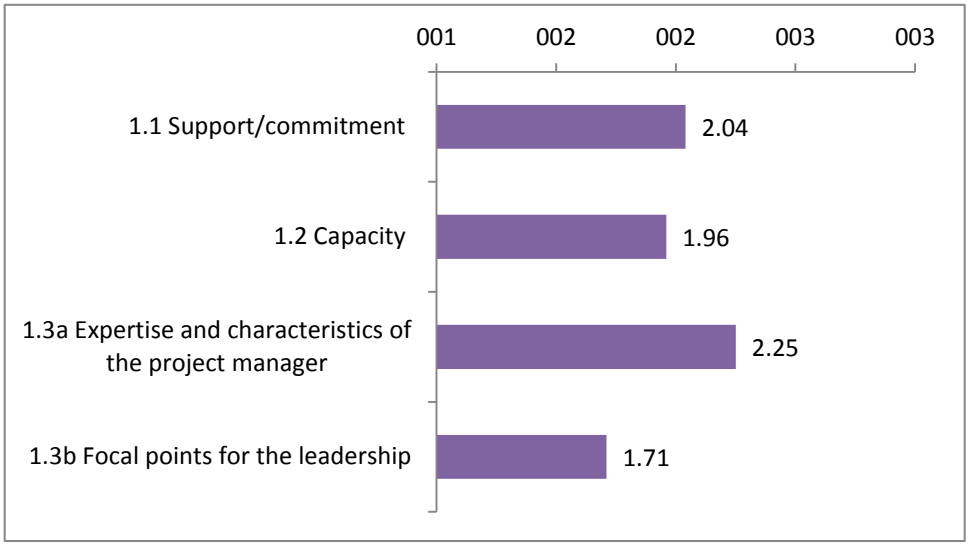


Figure 13.

Mean Scores on the First Preffi Cluster "Contextual conditions and feasibility" Subscales

The subscale *"Expertise and characteristics of the project manager"* achieved the highest level of quality (M=2.25) on this cluster. It indicates whether the person who is ultimately responsible has the necessary expertise to implement the project, whether his/her work style is compatible with the specific stage and the nature of the program, and whether the person who is ultimately responsible has the right personal characteristics to implement

the project. The subscale *"Support and commitment"* has also achieved a mean score higher than 2 (M=2.04). This subscale reflects if it has been established which internal and external partners are required for adequate support and commitment at each stage of the project, if there is sufficient support and commitment among the required partners and if agreements have been made about the involvement of internal and external partners. These two subscales represent certain strengths of the assessed programs. The other two subscales of this cluster – *"Focal points for the leadership"* and *"Capacity"* have also achieved mean scores higher than the theoretical mean.

Results presented in Figure 14 show that two of the three subscales on the second Preffi cluster "*Problem analysis*" have a mean score lower than the theoretical mean.

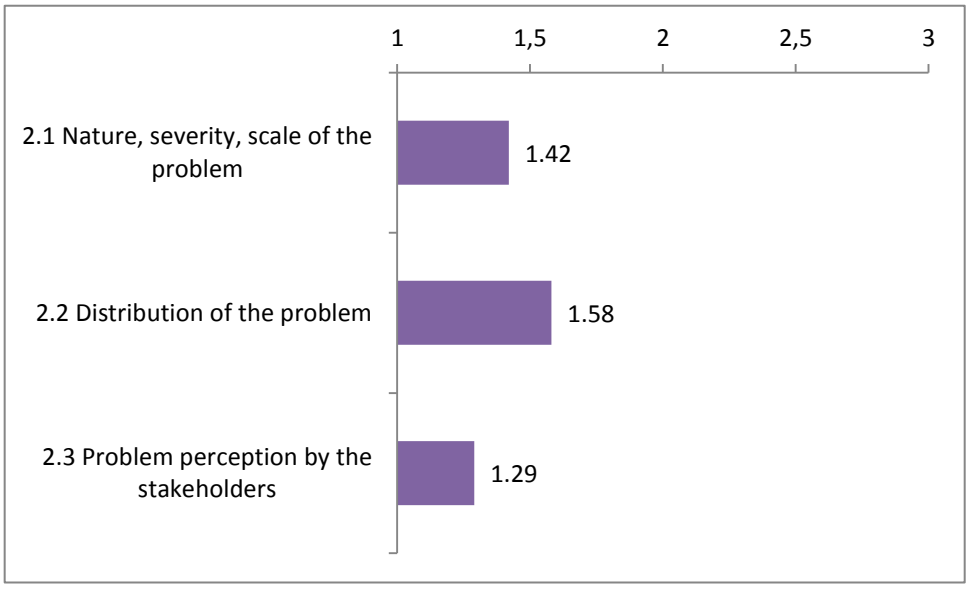


Figure 14.

Mean Scores on the Second Preffi Cluster "Problem analysis" Subscales

Only the subscale *"Distribution of the problem"* has a mean score (M=1.58) just above the theoretical mean on this cluster. The subscale *"Problem perception by the stakeholders"* achieved the lowest score on this cluster (M=1.29). It assesses if it is known to what extent the target group perceives the problem as a problem, if it has been established which persons, groups, agencies and social sectors are involved in perpetuating or solving the problem and to what extent these persons, groups, agencies and social sectors agree on the background and causes of the problem. It also assesses if it has been established how major social subgroups perceive the problem and if it has been checked whether there is interest or pressure from

politicians or public opinion to do something about the problem. The subscale *"Nature, severity and scale of the problem"* also achieved a low mean score (M=1.42) on this cluster. This subscale assesses if the problem or theme is clearly described, if it is clear whether the problem or theme is common in the group or community and if the problem is related to other social problems. It also assesses if it is clear what is known about the problem's immaterial and material costs.

Results from Figure 15 show that only one subscale of the third cluster, "Determinants of behaviour and environment", has a mean score higher than the theoretical mean while three other subscales have a mean score lower than the theoretical mean.

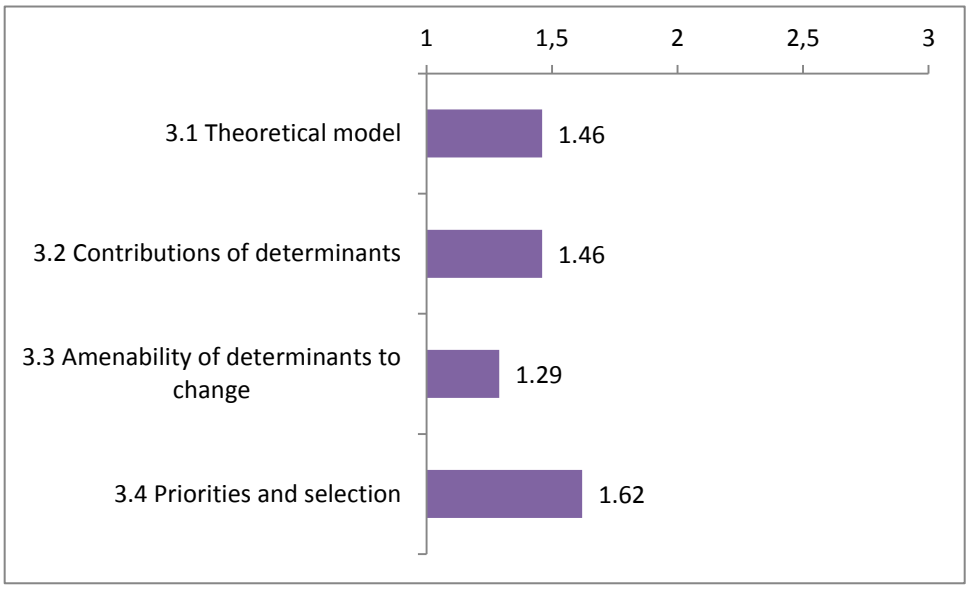


Figure 15.

Mean Scores on the Third Preffi Cluster

"Determinants of behaviour and environment" Subscales

The subscale with the highest mean score on this cluster is *"Priorities and selection"* (M=1.62). The lowest mean score was achieved on the subscale *"Amenability of determinants to change"* (M=1.29). It assesses if it has been estimated to what extent the determinants are amenable to change and if that estimation was based on theoretical and/or empirical knowledge about amenability of the determinants. The subscale *"Theoretical model"* has a mean score just below the theoretical mean (M=1.46). It reflects if the theoretical assumptions or models used to explain the problem are explicitly described in the proposal and if it has been made plausible that the model chosen is suitable for application to the problem. The

subscale *"Contributions of the determinants"* also achieved a lower mean score (M=1.46). It assesses if it is known which determinants affect the preferred or undesirable behaviour, the environmental factor or the problem. It also shows if it is clear which the most important determinants are, how reliable the evidence for the determinants is and to what extent the determinants apply to the relevant subgroups.

Results presented in Figure 16 indicate that on the fourth Preffi cluster, "Target group", two subscales achieved mean scores higher than the theoretical mean while one has a mean score lower than theoretical mean.

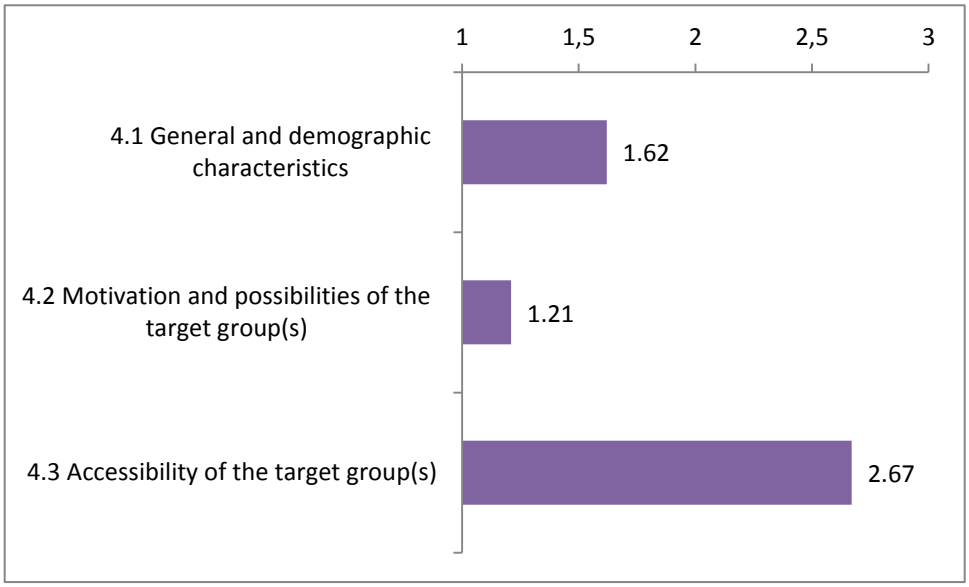


Figure 16.

Mean Scores on the Fourth Preffi Cluster "Target group" Subscales

Subscales with mean scores higher than the theoretical mean are *"General and demographic characteristics"* (M=1.62) and *"Accessibility of the target group(s)"* whose mean score is even higher than 2 (M=2.67). The Accessibility of the target group(s) subscale assesses if it is clear through which channels the target group can be reached and if the choice of channel(s) has been substantiated. The subscale with the lowest mean score on this cluster is *"Motivation and possibilities of the target group(s)"* (M=1.21). This subscale reflects if it is known to what extent the target group members are motivated to change, what factors affect the target group members' motivation to change and what wishes, needs, limitations and barriers to change the target group members themselves perceive.

Figure 17 shows that all subscales of the fifth cluster, "Objectives", have mean scores higher than the theoretical mean. It is noticeable that three of four subscales have a mean score higher than 2.

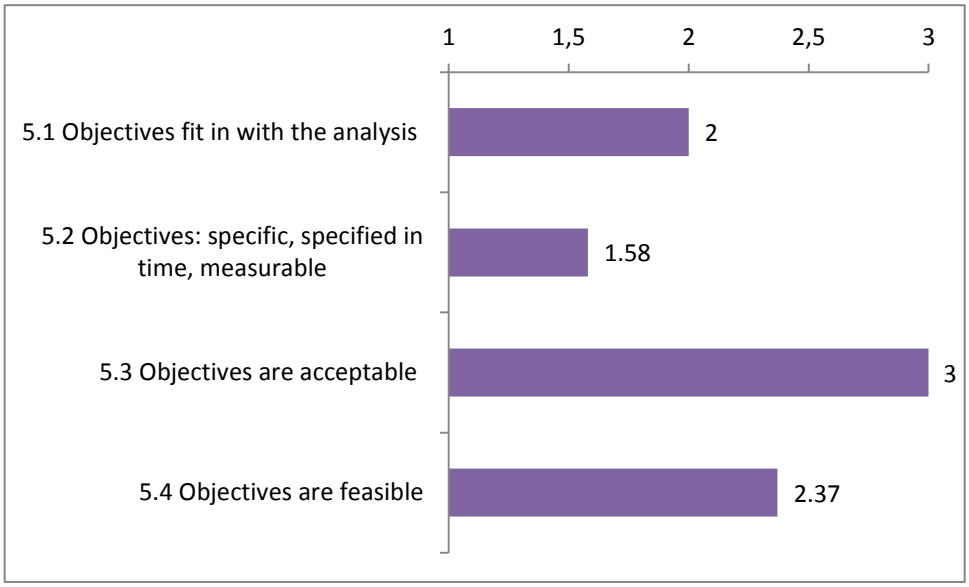


Figure 17.

Mean Scores on the Fifth Preffi Cluster "Objectives" Subscales

The subscale with the highest mean score is *"Objectives are acceptable"* which has achieved the maximal mean score (M=3). It assesses if the objectives fit in with the goals of the organisation that is proposing the program and if they are acceptable to the funding agencies and perhaps the medical ethics committee/institutional review board. It also reflects if the objectives are acceptable to possible partners and implementers and if they are acceptable to the target group. The other subscale with a high mean score is the *"Objectives are feasible"* subscale (M=2.37). It reflects if the staffing, money and time required to achieve the objectives have been estimated and if sufficient expertise, authority and partners are available to achieve the objectives. The third subscale with a high mean score is the *"Objectives fit in with the analysis"* subscale (M=2). This subscale assesses if the description of the objectives distinguishes various levels of objectives and if the objectives fit in with the problem analysis and the target group analysis. The lowest mean score was achieved on the subscale *"Objectives: specific, specified in time and measureable"* (M=1.58) which is still higher than the theoretical mean.

Figure 18 shows that on the "Intervention development" cluster eight subscales achieved a mean score higher than the theoretical mean, two of which have very high scores (M is higher than 2). Two subscales have a mean score lower than the theoretical mean.

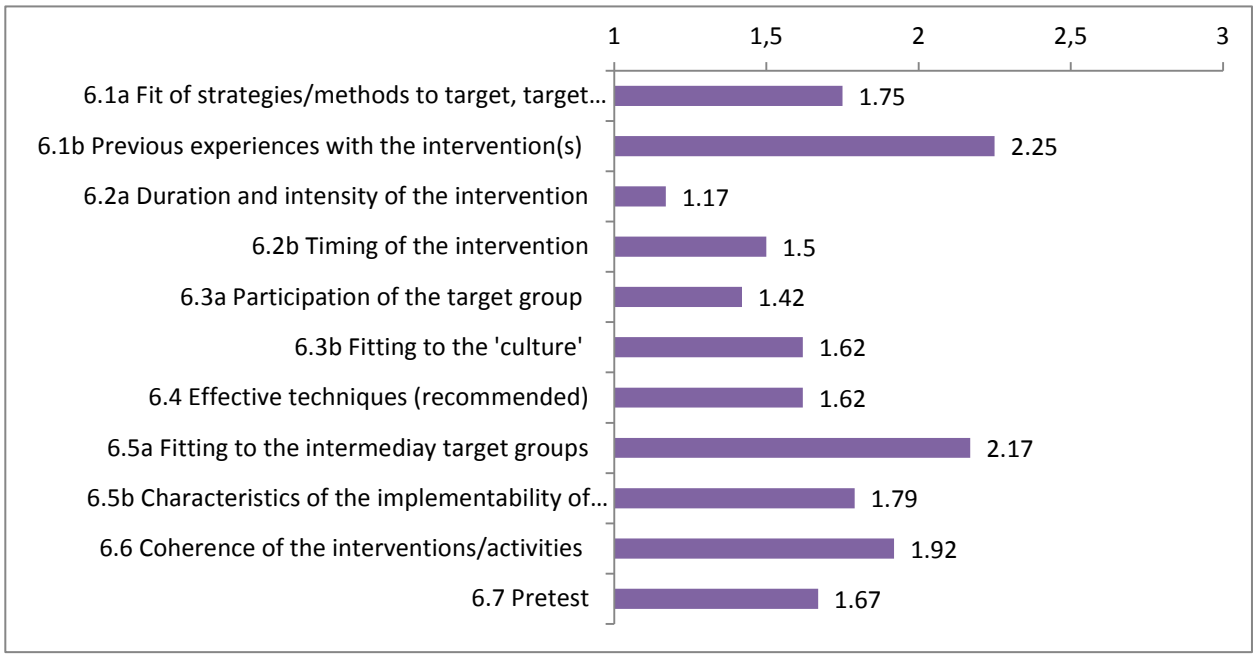


Figure 18.

Mean Scores on the Sixth Preffi Cluster "Intervention development" Subscales

The subscale *"Previous experiences with the intervention(s)"* shows a higher score than 2 (M=2.25). It reflects if any reports of successful or unsuccessful applications of the intervention by others are available, if program authors had any experience of successful or unsuccessful application of the intervention and if the proposed method appears to be potentially effective in the particular program. The subscale *"Fitting to intermediary target groups"* also has a mean score higher than 2 (M=2.17). It assesses if the members of the intermediate target group(s) have been consulted while the intervention was being developed and if the intervention is compatible with the modes of operation, procedures, standards and values of the intermediaries. Other subscales with mean scores higher than the theoretical mean are *"Fit of strategies/methods to the target group"* (M=1.75), *"Fitting to culture"* (M=1.62), *"Effective techniques"* (M=1.62), *"Characteristics of the implementability of the intervention(s)"* (M=1.79), *"Coherence of intervention/activities"* (M=1.92) and *"Pretest"*(M=1.67).

Two subscales have mean scores lower than the theoretical mean. The subscale *"Duration and intensity of the intervention"* has the lowest mean score (M=1.17) on this cluster. This subscale describes if any information from research or practice is available about the duration and intensity with which the intervention needs to be implemented to achieve the objectives and if that information has been used to decide upon the optimum duration and intensity of the intervention. The other subscale with a low mean score is the *"Participation of the target group"* (M=1.42). It assesses if in case the intervention has been developed elsewhere, the general target group was at least consulted while the intervention was being developed. It also reflects if the specific target group of the program was at least consulted while the intervention was being developed or before the intervention was selected. Also, this subscale assesses if the target group has participated sufficiently in the development or selection of the intervention.

Figure 19 shows that two subscales of the "Implementation" cluster achieved mean scores higher than the theoretical mean, one of which has a mean score higher than 2. Also, it is visible that three subscales have mean scores lower than the theoretical mean.

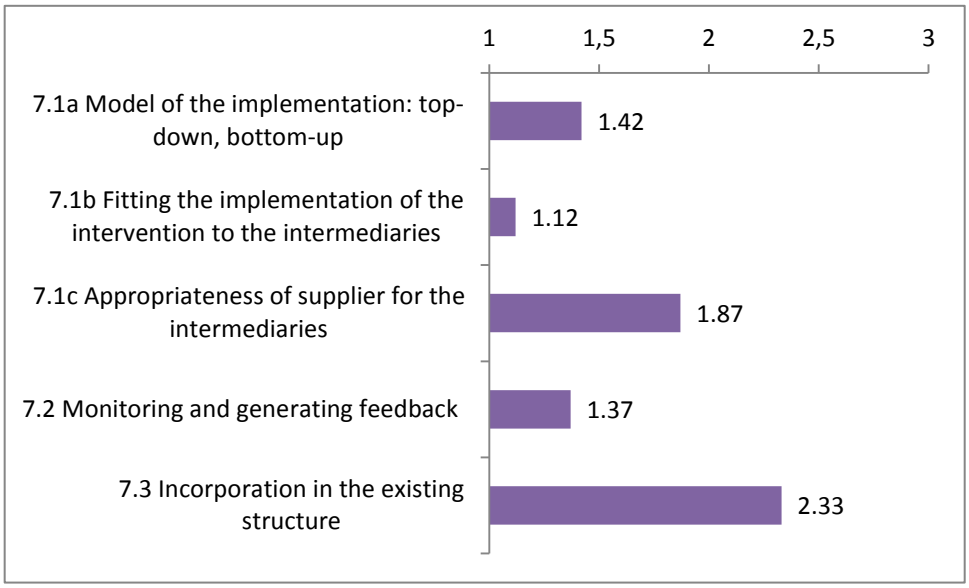


Figure 19.

Mean Scores on the Seventh Preffi Cluster "Implementation" Subscales

The subscale *"Incorporation in the existing structure"* shows the highest mean score on this cluster (M=2.33). This subscale reflects if the intervention has been incorporated in an existing structure, if the attempts have been made to incorporate the intervention in an

existing structure and if these activities or attempts carry enough weight, that is, if they have been addressed at the right hierarchical level. The other subscale with a mean score higher than the theoretical mean is the subscale "*Appropriateness of the supplier to the intermediaries"* (M=1.87). The subscale with the lowest mean score on this cluster is the "*Fitting the implementation of the intervention to the intermediaries"* (M=1.12). It describes if it is clear how the members of the intermediate target group(s) are distributed over the various stages of diffusion and use of innovations and if specific objectives have been set for each stage and for each intermediate target group. It also assesses if the implementation of the intervention fits in with the objectives set for each stage of diffusion and for each intermediate target group or target group segment, and if realistic objectives have been set in view of the fact that the intermediate target groups can be divided into 'innovators', 'early adopters', 'early majority', 'late majority' and 'laggards'. The *"Monitoring and generating feedback"* subscale also has a mean score lower than 1.5 (M=1.37). It reflects if the number of moments has been specified at which the progress of the diffusion and use of the intervention is to be assessed and if the assessment leads to active steps to adjust the process of diffusion and use of the intervention. A mean score lower than the theoretical mean was achieved on the subscale *"Model of the implementation: top-down, bottom-up"* (M=1.42). This subscale describes if the choice to use a particular model of implementation was made consciously. It also reflects if the intermediaries have the opportunity to adapt the intervention to their own situation and, if they have the opportunity to adapt it, if it is clear to them which elements must be retained.

Figure 20 shows that on the "Evaluation" cluster two subscales have achieved mean scores higher than the theoretical mean result while the remaining three subscales have mean scores below the theoretical mean.

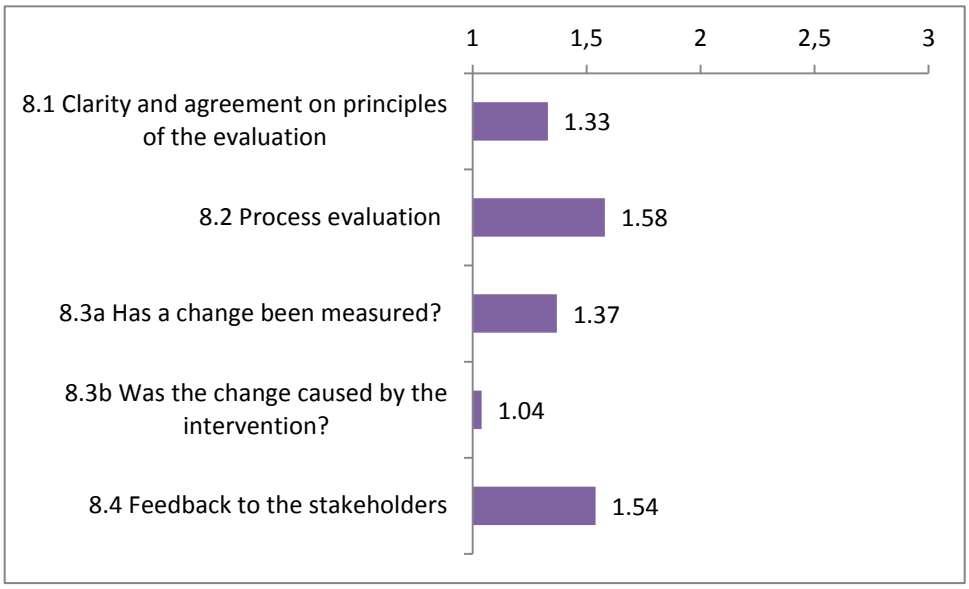


Figure 20.

Mean Scores on the Eight Preffi Cluster "Evaluation" Subscales

Subscales with mean scores just above the theoretical mean are *"Process evaluation"* (M=1.58) and *"Feedback to the stakeholders"* (M=1.54) subscales. There is no subscale with a mean score higher than 2 on this cluster.

"Was the change caused by the intervention?" subscale has the lowest mean score on this cluster (M=1.04). It reflects if it is clear which alternative explanations for the changes observed can be excluded and if the level of assertiveness of the conclusions being drawn is justified by the level of certainty offered by the study design. A lower mean score is also found for the subscale *"Clarity and agreement on principles of the evaluation"* (M=1.33). It describes if relevant persons, groups and/or organizations are involved in designing the evaluation. It assesses if all stakeholders have a clear idea about the questions that are to be answered by the evaluation and if they agree with these questions. It also reflects if it is clear what forms of evaluation are required to answer the questions and if the stakeholders agree on how 'hard' the evidence from the evaluation needs to be and whether that level of evidence is feasible. The assessed program proposals achieved lower scores on the subscale *"Has a change been measured?"* (M=1.37). This subscale reflects if it has been measured to what extent the objectives of the interventions have been achieved and if the evaluation methods used are valid and reliable.

Discussion of Results

Assessment of written program proposals with the Preffi 2.0 instrument has offered an insight into each program's quality level and into the average quality level of 24 programs from the County of Istria. The mean total Preffi score (MP) across all 24 Istrian programs is 5.68. This overall quality level of all assessed program proposals is below the chosen quality criterion, the theoretical mean (TM) of 6.65. The mean total score across all 24 programs which are representative for the County of Istria can be considered as an overall quality indicator of the state of the art in the field of mental health promotion and prevention in this County. This result suggests that the general quality level of the assessed program' proposals from the County of Istria is rather low. In the group of assessed programs, only four of the 24 program proposals are of higher quality.

When discussing these results it is important to stress that the Preffi instrument is a very comprehensive tool and that even lower total Preffi scores could implicate the presence of a certain quality in the assessed programs and projects. Program proposals assessed in this research were written by professionals of different backgrounds many of which never received training in program development. Without systematic and sustainable investments into knowledge and skills of NGO members who are proposing mental health promotion and prevention programs, it is not reasonable to expect higher quality of written program proposals. These results might also be attributed to characteristics of the Application form for financing programs developed by the Department of Health and Social Welfare of the County of Istria. It is possible that programs achieved lower total Preffi scores because some programs' aspects contained in the Preffi were not contained in the Application form and for that reason it was not possible to assess them. Although the 24 included programs are recognized in the County of Istria as the relatively better ones, this study has shown that most of the assessed programs need strategic investments in their quality improvement. There is also a need for further development of the criteria for financing programs and improvement of the Department's Application form.

This study has also provided information about the programs' quality level reflected in eight different Preffi clusters. The distribution of mean scores across the eight Preffi clusters shows which programs' concepts achieve low quality and need further improvements and which are of high quality. According to these results, there is a need for quality improvement of programs' concepts which are reflected in the seven Preffi clusters whose quality level is lower than the theoretical quality mean - *Contextual conditions and feasibility, Problem*

analysis, Determinants of behaviour and environment, Target group, Intervention development, Implementation and Evaluation. According to the results, only the cluster *"Objectives"* has achieved mean score higher that the theoretical mean.

Results suggest that the authors of mental health promotion and prevention programs from the County of Istria are most successful in defining and describing their program objectives. The cluster *"Objectives"* reflects if program objectives are fitting in with the problem analysis, if they are specific, specified in time and measureable, but also if they are acceptable to the main stakeholders and feasible. It also describes whether objectives are considered achievable given the available resources, contextual conditions and intended period of time. This result can be explained with the fact that members of most NGOs and other institutions who were involved in this study apply for more than one source of financing for their program proposals. In many application forms for financing programs, the section on objectives represents the central part of the application form. For that reason, it is possible that program developers already learned how to write them in a quality way.

Regarding the quality level, program objectives are followed by *"Contextual conditions and feasibility"* of the intervention being considered. This cluster describes the quality of support and commitment of internal and external partners, available capacities for the project, and leadership by the project manager including expertise and characteristics of the manager. In describing these program characteristics mostly technical information is needed which might be connected with the higher quality of this cluster compared to others. Still, as the mean score on this cluster is below the theoretical mean it would be advisable to improve this element of program proposals in the future.

The cluster with the lowest mean score is the *"Evaluation"* cluster. This result implies that the assessed programs have a very low quality of evaluation description in their program proposals. The evaluation cluster describes the quality of clarity and agreement on the principles of evaluation between different stakeholders and the quality of process and effect evaluation applied to the program. Effect evaluation refers to measuring planned changes in participants or target populations and if it is plausible that an observed change was caused by the intervention. The same cluster also assesses the quality level of the feedback on evaluation findings to the relevant stakeholders in the community. Assessed programs also show very low quality of the *"Problem analysis" cluster.* This cluster represents the quality level of the nature, severity and scale of the problem analysis, analysis of the distribution of the problem and the problem perception by stakeholders. Besides evaluation and problem analysis, low mean scores are present on the third cluster which reflects the *"Determinants of behaviour*

and environment". It refers to the quality level of the program's theoretical model, description of contributions of determinants to the problem, amenability of factors to change and the quality of how determinants are prioritized and selected.

It seems that for achieving higher quality levels of program's concepts - *Contextual conditions and feasibility, Problem analysis, Determinants of behaviour and environment, Target group, Intervention development, Implementation and Evaluation* program developers need additional education and training.

For a more profound understanding of which elements of program proposals achieved highest (score higher than 2) and lowest (score lower than 1.5) scores, mean scores on subscales of each of the eight Preffi clusters were analyzed. These results helped in detecting the strengths and weaknesses of the mental health promotion and prevention programs from the County of Istria more precisely and they are presented in Figure 21.

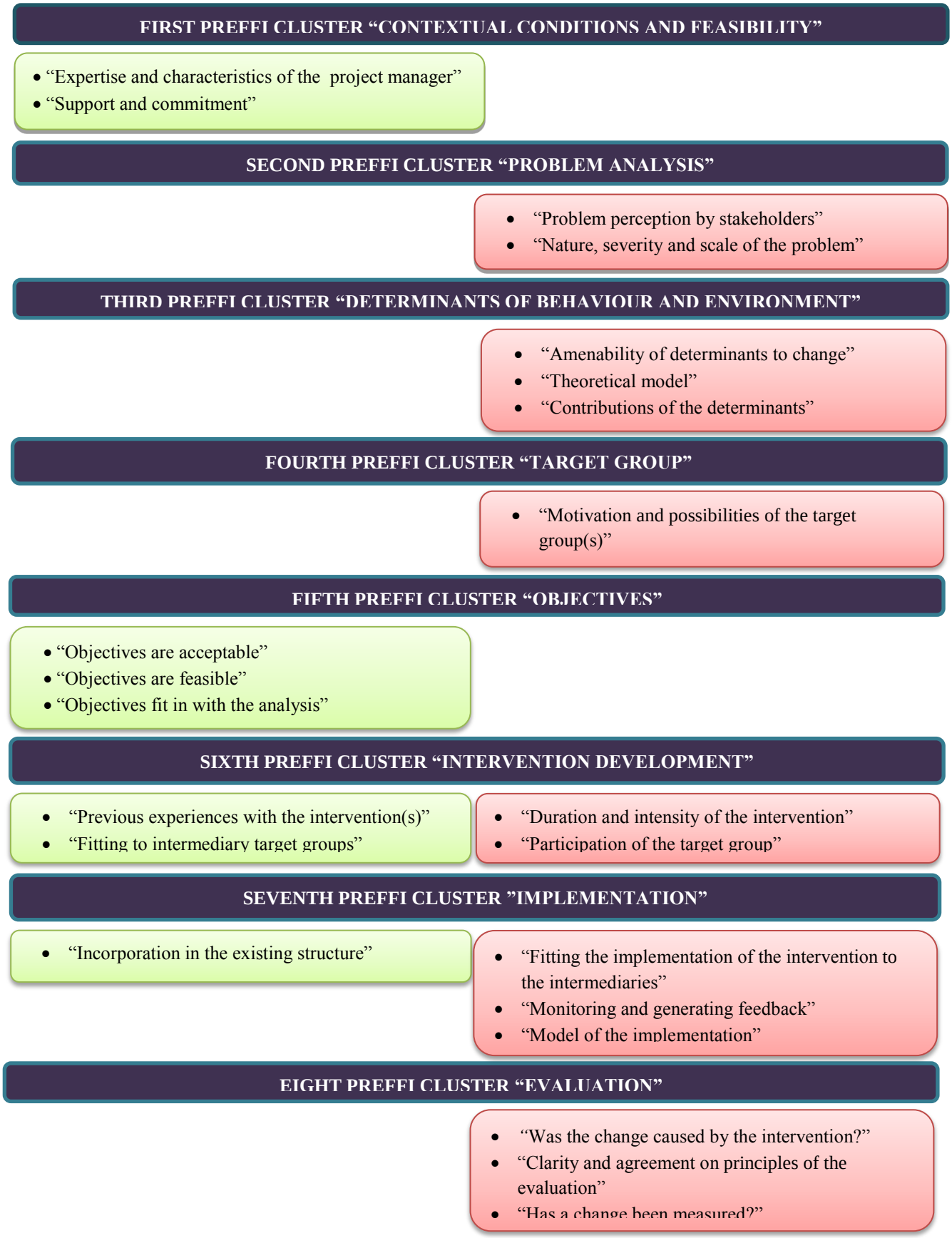


Figure 21.

Strengths and Weaknesses of Mental Health Promotion and Prevention Programs from Istria

Subscales with mean scores above 2 were considered program elements of higher quality representing strengths of the assessed programs from the County of Istria. These subscales are presented in green squares of the figure. Subscales with mean scores below the theoretical mean were considered as program elements which have a low quality and represent weak points of the assessed programs. These subscales are presented in red squares of the figure. Findings from this figure can serve as a base for planning the strategy for improvement of mental health promotion and prevention programs from the County of Istria.

4.3. Impact of the Training for Prevention on Mental Health Promotion and Prevention Programs' Effectiveness and Quality

4.3.1. Impact of the Training for Prevention on Mental Health Promotion and Prevention Program Effectiveness

The impact of the Training for Prevention on program effectiveness has been tested within a study described in Section 3.3. (p. 62). The study was conducted to test the following hypothesis of the second research task:

> *Hypothesis 2.1.* Programs whose managers and deliverers were involved in the Training for prevention achieve significantly higher scores on effectiveness, i.e. higher effect sizes, than programs whose managers and deliverers were not involved in the Training.

In order to assess the impact of the Training for prevention on program effectiveness it was planned to apply the outcome measures at two time points for all programs from control and experimental conditions. In two programs outcome measures were not administrated to the program participants so these programs were excluded from the analysis of the Training for Prevention impact on program effectiveness. These were program number 2 - Mentor program and program number 3 - Parenting program I. Therefore, the study whose results will be presented in this section was conducted on 10 programs from the experimental and 12 programs from the control group.

To assess the effectiveness of 22 programs, standardized effect size (ES) estimates were calculated using **Cohen's d** (Cohen, 1988).

Cohen's d is calculated by subtracting the mean score at post-test and the mean score from the pre-test, divided by the pooled standard deviation:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s},$$

For most of the programs several outcome variables were assessed depending on program aims. Within each program, for each outcome variable Cohen's d value was calculated by subtracting the mean score on the outcome variable at post-test and the mean score on the outcome variable at pre-test, divided by the pooled standard deviation. This resulted with several effect sizes within some programs (Appendix F, p. 236). In these cases an average Cohen's d value was calculated by computing mean effect sizes per program as a measure of a program's general effect size (Table 4.6.).

After the average effect size was calculated for each program, the variance (Var) of average Cohen's d value was then calculated for each program using the following equation:

$$\text{Var } d = \left[\frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}\right]$$

where (n) are the sample sizes for each group (in the pre-test and in the post-test measurement).

The **method of random-effects meta-analysis** was applied to test the differences in the effectiveness of programs whose managers and deliverers were involved in the Training for Prevention and those programs whose managers and deliverers have not participated in the Training. Also, a mixed-effects meta-analysis was used in assessing the influence of the different moderators on the programs' effect sizes, including the impact of the Training for Prevention. Meta-analyses were conducted utilizing the metafor package in R (Viechtbauer, 2010).

Results of the Study

The average effect sizes of each assessed program are presented in Table 4.6.

Table 4.6.

Effect Sizes of Intervention and Control Groups of Programs

INTERVENTION GROUP		
CODE OF THE PROGRAM	**Effect size (d)**	**Var**
(1) MH promotion through the theatre	0.27	0.20
(4) *Media literacy*	0.44	0.01
(5) *Training for group leaders*	1.32	0.27
(6) *Substance abuse prevention for parents*	0.53	0.26
(7) *Substance abuse prevention for teachers*	0.34	0.05
(8) *Parenting program II*	0.52	0.09
(16) *Self-confidence training*	0.64	0.06
(22) *Creative free time program II*	0.05	0.07
(23) *Parenting program VII*	0.39	0.17
(24) *Parenting program VIII*	0.25	0.21
CONTROL GROUP		
CODE OF THE PROGRAM	**Effect size (d)**	**Var**
(9) *Parenting program III*	0.65	0.20
(13) *Free time for children in foster care*	0.04	0.25
(10) *Parenting program IV*	0.39	0.05
(11) *Substance abuse prevention in the community*	0.22	0.27
(12) *Creative free time program I*	0.25	0.20
(14) *Parenting programme V*	0.75	0.27
(15) *Peer-violence prevention program*	0.11	0.04
(17) *Substance abuse prevention*	0.32	0.01
(18) *Parenting program VI*	0.75	0.27
(19) *Underage drinking prevention*	0.16	0.08
(20) *MH promotion through volunteerism*	0.58	0.08
(21) *MH promotion through dance*	0.10	0.02

Positive standardized effect size estimates indicate that the group improved in the post measurement compared with the results of measurement conducted before the implementation of a program. It is noticeable from Table 4.6. that all assessed programs in this study have a positive values of effect sizes. However, this does not mean that in all programs a positive effect was found. In five programs (13, 22, 21, 15, 19) effect sizes range from 0 to .20. According to Cohen (1988), effect sizes between 0.20 and 0.30 may be considered a "small" effect, around 0.50 a "medium" effect and 0.8 to infinity, a "large" effect. It seems that five programs had no effect in reaching all their desirable outcomes. Programs with the smallest Cohen's d values are: (13) Free time for children in foster care (d=.04), (22) Creative free time program II (d=.05), (21) MH promotion through dance (d=0.10), (15) Peer-violence prevention program (d=0.11) and (19) Underage drinking prevention (d=0.16). Programs with the largest effect sizes are program (5) Training for group leaders (d=1.32) and three parenting programs - Parenting program V (d=0.75), Parenting program VI (d=0.75) and Parenting program III (d=0.65).

In order to compare the effect sizes of programs from the intervention and control group, a meta-analysis was conducted. By conducting a meta-analysis, the results from different studies are quantified in such a way that the resulting values can be further aggregated and compared (Lipsey and Wilson, 2001; Rosnow and Rosenthal, 1996). The bases for most meta-analyses are the meta-analytic fixed-effects and random/mixed-effects models (Berkey et al. 1995; Raudenbush, 2009).

The first step in conducting a meta-analysis is testing the homogeneity of effect sizes and selecting the meta-analytic model (Viechtbauer, 2010). The selection of a model should be made mainly according to theoretical expectations about the source of variability in effect sizes (Vukasović, Bratko and Butković, 2012). Besides theoretical expectations, the result of a homogeneity test should also be considered in the process of selecting the most appropriate meta-analytic model.

The homogeneity of effect sizes is examined to determine whether the studies shared a common effect size. Testing for homogeneity in this study required the calculation of a homogeneity statistic, **Q-test**. If a Q-test is not statistically significant it is recommended to use a fixed-effects model. In cases when the Q-test is statistically significant, statisticians suggest to use a random-effects model. Borenstein, Hedges, Higgins and Rothstein (2009) emphasize that even when the Q-test is not statistically significant it is allowed to use the random-effects model if there is small sample of studies involved in the meta-analysis. The

same authors also stress that the decision on the meta-analytical model should be made in accordance with both theoretical assumptions of each model and Q-test results.

τ2 in the random-effects model estimates the total amount of heterogeneity in effect sizes and its precision depends mostly on k (number of studies) which is typically small in meta-analysis. Homogeneity testing in this study indicated the lack of variability among effect sizes of 22 assessed programs since the Q-test was not statistically significant (k=22, Q(df=21)=18.19, p=0.64). Even though this finding would suggest the usage of the fixed-effects model, according to theoretical assumptions of both models and the characteristics of this particular study it was decided that it is more appropriate to use the random-effects model in this research.

To understand this decision, it is important to explain the differences between two meta-analytical models. Borenstein and colleagues (2009) explain that the *fixed effects model* assumes that all studies in the meta-analysis are drawn from a common population. In other words, it assumes that all factors which could influence the effect size are the same in all study populations, and therefore the effect size is the same in all study populations. It follows that the observed effect size varies from one study to the next only because of the random error inherent in each study. However, this is a difficult assumption to make in many (or most) systematic reviews. When we decide to incorporate a group of studies in a meta-analysis we assume that the studies have enough in common that it makes sense to synthesize the information. However, there is generally no reason to assume that they are "identical" in the sense that the true effect size is exactly the same in all the studies. In *the random effects model* the goal is not to estimate one true effect, but to estimate the mean of a distribution of effects. Since each study provides information about an effect size in a different population, we want to be sure that all populations captured by the various studies are represented in the combined estimate. By contrast, when the researcher is accumulating data from a series of studies that had been performed by other people, it would be very unlikely that all the studies were functionally equivalent. Almost invariably, the subjects or interventions in these studies would have differed in ways that would have impacted the results, and therefore we should not assume a common effect size. Additionally, the goal of this analysis is usually to generalize to a range of populations. Therefore, if one did make the argument that all the studies used an identical, narrowly defined population, then it would not be possible to extrapolate from this population to others, and the utility of the analysis would be limited. Lipsey and Wilson (2000) argue that a fixed model assumes that variability between effects

sizes is due to sampling error, whilst a random effects model assumes that the variability between effects sizes is due to sampling error plus the variability in the population of effects (i.e. each study is estimating a slightly different population effect size).

The calculation of a random-effects meta-analytical model in 22 effectiveness studies has shown that the mean weighted effect size (d.) of all programs is 0.34 (SD=0.05). A confidence interval calculated for a measure of program's effect shows the range within which the true program's effect is likely to be. The 95% confidence interval about the mean includes a lower bound of 0.25 to a high bound of 0.44, indicating a small to moderate overall effect size. This result suggests that the 34% of variance in the programs' effects can be attributed to the programs.

In order to compare the effect sizes of programs in the experimental and control conditions, a random-effects model was conducted separately for both groups of programs. A graphical overview of the meta-analysis results can be obtained in a forest plot (Lewis and Clarke, 2001). The results of a random-effects model conducted on 10 programs from the intervention condition are presented in Figure 22.

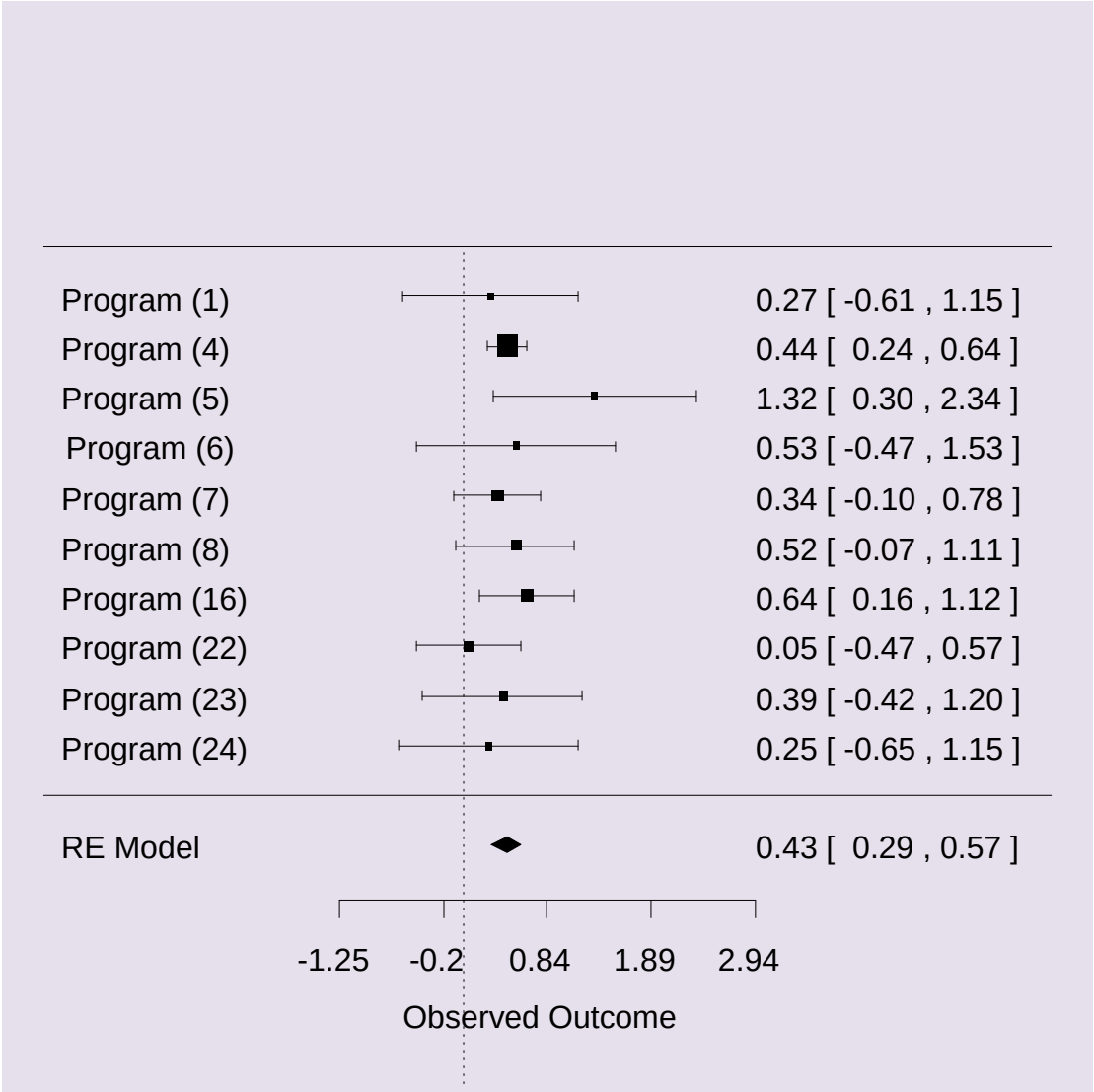


Figure 22.

Forest Plot of Effect Sizes, 95% Confidence Intervals and the Overall Mean Weighted Effect Size of Programs from the Intervention Condition

The overall mean weighted effect size (d.) of 10 programs whose managers and deliverers were involved in the Training for Prevention is 0.43 (graphically shown as a rhomb). The 95% confidence interval about the mean included a lower bound of 0.29 to a high bound of 0.57, indicating a moderate overall effect size of this group of programs. If the confidence interval is narrow, capturing only a small range of effect sizes, we can be quite confident that any effects far from this range have been ruled out by the study. This situation usually arises when the size of the study is quite large and, hence, the estimate of the true

effect is quite precise. Another way of saying this is to note that the study has reasonable 'power' to detect an effect. However, if the confidence interval is quite wide, capturing a diverse range of effect sizes, we can infer that the study sample was probably quite small.

The results of a random-effects model conducted on 12 programs from the control condition are presented in Figure 23.

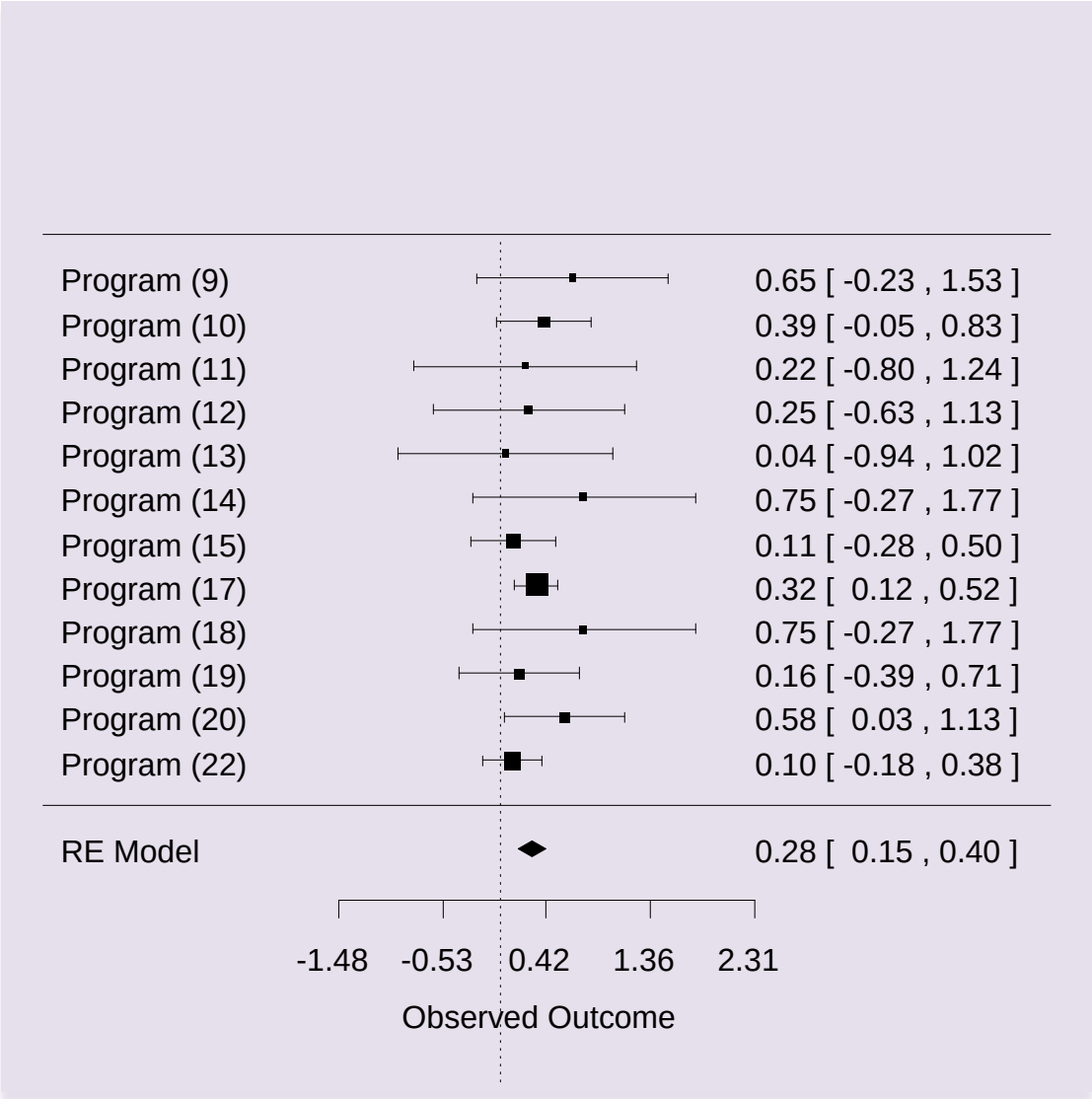


Figure 23.

Forest Plot of Effect Sizes, 95% Confidence Intervals and the Overall Mean Weighted Effect Size of Programs from the Control Condition

The overall mean weighted effect size estimated from the studies conducted on 12 programs whose managers and deliverers were not involved in the Training for Prevention is 0.28. The 95% confidence interval about the mean included a lower bound of 0.15 to a high bound of 0.40, indicating a small overall effect size of this group of programs.

If we compare overall mean weighted effect sizes of programs from the intervention and control condition it is noticeable that the programs whose managers and deliverers were involved in the Training have higher overall mean weighted effect size (d.=0.43) compared to the overall effect size of control group of programs (d.=0.28). However, since the 95% confidence interval about the mean of these two groups' overall effect sizes overlaps, the difference between these groups' effect sizes is not statistically significant.

In order to examine the influence of one or more moderator variables on the programs' effects, **mixed-effects model of meta-analysis** was conducted. It is a random-effects model which includes moderator testing (Viechtbauer, 2010). The exploration of study characteristics or features that might be related to variations in the magnitude of effect sizes across studies is referred to as moderator analysis. A moderator variable is one that informs about the circumstances under which the magnitude of effect sizes vary (Miller and Polloc, 1994). In terms of moderators, features coded in this study were:

- participation of the program managers and deliverers in the Training for Prevention (whether they were in the intervention or in control conditions),
- average Alpha value of outcome measures (lower or higher than .07),
- aim of the program (whether the program was focused on increasing knowledge or on developing skills and behaviours),
- program's intensity (whether the number of sessions is lower or higher than 5) and
- beginning of the program (whether the program started during the Training for Prevention or after the Training had finished).

Moderators were examined in separate univariate regression models to investigate the bivariate relations between moderators and effect sizes (Table 4.7.).

Table 4.7.

Univariate Effects for Moderators

MODERATORS	β	SE	z	p-value	95% CI
Training for prevention	0.15	.10	1.58	.11	(-0.04 - 0.34)
Alpha	0.01	.10	0.13	.90	(-0.18 – 0.21)
Aim of program	-0.02	.11	-0.21	.83	(-0.23 - 0.18)
Program's intensity	-0.11	.10	-1.08	.28	(-0.30 - 0.09)
Beginning of the program	-0.13	.14	-0.89	.37	(-0.40 - 0.15)

Among the set of predictors studied, none of them has been shown to be statistically significant in moderating the programs' effect sizes. However, it is noticeable that the Training for Prevention influence on programs' effect sizes is marginally statistically significant at the level of 10% (p=.11).

Discussion of Results

Homogeneity testing indicated a lack of variability among effect sizes of programs assessed in this study (k=22, Q(df=21)=18.19, p=0.64). This finding might be caused by the fact that all programs included in the study are at the universal level of interventions' spectrum. It is possible that the effect sizes of studied programs would vary more if in the process of meta-analysis selective and indicated prevention interventions would be included. Another explanation for the homogeneity of effect sizes could be that all of the assessed programs were developed and delivered by professionals from non-governmental organizations and most of them had a similar level of knowledge and skills in developing effective programs.

The calculation of a random-effects meta-analytical model on 22 effectiveness studies has shown that the mean weighted effect size (d.) off all assessed programs is 0.34 (SD=0.05). The 95% confidence interval about the mean includes a lower bound of 0.25 to a high bound of 0.44, indicating a small to moderate overall effect size. Many randomized controlled trials (RCT) have found average effect sizes of mental health promotion and prevention programs around .20 and .30, but researchers have also noticed that effect sizes can range from .10 for individual programs to .60 for preventive programs which combine several interventions (Brown et al., 2000; Jane-Llopis et al. 2003; Stice and Shaw, 2004; Tobler and Stratton,

1997). Hindshaw (2002) has found that most prevention trials have generated modest effects (i.e., effect sizes of .20–.30), which can be partly attributed to the weakness of the intervention and partly to flaws in the design (e.g., selection of participants, random allocation), the implementation (e.g., failure to control contextual variables), and outcome assessment (e.g., narrow perspective, absence of several follow up assessments). It is important to stress that the average effect sizes of assessed programs from the County of Istria would probably be different if RCT would be conducted in measuring programs' outcomes.

The results of this study are showing that some of the assessed programs from Istria have moderate or even strong positive impact on their participants. According to Cohen's guidelines for interpreting d values, some of the programs assessed in this study had no or had very small effects – program (13) Free time for children in foster care (d=0.04), (22) Creative free time program II (d=0.05), (21) MH promotion through dance (d=0.10), (15) Peer-violence prevention program (d=0.11), (19) Underage drinking prevention (d=0.16), (11) Substance abuse prevention in the community (d=0.22), (12) Creative free time program I (d=0.25), (24) Parenting program VIII (d=0.25) and (1) MH promotion through the theatre (d=0.27). It is important to stress that this does not mean that none of these programs succeed in achieving some positive outcomes, which is noticeable from Appendix F. Some of these programs were effective in reaching positive effects on some outcome variables, but their average effect sizes are rather low. This is probably caused by the intention of program proposal authors to set too many unrealistic desirable outcomes which cannot be achieved with the program. The program with the highest effect size is program (5) Training for group leaders (d=1.32) focused on promoting knowledge about group processes. The Cohen's d value of this program differs a lot from other effect sizes assessed in this study. A separate meta-analysis in which this program was excluded from analysis was therefore conducted but it did not change the results of meta-analysis significantly. It is interesting to notice that programs with the highest effect sizes are mostly parenting programs: (18) Parenting program VI (d=0.75), (14) Parenting program V (d=0.75), (9) Parenting program III (d=0.65) and (16) Self-confidence training (d=0.64). A possible explanation of these results could be that parenting programs from Istria are almost always theory-driven. In most cases the theoretical background of these interventions is a Choice theory developed by William Glasser (1998). Also, parenting programs from Istria are usually delivered by professionals who are trained in psychotherapy.

The comparison of the mean weighted effect sizes of programs from intervention and control conditions showed that the managers and deliverers who were involved in the Training achieved higher mean effect size (d.=0.43) compared to managers and deliverers who were not involved in the Training (d.=0.28). According to Cohen (1988), an effect size of 0.43 might be considered medium while an effect size of 0.28 can be considered small. This descriptive finding suggests that programs from intervention conditions were more effective in achieving their outcomes. However, since the 95% confidence intervals about the mean of this two groups' overall effect sizes overlaps, this difference between two groups' effect sizes is not statistically significant. Because of that we must conclude that **Hypothesis 2.3 was not confirmed in this research** since programs whose managers and deliverers were involved in the Training for Prevention did not achieve significantly higher scores on effectiveness, i.e. higher effect sizes, than programs whose managers and deliverers were not involved in the Training. The assumption of the Training for Prevention developers was that the transfer of knowledge about effective mental health promotion and prevention to program managers and deliverers would have a positive impact on the effectiveness of their programs. There are some possible explanations why this assumption was not confirmed in this study.

It could be that the Training for Prevention was not successful enough in transferring the knowledge to the program managers and deliverers who participated in it. However, the results presented in Section 4.3.1 show that the Training for Prevention did had a positive and significant impact on program managers' and deliverers' knowledge and skills in writing quality program proposals. The results have shown that the Training for Prevention significantly improved the level of quality in which program managers and deliverers select and describe the determinants of behaviour and environment which they want to approach with their programs. The Training also improved the level of quality in which they select and define the objectives of their programs. According to the results, participants of the Training for Prevention also plan and describe the evaluation of their programs on a higher quality level then those who were not involved in the Training. This meta-analysis has shown that increased knowledge and skills of program managers and deliverers for writing quality program proposals is not sufficient for achieving more effective outcomes.

The Theory of Planned Behaviour (TPB) developed by Ajzen (1991) and the model presented in Figure 24 could help in understanding this finding. TPB is the social cognition model that has been widely used to predict individual behaviours and has been one of the theories used most often when exploring the determinants of professional behaviour (Godin,

Bélanger-Gravel, Eccles and Grimshaw, 2008). According to this theory, human action is guided by three kinds of considerations: (1) behavioural beliefs, (2) normative beliefs, and (3) control beliefs.

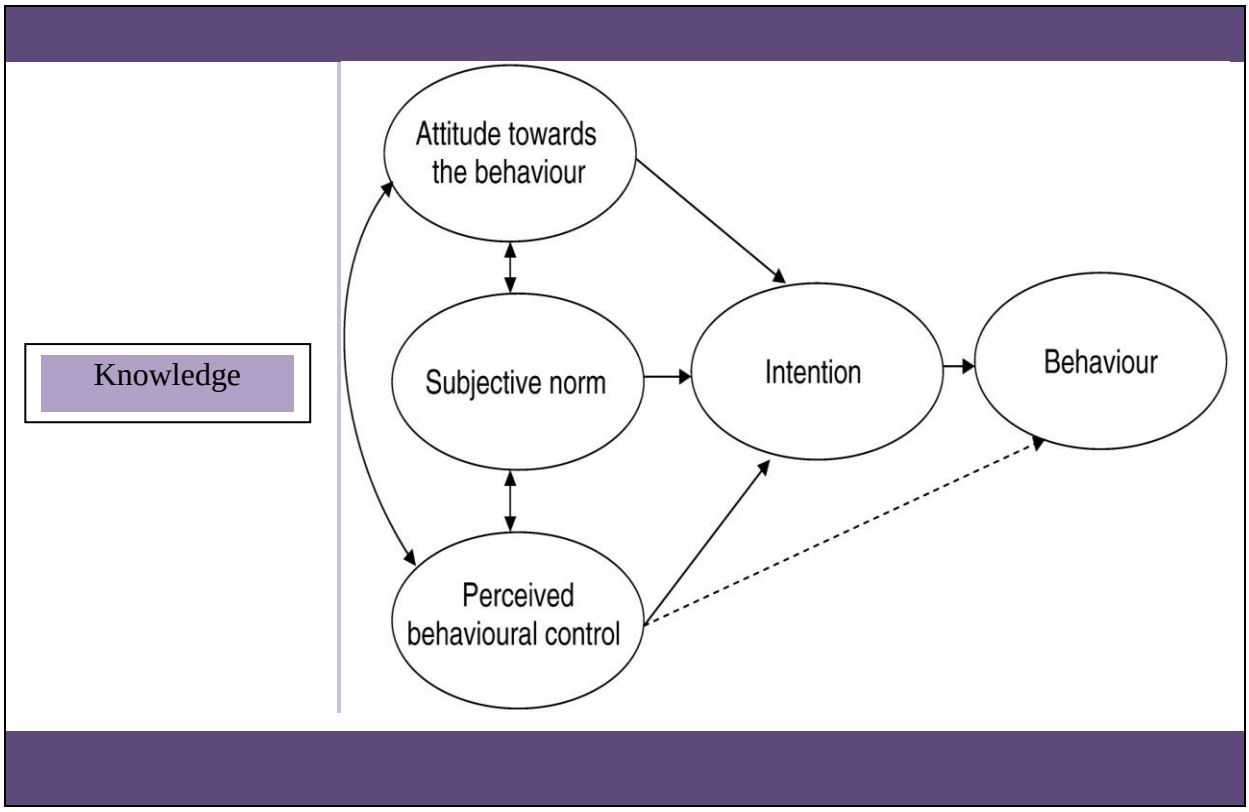


Figure 24.

The Theory of Planned Behaviour Model (Ajzen, 1991)

The Theory states that an individual's intention to perform behaviour is the proximal predictor of behaviour. In turn, intention is predicted by:

- attitude (a person's overall evaluation of the behaviour),
- subjective norm (a person's own estimate of the social pressure to perform or not perform the target behaviour), and
- perceived behavioural control (the extent to which a person feels able to enact the behaviour).

Perceived behavioural control has two aspects: (1) how much a person has control over the behaviour and (2) how confident a person feels about being able to perform or not perform the behaviour. Perceived behavioural control also has a direct effect on behaviour. A general

rule is that the more favourable the attitude, the subjective norm and the perceived control, the stronger should the person's intention to perform the behaviour in question. It is possible that the subjective norms and perceived behavioural control of programs' authors and developers influenced their readiness to incorporate gained knowledge into their mental health promotion and prevention practice. Sometimes there is a lack of control over the implementation quality of financed programs and because of that reason it is possible that in particular programs some program's components were not delivered to the participants or they were delivered poorly. High quality of implementation is found to be a core effect predictor, associated with positive intervention outcomes (Durlak et al, 2011). Jane-Llopis and Barry (2005) stress that high quality implementation, including training and supervision of program providers predicted higher program effectiveness. We can also assume that some Training participants did not feel confident in applying the gained knowledge into their practice. It is certain that in the future Training for prevention should involve more practical activities with the program managers and deliverers in order to increase their readiness to incorporate what they learned into their mental health promotion and prevention practice.

Another possible explanation of the study results could be the power of the study. As it was explained at the beginning of this section, two programs from the intervention condition did not apply outcome measures at two time points. Because of that reason the comparison between the effect sizes of programs was made on 10 programs from the intervention and 12 programs from the control conditions. It is possible that the differences between these two groups' effect sizes visible on the descriptive level would be statistically significant if more studies were involved in the research.

In order to assess the conditions which might have an impact on program effectiveness, several moderators were examined in separate univariate regression models. Among the set of five predictors studied, none of them has been shown to be statistically significant in moderating the programs' effect sizes. However, the analysis has shown that the Training for Prevention influence on the programs' effect sizes is marginally significant at the level of 10% (p=.11). As was stated before, it is possible to expect that this effect would be significant if the analysis included a bigger sample of programs. It is interesting to notice that according to the results of this study, the program's intensity did not appear to have an impact on the program's effects.

These study results show that the aim of the program was not significant in predicting the program's effects. The expectation was that the programs focused on increasing

knowledge achieve more effective outcomes than those focused on developing skills and changing behaviour, since only short-term programs' outcomes were measured in this study. It might be that the quality and intensity of programs focused on developing specific skills are more comprehensive and intensive compared to programs focused just on increasing knowledge, and require more effort on the part of program deliverers. The greater efforts of program deliverers from behaviour-change programs might balance the precedence which is usually held by deliverers from programs focused on increasing knowledge.

It was also assumed that the condition that some programs started with the implementation before or during the Training for Prevention will affect programs' outcomes negatively. However, the results of this study have shown that this was not the case. This finding is in accordance with the one that the Training for Prevention had only a marginally significant impact on program effectiveness.

According to this study, the intensity of a program did not have an impact on program effectiveness. Cuijpers (2002) stresses that there is no definite evidence that intense programs are more effective than less intense programs. However, it is important to stress that some authors (Stice et al., 2007; Nation et al., 2003; Connor, 2002; Yokishawa, 1994) have found that the dosage might be a key predictor of program effectiveness. Dosage is defined as the degree of exposure to the program or participants' attendance to the program sessions (Carroll, Patterson, Wood et al., 2007). Rorbach, Graham, and Hansen (1993) suggested that important moderators of intervention effects are: (1) conformity of the delivered program with the initial plan and the quality of delivery (i.e., staff's attitude and capacity to establish a positive working relationship with participants); (2) participants' adherence to the program (i.e., participants' involvement in program activities); and (3) dosage (i.e., participants' attendance to the training sessions). In our moderator analysis, we only tested for the possible moderator effect of the 'theoretical' dosage of the program according the written program proposal, we have not tested the impact of the actual dosage that might have been influenced by drop-out or varying levels of individual participation to the standardized training program.

Regarding the limited capacities in conducting this meta-analysis and since the moderators' analysis was not the main research task of this study, it was not possible to examine all possible moderators of effectiveness. In future studies, it would be very interesting to include implementation characteristics as potential moderators of program effectiveness. The results of the meta-analysis on mental health promotion and mental

disorder prevention interventions found that high-quality implementation, including training and supervision of programme providers and high participation in the programme sessions predicted higher programme efficacy and effectiveness (Jané-Llopis, 2002). It has been demonstrated in many studies that the quality with which an intervention is implemented affects how well it succeeds (Dusenburyet al., 2003; Dane and Schneider, 1998; Elliot and Ainsworth, 2012; Mihalic, 2004). For instance, two studies examining programmes to help people with mental health issues obtain employment found that employment outcomes among their study groups were weakest for those in poorly implemented programmes (McGrew and Griss, 2005). In the same way, a study of a parent training program found that when the program was implemented with high fidelity, the parenting practices improved significantly, but the effect was much less when implementation quality was low (Forgatch, Patterson and DeGarmo, 2005). It is more and more certain that the implementation characteristics are crucial in achieving desirable outcomes in mental health promotion and prevention programs.

4.3.2. Impact of the Training for Prevention on Mental Health Promotion and Prevention Programs' Quality

As described in Section 3.3. (p. 62), in order to examine the impact of the Training on programs' quality, written program proposals were assessed with Preffi 2.0 at two time points. During December 2010 and January 2011, 12 program proposals from the experimental condition and 12 program proposals from the control condition were assessed with Preffi 2.0 instrument in the baseline assessment. After the program managers and deliverers from the experimental condition were involved in the Training for Prevention, program managers and deliverers of all 24 programs were asked to write new proposals of their programs which were implemented in the period between February and December 2011. Because two organizations did not submit new program proposals, three program proposals were missing in the post-test assessment of written proposals in December 2011 and January 2012. Because of that reason, in the post-test assessment with the Preffi 2.0 instrument 21 new program proposals were assessed instead of 24. The main aim of this study was to test two hypotheses of the second research task of this doctoral study:

> *Hypothesis 2.2.* There is a difference between the experimental and control group on the total Preffi 2.0 score, i.e. programs involved in the Training for Prevention achieve significantly higher total scores on Preffi 2.0 after the Training compared to the programs that were not involved in the Training.

Hypothesis 2.3. There is a difference between experimental and control group on specific Preffi 2.0 cluster scores in a way that programs involved in the Training for Prevention achieve significantly higher scores on individual Preffi 2.0 clusters after the Training compared to the programs that were not involved in the Training.

Results of the Study

Table 4.8. represents the scores on eight Preffi cluster scores and total Preffi scores of programs in experimental and control groups assessed at two measurements. The table includes a number of assessed program proposals, mean scores, standard deviations, and minimal and maximal scores on each variable in both measurements for both groups (experimental and control). The Kolmogorov-Smirnov test (K-S) was conducted to test the normality of results' distribution on Preffi 2.0 instrument in two measurements. The analysis has shown that the results gathered during the first and second assessment of program proposals with the Preffi 2.0 instrument are normally distributed (Appendix G, p. 248).

Table 4.8.

Scores on Eight Preffi Clusters and Total Preffi Scores of Programs in Experimental and Control Groups Assessed at Two Measurements with Preffi 2.0

FIRST PREFFI CLUSTER "Contextual conditions and feasibility"										
	BASELINE					**POST-TEST**				
	N	**M**	**SD**	**MIN**	**MAX**	**N**	**M**	**SD**	**MIN**	**MAX**
EXPERIMENTAL GROUP	12	7.08	1.49	5	9.17	10	**8.42**	1.86	5.83	10
CONTROL GROUP	12	6.18	1.57	4.17	10	11	**8.09**	1.33	5.83	10
SECOND PREFFI CLUSTER "Problem analysis"										
	BASELINE					**POST-TEST**				
	N	**M**	**SD**	**MIN**	**MAX**	**N**	**M**	**SD**	**MIN**	**MAX**
EXPERIMENTAL GROUP	12	4.53	1.11	3.33	6.67	10	**6.33**	2.03	3.33	10
CONTROL GROUP	12	4.99	1.74	3.30	7.78	11	**5.45**	1.53	3.33	7.78
THIRD PREFFI CLUSTER "Determinants of behaviour and environment"										
	BASELINE					**POST-TEST**				
	N	**M**	**SD**	**MIN**	**MAX**	**N**	**M**	**SD**	**MIN**	**MAX**
EXPERIMENTAL GROUP	12	4.72	1.56	3.33	8.33	10	**6.67**	2.19	3.33	10
CONTROL GROUP	12	4.99	1.42	3.33	7.50	11	**5.08**	1.20	3.33	7.50

FOURTH PREFFI CLUSTER "Target group"										
	BASELINE					POST-TEST				
	N	M	SD	MIN	MAX	N	M	SD	MIN	MAX
EXPERIMENTAL GROUP	12	6.02	1.11	4.44	7.78	10	**7.56**	1.26	5.55	8.89
CONTROL GROUP	12	6.20	1.68	3.30	8,89	11	**6.46**	.97	4.44	7.78
FIFTH PREFFI CLUSTER "Objectives"										
	BASELINE					POST-TEST				
	N	M	SD	MIN	MAX	N	M	SD	MIN	MAX
EXPERIMENTAL GROUP	12	7.36	1.17	5.83	9.17	10	**8.58**	1.97	4.17	10
CONTROL GROUP	12	7.59	.97	5.83	9.17	11	**7.35**	1.28	5.83	9.17
SIXTH PREFFI CLUSTER "Intervention development"										
	BASELINE					POST-TEST				
	N	M	SD	MIN	MAX	N	M	SD	MIN	MAX
EXPERIMENTAL GROUP	12	5.79	1.46	3.40	7.57	10	**7.46**	1.74	4.24	9.70
CONTROL GROUP	12	5.53	1.30	3.94	8.48	11	**6.48**	1.21	5.15	9.09
SEVENTH PREFFI CLUSTER "Implementation"										
	BASELINE					POST-TEST				
	N	M	SD	MIN	MAX	N	M	SD	MIN	MAX
EXPERIMENTAL GROUP	12	6.06	1.25	3.40	7.57	10	**7.40**	1.42	5.33	9.33
CONTROL GROUP	12	4.78	1.27	3.33	8	11	**6.85**	1.04	5.33	8.67

EIGHTH PREFFI CLUSTER "Evaluation"										
	BASELINE					POST-TEST				
	N	M	SD	MIN	MAX	N	M	SD	MIN	MAX
EXPERIMENTAL GROUP	12	4.33	.83	3.33	6	10	**5.93**	1.90	3.33	8.67
CONTROL GROUP	12	4.83	1.69	3.33	8.67	11	**4.36**	.55	3.33	5.33
TOTAL PREFFI RESULT										
	BASELINE					POST-TEST				
	N	M	SD	MIN	MAX	N	M	SD	MIN	MAX
EXPERIMENTAL GROUP	12	5.74	.98	4.17	7.40	10	**7.29**	1.46	4.75	8.90
CONTROL GROUP	12	5.63	1.24	4.19	8.56	11	**6.26**	8.04	5	8.03

A **repeated measures analysis of variance** was conducted to compare the scores on eight Preffi 2.0 clusters and total scores on Preffi 2.0 at the baseline and post-test assessment of programs in experimental and control conditions. The results of a repeated measures analysis of variance are reported in Table 4.9. In the analysis, a number of measurements were used as a within sample variable (baseline and post-test) while the type of group (experimental or control) was used as a between sample variable.

Table 4.9.

Results of the Repeated Measures Analysis of Variance on Preffi Scores Regarding the Measurement Effect and Interaction of Measurement and Group Effects

FIRST PREFFI CLUSTER - "Contextual conditions and feasibility"			
SOURCE OF VARIABILITY	F	df	p
Measurement	24.385	1	.000**
Measurement x Group	.799	1	.383
SECOND PREFFI CLUSTER - "Problem analysis"			
SOURCE OF VARIABILITY	F	df	p
Measurement	5.769	1	.027*
Measurement x Group	2.877	1	.106
THIRD PREFFI CLUSTER – "Determinants of behaviour and environment"			
SOURCE OF VARIABILITY	F	df	p
Measurement	5.562	1	.029*
Measurement x Group	6.457	1	.020*
FOURTH PREFFI CLUSTER - "Target group"			
SOURCE OF VARIABILITY	F	df	p
Measurement	4.934	1	.039*
Measurement x Group	3.683	1	.070
FIFTH PREFFI CLUSTER - "Objectives"			
SOURCE OF VARIABILITY	F	df	p
Measurement	1.203	1	.286
Measurement x Group	5.905	1	.025*

SIXTH PREFFI CLUSTER - “Intervention development”			
SOURCE OF VARIABILITY	F	df	P
Measurement	25.106	1	.000**
Measurement x Group	19.000	1	.141
SEVENTH PREFFI CLUSTER - ”Implementation”			
SOURCE OF VARIABILITY	F	df	P
Measurement	49.517	1	.000**
Measurement x Group	.964	1	.339
EIGHTH PREFFI CLUSTER – “Evaluation”			
SOURCE OF VARIABILITY	F	df	P
Measurement	1.708	1	.207
Measurement x Group	7.547	1	.013*
TOTAL PREFFI RESULT			
SOURCE OF VARIABILITY	F	df	P
Measurement	16.573	1	.001**
Measurement x Group	4.182	1	.055

Note: * $p<.05$; ** $p<.01$

Regarding the *total Preffi scores***,** from Table 4.9. it is evident that only the main effect of the measurement showed to be significant (F=16.573, df=1, p=<.01). If we compare the mean total Preffi scores presented in Table 4.8., it is noticeable that the program managers and deliverers from both groups, experimental and control, have generally improved the quality of their program proposals at the post-test assessment and that the programs from the experimental group achieve higher total Preffi mean scores. At the same time, it is visible that the interaction of measurement and group effect on the total Preffi scores is not significant (F=4.182, df=1, p=.055). It is important to notice that the *p* value of interaction of effects is marginally significant and that this effect would probably be significant if the analysis included a larger sample of programs. These results suggest that all program proposals achieved higher total Preffi scores in the post-test assessment, but that there is no significant difference on the total Preffi scores between experimental and control groups of programs. The marginal significance points to a positive direction of the Training for Prevention impact on programs’ quality.

If we analyze the scores of the eight Preffi clusters individually, from Table 4.9 is noticeable that there are three Preffi clusters on which the interaction between measurement and group effects is significant (the third, fifth and the eighth Preffi cluster).

One of the Preffi clusters on which the interaction between effects of measurement and group is significant is the third Preffi cluster - *"Determinants of behaviour and environment"* (F=6.457, df=1, p=<.05). If we compare the mean results presented in the Table 4.8., it is visible that the programs in the experimental group achieve higher scores on this cluster in the post-test assessment (M=6.67, SD=2.19) than programs in the control group (M=5.08, SD=1.20). For the same cluster it is also noticeable that the effect of measurement is significant (F=5.562, df=1, p=<.05). It means that all programs, regardless of whether they were in the experimental or control condition, achieved higher scores on this cluster at the post-test compared to the baseline measurement.

Preffi clusters, on which the interaction of measurement and group effects is significant while the effect of measurement is not, are the fifth and the eighth Preffi cluster.

In case of the fifth Preffi cluster - *"Objectives",* the significance of the interaction between effects of the measurement and the group (F=5.905, df=1, p=<.05) suggests that there is a difference in this cluster's results between experimental and control groups of programs. According to the results presented in Table 4.8., program managers and deliverers from the experimental group significantly improved the quality of objectives of their programs (M=8.58, SD=1.97) at the post-test assessment compared to those in the control group (M=7.35, SD=1.28). Since the effect of measurement is not significant in the case of this Preffi cluster (F=1.203, df=1, p=.286), it means that only program managers and deliverers who participated in the Training for Prevention significantly improved this part of program proposals in the second assessment.

Another Preffi 2.0 cluster in which the interaction of measurement and group effect is significant while the effect of measurement is not, is the eighth cluster – *"Evaluation".* The significance of the interaction between effects of the measurement and the group (F=7.547, df=1, p=<.05) suggests that there is a difference in this cluster's scores between experimental and control groups of programs in the post-test assessment. According to the results presented in Table 4.8., program managers and deliverers involved in the Training for Prevention significantly improved the quality of evaluation description in their programs (M=7.29, SD=1.46) in the post-test assessment compared to the program managers and deliverers who were not involved in the Training (M=6.26, SD=1.46). It is important to stress that the level of

a result's significance is the strongest in this cluster (p=.013) if we compare it with two other clusters on which the interaction of measurement and group effects was significant ("Determinants of behaviour and environment", p=.020 and "Objectives", p=.025). Since the effect of the measurement is not significant in the case of this Preffi cluster (F=1.708, df=1, p=.207), it means that only program managers and deliverers who participated in the Training for Prevention significantly improved this part of program proposals in the post-test assessment.

It is important to add that in the fourth Preffi cluster, *"Target group",* the interaction of measurement and group effects is marginally significant at the level of 10% (F=3.683, df=1, p=.070). It is reasonable to assume that this interaction would probably be more significant if the analysis included a larger sample of programs.

From Table 4.9. it is also noticeable that for five Preffi clusters only the effect of time showed to be significant. These are the first Preffi cluster *"Contextual conditions and feasibility",* the second cluster *"Problem analysis",* the fourth cluster *"Target group",* the sixth cluster *"Intervention development"* and the seventh cluster *"Implementation".* These results suggest that managers and deliverers of all programs, regardless of wether they were in experimental or control conditions, had significantly different scores in these clusters in the post-test assessment.

For the first Preffi cluster - *"Contextual conditions and feasibility"*, the significant effect of measurement (F=24.385, df=1, p=<.01) implies that there is a significant difference in the score in this Preffi cluster in both groups of programs if we compare the baseline and post-test assessment results. According to the results presented in Table 4.8., it is visible that all programs achieved higher scores in this cluster in the post-test assessment. This cluster describes the quality of support and commitment of internal and external partners, capacities for the program and leadership by the program manager. It also includes expertise and characteristics of the manager.

The effect of measurement was also significant for the second Preffi cluster - *"Problem analysis***"** (F=5.769, df=1, p=<.05). According to the results presented in Table 4.8., all programs achieved higher scores in this cluster in the post-test assessment, regardless of wether they were in experimental or control conditions. This cluster reflects a quality level of nature, severity and scale of the problem analysis, analysis of distribution of the problem and problem perception by the stakeholders.

The significant effect of measurement was also detected for the fourth Preffi cluster - *"Target group"* (F=4.934, df=1, p=<.05). Again, results from Table 4.8. show that all programs achieved higher scores in this cluster in the post-test assessment, regardless of whether they were in experimental or control conditions. This cluster describes a quality level of general and demographic characteristics of the target group, motivation and opportunities of the target group to change and accessibility of the target group.

Also, the effect of measurement was significant for the sixth Preffi cluster - *"Intervention development"* (F=25.106, df=1, p=<.01). From the results presented in Table 4.8. it is noticeable that all programs achieved higher scores in this cluster in the post-test assessment. The sixth cluster reflects the rationale of the intervention strategy, previous experience with the intervention, duration, intensity and timing of the intervention, fitting to the target group and to the culture, participation of the target group and usage of effective techniques. It also describes the feasibility in existing practice, characteristics of implementability of the intervention and coherence of the interventions/activities.

The last cluster on which the effect of measurement was significant is the seventh Preffi cluster -*"Implementation"* (F=49.517, df=1, p=<.01). Results from Table 4.8. show that all programs achieved higher scores in this cluster in the post-test assessment. This cluster reflects the model of implementation, the fit of implementation interventions to intervention deliverers, appropriateness of the supplier for intermediating intervention deliverers, monitoring and generating feedback, and incorporation of the intervention in an existing organizational structure.

Discussion of Results

This study has shown that program managers and deliverers from both groups, experimental and control, have generally improved the quality of their program proposals at the post-test assessment and that the programs from the experimental group achieve higher total Preffi mean scores. Still, it was found that the interaction of measurement and group effects on the total Preffi scores is only marginally significant (F=4.182, df=1, p=.055). These results suggest that there is no significant difference on the total Preffi scores between experimental and control groups of programs after the Training for Prevention. Although it is important to emphasize that the marginal significance of the interaction of measurement and group effects on the total Preffi scores points to a positive direction of the Training for Prevention influence on programs' quality.

A possible explanation of these results could be the awareness of program managers and deliverers from the control condition that their written programs' quality will be assessed at the post-test assessment and compared with the programs in the experimental condition. Because of that reason, it is possible that they were motivated to write program proposals of a higher quality for the second assessment with Preffi 2.0 just by participating in the study and the measurements. The results might reflect that the motivation helped them to improve the general quality of their program proposals, which is reflected on the total Preffi score. This is effect is known in the research literature as the *Hawthorne effect*. It is important to stress that the program managers and deliverers from the control group did not get any feedback on their program quality after the first assessment with Preffi 2.0. Also, the post-test assessment of program proposals was conducted during December 2011, which was nine months after the Training for Prevention delivery. It is possible that this period affected the knowledge and skills of the program managers and deliverers who participated in the Training. It could be that the original impact of the Training for Prevention was partly lost over the period of nine months after the Training.

Based on these results, we can conclude that **Hypothesis 2.2. has not been confirmed** in this research. There is no difference between the experimental and control groups on the total Preffi 2.0 score, i.e. programs involved in the Training for Prevention did not achieve significantly higher total scores on Preffi 2.0 compared to the programs that were not involved in the Training.

The results suggest that the Training for Prevention significantly improved the level of quality in which program managers and deliverers select and describe the determinants of behaviour and environment (third Preffi 2.0 cluster) which they want to approach with their programs. The Training also improved the level of quality in which they select and define the objectives (fifth Preffi 2.0 cluster) of their programs. According to the results, participants of the Training for Prevention also plan and describe the evaluation (eight Preffi 2.0 cluster) of their programs on a higher quality level then those who were not involved in the Training. It seems that the knowledge about these three broad concepts was very effectively transferred to the participants of the Training for Prevention.

The *third Preffi cluster* reflects the quality level of the program's theoretical model, description of contributions of determinants to the problem, amenability of factors to change and the quality of how determinants are prioritized and selected. One of the crucial parts of the Training for Prevention was Logic Modelling (see Section 3.3.1.), which was delivered to

the Training participants within 8 hours of education and training. This topic was very broadly elaborated during the Training and involved a lot of practical work. Training participants were instructed how to develop Logic Models of their programs in a quality way. Participants were educated about the principles of internal theory of change inherent to each program and understanding the causal pathways of the problems they want to prevent with the program. Within this topic, Training participants were also taught how to precisely select and describe the target group which they want to include in their program. Because of these reasons, higher scores of the Training participants in this cluster compared to the results of the control group is highly understandable and expectable. For the same cluster it is also noticeable that the effect of measurement is significant ($F=5.562$, $df=1$, $p=<.05$). It means that all programs, regardless of whether they were in the experimental or control condition, achieved higher results in this cluster at the post-test compared to the baseline measurement. As explained earlier, the awareness of program managers and deliverers from the control group that their program proposals will be assessed in the post-test assessment and compared with the proposals from the control group, has very probably positively affected their results in the second assessment. It is reasonable to assume that they were motivated to improve their program proposals for the post-test assessment. Also, results on this cluster suggest that the program managers and deliverers from both groups managed to improve this part of their program proposals, but those who participated in the Training for Prevention improved them better. In other words, program managers and deliverers from the control group had a capacity to significantly improve these programs' concepts, but the improvement would be even higher if they had participated in the Training.

The fifth Preffi cluster describes if a program's objectives fit in with the problem analysis, if they are specific, specified in time and measurable, but also if they are acceptable to the main stakeholders and feasible. It also describes if objectives are considered achievable given the available resources, contextual conditions and intended period of time. The theme of Objectives was delivered to the Training participants within the topic of Logic Modelling (see Section 3.3.1.). Training participants were instructed and trained how to define specific and quality project goals based on conducted problem analysis, need assessment and available analysis of resources needed for programs' realization. They had a chance to practice a process of selection and definition of their programs' objectives through defining the objectives of their programs. Because of that reason, this significant improvement in the experimental group is understandable. According to the results in this cluster, we can

conclude that the program managers and deliverers from the control group were not able to improve their programs' objectives significantly only because they were motivated to write better proposals in the second measurement. It seems that selecting and describing a quality program's objectives requires specific knowledge and skills which cannot be developed without additional training or education.

Besides improving the quality of the third and fifth Preffi 2.0 clusters, the Training improved the quality of the eighth Preffi 2.0 cluster "Evaluation". Since the effect of the measurement is not significant in the case of this Preffi cluster ($F=1.708$, $df=1$, $p=.207$), it means that only program managers and deliverers who participated in the Training for Prevention significantly improved this part of program proposals in the post-test assessment. This cluster describes the quality and clarity of agreement on the principles of evaluation between different stakeholders. It also describes the quality of process and effect evaluation. The effect evaluation part assesses changes which are planned to be measured and if it is plausible that the change was caused by the intervention. The same cluster also assesses the quality level of the feedback on evaluation findings to the relevant stakeholders in a community. The topic of program evaluation was very elaborately described to the program managers and deliverers involved in the Training for Prevention (see Section 3.1.1.). During the eight hours of training, the evaluation process was thoroughly described starting with definition, theoretical overview of qualitative and quantitative indicators of program effectiveness, research methods of data collection and data sources. Using a logic model as a starting point, participants were taught which steps they have to follow during the evaluation process. They were informed about different types of evaluation and they have developed an evaluation plan for their program. Because of that reason, the significant improvement of the results in this cluster in the experimental group of programs is expectable. According to the results in this cluster also, we can conclude that the program managers and deliverers from the control group were not able to improve their programs' evaluation significantly only because they were motivated to write better proposals in the post-test assessment. It seems also that the knowledge about program evaluation is very specific and requires additional training or education.

For the purpose of improving the Training for Prevention effectiveness, it is important to notice that the experimental and control groups did not significantly differ in the level of quality in five Preffi clusters after the Training - Contextual conditions and feasibility, Problem analysis, Target group, Intervention development and Implementation. Based on

these findings, recommendations for the Training improvement will be presented in Section 5.3. Results from these five clusters suggest that the program managers and deliverers from both groups managed to significantly improve their program proposals at the post-test assessment. This finding could also be explained with the fact that the program managers and deliverers from the control condition were aware that their programs' quality will be assessed in the post-test assessment and compared with the programs which were in the experimental condition. It seems that the quality of these five concepts can be improved already through the increased motivation of program managers and deliverers to write the proposals of higher quality. This motivation might already be triggered by the message that their quality level well be measured and evaluated. There is a possibility that in the period between the first and second assessment with Preffi 2.0, managers and deliverers from the control group improved their skills of writing these parts of program proposals through conditions other than the Training for Prevention. In a way, it can be expected that all program managers and deliverers involved in this project could improve these skills spontaneously through the work they do in their organizations or through the experience they had with the programs which they implemented between the two assessments.

Regarding the results in specific Preffi clusters, we can conclude that **Hypothesis 2.3. is partially confirmed** in this research. There is a difference between experimental and control group scores in three of eight Preffi 2.0 clusters such that programs involved in the Training for Prevention achieve significantly higher scores in the third, fifth and eighth Preffi 2.0 clusters after the Training in comparison with programs which were not involved in the Training.

4.4. Predictive Validity of the Preffi 2.0 Instrument

In addition to analyzing the content validity and reliability of Preffi 2.0, predictive validity of the instrument was also tested in this doctoral research. A predictive validity study was conducted to test the following hypothesis of the first research task:

> *Hypothesis 1.3.* Programs that accomplish higher total scores on Preffi 2.0 achieve more effective outcomes than programs that accomplish lower total scores on Preffi 2.0.

Even though the main aim was to analyse the predictive validity of the whole instrument, predictive validity of individual Preffi 2.0 clusters was also explored. It was assumed that the programs which accomplish higher scores on particular Preffi 2.0 clusters

achieve more effective outcomes than programs that accomplish lower results in these clusters. Since it is an explorative research task, no directive hypothesis was defined.

Partial correlation analysis was used to analyse the linear relationship between scores on Preffi 2.0 and programs' effect sizes (Cohens' d), whose calculation was explained in Section 4.3.2. Preffi scores of 22 programs estimated during the first application of Preffi 2.0 in December 2010 and January 2011 were used as independent variables, while the effect sizes of programs were a dependent variable in this study (Table 4.10.). In the procedure of partial correlation analysis, participation in the Training for Prevention was one of the control variables because it was assumed that the involvement of program managers and deliverers from the experimental group in the Training could have a positive impact on programs' effects and increase them. The second control variable was the size of Alpha reliability coefficients of outcome measures applied in assessing the 22 program effectiveness. It was expected that the measures with lower Alpha coefficients (<.70) are less sensitive in reflecting the programs' outcomes. Another control variable was the intensity of the program because it was assumed that programs which have more sessions with the participants are more effective then programs that have less sessions with the participants.

Results of the Study

The results of the partial correlation analysis between 22 mental health promotion and prevention programs' scores on Preffi 2.0 and programs' effect sizes are presented in Table 4.11.

Table 4.10.

Preffi Scores of Programs Assessed in the First Assessment with Preffi 2.0 and Programs' Effect Sizes

CODE OF THE PROGRAM	1st C	2nd C	3rd C	4th C	5th C	6th C	7th C	8th C	TOTAL PREFFI 2.0	EFFECT SIZE (d)
(1) MH promotion through the theatre	6.67	4.44	4.17	4.44	7.50	5.45	6	4	5.33	0.27
(4) *Media literacy*	8.33	5.55	6.67	6.67	8.33	7.27	8	6	7.10	0.44
(5) *Training for the group leaders*	8.33	6.67	8.33	7.78	9.17	7.57	6	5.33	7.40	1.32
(6) *Substance abuse prevention for parents*	5.83	4.44	3.33	6.67	6.67	5.15	4.67	3.33	5.01	0.53
(7) *Substance abuse prevention for teachers*	5.83	4.44	4.17	6.67	5.83	4.54	6.67	3.33	5.18	0.34
(8) *Parenting programme II*	7.50	4.44	5.83	4.44	6.67	6.06	6.67	4.67	5.78	0.52
(9) *Parenting programme III*	4.17	3.30	4.16	3.30	7.50	3.94	4	3.33	4.21	0.65
(10) *Parenting programme IV*	6.67	4.44	5	7.78	8.33	6.36	4.67	6	6.16	0.39
(11) *Substance abuse prevention in the community*	6.67	6.67	4.17	5.55	6.67	4.85	4	4.67	5.41	0.22
(12) *Creative free time programme I*	5.83	3.33	3.33	4.44	7.50	4.85	5.33	3.33	4.74	0.25
(13) *Free time for children in foster care*	6.67	7.78	6.67	7.78	8.33	5.76	4	3.33	6.29	0.04
(14) *Parenting programme V*	5	3.33	3.33	4.44	5.83	4.24	3.33	4	4.19	0.75
(15) *Peer-violence prevention programme*	5	3.33	3.33	6.67	6.67	4.54	4	4	4.69	0.11
(16) *Self-confidence training*	5	3.33	3.33	6.67	5.83	3.40	4	4.67	4.53	0.64
(17) *Substance abuse prevention*	5	5.55	5	6.67	6.67	4.54	4	3.33	5.09	0.32
(18) *Parenting programme VI*	5	3.33	5.83	5.55	8.33	6.36	5.33	4.67	5.55	0.75

(19) *Underage drinking prevention*	6.67	5.55	5	5.55	7.50	5.76	4.67	6	5.84	0.16
(20) *MH promotion through volunteerism*	10	7.78	7.50	8.89	9.17	8.48	8	8.67	8.56	0.58
(21) *MH promotion through dance*	7.50	5.55	6.67	7.78	8.33	6.67	6	6.67	6.90	0.10
(22) *Creative free time programme II*	9.17	5.55	5	5.55	8.33	6.70	8	4.67	6.62	0.05
(23) *Parenting programme VII*	6.67	5.55	3.33	5.55	7.50	6.06	5.33	4	5.50	0.39
(24) *Parenting programme VIII*	9.17	3.33	4.17	6.67	8.33	6.36	6	4	6	0.25

Table 4.11.

Partial Correlations between Programs' Scores on Preffi 2.0 and Programs' Effect Sizes with Control Variables

(Participation in the Training for Prevention, Average Alpha of Outcome Measures and Program's Intensity)

		CORRELATIONS									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	FIRST CLUSTER "Contextual conditions and feasibility"										
2	SECOND CLUSTER "Problem analysis"	.65**									
3	THIRD CLUSTER "Determinants of behaviour and environment"	.70**	.67**								
4	FOURTH CLUSTER "Target group"	.74**	.61**	.61**							
5	FIFTH CLUSTER "Objectives"	.73**	.51*	.77**	.68**						
6	SIXTH CLUSTER "Intervention development"	.85**	.60**	.83**	.72**	.88**					
7	SEVENTH CLUSTER "Implementation"	.82**	.44	.65**	.55*	.59**	.82**				
8	EIGHTH CLUSTER "Evaluation"	.75**	.44	.69**	.58*	.64**	.80**	.90**			
9	TOTAL PREFFI 2.0	.92**	.74**	.88**	.80**	.84**	.95**	.84**	.84**		
10	EFFECT SIZE	.03	.05	.45*	-.08	.39	.28	.06	.09	.18	

* $p<.05$; ** $p<.01$

According to the results of this study, there is no significant correlation between *total scores on Preffi 2.0* and the effect sizes of programs. If we analyse the correlations between scores on *individual Preffi 2.0 clusters* and effects of programs presented in Table 4.11. it is noticeable that there is a moderate, positive linear relationship between the third Preffi cluster scores "Determinants of behaviour and environment" and the effect sizes of programs (r=.45, p<.05). Also important to emphasise is that there is a moderate, positive linear relationship between scores on the fifth Preffi 2.0 cluster "Objectives" and the effect sizes of programs significant at the level of 10% (r=.39, p=09.). It is reasonable to assume that this correlation would probably be more significant if the analysis included a larger sample of programs.

Discussion of Results

The results of a partial correlation analysis have shown that there is no significant correlation between total scores on Preffi 2.0 and the effect sizes of programs which means that **Hypothesis 1.3 was not confirmed** in this research. According to the conducted study, programs that accomplish higher total scores on Preffi 2.0 do not achieve more effective outcomes than programs that accomplish lower results on Preffi 2.0.

It seems that at this point, Preffi 2.0 as a whole is not valid in predicting the effectiveness of assessed written program proposals. As was explained in the introductory chapter of this dissertation, the effectiveness of a mental health promotion and prevention program is a very complex and dynamic construct. Mental health promotion and prevention scientists are ever more interested in understanding and explaining the characteristics and conditions needed for program effectiveness (Marchand et al., 2011, Neil and Christensen, 2009, Browne et al., 2004, Nation et al., 2003, Cujipers, 2002, Brown et al., 2000, Greenberg, Domitrovich and Bumbarger, 1999; Tilford, Delaney and Vogles, 1998; Kok et al. 1997, Hodgson, Abbasi and Clarkson, 1996 and many others). This study was focused on analysing the basic stage in one program's process – the analysis of written program proposal. The main intention was to examine if it is possible to predict the effectiveness of a program based on the written proposal's quality. Preffi 2.0 is a multidimensional instrument which incorporates most of the available knowledge on effect predictors or effect moderators in the field of mental health promotion and prevention. However, the assumption that those programs which were well developed and elaborated according to Preffi 2.0 would also be implemented in a quality way and would consequently achieve more effective outcomes was not confirmed in this research.

There are several possible explanations of this result which are related to both the effectiveness study and to the Preffi 2.0 instrument and its application on 24 programs.

It could be that the used outcome measures for assessing program effectiveness were not sufficiently valid or reliable, or did not represent the targeted core effects of the programs. However, this argument is not valid in case of this study since the possible effect of unreliable measurement was excluded from the correlation analysis. Also, the effectiveness of assessed programs was measured within the before-after research design study. There are certain threats to the internal validity of that kind of studies which are presented and described in Section 5.2. in more detail. It is also important to add that there is a possibility that the time span for finding programs' effects might have been too short, because of possible "sleeper" effects. Hawe, Degeling and Hall (1990) describe this effect as the situation when the intervention impact is not detected until sometime after the implementation of the intervention. In this case the effect can be missed if evaluation is undertaken immediately or only after a short time. Bry and Krinsley (1992) have repeatedly found evidence of delayed or "sleeper" effects on youth substance abuse precursors as a result of the researchers' family based prevention interventions. For this reason, longitudinal effectiveness studies in the field of mental health promotion and prevention are very much needed. This dissertation study was unfortunately time restricted and it was not possible to perform longitudinal observations of the programs' effects.

It might also be that Preffi 2.0 as a whole was not sensitive enough in assessing a program's quality based only on written program proposals. Molleman and his colleagues (2005b) used additional interviews with the programs' developers in order to assess their programs with Preffi 2.0, since sometime written programs proposals were not sufficient for programs' quality assessment. In the research whose results are presented in this dissertation, three independent assessors were evaluating only written program proposals and in cases when the description of a program's elements were missing, a score 1 - weak was marked on Preffi 2.0. In future applications of this instrument, it would be important to have an insight into a number of cases where some program elements were not assessable with Preffi 2.0 from the written program proposals only. This could also explain why some other Preffi 2.0 clusters were not valid in predicting the program effectiveness. The solution to this situation could be to ask the program developers for additional, written explanation of those elements which were missing in their program proposals. Another approach in ensuring that the

program developers describe all concepts contained in Preffi 2.0, could be to provide them with an electronic form of Preffi 2.0 so that they can fill it in directly while describing their programs.

It is also important to stress that there is a possibility that Preffi 2.0 is not covering all relevant variables for assessing a program's quality. The study presented in this dissertation has shown that the current Preffi 2.0 items are theoretically connected with specific quality indicators and are essential for assessing those indicators. Still, it might be that Preffi 2.0 needs to be revised in a way that recent findings about the effect predictors should be incorporated in it. Preffi 2.0 reflects the knowledge on effect predictor research up to around 10 years ago (Molleman, 2005, Molleman et al., 2003). There is a certain need for Preffi 2.0 update to a 3.0 version, which would better reflect recent research outcomes on effect predictors and moderators.

The results of this study also suggest that there is a missing link between high-quality written program proposals and effective outcomes at the end of a program's implementation. It is possible that even though programs' authors and developers are aware of programs' characteristics necessary for effective outcomes and incorporate them in high-quality written program proposals; they do not implement them in practice.

The Theory of Planned Behaviour (TPB) by Ajzen (1991), whose model is presented in Section 4.3.2, could explain this finding. It is possible that the subjective norms and perceived behavioural control of programs' authors and developers are influencing their readiness to implement programs as they were firstly designed. In some cases there is a lack of control over the implementation quality of financed programs and because of that reason it is possible that some program components are not delivered to the participants or that they were delivered poorly. It is also important to stress that very often program proposals are written by those who are not directly involved in program implementation. In such cases, program deliverers might not feel ready to deliver some program components to the participants. Further research focused on the implementation quality and influence of the implementation process, and the environmental conditions on program outcomes are very much needed, among other reasons to test if the level of expertise of the program implementers has a moderating impact on the translation of program proposals to implementation quality.

This study has shown that two of eight Preffi 2.0 clusters are moderately and positively correlated with the effects of programs. The third Preffi cluster *"Determinants of behaviour and environment"* has a moderate, positive linear relationship with the programs' effect sizes ($r=.45$, $p<.05$). According to this finding, this cluster is valid in predicting the effects of mental health promotion and prevention programs. It reflects the quality level of the program's theoretical model, description of contributions of determinants to the problem, amenability of factors to change and the quality of how determinants are prioritized and selected. Many authors who have conducted studies on different effect predictors in mental health promotion and prevention field emphasize the importance of the theoretical background of the intervention in achieving desirable outcomes (Jane-Llopis and Barry, 2005, Nation and colleagues, 2003, Bartholomew et al., 2001, Brown et al., 2000, Kok et al., 1997, Tobler and Stratton, 1997). According to these authors, programs that are theory-based in terms of the targeted risk and protective factors and the mechanisms of change used in the intervention have a more positive impact. Theories are essential to the design of programs because they facilitate understanding and describe the mediating processes that might operate in interventions (Lochman, 2001). Nation and colleagues (2003) explain that two types of theories that play a role in prevention programming - etiological theories and intervention theories. Etiological theories focus on the causes (e.g., risk or protective factors and processes) of the targeted problem (Kumpfer, 1997). Intervention theories are focused on the best methods for changing these etiological risks. Once the causes are identified, effective prevention programs are then based on empirically tested intervention theories shown to produce the desired changes in the causes and ultimately in the behaviour being prevented. Yamada and colleagues (1999) have found that the theory-based programs of prevention of sexually transmitted diseases and teen pregnancy are more effective than those which are not theory driven. Thomas and colleagues (1999) stress that the theory base is crucial for school based prevention program effectiveness. Bartholomew and colleagues (2001) noticed that the potential effect of the intervention could be much higher when in the processes of intervention development and implementation available theoretical and empirical knowledge is systematically applied. Domitrovich and colleagues (2010) also emphasize that theory has to guide the content, process and structure of mental health promotion and prevention interventions. They stress that developmental ecological models maximize the efficacy of a public health model by informing the selection of the core components for inclusion to affect multiple outcomes of prevention. According to these authors, understanding these processes and theories behind the programs' activities are critical for developing an effective

intervention strategy. Rigorous theory is also crucial in supporting the flexibility of implementation, so that programs can be adapted to the needs of particular settings and populations without necessarily compromising their integrity (Dadds, 2001). Jané-Llopis and Barry (2005) stress that a theoretical basis should be used in the design and implementation of any intervention. Green (2000) explains that interventions which are not theory-based might easily:

- Address wrong or inappropriate variables or
- Tackle only a proportion of the combination of variables required to have the desired effect.

The fifth Preffi 2.0 cluster "Objectives" also has a moderate, positive linear relationship with the effect sizes of programs significant at the level of 10% (r=.39, p=09.). This cluster assesses if a project's objectives fit in with the problem analysis, if they are specific, specified in time and measureable, but also if they are acceptable to the main stakeholders and feasible. It also describes if objectives are considered achievable given the available resources, contextual conditions and intended period of time. Many studies have confirmed that programs that have clear goals and objectives, and that are theory-based both in terms of the targeted risk and protective factors and the mechanisms of change used in the intervention program, have a more positive impact (Jane-Llopis and Barry, 2005, Ader et al., 2001, Brown et al., 2000, Kok et al., 1997; Tobler and Stratton, 1997). Jané-Llopis and Barry (2005) explain that the development of a shared mission and clear goals and objectives for a given intervention are critical to its success. The goals of a given initiative need to be concrete, attainable, measurable and agreed by all members. An early assessment of participation readiness, such as community readiness, is crucial in determining the nature and timescale of a new program. This is also important for the implementation process. Most successful implementation partnerships take time to establish relationships, and to build strong links with key players locally in order to effectively engage with and mobilise key players in supporting the program. Goals and objectives need to be transparent for the partners to engage in this process. Some authors stress that a program's theory and goals are crucial in developing logic models of programs (Wyatt Knowlton and Philips, 2009). Rogers (2005) emphasizes that a program's logic model represents how a program is intended to produce particular outcome and be effective at the end. Since the theory of a program has a very important role in the process of developing a program, it is understandable that Preffi 2.0 cluster which reflects the

theory behind the program also predicts its effects. The correlation of the programs' goals and effect sizes could be explained with the fact that clear goals are needed for selecting appropriate, theory-based activities to achieve intended mental, emotional or behavioural changes in participants. Also, it is important to remember that the evaluation of programs' 'outcomes is mostly designed based on the programs' goals and aims which certainly explains this correlation between programs' goals and effect sizes.

Lack of correlation between five Preffi 2.0 clusters and effect sizes can be explained with the fact that the scores on several Preffi 2.0 dimensions have generally been rather low (e.g. Evaluation). As a result of the low scores, the variation in scores in these clusters has also been rather low. The low variation on predictor indicators results with the low predictability of such indicators. This might partly explain the low predictive validity of Preffi 2.0. It could be expected that after the training of a part of the program managers and deliverers, the variation in Preffi scores increases, just like the predictive power of Preffi.

5. CONCLUSIONS AND RECOMMENDATIONS

The last chapter of the dissertation starts with an overview of conclusions from all studies conducted within this doctoral research. The limitations of those conclusions are also elaborated. Special attention will be given to recommendations for improving Preffi 2.0 and the Training for Prevention. They were formulated according to the results gained in this research, but also on findings from earlier studies on Preffi 2.0 which could be useful for instrument adaptation in Croatia. Directions for further research on Preffi 2.0 are also presented. The final part of this dissertation describes possible perspectives of Preffi 2.0 and the Training for Prevention in improving mental health promotion and prevention capacities in Croatia.

5.1. Summary of the Results

The first research task of this doctoral study was to assess **metric characteristics of Preffi 2.0** – content validity, reliability and predictive validity.

The results of the content validity study have **confirmed Hypothesis 1.1 -** the items of the Preffi 2.0 are theoretically connected with specific science-based quality indicators and essential for assessing those indicators. The Content Validity Ratio (CRV) of the whole Preffi 2.0 is 0.90, which is higher than the required minimum of 0.70. Preffi 2.0 items are considered as very essential and in accordance with the theory concepts on which the instrument is founded. The analysis of the mean CVR of eight Preffi clusters has shown that all eight Preffi clusters have a CVR between 0.83 and 0.98, which is generally a very high level of content validity. A correlation analysis between individual Preffi 2.0 clusters and the total score on Preffi has shown that Preffi 2.0 clusters are strongly and positively correlated with the total Preffi score. A strong and positive linear interrelationship between different Preffi 2.0 clusters was also found in this study. This result is not surprising since Preffi 2.0 consists of effect predictors distributed within eight clusters, which all reflect a broader concept of programs' quality.

Preffi 2.0 as a whole was found to be reliable in assessing the quality of programs during the two measurements of written program proposals (G1=.79, SEM1=.44 and G2=.85, SEM2=.43). According to the results of this study, **Hypothesis 1.2 was confirmed** - Preffi 2.0 as a whole is a reliable instrument with at least a medium or high value of concordance between three assessors (G= 0.70 or higher). However, SEM values higher than 0.26 found in both assessments with Preffi 2.0 imply that the accuracy of Preffi 2.0 is not yet sufficient

although Molleman and colleagues (2005b) emphasize that 3 assessors are enough for accurate assessment with Preffi. Regarding the source of variance for the total Preffi 2.0 score, the study has shown that the program is the largest source of variance for the total score in both assessments. Variance attributable to the differences in views between the assessors was higher in the first assessment (30.5%) compared to the second (21.4%).

Regarding the predictive validity of the Preffi 2.0, the results of a partial correlation analysis have shown that there is no significant correlation between total scores on Preffi 2.0 and the effect sizes of programs, which means that **Hypothesis 1.3 was not confirmed** in this research. According to the conducted study, programs that accomplish higher total scores on Preffi 2.0 do not achieve more effective outcomes than programs that accomplish lower results on Preffi 2.0. However, the analysis of the correlations between scores on individual Preffi 2.0 clusters and effects of programs has shown that there is a moderate, positive linear relationship between the third Preffi cluster scores "Determinants of behaviour and environment" and the effect sizes of programs ($r=.45$, $p<.05$). Also, there is a moderate, positive linear relationship between scores on the fifth Preffi 2.0 cluster "Objectives" and the effect sizes of programs significant at the level of 10% ($r=.39$, $p<.10$). This means that program proposals that provide a better theoretical base, research information on evidenced-based determinants of the targeted problem, an elaborated goal analysis and well-selected and well-defined intervention objectives that are in agreement with the problem analysis and theoretical model, will result in higher program effectiveness.

The second research task of this doctoral study was to examine **the impact of the Training for Prevention on the effectiveness (outcomes) and quality (measured with Preffi 2.0) of mental health promotion and prevention programs**.

A method of random-effects meta-analysis was applied to test the differences in the effectiveness of programs whose managers and deliverers were involved in the Training for Prevention and those programs whose managers and deliverers have not participated in the Training. **Hypothesis 2.3 was not confirmed in this study** as far as it concerns the predictive value of the total score on the whole Preffi 2.0**,** since programs whose managers and deliverers were involved in the Training did not achieve significantly higher scores on effectiveness, i.e. higher effect sizes than programs whose managers and deliverers were not involved in the Training. The comparison of the mean weighted effect sizes of programs from the intervention and control conditions showed that the managers and deliverers involved in the Training achieved a higher mean effect size ($d.=0.43$) compared to the managers and

deliverers who were not involved in the Training (d.=0.28). On a descriptive level, the results suggest that programs from the intervention conditions were more effective in achieving their outcomes. However, since that the 95% confidence intervals about the mean of these two groups' overall effect sizes overlap, the difference between two groups' effect sizes is not statistically significant. The calculation of a random-effects meta-analytical model on 22 effectiveness studies has shown that the mean weighted effect size (d.) off all assessed programs is 0.34 (SD=0.05). The 95% confidence interval about the mean includes a lower bound of 0.25 to a high bound of 0.44, indicating a small to moderate overall effect size. This is in line with findings from other meta-analytic reviews on the effectiveness of preventive interventions in the field of mental health (e.g. Llopis, Hosman, Jenkins & Anderson, 2003).

The homogeneity testing indicated a lack of variability among effect sizes of programs assessed in this study (k=22, Q(df=21)=18.19, p=0.64). In order to examine the conditions which may have an impact on program effectiveness, several moderators were examined in separate univariate regression models. Among the set of five predictors studied, none of them have been shown to be statistically significant in moderating the programs' effect sizes. However, the analysis has shown that the Training for Prevention influence on the programs' effect sizes is marginally significant at the level of 10% (p=.11). While the homogeneity test showed a low level of variability in effect sizes between the studied programs, it would anyway to be hard to find moderators that have a significant impact on program effects.

The analysis of the Training for Prevention impact on programs' quality has shown that **Hypothesis 2.2. was not confirmed** in this research. No difference between the experimental and control group on the total Preffi 2.0 score was found when measured for the second series of program proposals. Program managers and deliverers who were involved in the Training for Prevention did not achieve significantly higher total scores on Preffi 2.0 comparing to the managers and deliverers who were not involved in the Training. Nevertheless, it is important to emphasize that the marginal significance of the interaction of measurement and group effects on the total Preffi scores points to a positive direction of the Training for Prevention influence on program quality. **Hypothesis 2.3. was partialy confirmed** in this resarch when studied for the predictive value of each of the Preffi clusters separately. The results have shown that there is a difference between experimental and control group scores on three of eight Preffi 2.0 clusters in a way that programs involved in the Training for Prevention achieve significantly higher scores on these Preffi clusters after the Training in comparison with programs which were not involved in the Training. The Training significantly improved

the level of quality in which program managers and deliverers select and describe the determinants of behaviour and environment which they want to approach with their programs, the level of quality in which they select and define the objectives and how they plan and describe the evaluation process of their programs.

The third research task of this doctoral study was to identify **weaknesses and strengths of the programs of mental health promotion and prevention from the County of Istria** involved in the study.

An assessment of written program proposals with the Preffi 2.0 instrument has offered an insight into each program's quality level and into the average quality level of 24 programs from the County of Istria. The mean total Preffi score (MP) across all 24 Istrian programs is 5.68 on a scale ranging from 3.0 to 10.0. This overall quality level of all assessed program proposals is below the chosen quality criterion, the theoretical mean (TM) of 6.65. The results suggest that the general quality level of the assessed program proposals from the County of Istria is rather low. In the group of assessed programs, only four of the 24 program proposals are of higher quality; i.e. have reached a quality level above the theoretical mean. The distribution of mean scores across eight Preffi clusters showed that there is a need for quality improvement of the programs' concepts which are reflected in the seven Preffi clusters whose quality level was found to be lower than the theoretical quality mean – "Contextual conditions and feasibility", "Problem analysis", "Determinants of behaviour and environment", "Target group", "Intervention development", "Implementation" and "Evaluation". According to the results, only the cluster "Objectives" has achieved, averaged across all the studied program proposals, a mean score higher that the theoretical mean. For a more specific understanding of which elements of program proposals achieved the highest (higher than 2) and the lowest (lower than 1.5) scores, mean scores on subscales of each of the eight clusters were also measured.

This doctoral study has offered valuable experience in assessing the quality and effectiveness of mental health promotion and prevention programs. Based on the results of this study, many directions for further research in this filed could be defined. In addition to their contribution to the mental health promotion and prevention science, the results also represent important implications for improvement of mental health promotion and prevention practice and policy.

5.2. Limitations of the Study

The study presented in this dissertation represents a unique attempt to assess the quality and effectiveness of mental health promotion and prevention programs within one of the Croatian counties (Istria) in real-life conditions. Because of that reason and limited capacities, which will be elaborated in this section, certain limitations have been noticed.

Different conditions might affect the strengths of the conclusions from the *mental health promotion and prevention programs' quality study*, *study on reliability of Preffi 2.0* and *study on Training for Prevention impact on programs' quality*. Conclusions from all three studies were more or less based on Preffi 2.0 scores. Because of that it is important to emphasize that some Preffi 2.0 clusters and subclusters achieved lower reliability levels (see Section 4.1.), which is certainly limiting the statistical power of those studies' conclusions. It is very much possible that lower reliability was caused by the small sample of assessed programs (N=24). In future studies on Preffi 2.0, a larger number of program proposals should be included. It is also possible that lower reliability levels can be avoided by training the assessors on using Preffi 2.0. It is important to stress that both baseline and post-test assessments of program proposals with Preffi 2.0 were based on an application form designed by the Department of Health and Social Care at the County of Istria. Although this application form is comprehensive, it does not involve all areas which are incorporated in Preffi 2.0. Because of that, some areas were difficult to assess from written materials and it would have been better if individual interviews with each program's managers and deliverers had been conducted as an addition to the written project proposal assessment.

Regarding the *study on Training for Prevention impact on program quality* it is important to notice that the second measurement of program proposals was conducted nine months after the Training for Prevention delivery. In the interest of this study, it would have been better if managers and deliverers of all 24 programs had written new program proposals immediately after the Training. It is also important to stress that the readiness for prevention, attitudes about prevention and the level of knowledge of program managers and deliverers involved in the Training were not directly measured. It is possible that those variables might have affected the impact of the Training for Prevention. Programs included in this research were suggested by the County of Istria's Department of Health and Social Services and had to be a part of this study as one of the criteria for getting a project grant, which could have had

an influence on their motivation. A higher level of readiness and motivation could lead to a more significant impact of the Training for Prevention.

Concerning the *study on the impact of the Training for Prevention on program effectiveness and predictive validity of Preffi 2.0,* several possible limitations should be considered. In order to assess program effectiveness, different measures were applied depending on the programs' aims. As it is noticeable from Appendix F (p. 236), some of the applied measures achieved lower reliability levels. It is possible that lower measurement reliability was caused by very small sample sizes in some programs (e.g. some of the parenting programs had only 8 participants). It is important to stress that the results of these measures were also considered in calculating effect sizes of programs, since they were found to be reliable in some previous studies conducted by other researchers. It should be emphasized that the analysis of moderators, conducted in this study, shows that the Alpha values of outcome measures did not have a significant effect on programs' effect sizes.

Another possible limitation of the study conclusions could be the fact that in some programs from the experimental conditions, implementation of program activities started during the Training for Prevention (Appendix C, p. 218). A moderator analysis has shown that this condition did not affect the impact of the Training on program outcomes. It should also be considered that the impact of the Training on program effectiveness might be significant if the research design was different. In this research, the outcome evaluation was conducted for programs which the Training participants started to implement during or shortly after the Training. It is possible to assume that the results of this study would have been different if they first had to develop new program proposals according to the knowledge and skills gained in the Training and then implement improved versions of programs. In that case, new programs would be evaluated.

As another possible limitation of this study it should be noted that when comparing the effect sizes of programs in the experimental and control conditions, effect sizes of programs in those two groups were averaged. Programs differed in their aims. Some programs were focused on changes in the level of knowledge, while others were focused on changing behaviours or skills. Averaging the effect sizes of different programs may have had an unwanted impact on the results of this study. However, conducted moderator analysis has shown that the type of program did not affect the effect sizes of programs. Regarding the study limitations, it should also be explained that the pre-test and post-test evaluation design, which was used in assessing the program's effectiveness, has certain limitations since it does

not include any control or comparison group. The key threats to internal validity for this design are history, maturation, testing, instrumentation, and regression artefacts (Milas, 2005). This kind of evaluation design was used in this doctoral study because of the available human and financial capacities, and time limits. It was the most rigorous design which could be conducted in assessing the effectiveness of so many programs in the mentioned conditions. However, it is important to stress that the results of that kind of evaluation studies do represent valuable and important information which can be used as a base for further, more rigorous studies. Finally, only post-test measures were used to study outcomes in effectiveness and no long-term follow-up measurement was included. As a consequence, delayed effects or 'sleeper-effects' might have been missed in the effects' measurement.

5.3. Recommendations for Improving Preffi 2.0 and Training for Prevention

A study on the metric characteristics of the Preffi 2.0 instrument conducted within this doctoral research has provided important information on its capacities and advantages, but also on its deficiencies in assessing written program proposals. These study results should be considered in the process of adaptation of the instrument in Croatia, but they can also serve as a direction for future research on Preffi 2.0 and other quality assessment and quality assurance instruments.

Improvement of Preffi 2.0 Content Validity

Terminology which appears in Preffi 2.0 should be understandable to the mental health promotion and prevention experts for whom Preffi 2.0 is intended. Problems with understanding are especially noticeable with the terminology which refers to the implementation and evaluation process. This could be achieved by developing a dictionary of Preffi 2.0 terminology which would be provided to Preffi 2.0 users. The results of the content validity study have shown that some Preffi 2.0 items should be formulated more precisely in future versions of the instrument (item 6.2.b.4., item 7.1.b.2. and item 7.1.2.3.). In addition to these items, several other Preffi 2.0 items should be considered to be reformulated since the estimators did not mark their scores while assessing their content validity (item 1.3.b.3., item 2.1.3., item 3.4.3., item 6.2.a.1., item 6.2.a.2., item 6.5.b.2., item 7.1.b.1., item 7.1.b.2. and 7.1.b.3.). For the purpose of Preffi 2.0 adaptation in Croatia, future studies should involve experiences of experts and practitioners in using Preffi 2.0 – their reflections on the instrument's length, usefulness in assessing programs and intelligibility of items.

Since the results of the study on predictive validity of Preffi 2.0 have shown that it is not yet valid enough in predicting program effectiveness, it is possible that the instrument should be upgraded. The content validity of Preffi 2.0 was confirmed by a study in the context of the European project "Getting Evidence into Practice", in which national health promotion agencies from various EU countries tried to develop a joint European Quality Assessment Tool (Molleman, 2005). Preffi was found to compare very favourably with all other quality assessment tools in the world. Other tools included very few, if any, effect management criteria that are not included in the Preffi, whereas many Preffi items not included in other instruments were deemed relevant by an international panel of experts (Vermeulen et al., 2005). However, this was almost 10 years ago and it is possible to assume that new insights should be incorporated into the instrument better reflecting recent research outcomes on effect predictors and moderators provided by science and practice. A crucial recommendation on improvement of this instrument content validity is that new findings from the field of mental health promotion and prevention effectiveness should be incorporated into it. Quality management instruments are dynamic entities, requiring continuous adjustment to the rapidly developing knowledge about effective mental health promotion and prevention programs and their successful implementation. Regular updates of the instrument are required to incorporate the most recent research findings and experiences by practitioners.

In developing the next Preffi version, scientists will have to review the evidence supporting various Preffi criteria, identify aspects for which new evidence has become available and decide how this can be used to further improve the content validity of the instrument. Literature research and suggestions made by experts will have to be used to decide what new effect predictors need to be incorporated in Preffi 3.0. Some authors stressed (Hawe, Noort, King and Jordens, 1997; Hawe et al., 2000; Van den Broucke, 2003) that greater attention should be given to gender-specificity as part of fitting interventions to target groups. Molleman (2005) suggested that the attention paid to monitoring and evaluation, which has been incorporated in Preffi 2.0, and which is beginning to take shape in practice as well, will have to be supplemented with the cost-effectiveness aspect in a later version of Preffi. He also stresses that the theme of ethics should be included in the next version of Preffi.

Although this doctoral study was not focused on a direct analysis of the level in which Preffi 2.0 items are in accordance with the latest findings on effect predictors, some general recommendations can be given according to the literature review. Assessment of factors which can ensure sustainability of programs should also be considered in Preffi 2.0 –

involvement of volunteers, partnership with other organizations, incorporation of a program in existing structures, and projection of fund raising activities. Also, a media plan forprogram promotion and involvement of media campaigns in achieving desirable outcomes on a public level can ensure programs effectiveness. There are a lot of recent findings confirming that (Serrat, 2010; Friedli, 2007; Keleher and Armstrong, 2005). It is certain that a comprehensive study on this aspect of Preffi 2.0 content validity is needed and should be done in the near future.

A summary of recommendations for improvement of the content validity of Preffi 2.0 are presented in Box 2.

- Development of a dictionary which would contain the Preffi 2.0 terminology and which would be provided to Preffi 2.0 users.
- Revision of some Preffi 2.0 items (item 1.3.b.3., item 2.1.3., item 3.4.3., item 6.2.a.1., item 6.2.a.2., item 6.5.b.2., item 6.2.b.4., item 7.1.b.1., item 7.1.b.2., item 7.1.b.3. and item 7.1.2.3).
- Future studies should involve experts' and practitioners' experiences in using Preffi 2.0 – their reflections on the instrument's length, usefulness and intelligibility of items.
- An improved version of Preffi 2.0 should reflect new findings from the field of mental health promotion and prevention effectiveness.

Box 2.

A Summary of Recommendations for Improving Preffi 2.0 Content Validity

Improvement of Preffi 2.0 Reliability

It is important to stress that in this doctoral study Preffi 2.0 was used in assessing the quality of others' programs and not as a tool used by program developers while developing their programs.

This study has shown that on some Preffi items the level of concordance between assessors while assessing programs was not high enough. Because of that, it is very important to provide clear instructions to Preffi users through the user manual and to formulate Preffi criteria in the most unambiguous way. Also, it seems that there is a need for greater

specification of Preffi 2.0 items. Heuvelmans and Sanders (1993, according to Molleman, 2005) stress the importance of unambiguous and manageable instructions for the assessors using Preffi. Molleman (2005) has also found that the criteria (items) of Preffi are sometimes interpreted in various ways, which clashes with the desire to achieve the most reliable and valid assessment of those criteria.

This doctoral study has shown that the assessment with Preffi 2.0 should not be based only on written program proposals, but also on a discussion with program developers and deliverers. One of the main sources of problems that contributed to the relatively low reliability found in the previous studies of Preffi 2.0 (Molleman, 2005) were the five-page program descriptions to which they asked their respondents to apply the Preffi. This became evident from the numerous comments by correspondents stating that the descriptions offered too little information to adequately answer the questions asked in the operationalization. In other cases, assessors had to conclude from a single sentence in the program description that a particular criterion had been met. It may be assumed that the reliability of the assessments would be greatly improved if the assessment is based not only on a program description or program plan, but also on a discussion with the program manager to obtain additional information. In the study presented in this dissertation, three assessors who were using Preffi 2.0 have noticed the same problem while assessing only written program proposals. In the future usage of Preffi in Croatia, the assessment should involve discussions with the program developers. One other approach could be to ask program developers to describe their programs in a written form based on Preffi 2.0 questions. This approach would demand the provision of training and education to program developers on using Preffi 2.0.

It was also found that when several assessors evaluate a program with Preffi 2.0, the final Preffi 2.0 scores should be defined through their discussion and consensus of assessors. Molleman (2005) has also found that the idea of including discussions with colleagues, project managers or directors in the assessment procedure was frequently brought up in the interviews with Preffi users. He also noticed that when different assessors have assigned different scores to a program, discussions usually quickly result in consensus about the definitive score for a particular criterion.

Calculation of the final score on each Preffi 2.0 subcluster is time-consuming, which may lead to a lower quality of assessment. This could be avoided through developing a digital version of the instrument in which marked scores would be automatically computed and assigned to appropriate norm value – weak, moderate or strong. Molleman (2005) also stresses that digital version would contribute substantially to Preffi's reliability. He assumed

that the digital version could offer more specific explanations for each criterion, as well as suggestions that might lead to more objective assessments.

A summary of suggestions for improvement of Preffi 2.0 reliability based on the results of this doctoral study are presented in Box 3.

- Provision of training and clear instructions about Preffi 2.0 application for users of the instrument.
- Greater specification of general effect predictors in Preffi 2.0.
- The assessment with Preffi 2.0 shouldn't be based only on written program proposals, but also on a discussion with program developers and deliverers.
- When several assessors evaluate a program with Preffi 2.0, the final Preffi 2.0 scores should be defined through their discussion and consensus of assessors.
- Provision of a digital version of Preffi 2.0.

Box 3.

A Summary of Recommendations for Improving Preffi 2.0 Reliability

Improvement of Preffi 2.0 Predictive validity

The predictive validity of Preffi 2.0 is very much related to its content validity and reliability. Because of that, all recommendations presented in Box 2 and Box 3 are also applicable for improvement of the instrument's predictive validity. The study conducted within this dissertation did provide some interesting findings which should be considered in planning future studies on predictive validity of Preffi 2.0, but they should also be considered in improving the Training for Prevention.

Recommendations for Further Research on Preffi 2.0 in Croatia

Based on findings from this doctoral study and the current state of the art in the field of mental health promotion and prevention in Croatia, directions for future research on Preffi 2.0 can be formulated. All recommended studies should be conducted on the version of Preffi 2.0 after it is improved, according to the results of this doctoral study (content validity study).

The next step in Preffi 2.0 application in Croatia should be a study on experiences by mental health promotion and prevention experts and practitioners in using Preffi 2.0 – their reflections on the instrument's length, usefulness and items intelligibility. Preffi 2.0 could be further improved by learning from user experiences. That kind of study could also help in assessing the nature of the instrument, that is, whether it is a quality assurance instrument intended to improve the quality of programs, or an assessment instrument that allows the programs with higher quality to be selected for implementation and financing. The Preffi development team (Molleman, 2005) emphasizes the use of the Preffi 2.0 as a quality assurance instrument to improve the effectiveness of existing programs. According to these authors, if Preffi is to be used in assessments for selection purposes, the instrument and its application will need to be further developed, to allow reliable assessments to be achieved with only a few assessors. This doctoral study has shown that Preffi 2.0 has a potential to be used as a quality assessment instrument. It is possible that in the next studies Preffi 2.0 should be used as both a quality assurance and a quality assessment instrument for selecting programs to be financed. More specifically, it means that Preffi 2.0 could be used by program practitioners during the development of a program to improve its quality and then the quality of programs may be assessed with Preffi 2.0 by independent assessors.

Because Preffi 2.0 is a very comprehensive instrument (121 items), useful information may become available through the application of a factor analysis study. Information gained about the interdependencies between observed variables can be used later to reduce the set of variables in the instrument.

A study on the reliability of an improved version of Preffi 2.0 should be conducted on a larger sample of program proposals with trained assessors.

From the policy perspective, future research on predictive validity of Preffi 2.0 is very much needed. Molleman (2005) has also stressed a need for further research on Preffi 2.0 use as a diagnostic instrument. The author has noticed that many national and local agencies in the EU would be very much interested in the results of such studies (ECDC - European Centre for Disease Control and Prevention, IUHPE - Global Health Promotion Effectiveness Program of the International Union for Health Promotion and Education). An important question is whether Preffi 2.0 should be used to assess programs in a procedure of selection for financing and, if it is used in that way, how to define norms for financing on particular clusters and the whole instrument? A study on predictive validity of Preffi 2.0 conducted in this doctoral research showed that concepts of "Determinants of behaviour and environment" and "Objectives" do have certain validity in predicting program effectiveness. It might be that

those two clusters should be pondered differently while making a final decision on selecting programs which will be financed and implemented in practice. The doctoral study provided interesting information on predictive validity of Preffi 2.0, but future research on this topic is needed.

All of the recommended studies should be conducted before making decisions for large scale dissemination and implementation of Preffi 2.0 as a quality assessment and quality assurance instrument in Croatia. Also, it would be interesting to develop a digital version of Preffi 2.0 in Croatia and use it in future Preffi studies. International collaboration studies on Preffi 2.0 would also be significant and useful. A summary of recommendations for further research on Preffi 2.0 in Croatia is presented in Box 4.

- Study on experiences by mental health promotion and prevention experts and practitioners in using improved version of Preffi 2.0 – their reflections on the instrument's length, usefulness and intelligibility of items.
- Factor analysis of Preffi 2.0.
- Study on improved Preffi 2.0 reliability on a larger sample of program proposals.
- Study on predictive validity of improved Preffi 2.0.

Box 4.

A Summary of Recommendations for Further Research on Preffi 2.0 in Croatia

Improvement of Training for Prevention

Many important conclusions were derived from this doctoral study which could be used in increasing the Training for Prevention impact on the quality of the written program proposals and program effectiveness. There are several suggestions on how to improve the effectiveness of the Training.

Application of Preffi 2.0 on the second series of program proposals in this study has shown that the themes of "Contextual conditions and feasibility", "Problem analysis", "Target group", "Intervention development" and "Implementation" should be more thoroughly and intensively elaborated during the Training. According to the results of the study on predictive validity of Preffi 2.0, it seems that special attention in the Training should be paid to the themes of "Determinants of behaviour and environment" and programs' "Objectives", since these two Preffi 2.0 clusters were shown to be correlated with program effectiveness. In order

to maximize the impact of the Training, a booklet with all Training presentations and materials should be provided to the participants. Training should also involve individual consultations with the Training participants on the quality of their programs and more practical work with participants and examples from practice. That might improve the impact of Training. In order to increase the Training participant's motivation for learning, a certificate of attendance should be given to those who have participated in the whole Training and who have improved the quality of their program proposals. Experience in delivering the Training for Prevention in this study has shown that there should be two levels of the Training depending on participants' knowledge and skills in the field of mental health promotion and prevention practice – a basic level and an advanced level.

The results of meta-analysis have shown that increased knowledge and skills of program managers and deliverers on writing a high-quality program proposals in itself does not lead to more effective outcomes. This result suggests that the current version of the Training for Prevention is not yet sufficient in improving program effectiveness. Because of that, the content of the Training should be continuously updated with new scientific knowledge on effect predictors. A summary of recommendations for improvement of the Training for Prevention are presented in Box 5.

- Themes – “Contextual conditions and feasibility”, “Problem analysis”, “Target group”, “Intervention development” and “Implementation” should be more thoroughly and intensively elaborated during the Training for Prevention.
- Special attention in the Training should be given to the themes of “Determinants of behaviour and environment” and programs’ “Objectives”.
- A booklet with all Training materials should be provided to participants.
- Training should also involve individual consultations with the Training participants.
- Training should involve more practical work with participants and examples from practice.
- A certificate of attendance should be given to those who have participated in the whole Training and who have improved their program proposals quality at the end of the Training.
- Two levels of the Training depending on participants’ knowledge and skills – basic and advanced level
- The content of the Training should be continuously updated with new scientific knowledge on effect predictors.

Box 5.

A Summary of Recommendations for Improvement of Training for Prevention

As it is noticeable from Figure 8 presented in the introductory chapter (p. 36), the Training for Prevention is aimed at enhancing the expertise of organization’s managers, program developers and deliverers. From the Training for Prevention model it is evident that the trained expertise is not the only condition needed for comprehensive program development and quality, implementation quality and, finally, program’s effects. This doctoral study has confirmed that the process of improvement of mental health promotion and prevention program effectiveness is a policy process which requires collaboration of different entities responsible for certain parts of that process.

5.4. Perspectives of Preffi 2.0 and Training for Prevention in improving Prevention Capacities in Croatia

There is a need to improve the effectiveness of mental health promotion and prevention in Croatia, and this research has shown that there are indeed considerable opportunities for improvement. Both Preffi 2.0 and Training for Prevention represent a significant capacity in improving mental health promotion and prevention practice in Croatia. But as it was stated in previous sections, before their wider dissemination, Preffi 2.0 and Training should be further studied and improved during the upcoming years. The process of mental health promotion and prevention improvement is a stepwise and longitudinal process.

In the future, national and/or local authorities could offer the Training for Prevention to NGOs and institutions which are implementing mental health promotion or prevention programs. Continuous and planned investments in the knowledge and skills of developers and deliverers of mental health promotion and prevention programs could lead to changes on the public health level. Also, national and/or local authorities who are financing mental health promotion and prevention programs could organize training for mental health promotion and prevention program developers on the usage of a digital version of Preffi 2.0 in developing and writing their programs. In the future, an improved version of Preffi 2.0 could be used in assessing the quality of written program proposals as a base for selecting the programs that would be financed, after the program developers have been educated on how to use Preffi in developing their programs.

After screening programs with Preffi 2.0 and developing quality norms, a database of quality programs can be developed on the national or local level. Depending on detected strengths and weaknesses of programs, appropriate initiatives can be conducted in order to increase their quality. The database could also help in identifying capacities and gaps of existing programs in approaching the needs and problems in the community.

Also, a special, adapted version of Training for Prevention could be offered to the stakeholders on the national and local levels who are making decisions regarding the mental health promotion and prevention practice. This could improve their sensibility and awareness about the importance of evidence-based mental health promotion and prevention, and initiate changes on the policy level. The Training for Prevention could be much more effective if it were accompanied by measures like supervision of implementation or provision of external evaluation of financed programs. These conditions mostly depend on decisions on the policy level. All of this requires continuous research of Preffi 2.0 and Training for Prevention, and

communication about the research outcomes with national and local policymakers, service managers and practitioners.

In a certain way, the Preffi 2.0 instrument represents an innovative approach in improving the quality of mental health promotion and prevention practice since no such instrument was previously available in Croatia. As Rogers (2003) has stated, an innovation is an idea, practice, or object that is perceived as new by an individual or other unit of adoption. The innovation-decision process consists of all the decisions, activities and their impacts that occur from recognition of a need or a problem, through research, development, and commercialization of an innovation, through diffusion and adoption of the innovation by users, to its consequences (Figure 25).

Figure 25.

Main Stages in the Innovation-Decision Process

As is evident from Figure 25**,** the innovation-development process often begins with *recognition of a problem or need*, which stimulates research and development activities designed to create an innovation to solve the problem or need. In the case of Preffi 2.0, this would mean that mental health promotion and prevention scientists and practitioners have recognized a need for quality and effectiveness improvement. It is possible to notice that the

stakeholders from the County of Istria did recognize a need for improvement of mental health promotion and prevention practice. They are also aware of a need to improve their criteria for assessing and selecting programs while making decisions on their financing.

Once problems and needs are detected, a *research* phase follows. Most innovations are created by scientific research, although they often result from an interplay between the scientific methods and practical problems. The study presented in this dissertation represents a kind of research within which some characteristics of the innovation (Preffi 2.0) were examined. However, further studies on Preffi 2.0 usefulness should be conducted. Within this stage it should be researched if Preffi 2.0 should be used and developed as a quality assessment or quality assurance instrument or maybe both.

Development is the process of putting a new idea in a form that is expected to meet the needs of an audience of potential adopters. In that phase of innovation-decision process, experiences of potential adopters in using Preffi 2.0 should be examined. Also, based on their experiences and needs, Preffi 2.0 should be adjusted. This stage depends on decisions made in the previous one about the intended purpose of Preffi 2.0.

Commercialization is the production, manufacturing, packaging, marketing and distribution of a product that embodies an innovation. If a digital version of Preffi 2.0 were developed, this stage of the innovation-decision process would involve activities regarding that product.

A particularly crucial point in the innovation-development process is the decision to begin diffusing an innovation to potential adopters - a stage of *diffusion and adoption.* Before doing that, the innovation should be tested for its effectiveness and efficacy under real life conditions. As it was already stated, before broader dissemination of Preffi 2.0 in Croatia, further studies on Preffi 2.0 should be done as recommended in previous sections of this chapter.

Consequences, the final stage of the innovation-decision process are changes that occur to an individual or to a social system as a result of adoption or rejection of an innovation. It would be expected that the continuous usage of Preffi 2.0 as a quality assessment and quality improvement instrument would result in improved mental health promotion and prevention practice in Croatia. It is important to remember that this is a long-term process.

It is also important to mention that the characteristics of an innovation, as perceived by the members of a social system, determine its rate of adoption. According to Rogers (2005), five attributes of innovations are:

(1) Relative advantage – the degree to which an innovation is perceived as better than the idea it supersedes. The greater the perceived relative advantage of an innovation, the more rapid its rate of adoption will be. Results from this doctoral study and future studies on Preffi 2.0 could serve in promoting its usefulness and advantages.

(2) Compatibility – the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters. An idea that is incompatible with the values and norms of a social system will not be adopted as rapidly as an innovation that is compatible. It is certain that Preffi 2.0 would be adopted more easily in those counties which are already sensible and aware of the importance of theory-based mental health promotion and prevention practice. It could be stressed that the use of Preffi 2.0 is perfectly in line with current national policies like the National Drug Abuse Prevention Strategy in Croatia for the Period from 2012 to 2017 (http://www.uredzadroge.hr/upload/File/Dokumenti/Nacionalna_strategija_2012_2017/Nacionalna_strategija_suzbijanja_zloup.droga_2012.-2017.pdf) and National strategy for mental health care for the period from 2011 to 2016 (www.vlada.hr).

(3) Complexity – the degree to which an innovation is perceived as difficult to understand and use. New ideas that are simpler to understand are adopted more rapidly than innovations that require the adopter to develop new skills and knowledge. Although Preffi 2.0 is a comprehensive instrument, potential adopters could be trained and prepared for its usage. In that case, a digital version of Preffi 2.0 may be helpful. Also the one page overall scheme of Preffi, designed by its original makers, can be used as a tool to make Preffi easily understandable for potential users (Appendix H, p. 249).

(4) Trialability – the degree to which the innovation may be experimented with on a limited basis. An innovation that is triable represents less uncertainty to the individual who is considering it for adoption, as it is possible to learn by doing. The recommended development of a digital version of Preffi, and making this version

easily accessible on the internet for use by interested managers or practioners around the country could facilitate its triability.

(5) Observability – the degree to which the results of an innovation are visible to others. The easier it is for individuals to see the results of an innovation, the more likely they are to adopt it. Benefits of using Preffi 2.0, like the development of better quality program proposals or financing more effective programs would certainly motivate others for using it.

In addition to these five perceived attributes of an innovation, other variables such as the type of innovation decision, the nature of communication channels diffusing the innovation at various states in the innovation-decision process, the nature of the social system in which the innovation is diffusing, and the extent of promotion efforts on the part of change agents in diffusing the innovation, affect an innovation's rate of adoption. All of the numbered conditions should be considered in the process of improving and disseminating Preffi 2.0.

In conclusion, from the results presented in this doctoral study it is evident that Preffi 2.0 and the Training for Prevention represent a promising approach in improving the quality and effectiveness of mental health promotion and prevention practice and science in Croatia. However, findings from this research also imply that both Preffi 2.0 and Training for Prevention should be continuously studied in the future and accordingly adjusted and improved. This doctoral study has also shown that the whole process of improvement of mental health promotion and prevention quality and effectiveness is a long-term process requiring continuous research, communication with stakeholders and a comprehensive implementation approach. It also requires a policy support and engagement of stakeholders involved in making decisions on national and local levels which concern mental health promotion and prevention practice. Because of that, knowledge on the development and diffusion of innovations proposed by Rogers (2003) should be considered in the process of improving Preffi 2.0 and Training for Prevention, and using them in increasing the quality and effectiveness of mental health promotion and prevention practice and science in Croatia.

The research team of this doctoral project is confident that this study will stimulate further research in the field of mental health promotion and prevention effectiveness and quality in Croatia and abroad. It will also encourage the establishment of scientific collaboration of Croatian and foreign universities and scientists involved in the Preffi 2.0 instrument studies.

REFERENCES

Ader, M., Berensson, K., Carlsson, P., Granath, M. and Urwitz, V. (2001). Quality Indicators for Health Promotion Programmes. *Health Promotion International, 16 (2),* 187-195.

Ajzen, I. (1991). The Theory of Planned Behaviour. *Organizational Behaviour and Human Decision Processes, 50,* 179-211.

Alkin, M. (2004). *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks. CA: Sage.

Aro, A., Van den Broucke, S. and Räty, S. (2005). Toward European Consensus Tools for Reviewing the Evidence and Enhancing the Quality of Health Promotion Practice. *Promotion & Education, 12 (1),* 10-14.

Aveline, M.O. (1997). The Limitations of Randomized Control Trials as Guides to Clinical Effectiveness with Reference to the Psychotherapeutic Management of Neurosis and Personality disorders, *Current Opinion in Psychiatry, 10,* 113–115.

Barry, M.M. (2001). Promoting Positive Mental Health: Theoretical Frameworks for Practice. *International Journal of Mental Health Promotion, 3 (1),* 25-34.

Bartholomew, L.K., Parcel, G.S., Kok, G. and Gottlieb, N.H. (2001). *Intervention Mapping: Designing Theory and Evidence-based Health Promotion Programs*. Mayfield: Mountain View.

Bartholomew, L. K., Parcel, G. S., Kok, G., and Gottlieb, N. H. (2006). *Planning Health Promotion Programs: An Intervention Mapping Approach (2nd ed.)*. San Francisco, CA: Jossey-Bass.

Bašić, J. (2009). *Teorije Prevencije: Prevencija Poremećaja u Ponašanju i Rizičnih Ponašanja Djece i Mladih*. Zagreb: Školska knjiga.

Bašić, J., Ferić Šlehan, M., Kranželić Tavra, V. (ur.) (2007a). *Zajednice Koje Brinu - Model Prevencije Poremećaja u Ponašanju: Strategijska promišljanja, Resursi i Programi Prevencije u Istarskoj Županiji*. Pula: Istarska županija.

Bašić, J., Ferić Šlehan, M., Kranželić Tavra, V. (ur.) (2007b). *Zajednice koje Brinu - Model Prevencije Poremećaja u Ponašanju: Epidemiološka studija – Mjerenje Rizičnih i Zaštitnih Čimbenika u Istarskoj Županiji*. Pula: Istarska županija.

Bašić, J., Grozić-Živolić, S. (ur.) (2010). *Zajednice Koje Brinu - Model Prevencije Poremećaja u Ponašanju Djece i Mladih: Razvoj, Implementacija i Evaluacija Prevencije u Zajednici*. Zagreb i Pula: Edukacijsko-rehabilitacijski fakultet Sveučilišta u Zagrebu i Istarska županija.

Bašić, J., Koller-Trbović, N. and Žižak, A. (1993.) *Integralna Metoda u Radu s Predškolskom Djecom i Njihovim Roditeljima: Teorijski Pristup*. Zagreb: Alinea.

Bašić, J., Mihić, J. and Novak, M. (2010). Risk Analysis in the Period of Growing-up of Children and Youth: Starting Point for Effective Prevention. *Journal of Public Health, 19, (1),* 3-11.

Berkey C.S., Hoaglin, D.C., Mosteller, F. and Colditz, G.A. (1995). A Random-Effects Regression Model for Meta-Analysis. *Statistics in Medicine, 14(4*), 395-411.

Borenstein, M., Hedges, L. V., Higgins, L. V. and Rothstein, H. R. (2009.), *Introduction to Meta-Analysis*, Chichester, UK, John Wiley & Sons, Ltd. doi:10.1002/9780470743386.

Botvin, G., Griffin, K.W. and Nichols, D. (2006). Preventing Youth Violence and Delinquency through a Universal School-Based Prevention Approach. *Prevention Science, 7,* 403-408.

Botvin, G. J., Griffin, K. W., Diaz, T., Scheier, L. M., Williams, C. and Epstein, J. A. (2000). Preventing Illicit Drug Use In Adolescents: Long-term Follow-up Data from a Randomized Control Trial of a School Population. *Addictive Behaviours, 25,* 769-774.

Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlag.

Brown, C.H., Berndt, D., Brinales, J.M., Zong, X. and Bhagwat, D. (2000). Evaluating the Evidence of Effectiveness for Preventive Interventions: Using a Registry System to Influence Policy through Science. *Addictive Behaviours, 25 (6),* 955-964.

Browne, G., Gafni, A., Roberts, J., Byrne, C. and Majumdar, B. (2004). Effective/efficient Mental Health Programs for School-age Children: A Synthesis of Reviews. *Social Science and Medicine, 58,* 1367-1384.

Bry, B.H., and Krinsley, K.E. (1992). Booster Sessions and Long-term Effects of Behavioural Family Therapy on Adolescent Substance Use and School Performance. *Journal of Behaviour Therapy and Experimental Psychiatry, 23,* 183-189

Carroll, C., Patterson, M., Wood, S., Booth, Rick, J. and Balain, S. (2007). A Conceptual Framework for Implementation Fidelity. *Implementation Science, 2,* 40-53.

Catalano, R.F., Berglund, M.L., Ryan, J.A.M., Lonczak, H.S., and Hawkins, J.D. (2004). Positive Youth Development in the United States: Research Findings on Evaluations of Positive Youth Development Programs. *Annals of the American Academy of Political and Social Science, 591,* 98-124.

Catalano, R.F., Hawkins, J.D., Berglund, L., Pollard, J.A., and Arthur, M.W. (2002). Prevention Science and Positive Youth Development: Competitive or Cooperative Frameworks? *Journal of Adolescent Health, 31*, 230-239.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioural Sciences* (2nd ed.), New Jersey: Lawrence Erlbaum Associates.

Commonwealth Department of Health and Aged Care (2000). P*romotion, Prevention and Early Intervention for Mental Health — A Monograph*. Canberra: Mental Health and Special Programs Branch, Commonwealth Department of Health and Aged Care.

Connor, D. F. (2002). *Aggression and Antisocial Behaviour in Children and Adolescents: Research and Treatment*. New York, NY: The Guilford Press.

Cronbach, L.J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.

Cuijper, P. (2002). Effective Ingredients of School-based Drug Prevention Programs. A Systematic Review. *Addictive Behaviours 27,* 1009–1023.

Cuijpers, P., van Straten, A., Smit, F., Mihalopoulos, C. and Beekman, A. (2008). Preventing the Onset of Depressive Disorders: A Meta-Analytic Review of Psychological Interventions. *American Journal of Psychiatry, 165,* 1272–1280

Dadds**,** M. (2001). Politics of Pedagogy: the State and Children. *Teachers and Teaching: Theory and Practice, 7 (1),* 49-53.

Dane, A. and Schneider, B. (1998). Program Integrity in Primary and Early Secondary Prevention: Are Implementation Effects out of Control. *Clinical Psychology Review, 18,* 23-45.

DeVellis, R.F. (1991). *Scale development*. Newbury Park, NJ: Sage Publications.

Domitrovich, C., Bradshaw, C.P., Greenberg, M.T., Embry, D., Poduska, J.M. and Ialongo, N.S. (2010). Integrated Models of School-Based Prevention: Logic and Theory. Psychology in the Schools, Vol. 47(1). Published online in Wiley Inter Science (www.interscience.wiley.com) DOI: 10.1002/pits.20452.

Dryfoos, J. G. (1990). *Adolescents at risk: Prevalence and prevention.* New York: Oxford University Press.

Dryfoos, J. G. (1998). *Safe Passage: Making it Through Adolescence in a Risky Society*. New York: Oxford University Press.

Durlak, J.A. (1998). Why Program Implementation is Important. *Journal of Prevention and Intervention in the Community, 17,* 5–18.

Durlak, J.A. and DuPre, E.P. (2008). Implementation Matters: a Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *American Journal of Community Psychology, 41,* 327–50.

Durlak, J., Weisberg, R., Dymnicki, A., Taylor, R. and Schellinger, K. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development, 82 (1),* 405–432.

Durlak, J.A., and Wells, A.M. (1998). Evaluation of Indicated Preventive Intervention (secondary prevention) Mental Health Programs for Children and Adolescents. *American Journal of Community Psychology, 26,* 775-802.

Dusenbury, L., Brannigan, R., Hansen, W.B., Walsh, J. and Falco, M. (2005). Quality of Implementation: Developing Measures Crucial to Understanding the Diffusion of Preventive Interventions. *Health Education Research, 20,* 308–13.

Elliot, M. A. and Ainsworth, K. (2012). Predicting University Undergraduates' Binge-Drinking Behaviour: A Comparative Test of the One- and two-component Theories of Planned Behaviour. *Addictive Behaviours, 37,* 92-101.

Elliot, D. and Mihalic, S. (2004). Issues in Disseminating and Replicating Effective Prevention Programs. *Prevention Science, 5,* 47-53.

European Commission (1999). *The Evidence of Health Promotion Effectiveness. Shaping Public Health in a New Europe,* a report for the European Commission by the International Union for Health Promotion and Education, European Commission, Brussels-Luxembourg.

Forgatch, M., Patterson, G. and DeGarmo, D. (2005). Evaluating Fidelity: Predictive Validity for a Measure of Competent Adherence to the Oregon Model of Parent Management Training. *Behaviour Therapy, 36,* 3-13.

Friedli, L. (2007). *Social Marketing and Mental Health Briefing*. London: Care Services Improvement Partnership.

Gaviria, A. and Raphael, S. (2001). School – Based Peer Effect Sand Juvenile Behaviour. *Review of Economics and Statistics, 83(2),* 257–268.

Glasgow, R.E., Vogt, T.M. and Boles, S.M. (1999). Evaluating of Public Health Impact of Health Promotion Interventions: The RE-AIM Framework. *American Journal of Public Health, 89 (9),* 1322-1327.

Glasser, W. (1998). *Choice Theory: a New Psychology of Personal Freedom*. New York: Harper Perennial.

Glied, S., and Cuellar, A. (2003). Trends and Issues in Child and Adolescent Mental Health. *Health Affairs, 22(5),* 39-50.

Godin, G., Bélanger-Gravel, A., Eccles, M. and Grimshaw, J. (2008). Healthcare Professionals' Intentions and Behaviours: A Systematic Review of Studies Based on Social Cognitive Theories. *Implementation Science, 3,* 36 - 50.

Green, J. (2000). The Role of Theory in Evidence-based Health Promotion Practice, *Health Education Research, 15 (2),* 125-129.

Greenberg, M.T., Domitrovich, C.E, and Bumbarger, B. (2000). *Preventing Mental Disorders in School-aged Children: A Review of the Effectiveness of Prevention Programs* (revised). State College, PA: Prevention Research Center for the Promotion of Human Development, College of Health and Human Development, Pennsylvania State University.

Greenberg, M., Domitrovich, C.E., Graczyk, P. and Zins, J. (2001). *The Study of Implementation in School-Based Preventive Interventions: Theory, Research, & Practice.* Report to Center for Mental Health Services (CMHS), Substance Abuse Mental Health Services Administration, U.S. Department of Health and Human Services.

Greenberg, M.T., Weissberg, R.P., O'Brien, M.U., Zins, J.E., Fredericks, L., Resnik, H. and Elias, M.J. (2003). Enhancing School-based Prevention and Youth Development Through Coordinated Social, Emotional, and Academic Learning. *American Psychologist, 58,* 466-474.

Hawe, P., Degeling, D. and Hall, J. (1990). *Evaluating Health Promotion*. Sydney: MacLennan and Petty.

Hawe, P., King, L., Noort, M., Jordens, C. And Lloyd, B. (2000). *Indicators to Help With Capacity Building in Health Promotion*. Sydney: New South Wales Health Department.

Hawe, P., Noort, M., King, L. and Jordens, C. (1997). Multiplying Health Gains: the Critical Role of Capacity Building Within Health Promotion Programs. *Health Policy,* 29-42.

Hawkins, J.D., Kosterman, R., Catalano, R.F., Hill, K.G., and Abbott, R.D. (2005). Promoting Positive Adult Functioning Through Social Development Intervention in Childhood: Long-Term Effects from the Seattle Social Development Project. *Archives of Paediatrics and Adolescent Medicine, 159,* 25-31.

Herman, H., Saxena, S. and Moodie, H. (eds) (2005). *Promoting Mental Health: Concepts, Emerging Evidence and Practice*. Geneva: World Health Organization.

Hibell, B., Guttormsson, U., Ahlström, S., Balakireva, O., Bjarnason, T., Kokkevi, A. and Kraus, L. (2012). The 2011 ESPAD Report on Substance Use Among Students in European

Countries.http://www.uredzadroge.hr/upload/File/Istrazivanja/Espad/The_2011_ESPAD_Report_FULL_2012_06-08.pdf

Hindshaw, S. P. (2002). Prevention/intervention Trials and Developmental Theory: Commentary on the Fast track Special Section. *Journal of Abnormal Child Psychology, 30,* 50– 53.

Hodgson, R., Abbasi, T. and Clarkson, J. (1996). Effective Mental Health Promotion. A Literature Review. *Health Education Journal, 55,* 55-74.

Horowitz, J.L., and Garber, J. (2006). The Prevention of Depressive Symptoms in Children and Adolescents: A Meta-analytic Review. *Journal of Consulting and Clinical Psychology, 74,* 401-415.

Hosman, C.M.H. (1994). *Effectiveness and Effect Management in Mental Health Promotion and Prevention*. In D. Trent and C. Reed. Mental Health Promotion. Aldershot: Avebury.

Hosman, C. M.H. (1997). Conceptual Clarifications on Promotion and Prevention: A Plea for Functionally Related Fields. *Promotion of Mental Health. 6,* 179-203.

Hosman, C.M.H. (2008). *Evidence of Effectiveness and Improving Effectiveness in Prevention and Mental Health Promotion*. Published in: Textbook Inleiding Preventieve Geestelijke Gezondheidsz.

Hosman, C.M.H. and Engels, C. (1999). The Value of Model Programmes in Mental Health Promotion and Mental Disorder Prevention. *International Journal of Mental Health Promotion, 1, 2,* 1-14.

Hosman, C.M.H., Jane-Llopis, E., and Saxena, S. (eds.) (2004). *Prevention of Mental Disorders: Effective Interventions and Policy Options*. WHO Summary Report. Geneva: World Health Organization.

Huchting, K., Lac, A. and LaBrie, J. W. (2008). An Application of the Theory of Planned Behaviour to Sorority Alcohol Consumption. *Addictive Behaviours, 33,* 538-551.

ISO (1993). *Guide to the Expression of Uncertainty in Measurement*. International Organization for Standardization (ISO) and the International Committee on Weights and Measures (CIPM): Switzerland.

Itzak, A. (1991). The Theory of Planned Behaviour. *Organizational Behaviour and Human Decision Processes*, *50*, 179-211.

Jane-Llopis, E. (2002). *What Makes the Ounce of Prevention Effective? A Meta-analysis of Mental Health Promotion and Mental Disorder Prevention Programmes* (dissertation). Nijmegen*:* University of Nijmegen.

Jane-Llopis, E., and Anderson. P. (2005). *Mental Health Promotion and Mental Disorder Prevention: A policy for Europe*. Nijmegen: Radboud University Nijmegen.

Jane-Llopis, E. and Barry, M. M. (2005). What Makes Mental Health Promotion Effective? *Promotion and Education, 2,* 47-54.

Jane-Llopis, E., Hosman, C. Jenkins, R., and Anderson, P. (2003). A Meta-analysis of Depression Prevention Programmes: What Predicts Effect? *British Journal of Psychiatry, 183,* 384-397.

Jané-Llopis, E., Katschnig, H, McDaid, D. and Wahlbeck, K. (2010). *Evidence in Public Mental Health – Commissioning, Interpreting and Making Use of Evidence on Mental Health Promotion and Mental Disorder Prevention: an Everyday Primer*. Lisabon: Instituto Nacional de Saude Doutor Ricardo Jorge.

Kahan, B. and Goodstadt, M. (1999). Continuous Quality Improvement and Health Promotion: Can CQI Lead to Better Outcomes? *Health Promotion International*, *14 (1),* 83-91.

Kellam, S.G., Brown, C.H., Poduska, J.M., Ialongo, N.S., Wang, W., Toyinbo, P., Petras,H., Ford, C., Windham, A., and Wilcox, H.C. (2008). Effects of a Universal Classroom Behaviour Management Program in First and Second Grades on Young Adult Behavioural, Psychiatric, and Social Outcomes. *Drug and Alcohol Dependence, 95, 1*, S5-S28

Kellam, S. G. and Langevin, D. J. (2003). A Framework for Understanding "Evidence" in Prevention Research and Programs. *Prevention Science*, *4(3),* 137-153.

Keleher, H. and Armstrong, R. (2005). *Evidence-based Mental Health Promotion* Resource. Melbourne: Department of Human Services.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., and Walters, E.E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Co morbidity Survey Replication. *Archives of General Psychiatry, 62(6),* 593-602.

Kirby, D., Short, L., Colins, J., Rugg, D, Kolbe, L., Howard, M., Miller, B., Sonenstein, F. and Zabin, L.S. (1994). School-Based Programs to Reduce Sexual Risk Behaviours: A Review of Effectiveness. *Public Health Reports, 109(3)* ,339-360.

Kok, G., Lo, S. H., Peters, Y. G-J. and Ruiter, A.C.R. (2011). Changing Energy-related Behaviour: An Intervention Mapping approach. *Energy Policy, 39,* 5280-5286.

Kok, G.J., Van den Borne, B. and Dollan Meullen, P. (1997). Effectiveness of Health Education and Health Promotion: Meta-analysis of Effect Studies and Determinants of Effectiveness. *Patient Education and Counselling, 30,* 19-27.

Kumpfer, K. L. (1997). *What Works in the Prevention of Drug Abuse: Individual, School and Family Approaches.* DHHS, Center for Substance Abuse Prevention.

Kuzman, M., Pejnović Franelić, I., Pavić Šimetin, I. and Pejak, M. (2008). *Europsko istraživanje o pušenju, alkoholu i drogama među učenicima: Izvješće za Republiku Hrvatsku i Grad Zagreb*. Zagreb: Hrvatski zavod za javno zdravstvo i Grad Zagreb - Gradski ured za zdravstvo, rad, socijalnu zaštitu.

Lawshe, C.H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology, 28*, 563–575.

Lewis S. and Clarke M. (2001). Forest Plots: Trying to See the Wood and the Trees. *British Medical Journal, 322 (7300*), 1479-1480.

Lipsey, M.W. and Wilson, D.B. (eds.) (2001). *Practical Meta-Analysis.* CA: Sage, Thousand Oaks.

Lochman, J.E. (2001). Issues in Prevention with School-aged Children: Ongoing Intervention Refinement, Developmental theory, Prediction and Moderation, and Implementation and Dissemination. *Prevention and Treatment, 4.* Available on the Web: http://journals.apa.org/prevention/volume4/pre0040004c.html

Marchand, E., Stice, E., Rohde, P. and Black Becker, C. (2011). Moving from Efficacy to Effectiveess Trials in Prevention Research. *Behaviour Research and Therapy 49,* 32-41.

McGrew, J. and Griss, M. (2004). Concurrent and Predictive Validity of Two Scales to Assess the Fidelity of Implementation of Supported Employment. *Psychiatric Rehabilitation Journal, 29*, 41-47.

McQueen, D. and Jones, C.M. (2007). (eds) *Global Perspectives on Health Promotion Effectiveness*. New York: Springer.

Mihalic, S. (2004). The Importance of Implementation Fidelity. *Emotional & Behavioural Disorders in Youth, 4,* 83-86 and 99–105.

Milas, G. (2005). *Istraživačke Metode u Psihologiji i Drugim Društvenim Znanostima*. Zagreb: Naklada Slap.

Miller, N. and Polloc, V. (1994). *Meta-analytic Synthesis for Theory Development*. In Cooper, H. and Hedges, L. (eds.), *The Handbook of Research Synthesis*. Russell Sage Foundation, New York, 457-483.

Mittelmark, M.B., Jones, C.M., Lamarre, M-C., Perry, M., Van der Wel, M., and Wise, M. (2007). *Enhancing the Effectiveness and Quality of Health Promotion*. In: McQueen, D. and Jones, C.M. (eds.), *Global Perspectives on Health Promotion Effectiveness*. New York: Springer, 385-418.

Molleman, G.R.M. (2005). *Preffi 2.0: Health Promotion Effect Management Instrument. Development, validity, reliability and usability*. Woerden: Netherlands Institute for Health Promotion and Disease Prevention.

Molleman, G.R.M., Peters, L.W.H., Hommels, L.M. and Ploeg, M.A. (2003). *Assessment Package; Health promotion Effect Management Instrument Preffi 2.0*. Woerden: Netherlands Institute for Health Promotion and Disease Prevention.

Molleman, G.R.M., Peters, L.W.H., Hosman, C.M.H. and Kok, G. (2005a). Implementation of a Quality Assurance Instrument (Preffi 1.0) to Improve the Effectiveness of Health Promotion in The Netherlands. *Health Education Research, 20* (4), 410-422.

Molleman, G.R.M., Peters, L.W.H., Hosman, C.M.H., Kok, G., and Oosterveld, P. (2005b). Project Quality Rating by Experts and Practitioners with Preffi 2.0 as a Quality Assessment Instrument. *Health Education Research, 21 (2),* 219 – 229.

Montano, D. and Kasptzyk, D. (2002*). The Theory of Reasoned Action and the Theory of Planned Behaviour*. In: Glanz, K., Rimer, B.K. and Lewis, F.M. (eds.), *Health Behaviour and Health Education, 3rd edition*. Jossey-Bass, San Francisco CA, p. 67-98.

Mrazek, P.J. and Haggerty, R.J. (1994). *Reducing the Risks for Mental Disorders: Frontiers for Preventive Intervention Research*. Washington DC: National Academy Press.

Mrazek, P. J. and Haggerty, R. J. (2010). *Reducing Risk for Mental Disorders: Frontiers for Preventive*. In: B. L. Levin, K. Hennessy, & J. Petrila (Eds.), *Mental health services: A public health perspective* (3rd Ed.). (pp.355-356). New York: Oxford University Press.

Murray, C.J.L. and Lopez A.D. (1996). *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injury and Risk Factors in 1990 and Projected to 2020*. Geneva: World Bank, World Health Organization & Harvard School of Public Health.

Nation, M., Crusto, C., Wandersman, A., Kumpfer, K.L., Seybolt, D., Morrissey-Kane, E. And Davino, K. (2003). What Works in Prevention. Principles of Effective Prevention Programs. *American Psychologist, 58 (6),* 449-456.

Neil, A. and Christensen, H. (2009). Efficacy and Effectiveness of School-based Prevention and Early Intervention Programs for Anxiety. *Clinical Psychology Review 29,* 208–215.

Nunnally, J.C. and Berstein, I.H. (1994). *Psychometric Theory*. New York: McGraw-Hill.

O'Connell, M.E., Boat, T. and Warner, K.E. (ed.) (2009). *Preventing Mental, Emotional, and Behavioural Disorders Among Young People: Progress and Possibilities*. National Research Council and Institute of Medicine, Committee on Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults: Research Advances and Promising Interventions. Board on Children, Youth, and Families, Division of Behavioural and Social Sciences and Education. Washington, DC: The National Academies Press.

Petras, H., Kellam, S.G., Brown, C.H., Muthén, B.O., Ialongo, N.S., and Poduska, J.M. (2008). Developmental Epidemiological Courses Leading to Antisocial Personality

Disorderand Violent and Criminal Behaviour: Effects by Young Adulthood of a Universal PreventiveIntervention in First- and Second-grade Classrooms. *Drug and Alcohol Dependence*, 95, *1*, S45-S59.

Pregrad, J. (2007). *Priručnik Projekt za Sigurno i Poticajno Okruženje u Školama – Prevencija i Borba Protiv Nasilja Među Djecom*. Zagreb: Ured UNICEF-a za Hrvatsku.

Prohaska, J.O., Redding, C.A. and Evers, K.E. (2002). *The Transtheoretical Model and Stages of Change*. In: Glanz, K., Rimer, B.K. and Lewis, F.M. (eds), *Health Behaviour and Health Education, 3rd edition*. Jossey-Bass, San Francisco CA, p. 99-120.

Raphael, B. (1999). The Value of Model Programmes in Mental Health Promotion and Mental Disorder Prevention. *Journal of Public Mental Health, 1 (2),* 20-21.

Raudenbush, S.W. (2009). *Analyzing Effect Sizes: Random Effects Models*. In: Cooper, H., Hedges, L.V. and Valentine, J.C. (eds.), *The Handbook of Research Synthesis and Meta-Analysis,* 2nd edition, pp. 295-315. New York: Russell Sage Foundation.

Rogers, E. (2005a). *Diffusion of innovations*. New York: The Free Press.

Rogers, P.J. (2005). *Logic models*. In: Sandra Mathison (eds.) *Encyclopaedia of Evaluation*. Beverly Hills, CA: Sage Publications.

Rorbach, L. A., Graham, J. W. and Hansen, W. B. (1993). Diffusion of School-Based Substance Abuse Prevention Program: Predictors of Program Implementation. *Preventive Medicine, 22,* 237– 260.

Rosnow, R. L., and Rosenthal, R. (1996). Computing Contrasts, Effect Sizes, and Counter Nulls on Other People's Published Data: General Procedures for Research Consumers. *Psychological Methods*, *1,* 331-340.

Rossi, P.H., Lipsey, M.W. and Freeman, H.E. (2004). *Evaluation. A Systematic Approach*. London: Sage Publications.

Serrat., O. (2010). *The Future of Social Marketing*. Knowledge Solutions.

Shadish, W.R. (1998). Evaluation Theory is Who We Are. *American Journal of Evaluation*. *19(1),* 1-19.

Shaw, I.E., Green, J.C. and Melvin, M.M. (2007). *Handbook of Evaluation – Policies, Programs and Practices.* London: Sage Publications.

Shavelson, R. J., and Webb, N. M. (1991). *Generalizability Theory: A Primer*. Newbury Park, CA: Sage Publications.

Speller, V., Evans, D. and Head, M.J. (1997). Developing Quality Assurance Standards for Health Promotion Practice in UK. *Health Promotion International, 12 (3),* 215-224.

Spoth, R., Rohrbach, L., Hawkins, D., Greenberg, M., Pentz, M., Robertson, E. and Sloboda, Z. (2008). *Type 2 Translational Research: Overview and Definitions. Society for Prevention Research Mapping Advances in Prevention Science Task Force*. (http://www.preventionresearch.org/SPR_Type%202%20Translation%20Research_Overview%20and%20Definition.pdf).

Stice, E., Gau, J., Presnell, K., and Shaw, H. (2007). Testing Mediators of Intervention Effects in Randomized Controlled Trials: An Evaluation of Two Eating Disorder Prevention Programs. *Journal of Counselling and Clinical Psychology, 75 (1),* 20 – 32.

Stice, E., and Shaw, H. (2004). Eating Disorder Prevention Programs: A Meta-analytic Review. *Psychological Bulletin, 130(2),* 206-227.

Stice, E., Shaw, H., Bohon, C., Marti, N. and Rohde, P. (2009). A Meta-Analytic Review of Depression Prevention Programs for Children and Adolescents: Factors that Predict Magnitude of Intervention Effects. *Journal of Consulting and Clinical Psychology*. *77(3),* 486–503.

Taylor, J. (1997). *An Introduction to Error Analysis*, 2nd. ed. CA: University Science Books: Sausalito.

Thomas, H., Siracusa, I., Gross, G., Beath, I., Hanna, I., Michaud, M., Moore, P., Partingon, B., Tober, J., Voorberg, N. and Brunton, G. (1999). *Effectiveness of School-based Interventions in Reducting Adolescent Risk Behaviour: A Systematic Review of Reviews.* Ontario: Public Health Research Education & Development Program.

Tilford, S., Delaney, F. and Vogles, M. (1998). *Effectiveness of Mental Health Promotion Interventions – a Review*. Health Education Authority. www.york.ac.uk/inst/crd/ehs33.pdf

Tobler, N.S. and Stratton, H.H. (1997). Effectiveness of School-based Drug Prevention Programs: A Meta-analysis of the Research. *Journal of Primary Prevention, 18,* 71-128.

Tolan, P., and Dodge, K. (2005). Children's Mental Health as a Primary Care and Concern. *American Psychologist, 60(6),* 601-614.

Van den Broucke, S. (2003). *The Role of Academic Research in Health Promotion Capacity Buildin*g. Bangkok: In Interregional Consultancy on capacity building for Health Promotion WHO.

Vermeulen, T. R. N., Kok, H. H., Van den Broucke, S., Molleman, G. R. M. and Bollars, C. (2005). Health Promotion Quality Assurance Self-Assessment tools; Systematic Review and Qualitative Content Analysis. (*submitted*).

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the Metafor. *Journal of Statistical Software Package, 36 (3),* 1-48.

Vreeland, B. (2007). Bridging the Gap between Mental and Physical Health: A Multidisciplinary Approach. *The Journal of Clinical Psychiatry, 68, 4,* 26-33.

Vukasović, T., Bratko, D. and Butković, A. (2012). Genetski Doprinos Individualnim Razlikama u Subjektivnoj Dobrobiti: Metaanaliza. *Društvena Istraživanja, 1(115),* 1-17.

Webb, N., Shavelson, R.J. and Haertel, E.H. (2006). Reliability Coefficients and Generalizability Theory. *Handbook of Statistics, 26,* 4-45.

Wilson, S.J., and Lipsey, M.J. (2007). Effectiveness of School-based Intervention Programs on Aggressive Behaviour: Update of a Meta-analysis. *American Journal of Preventive Medicine, 33, 2,* 130-143.

Wood, C. and Wise, M. (1997). *Building Australia's Capacity to Promote Mental Health: Review of Infrastructure for Promoting Health in Australia*. Canberra: National Mental Health Strategy.

Wyatt Knowlton, L. and Philips, C.C. (2009). *The Logic Model Guidebook*. Better Strategies for Better Results. USA: Sage Publications.

Yamada, J., DiCenso, A., Feldman, I., Cormillott, P., Wade, K., Wignall, R. and Thomas, H. (1999). A Systematic Review of the Effectiveness of Primary Prevention Programs to Prevent Sexually Transmitted Diseases (STDs) in Adolescents. Public Health Research Education and Development. http://www.ephpp.ca/PDF/1999_Adolescent%20STD%20Prevention.pdf

Yokishawa, H. (1994). Prevention as Cumulative Protection: Effects of Early Family Support and Education on Chronic Delinquency and its Risks. *Psychological Bulletin, 115,* 28– 54.

Zhu, J. and Han, L. (2011). Analysis on the Main Factors Affecting the Reliability of Test Papers. *Journal of Language Teaching and Research, 2 (1),* 236-238.

Internet sources:

Centre for Prevention Research, http://www.erf.unizg.hr/CPI/CentarZaPrevencijskaIstrazivanja.html

Croatian National Institute of Public Health, http://www.hzjz.hr/

Croatian Healthy Cities Network, http://www.zdravi-gradovi.com.hr/

Ministry of Interior Affairs of the Republic of Croatia, http://www.mup.hr/

National Drug Abuse Prevention Strategy in Croatia for the Period From 2012 to 2017, http://www.uredzadroge.hr/upload/File/Dokumenti/Nacionalna_strategija_2012_2017/Nacionalna_strategija_suzbijanja_zloup.droga_2012.-2017.pdf

National Strategy for Mental Health Care for the Period from 2011to 2016, www.vlada.hr

National strategy for prevention of behavioural problems of children and youth for 2009-2012, http://www.mobms.hr/media/17218/strategija%20pup%20izvje%C5%A1ce.pdf

Rubin, R. J. (2004). Outcomes Measurement Glossary. California State Library. Retrieved from http://www.library.ca.gov/assets/acrobat/lsta/OutcomesMeasurementGlossary.pdf

Society for Prevention Research, http://www.preventionresearch.org/

Glossary of Evaluation Terms, http://pdf.usaid.gov/pdf_docs/PNADO820.pdf

APPENDIX A

Description of Questionnaire on Content Validity of Preffi 2.0

The questionnaire has two parts – the introductory part in which the theoretical background of Preffi 2.0 is explained to the assessors and the part which consist of all 121 Preffi 2.0 items for which assessors were asked to mark their scores on a five-point Likert scale. At the end of the questionnaire, estimators could write additional comments and elaborate their opinion about the accordance of the instrument's items with the theoretical background of the Preffi 2.0 instrument.

Since all Preffi 2.0 items are presented in Appendix B, only the introductory part of the Questionnaire on content validity of Preffi 2.0 will be presented in this Appendix.

INSTRUCTIONS FOR ASSESSING THE CONTENT VALIDITY OF PREFFI 2.0 INSTRUMENT

Dear participants,

thank you for your willingness to participate in this assessment and contribute to the development of prevention science and practice in Croatia. This questionnaire consists of several units. In the first part you need to answer several general socio-demographic questions which are mostly related to your professional work. That part is followed by a brief description of the theoretical background of Preffi 2.0 instrument. I kindly ask you to read it carefully because, according to your understanding of the theory on effect predictors and theoretical background of Preffi, you will estimate the content validity of each Preffi 2.0 item. The theoretical part is followed by Preffi 2.0 items which are distributed within eight sections/clusters. For each item you will need to estimate the level of the item, according to your opinion and understanding of the theoretical background of Preffi 2.0, relevant and essential for assessing the concept of quality indicators. 1 stands for "Completely not in accordance with the theory and not essential", 2 for "Not in accordance with the theory and not essential", 3 for "Partially in accordance with the theory and essential", 4 for "Mostly in accordance with the theory and essential", and 5 stands for "Completely in accordance with the theory and essential". At the end of the questionnaire there is space where you can qualitatively present and explain your opinion about content validity of the Preffi 2.0 instrument

SOCIO-DEMOGRAPHIC QUESTIONS

Gender:

Age:

Level of education:

Occupation:

Years of experience in the field of mental health promotion and prevention:

DESCRIPTION OF THEORETICAL BACKGROUND OF PREFFI 2.0

Growing evidence suggests that with high-quality interventions a wide variety of behavioural and mental health problems can be reduced, including violence and delinquency (Botvin et al., 2006), tobacco and alcohol use (Tobler and Stratton, 1997), risk sexual behaviour (Kirby et al. 1994), and other emotional problems (Hawkins et al. 2005; Hosman et al., 2004). Randomized control studies in mental health promotion and prevention science provided evidence that many interventions are effective in promoting mental health and preventing behavioural problems. Outcome and impact research provides the knowledge necessary to identify predictors of efficacy and effectiveness in mental health promotion and prevention programs. Nation and colleagues (2003) have found, just as Dryfoos (1998), that there is substantial overlap in the principles of effective programs across mental health promotion and prevention domains that allow us to identify general principles of effectiveness. Determinants of an intervention's impact or effect are referred to as "effect predictors" or "effect moderators" (Hosman & Engels, 1999; Raphael, 1999; Hosman, 1994).

The main assumption of the Dutch mental health promotion and prevention scientists was that the knowledge about effect predictors translated into practical guidelines that will be used by prevention practitioners in developing and implementing prevention programs could systematically increase the effectiveness of the intervention. With that intention the PREvention EFFect-management Instrument (Preffi 1.0) was developed (Molleman, 2005). Preffi 1.0 was designed for professionals in the field of mental health promotion and prevention in the form of guidelines that can be used in developing and improving their interventions to maximize their effectiveness. Preffi 1.0 was upgraded into a Preffi 2.0 version in order to assess if programs are designed in such a way to be maximally effective in reaching their goals. The Preffi 2.0 instrument consists of 39 quality criteria – effect predictors, variables that are demonstrably related to the program's intended output. These

criteria are distributed within eight clusters: (1) problem analysis, (2) determinants of behaviour and environment, (3) target group, (4) objectives, (5) intervention development, (6) implementation, (7) evaluation and contextual conditions, and (8) feasibility. Preffi 2.0 items reflect research findings on effect predictors, as well as insights into such predictors derived from critical discussions with practitioners. Table 1. presents effect predictors detected within different research studies in the field of mental health promotion and prevention.

Table 1.

An Overview of Effect Predictors Detected by Different Authors

Study Author/s and Year	***Effect Predictors***
Stice et al. (2007)	*- fit between the program and the population it targets* *- characteristics of the intervention itself (e.g., duration, methods, socio-cultural relevance)*
Bartholomew et al. (2001)	*- systematic application of available theoretical and empirical knowledge during the processes of intervention development and implementation*
Kok et al. (1997); Tobler and Stratton (1997); Brown et al. (2000); Jane-Llopis and Barry (2005)	*- clear goals and objectives* *- theoretical basis of the program*
Jane-Llopis and Barry (2005)	*- theoretical basis of the program, clear goals and objectives, high quality evaluation and research methods, infrastructural support from management, program fidelity and transferability to different countries and cultures* *- high quality implementation, training and supervision of program providers, high participation in the program sessions*
Durlak et al. (2011) *Dryfoos (1990)*	*- quality of implementation process* *- provision of intense individualized attention, multilevel interventions, early identification of a problem, training based on the skills development, engagement of peers and parents in the intervention*

Nation and colleagues (2003)	*- comprehensiveness, various teaching methods, sufficient dosage, theoretical basis, opportunities for positive relationships, appropriate timing, socio-cultural relevance, outcomes evaluation, well-trained staff*
Ader and colleagues (2001)	*- quality of program's structure - goals, target group, design, responsibility, resources, and organization* *- quality of program's process – network, commitment, exposure, participation* *- quality of program's outcomes – behavioural changes, environmental changes, epidemiological changes and maintenance*
Tobler and Stratton (1997)	*- quality of the research design (evaluation)*
Nation et al. (2003); Jane-Llopis, Hosman, Jenkins et al. (2003)	*- variety of intervention methods* *- appropriate timing*

Stice and colleagues (2007) stress that crucial effect predictors in mental health promotion and prevention are characteristics of the intervention itself (e.g., duration, methods, socio-cultural relevance) and the fit between the program and the population it targets. Bartholomew and colleagues (2001) noticed that the potential effect of the intervention could be much higher when available theoretical and empirical knowledge is systematically applied in the processes of intervention development and implementation . Programs that have clear goals and objectives, and that are theory-based both in terms of the targeted risk and protective factors and the mechanisms of change used in the intervention program have a more positive impact (Jane-Llopis and Barry, 2005; Brown et al., 2000; Kok et al., 1997; Tobler and Stratton, 1997). High quality of implementation is found to be a core effect predictor, associated with positive intervention outcomes (Durlak et al, 2011). Jane-Llopis and Barry (2005) stress that high quality implementation, including training and supervision of program providers and high participation in the program sessions predicted higher program effectiveness. Those authors gave a systematic review of the crucial factors identified in determining program success – theoretical basis of the program, clear goals and objectives, program provider training and support, evaluation and high quality research

methods, infrastructural support from management, program fidelity and transferability to different countries and cultures. Also, comprehensive programs that utilize a variety of methods and that are delivered at the appropriate time are more successful (Nation et al., 2003; Jane-Llopis, Hosman, Jenkins et al., 2003). Tobler and Stratton (1997) also found that programs rating higher in the quality of the research design were significantly more effective than programs that rated lower in quality of the program evaluation. Ader and colleagues (2001) have detected 14 quality indicators that have proved to be necessary and important in mental health promotion and prevention and need to be clarified during interventions' development. Those include:

(2) indicators that refer to the program's structure - goals, target group, design, responsibility, resources, and organization,

(2) indicators that refer to the program's process – network, commitment, exposure, participation and

(3) the indicators of the program's outcomes – behavioural changes, environmental changes, epidemiological changes and maintenance.

Dryfoos (1990) reviewed over 100 prevention programs related to substance abuse, teen pregnancy, school dropout and juvenile delinquency. Her review yielded several key characteristics associated with successful programs such as provision of intense individualized attention, multilevel interventions, early identification of a problem, training based on skills development, and engagement of peers and parents in the intervention. Nation and colleagues (2003) have identified 9 characteristics that were consistently associated with effective prevention programs across 4 areas – substance abuse, risky sexual behaviour, school failure, and juvenile delinquency and violence. According to their findings, effective programs were: (1) comprehensive, (2) included varied teaching methods, (3) provided sufficient dosage, (4) were theory driven, (5) provided opportunities for positive relationships, (6) were appropriately timed, (7) were socio-culturally relevant, (8) included outcomes evaluation, and (9) involved well-trained staff. Those nine characteristics could be related to the four broad areas of prevention programming:

1. *Program characteristics,*
2. *Matching programs to target population,*
3. *Implementation quality and*
4. *Evaluation of the interventions.*

Besides findings on effect predictors, authors of Preffi 2.0 as a theoretical starting point of the instrument also used the Intervention planning model (Bartholomew et al., 2006). Prefi 2.0 clusters follow the logical steps of that model:

- *analysing the problem, including its nature and scale, as well as its determinants,*
- *making successive decisions about objectives, target groups and suitable intervention types,*
- *paying special attention to the inclusion of effective elements, as derived mostly from social learning theory,*
- *pre-testing and implementing the interventions,*
- *evaluating in terms of both process and effect, and*
- *contextual conditions and project management, including the personal characteristics of the project manager.*

In accordance with the Intervention planning model, the authors have developed a model of Preffi 2.0, shown in Figure 1.

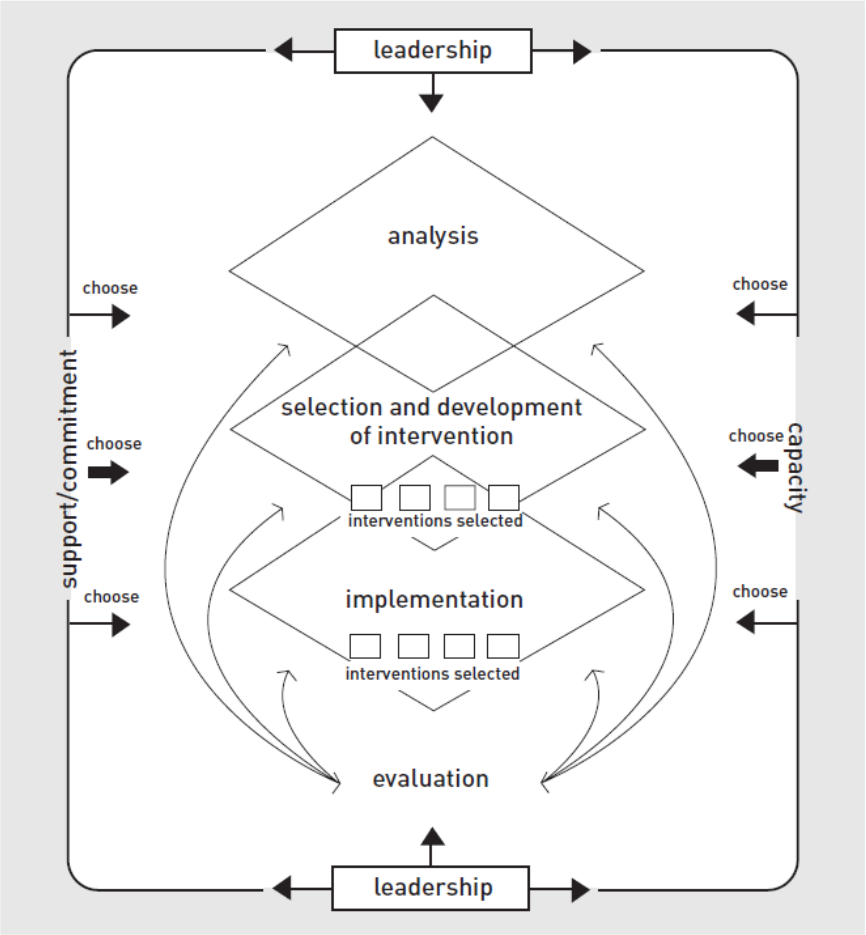


Figure 1.

Preffi 2.0 Model

The model illustrates the dynamics of all concepts needed for the development of a program. From the model it is evident that there are key elements for the development of a program presented in the center of the model - the problem analysis and analysis of a need for the program, the process of developing and planning the program and program implementation and evaluation. On the sides are those elements that represent contextual conditions for the development of a program - the quality of leadership in the organization that conceived the program, the capacity to implement the program, support organizations and deliverers of a program. Also, it is clear from the model that the different phases of program development are interrelated and interdependent. For example, the design of program evaluation should be synchronized with the objectives of a program, but it also depends on some contextual conditions, such as the financial resources necessary for evaluation .

APPENDIX B

PREFFI 2.0

OPERATIONALISATION AND NORMS

The purpose of this document is to help you fill out the Preffi 2.0 evaluation sheet. We recommend that you first read the user manual in the Preffi 2.0 assessment package. The document includes one or more "yes" or "no" questions for each Preffi criterion (unit). Based on the answers to those questions, the criterion (unit) can be categorized as „weak", „moderate" or „strong".

Some questions may be difficult to answer, especially if the project plan does not provide enough information or if you yourself lack knowledge related to certain fields of expertise. In any case, you should answer as many questions as possible. The answer „not available" is offered in a limited number of criteria, usually in the cases when the criteria are difficult to put in words in project descriptions or when they are not obvious to those who are not themselves included in project implementation (for example, „competence and characteristics of the project manager" and „adjusting to the culture"). If some criteria allow "not available" as an answer, this will be explicitly noted.

The document provides space for comments on every criterion; for example, your comments on why you answered a certain question with "yes" or "no". You may also specify and describe aspects you believe need improvement, and you can also transfer this to the Answer Sheet.

Criteria in the document are listed in the same order as in the Evaluation sheet. So the document starts with Problem analysis, cluster 2 and ends with Contextual conditions and feasibility, cluster 1. The User manual explains the rationale behind such an order.

Cluster 2. Problem analysis

2.1. Nature, severity and scope of the problem

Operationalisation:

1. Is the problem or the topic clear?
2. Is it clear whether the problem or the topic is frequent within the group or community?
   Additional questions:
   Is the prevalence of the problem known (=number of existing cases)?
   Is the incidence of the problem known (=number of new cases in a certain period of time)?
3. Is the interrelatedness of health and social problems clear? This includes indicators like rate of unemployment, income, fear of crime, racial discrimination, drug addiction, number of welfare cases and housing conditions.
4. Is what is known about immaterial costs of the problem clearly stated – such as mortality (mortality rate, life expectancy), diseases and disorders, limitations, disabilities, harmful impact, medicine use and absence from work?
5. Is what is known about material costs clearly stated – such as cost of services, health care costs, measurement costs, loss of revenue due to attempts to solve or contain the problem.

Norms:

- Weak: question 1 = no and /or question 2 = no and/or question 3 = no
- Moderate: questions 1 - 3 = yes and question 4 and/or 5 = no or not available
- Strong: questions 1 - 3 = yes and questions 4 and/or 5 = yes

2.2. Distribution of the problem

Operationalisation:

1. Is it clear how the problem is distributed regarding:
   - age?
   - sex?
   - socio-economic status?
   - ethnical background?
   - religious background?

- cultural or subcultural origin?
- time (seasons, days of the week, hours of the day)?

2. Is anything known about the geographical distribution of the problem, in terms of a certain County, city or area? (For example, the unusually high mortality from cancer in a certain County; traffic accidents on certain intersections; fear of crime in certain streets or buildings, etc.)
3. Are data available for a specific target area at which the project is aimed (designed for the whole country or a province, County, city, town district)? If not, has data been correctly extrapolated from general data?

Note: each question enumerates many points of interest, but not all of these need to be of importance for every project situation.

Norms:

- ❖ Weak: question 1 = no
- ❖ Moderate: question 1 = yes and question 2 = no and question 3 = no
- ❖ Strong: question 1 = yes and question 2 = yes and/or question 3 = yes

2.3. Perception of the problem by key people

Operationalisation:

1. Is it known to what extent the problem is actually perceived by the target group as a problem?
2. Has it been established which individuals, groups, agencies and parts of the social sector are involved in the process of tackling or solving problems?
3. Has it been established to what extent these individuals, groups, agencies and parts of the social sector agree about the source and cause of the problem?
4. Has it been established how major social subgroups, such as ethnic or cultural groups, men and women or different types of schools, perceive the problem?
5. Has it been checked whether politicians and the public opinion are interested in or pressure for certain steps to be taken for solving the problem?

Norms:

- ❖ Weak: question 1=no and/or question 2=no
- ❖ Moderate: at least question 1 = yes and question 2 = yes (with the possibility that the answer to some of the remaining questions is also yes)

- ❖ Strong: at least question 1 = yes and question 2 = yes and two more questions = yes

Note: In this cluster the 'target group' always implies the final target group to which the project refers to.

Cluster 3. Determinants of (psychological) problems, behaviour and environment

3.1. Theoretical model

Operationalisation:

1. Have the theoretical assumptions or the model used for explaining the (psychological) problem, risk and desired behaviour or environmental factors been clearly stated?
2. Has it been clearly shown that the selected model is most suitable for approaching these (psychological) problems, behaviour or environmental factor (for example, because the model has been specifically developed for a specific problem, behaviour or environmental factors, because the model has already been successfully applied or it has been discussed in a scientific journal or because its applicability can be supported by theoretical arguments)?
3. Has it been clearly described how factors affect each other, how they affect behaviour, environmental factors and/or the problem – favourably or unfavourably?

Norms:

- Weak: questions 1 and/or 2 = no
- Moderate: 1 = yes, 2 = yes, 3 = no
- Strong: all questions = yes

3.2. Contribution of determinants to psychological problems, behaviour or environmental factors

Operationalisation:

1. Is it known which determinants influence desired and undesired behaviour, environmental factors or the (psychological) problem (on a personal level, on the level of social environment and psychological environment)?
2. Is it clear which determinants are the most important?
3. Is it clear how reliable is the evidence of determinants?
4. Is it clear to what extent determinants can be applied to relevant subgroups (e.g. according to age, sex, ethnicity, religion, etc.)

Norms:

- Weak: question 1 = no (making the other questions irrelevant)
- Moderate: 1 = yes and at the most one more question = yes
- Strong: question 1 = yes and at least two more questions = yes

3.3. Susceptibility of determinants to change

Operationalisation:

1. Has it been estimated to what extent determinants are susceptible to change in the described situation (on the level of an individual and on the level of social and physical environment)?
2. Has this estimate been based on theoretical and/or scientific knowledge about the variability of determinants? (Suggestion: consult relevant literature, co-workers or experts, conduct preliminary testing)

Norms:

- Weak: question 1 = no (making the second question irrelevant)
- Moderate: question 1 = yes, question 2 = no
- Strong: question 1 = yes and question 2 = yes

3.4. Priorities and selection

Operationalisation:

1. Have the target behavioural or environmental factors or (psychological) problems been specified?
2. Has it been explained to which health problem(s) or life quality problem(s) these factors are related?
3. How have the target determinants for behavioural or environmental factors or (psychological) problems been explained?
4. Have the groups in risk and/or target groups been mentioned and specified?

Norms:

- Weak: maximum of two questions = yes
- Moderate: three questions = yes
- Strong: all questions = yes

Cluster 4. Target group

Note: In this cluster, the expression 'target group' always refers to the final target group.

4.1. General and demographic characteristics of the target group

Operationalisation:
Suggestion: Much of the data collected during problem analysis is also probably relevant in this chapter.

1. Is it clear which general and demographic characteristics are relevant for this specific project? An affirmative answer requires that at least the first five characteristics from the following list apply:
   - the size of the target group
   - age
   - sex
   - socio-economic status (level of education, income, profession, work status)
   - ethnical background
   - cultural background
   - religious background
   - marital status, housing conditions
   - number of family members
   - geographic position
   - language (spoken and written), illiteracy
2. Are concrete figures available about relevant characteristics of the target group in this project?

Norms:

- Weak: both questions = no
- Moderate: question 1 = yes, question 2=no
- Strong: both questions = yes

4.2. Motivation and options of the target group

Operationalisation:

1. Is it known to what extent the target group is ready for change?

2. Is it known which factors influence the motivation of members of the target group to change? (These can include awareness of the problem, attitude, self-efficiency, obstacles, etc. Suggestion: see also cluster 3)
3. Is it known for the purpose of this specific project, which desires, needs, limitations and obstacles for change the group is aware of?

Norms:

- Weak: question 1 = no, regardless of the answers to questions 2 and 3
- Moderate: question 1 = yes and question 2 or 3 = no
- Strong: all questions = yes

4.3. Accessibility of the target group

Operationalisation:

1. Is it clear by what means the target group can be covered? (Suggestion: think about locations, media, intermediary persons)
2. Is the selection of the means (locations, media, intermediaries) corroborated by the project?

Norms:

- Weak: both questions = no
- Moderate: question 1 = yes and question 2 = no
- Strong: both questions = yes

Cluster 5. Objectives

5.1. Objectives are adjusted to the analysis

Operationalisation:

1. Does the objective description make a clear distinction between different objective levels? The levels may refer to health/life quality, behaviour/environment/problems and determinant's level as well as the level of objectives for creating preconditions.
2. Do the objectives adjust and are they in accordance with the analysis conducted in previous clusters? (see clusters 2 and 3)

Norms:

- Weak: both questions = no
- Moderate: one question = yes and one question = no

- Strong: both questions = yes

5.2. Objectives are specific, time-limited and measurable

Operationalisation:

1. Do objectives specify factors that need to be changed? (Suggestion: This question has been analysed in 5.1)
2. Has for the objectives a target group been specified in which these objectives need to be achieved?
3. Do objectives specify the desired magnitude of effects that wants to be achieved (e.g.: 10% decrease)?
4. Do objectives specify the time period in which they need to be realised?

Norms:

- Weak: questions 1. and/or 2.=no
- Moderate: question 1=yes, question 2=yes, 3. question=no, 4. question=no
- Strong: question 1=yes, question 2=yes and questions 3 and/or 4=yes

5.3. Objectives are acceptable

Operationalisation:

1. Are the project theme and the set objectives in accordance with the objectives of your organisation?
2. Are the intervention objectives acceptable (or can they become acceptable) for financing/to the evaluation board or maybe to the medical ethical board/institutional board for evaluation?
3. Are the objectives of the intervention acceptable (or can they become acceptable) to possible partners and implementers?
4. Are the intervention objectives acceptable (or can they become acceptable) to the target group?

We are aware that questions related to this criterion are not easy to answer. If you can answer them use the given norm. If you cannot give an answer, mark “not available” on the Answer Sheet.

Norms:

- Weak: at least one negative answer in questions 1-3
- Moderate: questions 1 - 3 = yes
- Strong: questions 1 – 4 = yes

5.4. Objectives are achievable

Operationalisation:

1. Has the necessary personnel, money and time for achieving the set objectives been estimated? (Suggestion: data from criterion 3.3 can be useful here)
2. Is there a sufficient number of available experts, competent persons and partners for achieving the set objectives?

We are aware that questions related to this criterion are not easy to answer. If you can answer them use the given norm. If you cannot give an answer, mark "not available" on the Answers Sheet.

Norms:

- Weak: question 1 = no, regardless of the answer to question 2
- Moderate: question 1 = yes and question 2 = no
- Strong: both questions = yes

Cluster 6. Intervention development

Note: If the project includes more interventions, you can answer the questions in general. However, if you are interested in assessing each specific intervention, it is possible to answer each question separately. (see User Manual, section 3.3)

6.1. Rationale for the intervention strategy

6.1.a. Adjusting the strategies and methods to objectives and target groups

Operationalisation:

1. Have the intervention methods been specified?
2. Has it been established how intervention methods are appropriate and adequate for achieving the desired objectives (e.g. through research or theoretical considerations)?

Norms:

- Weak: both questions = no
- Moderate: question 1 = yes and question 2 = no

- Strong: both questions = yes

6.1.b. Previous experience with intervention(s)

Operationalisation:

1. Are any reports available about a successful or unsuccessful use of this intervention by someone else (in literature or through other experts)?
2. Do you as an individual have some experience of a successful or unsuccessful application of intervention?
3. Does the suggested method seem potentially efficient for this specific situation? (Suggestion: you have to consider the extent to which your situation can be compared to other situations where some experience has already been gathered, especially concerning objective terms/determinants, themes/problems, target groups and contextual conditions)

Norms:

- Weak: question 1 = no and question 2 = no
- Moderate: question 1 and/or 2 = yes and question 3 = no
- Strong: question 1 and/or 2 = yes and question 3 = yes

6.2. Duration, intensity and chronology

6.2.a Duration and intensity of the intervention

Operationalisation

1. Are some research data or practical experiences available about the duration and intensity in which the intervention should be implemented in order to achieve the set objectives?
2. Has this data been used in deciding on the optimum duration and intensity of the proposed intervention?

Norms:

- Weak: question 1 = no (making the answer to the second question irrelevant)
- Moderate: question 1 = yes and question 2 = no
- Strong: both questions = yes

6.2.b. Intervention chronology

Operationalisation:

1. Has it been established whether target groups react better to the intervention in a certain time of year? (This can for example refer to education about sunbathing and skin cancer during summer. Religious and state holidays should be taken into account or periods like Ramadan).
2. Has it been established whether the chronology of the intervention is compatible with specific important experiences of target group members? (This can for example refer to care interventions for mental health in crisis situations or to the level of sexual experience of young people included in the AIDS prevention program).
3. Has it been established to what extent the time period of the intervention agrees with the age or development stage of the target group? (This can for example refer to the information that interventions for preventing aggression with children are most effective if they are conducted when the children are 3 or 4 years of age).
4. In the case when the intervention is to be implemented with the help of intermediary persons, has the chronology of the intervention been adjusted to these persons?

Norms:

- Weak: maximum one question = yes
- Moderate: maximum two questions = yes
- Strong: at least three questions = yes

6.3. Adjusting to the target group

6.3.a. Participation of the target group

Operationalisation:

1. In the case when the intervention has been developed somewhere else (for example, on the national level): has the general target group been at least consulted during intervention development?
2. For any project: has the specific target group (e.g. residents of a target district) for the ongoing project at least been consulted during intervention development or before selecting the intervention model?
3. For any project: regarding the project's characteristics, has the target group been sufficiently involved in development and intervention selection?

Norms:

- Weak: question 1 = no or not available and question 2 = no (making the third question irrelevant)

- Moderate: questions 1 and/or 2 = yes, question 3 = no
- Strong: question 1 and/or 2 = yes, question 3 = yes

6.3.b. Adjusting to 'culture'

Note: The term 'culture' is used in the broadest sense; it can include adapting to age, sex, socio-economic status, etc. For example, it might be necessary to address young people differently than adults and older people.

Operationalisation:

1. Is the content (message) in accordance with knowledge, views, customs, roles and capacities of members of the cultural or subcultural groups?
2. Is the way of reaching members of the cultural or subcultural group adequate and does it adequately convey the messages? Is the medium for communication frequently used and attractive?
3. Is the target group accessible to the source or message transmitter (e.g. intermediary)?
4. Has the source or message transmitter shown proof of sufficient understanding and knowledge about culturally determined customs and social norms of the target group?
5. Does the target group perceive the intervention as being in accordance with their culture?

We are aware that questions related to this criterion are not easy to answer. If you can answer them use the given norm. If you cannot give an answer, mark “not available” on the Evaluation sheet.

Norms:

- Weak: question 1, 2, 3 and/or 4 = no
- Moderate: questions 1 - 4 = yes and 5 question = no
- Strong: all questions = yes

6.4. Effective techniques

Operationalisation:

1. Have the following techniques been used in the project, considering the importance they have for the project to be assessed?

Effective techniques

- a room for personalised approach
- feedback (about the situation in the target group, behaviour or intervention effects)
- use of rewarding strategies
- removal of obstacles towards the desired behaviour
- mobilising social support/commitments, involving the social environment
- training skills
- ensuring follow-up
- defining objectives and implementation intentions
- interactive approach

We are aware that questions related to this criterion are not easy to answer. If you can answer them, use the given norm. If you cannot give an answer, mark "not available" on the Answer Sheet.

Norms:

- Weak: none or few of the effective strategies have been used
- Moderate: some effective strategies have been used
- Strong: many effective strategies have been used

6.5. Feasibility in existing practice

6.5.a. Adjusting to the intermediary target group

Operationalisation:

1. Have the members of the intermediary target group been consulted during the development process of the intervention (for the final target group)?
2. Is the intervention in accordance with ways of operating, procedures, standards and values of intermediaries and their organisation?

We are aware that questions related to this criterion are not easy to answer. If you can answer them, use the given norm. If you cannot give an answer, mark "not available" on the Answer Sheet.

Norms:

- Weak: both questions = no

- Moderate: one of two questions = no
- Strong: both questions = yes

6.5.b. Intervention/s feasibility characteristics

Operationalisation:
The following questions can be answered for every intermediary group separately:

1. Has it been established/recorded to what extent intermediaries feel that the use/implementation of the intervention will improve their current practice?
2. Has it been established/recorded to what extent intermediaries feel that the new intervention is in accordance with the current procedure?
3. Has it been established/recorded to what extent intermediaries possess the necessary skills for implementing the intervention?
4. Has it been established/recorded whether the intervention procedure is clear to the intermediaries, i.e. whether they know what is expected of them?
5. Has it been established/recorded whether the intermediaries think that the new intervention gives them enough space for experimenting? Can intermediaries test the intervention without being strictly bound by the content of the intervention?
6. Has it been established/recorded whether the intermediaries feel they can immediately notice the effects of the intervention?
7. Has it been established/recorded to what extent intermediaries feel the intervention to be affordable?

We are aware that questions related to this criterion are not easy to answer. If you can answer them, use the given norm. If you cannot give an answer, mark "not available" on the Answer Sheet.

Norms:

- Weak: 0 - 2 questions = yes
- Moderate: 3 - 5 questions = yes
- Strong: 6-7 questions = yes

6.6. Coordination between intervention/activity

Operationalisation:

1. Is the program comprehensive enough to reach the set objectives? In other words, does it make sufficient use of available segments of intervention methods, ways and determinants of the target group?

2. If the program/project includes multiple interventions (segments of intervention methods, ways and determinants of the target group), are these different interventions coordinated in a satisfying manner?

Norms:
- Weak: both questions = no
- Moderate: one of two questions = yes
- Strong: both questions = yes

6.7. Preliminary testing

Operationalisation:
1. Has preliminary testing been used?
2. Have conclusions been made and steps taken in accordance, in terms of communication and/or effects, based on preliminary testing? In other words, has the intervention been adjusted where necessary?

Norms:
- Weak: question 1 = no (making the second question irrelevant)
- Moderate: question 1 = yes and question 2 = no
- Strong: both questions = yes

Cluster 7. Implementation

7.1. The selection of the implementation strategy adjusted to intermediaries

7.1.a. Implementation model: top down and/or bottom-up

Operationalisation:
1. Have certain implementation models been selected deliberately?
2. Do intermediaries have the chance of adjusting the intervention to their own situation?
3. If intermediaries have the chance of adjusting the intervention, is it clear which parts of the intervention need to be preserved?

Norms:

- Weak: question 1 = no or question 1 = yes and question 2 = no (making the third question irrelevant)
- Moderate: question 1 = yes, question 2 = yes and question 3 = no
- Strong: all questions = yes

7.1.b. Adjusting intervention implementation to intermediaries

Operationalisation:

1. Is it clear how members of the intermediary group are distributed during different expansion and innovation application phases (awareness of innovation; decision to apply the innovation; reporting the innovation; continued innovation application)
2. Have specific objectives for each expansion and innovation application phase been set, for every segment of the intermediary or target group?
3. Do the implementation interventions fit in with the objectives that have been set for each stage of diffusion and use and for each intermediary target group or target groups segment?
4. Are the set objectives realistic considering the fact that the intermediary group can be divided into 'innovators', 'early adopters', 'early majority', 'late majority' and 'laggards'?

Norms:

- Weak: question 1 = no and question 2 = no
- Moderate: question 1and/or 2 = yes
- Strong: at least questions 1-3 = yes

7.1.c. Appropriateness of the supplier for intermediaries

Operationalisation:

1. Is it known whether the planned supplier is appropriate in the eyes of the intermediary target group? Aspects of appropriateness include:

- support/commitment
- authority
- competence
- image
- the size of the supplier agency
- position within the network

- financial capacity and other available resources

2. Are different contact persons used, when appropriate, for different segments of the intermediary target group?

We are aware that questions related to this criterion are not easy to answer. If you can answer them, use the given norm. If you cannot give an answer, mark "not available" on the Answer Sheet.

Norms:
- Weak: question 1 = no (characteristics have not been taken into consideration)
- Moderate: question 1 = yes (some characteristics have been take into consideration) and question 2 = no (different persons were looked for but were not included)
- Strong: question 1 = yes (some characteristics have been take into consideration) and question 2 = yes or it is irrelevant

7.2. Monitoring and generating feedback

Operationalisation:
1. Has it been specified in how many points of time the expansion progress and intervention implementation will be assessed, e.g. by collecting feedback from intermediaries and the final target group?
2. Does the assessment lead to an active adjustment of the expansion process and intervention implementation?

Norms:
- Weak: both questions = no
- Moderate: question 1 = yes and question 2 = no
- Strong: both questions = yes

7.3. Incorporation into existing structures

Operationalisation:
1. Has the intervention been incorporated into the existing structure?
2. Has an effort been made, or is it made right now, to fit the intervention into already existing structures?
3. Are these activities and attempts strong enough, i.e. are they aimed at the right hierarchical level? (e.g. it is easier to influence business people through other business people)

Norms:

- Weak: question 1 = no and question 2 = no (making the third question irrelevant)
- Moderate: question 1 = no, question 2 = yes and question 3 = no
- Strong: question 1 = yes or question 2 = yes and question 3 = yes

Cluster 8. Evaluation

Note: In the case of the project including more interventions and/or evaluations, the questions can be answered generally. However, if you are interested in evaluating each specific intervention or evaluation, it is possible to provide answers for each intervention or evaluation separately, for example with the help of the matrix. (see User Manual, section 3.3)

8.1. Explicitness and agreement on evaluation principles

Operationalisation:

1. Have important individuals, groups and/or organisations been included in designing the evaluation? This refers to commission organisations, the ones who need to implement the intervention, members of the target group and potential external experts.
2. Do all key people have a clear idea about the questions that the evaluation must answer and do they agree on these questions?
3. Is it clear which form/s of the evaluation is/are necessary in order to answer the questions?
4. Do key people agree about the strength of proof that needs to be obtained through the evaluation and is this level of proof achievable?

Norms:

- Weak: question 2 = no
- Moderate: question 2 = yes and question 1 and/or question 3 = no
- Strong: at least questions 1, 2, 3 = yes

8.2. Process evaluation

Operationalisation:

1. Does the process evaluation allow insight into the degree to which the activities have been implemented according to plan?
2. Does the process evaluation allow insight into user's opinions (final and/or intermediary target group) about activities and materials?

3. Does the process evaluation allow insight into intervention coverage (which people have been included, how representative are they, who was excluded from the intervention and why)?
4. Does the process evaluation allow insight into the degree to which the objectives of creating preconditions for the project have been reached?
5. Does the process evaluation allow insight into possible unpredictable circumstances and side-effects?
6. Does the process evaluation reveal conditions for success?

Norms:
- Weak: maximum three questions = yes
- Moderate: three or four questions = yes
- Strong: at least five questions = yes

8.3. Effect evaluation

Note: We are aware that answering questions in this cluster requires certain professional knowledge about effect evaluation, which can make it more difficult for individuals to answer. It is a problem we are not able to solve at this moment. It is in this sense our goal to offer support through the Internet version of Preffi 2.0. which is to be developed in the future.

8.3.a. Has any change been measured or is being measured at this moment?

Operationalisation:
1. Has it been measured (or is it being measured now) to which degree the objectives of the intervention have been reached (or are reached)? It is necessary to take into account different objectives (emphasized in section 5.1), especially momentary (or intermediate) intervention objectives. This will mostly not include end objectives of the intervention on a public health level since their realisation requires a longer period.
2. Are the used measuring methods valid and reliable? This concerns questions referring to outcome measures, measuring methods, measuring instruments and the size and representative quality of the sample/group that is being studied.

Norms:
- Weak: both questions = no
- Moderate: question 1 = yes and question 2 = no or not available
- Strong: both questions = yes

8.3.b. Is it likely that the change was caused by intervention?

Operationalisation:

1. Is it clear which of the alternative explanations for noticed changes can be excluded (out of the six possible explanations listed in the criteria in the manual)? Special attention should be directed towards information about study design and use of multiple measuring methods and multiple sources (e.g. results of the process evaluation and effect evaluation) and to the degree of agreement between their findings.
2. Is the level of credibility of the made conclusions justified by the level of security offered by the study design? Conclusions must be aligned with the measure in which alternative explanations cannot be excluded.

We are aware that questions related to this criterion are not easy to answer. If you can answer them use the given norm. If you cannot give an answer, mark "not available" on the Answer Sheet.

Norms:

- Weak: both questions = no
- Moderate: question 1 = yes and question 2 = no
- Strong: both questions = yes

8.4. Feedback to key people

Operationalisation:

1. Have key people been introduced to important feedback acquired in the evaluation process? This includes the following aspects:

- Do the findings agree with problems noticed and/or questions asked by key people?
- Does the provided information include aspects the key people have the power to change? (Can this information be used to derive some policy recommendations?)
- Have any side-effects been clearly shown?
- Are the proposed measures acceptable to key people?
- Will findings be available within a reasonable time?

2. Is the manner of presenting the findings adjusted to key people (in terms of readability and conciseness)?

We are aware that questions related to this criterion are not easy to answer. If you can answer them, use the given norm. If you cannot give an answer, mark “not available” on the Answer Sheet.

Norms:

- Weak: question 1 = no
- Moderate: question 1 = yes and question 2 = no
- Strong: both questions = yes

Cluster 1. Contextual conditions and feasibility

1.1.Support/commitment

Operationalisation:

1. Has it been established which internal and external partners are necessary for ensuring adequate support and commitment during every phase of the project?
2. Is there sufficient support and commitment among the required partners?
3. Have agreements been made and confirmed about involving internal and external collaborators in the project?

We are aware that questions related to this criterion are not easy to answer. If you can answer them use the given norm. If you cannot give an answer, mark “not available” on the Answer Sheet.

Norms:

- Weak: question 1 = yes or no, question 2 = no and question 3 = no
- Moderate: question 1 = yes, question 2 = yes and question 3 = no
- Strong: all questions = yes

1.2.Capacity

Operationalisation:

1. Have available resources for the project been established?
2. Are the available resources in line with the objectives of the project?
3. Have the available resources in every phase of the project been used in the most efficient way?

Norms:

- Weak: all questions = no OR question 1 = yes or no, question 2 = no and question 3 = no or 'not available'
- Moderate: question 1 = yes, question 2 = yes and question 3 = no or 'not available'
- Strong: all questions = yes

1.3.Management by the project manager

1.3.1. Expertise and characteristics of the project manager

Operationalisation:

1. Is only one person responsible for the project?
2. Does the person with exclusive responsibility have the necessary competence for implementing the project?
3. Is the work style of the person with exclusive responsibility for the project compatible with the specific phase and peculiarities of the project?
4. Does the person with exclusive responsibility for the project have appropriate personal characteristics for implementing the project?

We are aware that questions related to this criterion are not easy to answer. If you can answer them, use the given norm. If you cannot give an answer, mark “not available” on the Answer Sheet.

Norms:

- Weak: question 1 = no or question 2 = yes and at least one more question yes
- Moderate: question 1 = yes and two questions out of 2, 3 and 4 = yes
- Strong: all questions = yes

1.3.b. Key points for management

Operationalisation:

1. Is the project being implemented in accordance with the project plan which includes clear moments for making decisions?
2. Is the communication plan being actively implemented?
3. Has the project manager the opportunity to use available resources in a flexible manner?

4. Is the project manager ensuring that his/her competence and the competence of other staff involved is up-to-date by organising additional training, intervision, etc.?

We are aware that questions related to this criterion are not easy to answer. If you can answer them, use the given norm. If you cannot give an answer, mark “not available” on the Answer Sheet.

Norms:

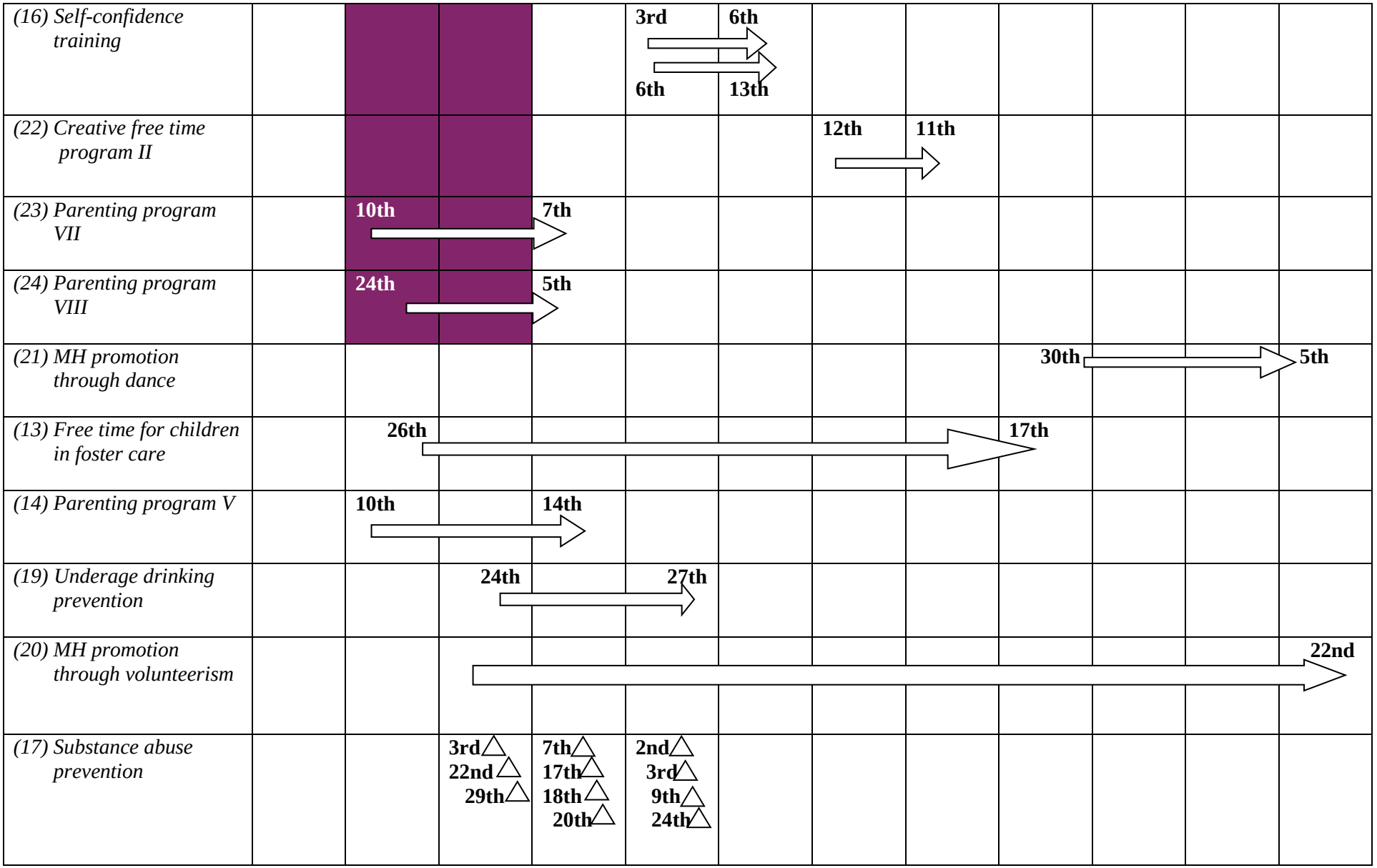
- Weak: question 1 = no; or question 2 = yes and at least one more question = yes
- Moderate: question 1 = yes and two of the remaining questions = yes
- Strong: all questions = yes

APPENDIX C

Dates of Assessments Conducted Within the Study on Effectiveness of Programs in Experimental and Control Conditions

Legend: *Intervention Training for prevention,* *Duration of interventions and dates of pre-test and post-test measurements*
Only one lecture in the program

CODE OF THE PROGRAM	JAN 2011	FEB 2011	MAR 2011	APR 2011	MAY 2011	JUN 2011	JUL 2011	AUG 2011	SEP 2011	OCT 2011	NOV 2011	DEC 2011
(1)MH promotion through the theatre		**19th**							**28th**			
(4)Media literacy										**17-27th**	**23rd** **2th** **24th** **6th** **28th** **5th** **30th** **5th** **2-21st** **7-21st** **7-23rd**	
(5)Training for group leaders			**9-30th**									
(6) Substance abuse prevention for parents												**1st**
(7) Substance abuse prevention for teachers											**15th**	
(8) Parenting program II	**9th**			**14th**								


(16) Self-confidence training
3rd
6th
6th
13th
(22) Creative free time program II
12th
11th
(23) Parenting program VII
10th
7th
(24) Parenting program VIII
24th
5th
(21) MH promotion through dance
30th
5th
(13) Free time for children in foster care
26th
17th
(14) Parenting program V
10th
14th
(19) Underage drinking prevention
24th
27th
(20) MH promotion through volunteerism
22nd
(17) Substance abuse prevention
3rd
22nd
29th
7th
17th
18th
20th
2nd
3rd
9th
24th

(11) Substance abuse prevention in the community										**28th**		**10th**
(18) Parenting program VI										**10th**	**21st**	
(15) Peer-violence prevention program			**11th** **14th** **15th** **22nd**	**13th** **13th**	**17th** **23rd**							
(12) Creative free time program I								**22nd**		**13th**		
(10) Parenting program IV									**13th**	**17th**	**9th** **17th** **7th**	**14th** **15th** **12th**
(9) Parenting program III			**8th**	**12th**							**8th**	**20th**

APPENDIX D

Review of Measures Used in the Evaluation of Mental Health Promotion and Prevention Program Outcomes

CODE OF THE PROGRAM	MEASURED CONSTRUCTS	NAME AND AUTHOR OF THE MEASURE	DESCRIPTION OF THE MEASURE
(1) MH promotion through the theatre	Self-esteem	Coopersmith's self-esteem questionnaire (Lacković-Grgin et al., 2002)	25 items, agreement and disagreement with the statement (1- correct, 2- incorrect)
	Social skills - cooperation, assertion and self-control	Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990)	3 subscales, 30 items, three-point Likert scale of likelihood (1- never, 2- sometimes, 3– often)
(4) Media literacy	Reactions on cyber bullying	Scale of reactions on cyber bullying (Unicef, Cro)	11 items, five-point Likert scale of likelihood (from 1- not true at all to 5- completely true)
	Attitudes towards cyber bullying	Scale of attitudes towards cyber bullying (Unicef, Cro)	6 items, five-point Likert scale of agreement (from 1- completely disagree to 5- completely agree)
(5) Training for group leaders	Knowledge about steps of group work process, techniques of group work, characteristics and roles of the group leader	Test of knowledge on group work (Mihić, 2011)	5 open-ended questions, answers were scored from 1 to 5 (1- no knowledge to 5-excellent knowledge)
(6) Substance abuse prevention for parents	Knowledge about behavioural, physical and psychological changes caused by drug abuse, knowledge about available services	Knowledge on drug abuse symptoms (Mihić, 2011)	2 open-ended questions, answers were scored from 1 to 5 (1-no knowledge to 5-excellent knowledge)
		Knowledge on services for drug abuse prevention and treatment (Mihić, 2011)	1 open-ended questions, answers were scored from 1 to 5 (1-no knowledge to 5-excellent knowledge)

CODE OF THE PROGRAM	MEASURED CONSTRUCTS	NAME AND AUTHOR OF THE MEASURE	DESCRIPTION OF THE MEASURE
(7) Substance abuse prevention for teachers	Knowledge about behavioural, physical and psychological changes caused by drug abuse, knowledge about available services	Knowledge on drug abuse symptoms (Mihić, 2011)	2 open-ended questions, answers were scored from 1 to 5 (1-no knowledge to 5-excellent knowledge)
		Knowledge on services for drug abuse prevention and treatment (Mihić, 2011	1 open-ended questions, answers were scored from 1 to 5 (1-no knowledge to 5-excellent knowledge)
(8) Parenting program II	Attachment to partner – anxiety, avoidance	Attachment to partner questionnaire (Kamenov and Jelić, 2003)	2 subscales, 18 items, seven-point Likert scale of likelihood (from 1 – completely not true of me to 7 – completely true of me)
	Parental stress, attachment to children, satisfaction with parental role, positive and negative emotions associated with parental role and difficulties associated with parenting	The parental stress scale (Berry and Jones, 1995)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Perception of competence for parental role	Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Acceptance of children, autonomy, psychological control, supervision, positive discipline, negative discipline, compliance	Parental behaviour questionnaire (Keresteš, 1999)	7 subscales, 29 items, four-point Likert scale of likelihood (1- not at all true of me to 4- true of me)
(9) Parenting program III	Parental behaviours towards children	Scale of parental behaviours (Pećnik, 2010)	4 items, four-point Likert scale of likelihood (1-several times per day, 2-every day once, 3-several times per week, 4 – not even once during one week)

CODE OF THE PROGRAM	MEASURED CONSTRUCTS	NAME AND AUTHOR OF THE MEASURE	DESCRIPTION OF THE MEASURE
(10) Parenting program IV	Symptoms of depression, anxiety and irritability	A clinical scale for the self-assessment of irritability (Snaith et al., 1978)	18 items, four-point Likert scale of likelihood (from 1- never to 4 – very often)
	Parental stress, attachment to children, satisfaction with parental role, positive and negative emotions associated with parental role and difficulties associated with parenting	The parental stress scale (Berry and Jones, 1995)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
(11) Substance abuse prevention in the community	Perceived harm associated with drug use	Scale of perceived harm associated with drug use (Mihić, 2011)	17 items, four-point Likert scale of attitudes (from 1- no risk for health to 4- great risk for health)
(12) Creative free time program I	Perception of competence for parental role	Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Acceptance of children	Parental behaviour questionnaire (Keresteš, 1999)	1 subscale, 4 items, four-point Likert scale of likelihood (from 1- not at all true of myself to 4- true of myself)
(13) Free time for children in foster care	Self-esteem	Coopersmith's self-esteem questionnaire (Lacković-Grgin et al., 2002)	25 items, agreement and disagreement with the statement (1- correct, 2- incorrect)
	Social skills - cooperation, assertion, responsibility, empathy, and self-control	Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990)	4 subscales, 39 items, three-point Likert scale of likelihood (1- never, 2– sometimes, 3– often)

CODE OF THE PROGRAM	MEASURED CONSTRUCTS	NAME AND AUTHOR OF THE MEASURE	DESCRIPTION OF THE MEASURE
(14) Parenting program V	Parental stress, attachment with children, satisfaction with parental role, positive and negative emotions associated with parental role and difficulties associated with parenting	The parental stress scale (Berry and Jones, 1995)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Perception of competence for parental role	Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Acceptance of children, autonomy, psychological control, supervision, positive discipline, negative discipline, compliance	Parental behaviour questionnaire (Keresteš, 1999)	7 subscales, 29 items, four-point Likert scale of likelihood (1- not at all true of myself to 4-true of myself)
	Communication between family members	Scale of family members' communication quality (adapted according to Brajša, Žižak, Mejovšek, Bašić, 1990)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
(15) *Peer-violence prevention program*	Reactions during a conflict situation	Scale of reactions during a conflict situation (Mihić, 2011)	13 items, yes or no answers
(16) Self-confidence training	Self-esteem	Coopersmith's self-esteem questionnaire (Lacković-Grgin, 2002)	25 items, agreement and disagreement with the statement (1- correct, 2- incorrect)
	Social skills	Social skills questionnaire (Ferić Šlehan, Kranželić, 2005)	42 items, five-point Likert scale of likelihood (from 1-not at all true of me to 5 – completely true of me)

CODE OF THE PROGRAM	MEASURED CONSTRUCTS	NAME AND AUTHOR OF THE MEASURE	DESCRIPTION OF THE MEASURE
(17) Substance abuse prevention	Perceived harm associated with drug use	Scale of perceived harm associated with drug use (Mihić, 2011)	10 items, four-point Likert scale of attitudes (from 1-no risk for health to 4-great risk for health)
	Attitudes towards drug use	Scale of attitudes towards drug use (Mihić, 2011)	7 items, four-point Likert scale of agreement (from 1 – completely disagree to 4- completely agree)
(18) Parenting program VI	Parental stress, attachment with children, satisfaction with parental role, positive and negative emotions associated with parental role and difficulties associated with parenting	The parental stress scale (Berry and Jones, 1995)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Perception of competence for parental role	Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Acceptance of children, autonomy, psychological control, supervision, positive discipline, negative discipline, compliance	Parental behaviour questionnaire (Keresteš**,** 1999)	7 subscales, 29 items, four-point Likert scale of likelihood (1- not at all true of me to 4-true of me)
	Communication between family members	Scale of family members' communication quality (adapted according to Brajša, Žižak, Mejovšek, Bašić, 1990)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)

CODE OF THE PROGRAM	MEASURED CONSTRUCTS	NAME AND AUTHOR OF THE MEASURE	DESCRIPTION OF THE MEASURE
(19) Underage drinking prevention	Self-esteem	Coopersmith's self-esteem questionnaire (Lacković-Grgin et al., 2002)	25 items, agreement with the statement (1-true, 2-not true)
	Social skills - cooperation, assertion, responsibility, empathy, and self-Control	Social skills rating system - grades 7-12 (Gresham and Elliott, 1990)	4 subscales, 39 items, three-point Likert scale of likelihood (1- never, 2 – sometimes, 3 – often)
	Attitudes about problem solving strategies	Scale of attitudes about problem solving strategies (Mihić, 2011)	8 items, five-point Likert scale of agreement (from 1- completely disagree to 5 – completely agree)
	Attitudes towards alcohol consumption	Scale of attitudes towards alcohol consumption (Mihić, 2011)	6 items, four-point Likert scale of agreement (from 1- completely disagree to 4 – completely agree)
	Reactions to alcohol offers	Scale of reactions to alcohol offers (Mihić, 2011)	8 items, four-point Likert scale of likelihood (from 1- not true for me to 4 – true for me)
	Perception of advertisements	Scale of perception of advertisements (Mihić, 2011)	5 items, five-point Likert scale of likelihood (from 1- not true for me to 5 – completely true for me)
(20) MH promotion through volunteerism	Social skills - cooperation, assertion, responsibility, empathy, and self-control	Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990)	4 subscales, 39 items, three-point Likert scale of likelihood (1- never, 2 – sometimes, 3 – often)
	Altruism	Scale of altruism (Raboteg-Šarić, 2002)	17 items, four-point Likert scale of likelihood (from 1 – never to 4 – very often)

CODE OF THE PROGRAM	MEASURED CONSTRUCTS	NAME AND AUTHOR OF THE MEASURE	DESCRIPTION OF THE MEASURE
(21) MH promotion through dance	Self-esteem	Coopersmith's self-esteem questionnaire (Lacković-Grgin et al., 2002)	25 items, agreement with the statement (1-true, 2-not true)
	Social skills - cooperation, assertion and self-control	Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990)	3 subscales, 10 items, three-point Likert scale of likelihood (1- never, 2 - sometimes, 3 – often)
	Social and emotional loneliness	Scale of social and emotional loneliness (Ćubela Adorić, 2004)	13 items, seven-point Likert scale of likelihood (from 1- completely disagree to 7- completely agree)
(22) Creative free time program II	Attitudes towards leisure time activities	Scale of attitudes towards leisure time activities (Mihić, 2011)	11 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Leisure time activities	Scale of leisure time activities (Mihić, 2011)	16 items, five-point Likert scale of likelihood (from 1 – not true to 5 – completely true)
(23) Parenting program VII	Parental stress, attachment to children, satisfaction with parental role, positive and negative emotions associated with parental role and difficulties associated with parenting	The parental stress scale (Berry and Jones, 1995)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Perception of competence for parental role	Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Acceptance of children, autonomy, psychological control, supervision, positive discipline, negative discipline, compliance	Parental behaviour questionnaire (Keresteš, 1999)	7 subscales, 29 items, four-point Likert scale of likelihood (1- not at all true of myself to 4-true of myself)

	Communication between family members	Scale of family members' communication quality (adapted according to Brajša, Žižak, Mejovšek, Bašić, 1990)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
(24) Parenting program VIII	Perception of competence for parental role	Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	18 items, five-point Likert scale of agreement (from 1 – completely disagree to 5 – completely agree)
	Acceptance of children, autonomy, psychological control, supervision, positive discipline, negative discipline, compliance	Parental behaviour questionnaire (Keresteš, 1999)	7 subscales, 29 items, four-point Likert scale of likelihood (1- not at all true of myself to 4-true of myself)
	Attachment to a partner – anxiety and avoidance	Attachment to a partner questionnaire (Kamenov and Jelić, 2003	2 subscales, 18 items, seven-point Likert scale of likelihood (from 1 – completely not true of me to 7 – completely true of me)

APPENDIX E

Content Validity Ratios of Preffi 2.0 Items

PREFFI 2.0 ITEMS		N	CVR
FIRST PREFFI CLUSTER "Contextual conditions and feasibility"			
1.1.1.	Has it been established which internal and external partners are required for adequate support and commitment at each stage of the project?	10	1.0
1.1.2.	Is there sufficient support and commitment among the required partners?	10	1.0
1.1.3.	Have agreements been made/confirmed about the involvement of internal and external partners?	10	1.0
1.2.1.	Has the available capacity for the project been assessed?	10	1.0
1.2.2.	Is the available capacity suitable for the project's objectives?	10	1.0
1.2.3.	Is the available capacity being used in the most efficient manner at each stage of the project?	10	1.0
1.3.a.1.	Is one person ultimately responsible for the project?	10	0.80
1.3.a.2.	Does the person who is ultimately responsible have the necessary expertise to implement the project?	10	1.0
1.3.a.3.	Does the person who is ultimately responsible have a work style that is compatible with the specific stage and the nature of the project?	10	1.0
1.3.a.4.	Does the person who is ultimately responsible have the right personal characteristics to implement the project?	10	0.80
1.3.b.1.	Is the project being implemented in accordance with a project plan which includes clear decision moments?	10	1.0
1.3.b.2.	Is a communication plan being actively implemented?	10	1.0
1.3.b.3.	Is the project manager able to utilize the available resources in a flexible manner?	9	1.0
1.3.b.4.	Does the project manager ensure that his/her own expertise and that of the other staff involved is being kept up to date by organizing additional training, intervision, etc.?	10	1.0
SECOND PREFFI CLUSTER "Problem analysis"			
2.1.1.	Is the problem or theme clear?	10	1.0
2.1.2.	Is it clear whether the problem or theme is common in the group or community?	10	1.0
2.1.3.	Is it clear whether the health problem is related to social problems?	9	0.78
2.1.4.	Is it clear what is known about the problem's immaterial costs, such as mortality (mortality rate, life expectancy), diseases and disorders, limitations, disabilities, nuisance, medical consumption, absenteeism?	10	0.80
2.1.5.	Is it clear what is known about the problem's material costs, such as: costs of facilities, health care costs, costs of measures, loss of income resulting from attempts to solve or limit the problem?	10	0.80

PREFFI 2.0 ITEMS		N	CVR
2.2.1.	Is it clear how the problem is distributed in terms of age, sex, socio-economic status, ethnic background, religious background, cultural or subcultural background and time (seasons, days of the week, hours of the day)?	10	1.0
2.2.2.	Is anything known about the geographic distribution of the problem, in terms of certain Countys, towns or districts?	10	0.80
2.2.3.	Are data available for the specific target area of your project (countrywide, province, County, town, district)? If not, have data been correctly extrapolated from more general data?	10	1.0
2.3.1.	Is it known to what extent the target group does indeed perceive the problem as a problem?	10	1.0
2.3.2.	Has it been established which persons, groups, agencies and social sectors are involved in perpetuating or solving the problem?	10	1.0
2.3.3.	Has it been established to what extent these persons, groups, agencies and social sectors agree on the background and causes of the problem?	10	0.60
2.3.4.	Has it been established how major social subgroups, such as ethnic or cultural groups, men and women or various types of schools, perceive the problem?	10	0.80
2.3.5.	Has it been checked whether there is interest or pressure from politicians or public opinion to do something about the problem?	10	0.80
THIRD PREFFI CLUSTER "Determinants of behaviour and environment"			
3.1.1.	Have the theoretical assumptions or models used to explain the (psychological) problems, risk behaviour and preferred behaviour or the environmental factor been made explicit?	10	1.0
3.1.2.	Has it been made plausible that the model chosen is suitable for application to the (psychological) problems, the behaviour or the environmental factor (e.g. because the model was specifically developed for a particular problem, behaviour or environmental factor, has been used successfully before or has been discussed in scientific journals, or because its applicability can be theoretically argued)?	10	1.0
3.1.3.	Is it clear how the factors influence each other, the behaviour, the environmental factor and/or the problem, favourably or unfavourably?	10	1.0
3.2.1	Is it known which determinants (at the personal, social environment and physical environment levels) affect the preferred or undesirable behaviour, the environmental factor or the 10(psychological) problem?	10	1.0
3.2.2.	Is it clear which are the most important determinants?	10	1.0
3.2.3.	Is it clear how reliable the evidence for the determinants is?	10	1.0
3.2.4.	Is it clear to what extent the determinants apply to the relevant subgroups (e.g., age, sex, background in terms of ethnicity, religion, etc.)?	10	1.0

	PREFFI 2.0 ITEMS	N	CVR
3.3.1.	Has it been estimated to what extent the determinants in the situation at hand are amenable to change (at personal level and at the social and physical environment level)?	10	0.80
3.3.2.	Was this estimation based on theoretical and/or empirical knowledge about amenability of the determinants?	10	1.0
3.4.1.	Have the targeted behavioural or environmental factor(s) or (psychological) problems been specified?	10	1.0
3.4.2.	Has it been specified to which health or quality-of-life problem(s) these factors relate?	10	1.0
3.4.3.	Have the targeted determinant(s) of behavioural or environmental factors or (psychological) problems been specified?	9	1.0
3.4.4.	Have the intended risk and/or target groups been specified?	10	1.0
FOURTH PREFFI CLUSTER "Target group"			
4.1.1.	Is it clear what general and demographic characteristics are relevant to this specific project?	10	1.0
4.1.2.	Are concrete figures available about the relevant characteristics of the project's target group?	10	1.0
4.2.1.	Is it known to what extent the target group members are motivated to change?	10	0.80
4.2.2.	Is it known what factors affect the target group members' motivation to change?	10	1.0
4.2.3.	Is it known, in terms of this specific project, what wishes, needs, limitations and barriers to change the target group members themselves perceive?	10	1.0
4.3.1.	Is it clear by what channels the target group can be reached?	10	1.0
4.3.2.	Has the choice of channel(s) (location, medium, person) been substantiated?	10	0.80
FIFTH PREFFI CLUSTER "Objectives"			
5.1.1.	Does the description of the objectives distinguish various levels of objectives?	10	1.0
5.1.2.	Do the objectives fit in with the analysis made in the previous clusters?	10	1.0
5.2.1.	Do the objectives specify the factors to be changed?	10	1.0
5.2.2.	Do the objectives specify the target group for which the intended objective is to be achieved?	10	1.0
5.2.3.	Do the objectives specify the intended magnitude of the effects (e.g., a 10% reduction)?	10	1.0
5.2.4.	Do the objectives specify the time within which the objectives are to be achieved?	10	1.0
5.3.1.	Do the theme and objectives fit in with the goals of your organisation?	10	0.80
5.3.2.	Are the objectives of the intervention acceptable (or can they be made acceptable) to funding/commissioning agencies and perhaps the medical ethics committee/institutional review board?	10	0.80
5.3.3.	Are the intervention objectives acceptable (or can they be made acceptable) to possible partners and implementers?	10	1.0
5.3.4.	Are the intervention objectives acceptable (or can they be made acceptable) to the target group?	10	1.0

	PREFFI 2.0 ITEMS	N	CVR
5.4.1.	Have the staffing, money and time required to achieve the objectives been estimated?	10	1.0
5.4.2.	Are sufficient expertise, authority and partners available to achieve the objectives?	10	1.0
SIXTH PREFFI CLUSTER "Intervention development"			
6.1.a.1.	Have the intervention methods been specified?	10	1.0
6.1.a.2.	Has it been established that the intervention methods are suitable to achieve the intended objectives (e.g. by research or theoretical considerations)?	10	1.0
6.1.b.1.	Are any reports of successful or unsuccessful applications of the intervention by others available (from the literature or from experts)?	10	1.0
6.1.b.2.	Have you yourself had any experience of successful or unsuccessful application of the intervention?	10	0.60
6.1.b.3.	Does the proposed method appear to be potentially effective in this particular situation?	10	0.80
6.2.a.1.	Is any information from research or practice available about the duration and intensity with which the intervention needs to be implemented to achieve the objectives?	9	1.0
6.2.a.2.	Has this information been used to decide upon the optimum duration and intensity of the intervention?	9	1.0
6.2.b.1.	Has it been established whether the target group's receptiveness to the intervention is linked to certain times of the year?	10	0.80
6.2.b.2.	Has it been established whether the timing of the intervention is compatible with specific relevant individual experiences of group members?	10	1.0
6.2.b.3.	Has it been established to what extent the timing of the intervention is compatible with the age or developmental stage of the target group?	10	1.0
6.2.b.4.	If the intervention is to be implemented by intermediate groups: has the timing of the interventions been adapted to the intermediate groups?	10	0.40
6.3.a.1.	In the case of interventions developed elsewhere (e.g. at national level): has the general target group at least been consulted while the intervention was being developed?	10	0.60
6.3.a.2.	For any project: has the specific target group for the present project (e.g., residents of the target district) been at least consulted while the intervention was being developed or before the model intervention was selected?	10	0.80
6.3.a.3.	For any project: in view of the nature of the project, has the target group participated sufficiently in the development or selection of the intervention?	10	0.60
6.3.b.1.	Does the content (the message) fit the knowledge, views, customs, roles and capacities of the members of the cultural or subcultural group?	10	0.80

	PREFFI 2.0 ITEMS	N	CVR
6.3.b.2.	Is the medium (the channel) suitable to reach the members of the cultural or subcultural target group and bring the message across? Is the medium in common use and attractive?	10	1.0
6.3.b.3.	Does the source or transmitter (e.g. the intermediary) have access to the target group?	10	0.80
6.3.b.4.	Has the source or transmitter shown evidence of sufficient understanding and knowledge of the culturally determined customs and social norms among the target group?	10	0.80
6.3.b.5.	Does the target group perceive the intervention as compatible with their culture?	10	1.0
6.4.1.	Have the techniques listed below been used, insofar as they are relevant to the present project? Effective techniques: room for personalised approach, feedback, use of reward strategies, removing barriers to preferred behaviour, mobilising social support/commitment, involving social environment, training skills, arranging follow-up, goal-setting and implementation intentions, interactive approach.	10	1.0
6.5.a.1.	Have members of the intermediate target group(s) been consulted while the intervention (for the ultimate target group) was being developed?	10	0.60
6.5.a.2.	Is the intervention compatible with the modes of operation, procedures, standards and values of the intermediaries and their organisation(s)?	10	0.80
6.5.b.1.	Has it been established/recorded to what extent the intermediaries feel that the use/implementation of the intervention will be an improvement for them, compared with current practice?	10	0.80
6.5.b.2.	Has it been established/recorded to what extent the intermediaries feel that the new intervention is compatible with current operational procedures?	9	0.78
6.5.b.3.	Has it been established/recorded to what extent the intermediaries possess the necessary skills to implement the intervention?	10	1.0
6.5.b.4.	Has it been established/recorded whether the intervention procedure is sufficiently clear to the intermediaries, that is, whether they know what is expected of them?	10	0.80
6.5.b.5.	Has it been established/recorded whether the intermediaries feel that the new intervention leaves them enough room to experiment? Can intermediaries try out the intervention without being strictly committed to it?	10	0.60
6.5.b.6.	Has it been established/recorded whether the intermediaries feel they can immediately notice the effects of the intervention?	10	0.80
6.5.b.7.	Has it been established/recorded to what extent the intermediaries feel the intervention is affordable?	10	0.80
6.6.1.	Is the programme sufficiently comprehensive to meet its objectives? In other words, does it make sufficient use of available intervention methods, channels, settings and target group segments?	10	1.0

	PREFFI 2.0 ITEMS	N	CVR
6.6.2.	If the programme/project involves multiple interventions (intervention methods, channels, settings, target group segments, etc), have these been sufficiently coordinated?	10	1.0
6.7.1.	Has a pre-test been used?	10	1.0
6.7.2.	Have conclusions been drawn from the pre-test results and have these conclusions been acted upon, in terms of communicability and/or effects? In other words: has the intervention been adjusted where necessary?	10	1.0
SEVENTH PREFFI CLUSTER ”Implementation”			
7.1.a.1.	Has a conscious choice been made to use a particular mode of implementation?	10	1.0
7.1.a.2.	Do the intermediaries have the opportunity to adapt the intervention to their own situation?	10	0.60
7.1.a.3.	If the intermediaries have the opportunity to adapt the intervention, is it clear to them which elements must be retained?	10	0.60
7.1.b.1.	Is it clear how the members of the intermediate target group(s) are distributed over the various stages of diffusion and use of innovations (awareness of an innovation; decision whether or not to use the innovation; application of the innovation; continued application).	9	0.78
7.1.b.2.	Have specific objectives been set for each stage of the process of diffusion and use and for each intermediate target group or target group segment?	9	0.55
7.1.b.3.	Do the implementation interventions fit in with the objectives that have been set for each stage of diffusion and use and for each intermediate target group or target group segment?	9	0.55
7.1.c.1.	Has it been established whether the intended supplier is suitable in the eyes of the intermediate target groups?	10	1.0
7.1.c.2.	Are different contact persons being used – where appropriate – for the various segments of the intermediate target group?	10	0.60
7.2.1.	Have a number of moments been specified at which the progress of the diffusion and use of the intervention is to be assessed, for instance by collecting feedback from the intermediate or ultimate target group(s)?	10	1.0
7.2.2.	Does the assessment lead to active steps to adjust the process of diffusion and use of the intervention?	10	1.0
7.3.1.	Has the intervention been incorporated in an existing structure?	10	0.80
7.3.2.	Have attempts been made, or are they being made, to incorporate the intervention in an existing structure?	10	0.80
7.3.3.	Do these activities or attempts carry enough weight, that is, have they been addressed at the right hierarchical level?	10	0.80
EIGHT PREFFI CLUSTER “Evaluation”			
8.1.1.	Are relevant persons, groups and/or organizations involved in designing the evaluation?	10	0.80

	PREFFI 2.0 ITEMS	N	CVR
8.1.2.	Do all stakeholders have a clear idea about the questions that are to be answered by the evaluation and do they agree with these questions?	10	0.80
8.1.3.	Is it clear what form(s) of evaluation is/are required to answer the questions?	10	0.80
8.1.4.	Do stakeholders agree how 'hard' the evidence from the evaluation needs to be and whether this level of evidence is feasible?	10	0.80
8.2.1.	Does the process evaluation provide information about the degree to which the activities have been implemented according to plan?	10	1.0
8.2.2.	Does the process evaluation provide information on the opinion of users (ultimate and/or intermediate target group(s) about the activities and materials?	10	1.0
8.2.3.	Does the process evaluation provide information on the coverage of the intervention (which people have been reached, how representative are they, who have dropped out and why)?	10	1.0
8.2.4.	Does the process evaluation provide information on the degree to which the preconditioned objectives have been achieved?	10	1.0
8.2.5.	Does the process evaluation provide information on possible unforeseen circumstances and side effects?	10	1.0
8.2.6.	Does the process evaluation reveal conditions for success?	10	1.0
8.3.a.1.	Has it been (or is it being) measured to what extent the objectives of the interventions have been (or are being) achieved?	10	1.0
8.3.a.2.	Are the methods used valid and reliable? This concerns matters like the choice of outcome measures, measuring methods, measuring instruments and the size and representativeness of the group being studied.	10	1.0
8.3.b.1.	Is it clear which alternative explanations for the changes observed can be excluded?	10	1.0
8.3.b.2.	Is the level of assertiveness of the conclusions being drawn justified by the level of certainty offered by the study design? Conclusions have to be toned down to the extent that alternative explanations cannot be excluded.	10	1.0
8.4.1.	Have the stakeholders been provided with relevant feedback from the evaluation?	10	1.0
8.4.2.	Is the mode of presentation of the findings suitable for the stakeholders (in terms of readability or conciseness)?	10	0.80

APPENDIX F

Effect Sizes of Mental Health Promotion and Prevention Programs Involved in the Study

PROGRAM (1) MH promotion through the theatre

SCALES	TIME POINT	N	M	SD	Alpha	d
Coopersmith's self-esteem questionnaire (Lacković-Grgin et al., 2002)	1st	13	16	4.79	.76	0.08
	2nd	8	16.37	4.98		
Social skills Questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Cooperation	1st	13	14.38	3.80	.89	0.27
	2nd	8	15.42	3.82		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Assertion	1st	13	12.50	3.30	.89	0.11
	2nd	8	12.87	3.44		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Self-Control	1st	13	12.09	3.51	.89	0.62
	2nd	8	16.08	9.50		
AVERAGE EFFECT SIZE:						**0.27**

PROGRAM (4) Media literacy

SCALES	TIME POINT	N	M	SD	Alpha	d
Scale of reactions on cyber bullying (Unicef, Cro)	1st	137	3.89	0.54	.61	0.79
	2nd	136	4.30	0.49		
Scale of attitudes towards cyber bullying (Unicef, Cro)	1st	136	3.78	0.66	.58	0.10
	2nd	136	3.85	0.70		
AVERAGE EFFECT SIZE:						**0.44**

PROGRAM (5) Training for the group leaders

SCALE	TIME POINT	N	M	SD	Alpha	d
Test of knowledge on group work (Mihić, 2011)	1st	9	1.78	0.83	.76	1.32
	2nd	9	2.66	0.44		
EFFECT SIZE:						**1.32**

PROGRAM (6) Substance abuse prevention for parents

SCALES	TIME POINT	N	M	SD	Alpha	d
Knowledge on drug abuse symptoms (Mihić, 2011)	1st	8	3.56	2.83	.85	0.75
	2nd	8	5.75	2.98		
Knowledge on services for drug abuse prevention and treatment (Mihić, 2011)	1st	8	2.12	1.12	.80	0.31
	2nd	8	2.62	1.99		
AVERAGE EFFECT SIZE:						**0.53**

PROGRAM (7) Substance abuse prevention for teachers

SCALES	TIME POINT	N	M	SD	Alpha	d
Knowledge on drug abuse symptoms (Mihić, 2011)	1st	42	3.08	1.26	.60	0.86
	2nd	36	4.50	2		
Knowledge on services for drug abuse prevention and treatment (Mihić, 2011)	1st	42	2.33	1.10	.55	-0.52
	2nd	36	1.80	0.89		
AVERAGE EFFECT SIZE:						**0.34**

PROGRAM (8) Parenting program II

SCALES	TIME POINT	N	M	SD	Alpha	d
Attachment to a partner questionnaire (Kamenov and Jelić, 2003) - ANXIETY	1st	21	27.04	7.38	.78	0.71
	2nd	21	21.74	7.60		
Attachment to a partner questionnaire (Kamenov and Jelić, 2003) - AVOIDANCE	1st	21	21.11	9.52	.78	0.33
	2nd	21	18.43	6.32		
The parental stress scale (Berry and Jones, 1995)	1st	21	40.10	8.46	.67	0.69
	2nd	21	34.72	7.14		
Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	1st	21	65.66	12.39	.75	0.69
	2nd	21	73.02	8.43		
Parental behaviour questionnaire (Keresteš, 1999) - Acceptance of children	1st	21	3.59	0.46	.86	0.55
	2nd	21	3.80	0.28		
Parental behaviour questionnaire (Keresteš, 1999)- Autonomy	1st	21	3.64	0.34	.86	0.48
	2nd	21	3.80	0.32		
Parental behaviour questionnaire (Keresteš, 1999) - Psychological control	1st	21	2.37	0.70	.86	0.45
	2nd	21	2.07	0.62		
Parental behaviour questionnaire (Keresteš, 1999) - Supervision	1st	21	3.18	0.75	.86	0.59
	2nd	21	3.56	0.51		
Parental behaviour questionnaire (Keresteš, 1999) - Positive discipline	1st	21	3.28	0.56	.86	0.12
	2nd	21	3.35	0.56		
Parental behaviour questionnaire (Keresteš, 1999)- Negative discipline	1st	21	2.26	0.69	.86	0.83
	2nd	21	1.80	0.36		
Parental behaviour questionnaire (Keresteš, 1999) - Compliance	1st	21	2.39	0.65	.86	0.26
	2nd	21	2.24	0.51		
AVERAGE EFFECT SIZE:						**0.52**

PROGRAM (9) Parenting program III

SCALES	TIME POINT	N	M	SD	Alpha	d
Scale of parental behaviours (Pećnik, 2010)	1st	13	3.13	0.38	.51	0.65
	2nd	9	3.36	0.31		
EFFECT SIZE:						**0.65**

PROGRAM (10) Parenting program IV

SCALES	TIME POINT	N	M	SD	Alpha	d
A clinical scale for the self-assessment of irritability (Snaith et al., 1978) - Depressive symptoms	1st	51	3.96	2.31	.89	0.39
	2nd	36	3.11	1.89		
A clinical scale for the self-assessment of irritability (Snaith et al., 1978) - Anxiety symptoms	1st	51	5.82	2.94	.89	0.50
	2nd	36	4.39	2.71		
A clinical scale for the self-assessment of irritability (Snaith et al., 1978) - Irritability towards others	1st	51	3.98	2.08	.89	0.33
	2nd	36	3.30	2.02		
A clinical scale for the self-assessment of irritability (Snaith et al., 1978) - Irritability towards itself	1st	51	1.78	1.75	.89	0.26
	2nd	36	1.36	1.31		
The parental stress scale (Berry and Jones, 1995)	1st	52	39.79	9.80	.87	0.49
	2nd	36	35.19	8.53		
AVERAGEAVERAGE EFFECT SIZE:						**0.39**

PROGRAM (11) Substance abuse prevention in the community

SCALES	TIME POINT	N	M	SD	Alpha	d
Scale of perceived harm associated with drug use (Mihić, 2011	1st	8	3.31	0.39	.90	0.22
	2nd	7	3.39	0.32		
EFFECT SIZE:						**0.22**

PROGRAM (12) Creative free time program I.

SCALES	TIME POINT	N	M	SD	Alpha	d
Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	1st	10	73.60	5.17	.40	0.32
	2nd	10	75.10	4.17		
Parental behaviour questionnaire (Keresteš, 1999) - Acceptance of children	1st	10	3.87	0.24	.63	0.19
	2nd	10	3.91	0.17		
AVERAGE EFFECT SIZE:						**0.25**

PROGRAM (13) Free time for children in foster care

SCALES	TIME POINT	N	M	SD	Alpha	d
Coopersmith's self-esteem questionnaire (Lacković-Grgin et al., 2002)	1st	8	14.75	5.36	.87	0.65
	2nd	8	18	5.13		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Cooperation	1st	8	12.50	4	.55	0.35
	2nd	8	13.87	3.83		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Assertion	1st	8	10.50	2.45	.55	- 0.18
	2nd	8	10.12	1.81		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) -Empathy	1st	8	13.51	2	.55	-0.26
	2nd	8	12.75	3.53		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) – Self-Control	1st	8	12.75	3.53	.55	-0.37
	2nd	8	11.47	3.48		
AVERAGE EFFECT SIZE:						**0.04**

PROGRAM (14) Parenting program V

SCALES	TIME POINT	N	M	SD	Alpha	d
The parental stress scale (Berry and Jones, 1995)	1st	8	36.12	5.77	.91	0.78
	2nd	8	29.37	10.76		
Parental behaviour questionnaire (Keresteš, 1999) - Acceptance of children	1st	8	3.59	0.19	.59	1.72
	2nd	8	3.87	0.13		
Parental behaviour questionnaire (Keresteš, 1999) - Autonomy	1st	8	3.72	0.34	.59	0.63
	2nd	8	3.91	0.26		
Parental behaviour questionnaire (Keresteš, 1999) - Psychological control	1st	8	2.16	0.56	.59	0.20
	2nd	8	2.03	0.72		
Parental behaviour questionnaire (Keresteš, 1999) - Supervision	1st	8	3.44	0.26	.59	0.75
	2nd	8	3.65	0.30		
Parental behaviour questionnaire (Keresteš, 1999) - Positive discipline	1st	8	3.17	0.34	.59	0.72
	2nd	8	3.40	0.30		
Parental behaviour questionnaire (Keresteš, 1999) - Negative discipline	1st	8	2.20	0.51	.59	0.45
	2nd	8	1.97	0.52		
Parental behaviour questionnaire (Keresteš, 1999) - Compliance	1st	8	2.67	0.47	.59	0.51
	2nd	8	2.29	0.93		
Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	1st	8	63.25	5.75	.79	0.88
	2nd	8	70.37	9.90		
Scale of family members' communication quality (adapted according to Brajša, Žižak, Mejovšek, Bašić, 1990)	1st	8	72	7.05	.88	0.83
	2nd	8	78.5	8.52		
AVERAGE EFFECT SIZE:						**0.75**

PROGRAM (15) Peer-violence prevention program

SCALE	TIME POINT	N	M	SD	Alpha	d
Scale of reactions during a conflict situation (Mihić, 2011)	1st	55	9.20	1.98	.58	0.11
	2nd	44	9.41	1.78		
EFFECT SIZE:						**0.11**

PROGRAM (16) Self-confidence training

SCALES	TIME POINT	N	M	SD	Alpha	d
Coopersmith's self-esteem queistionnaire (Lacković-Grgin, 2002)	1st	31	18.39	4.47	.84	-0.09
	2nd	33	17.97	5.15		
Social skills questionnaire (Ferić Šlehan, Kranželić, 2005)	1st	31	4.16	0.44	.94	0.73
	2nd	33	4.51	0.51		
AVERAGE EFFECT SIZE:						**0.64**

PROGRAM (17) Substance abuse prevention

SCALES	TIME POINT	N	M	SD	Alpha	d
Scale of perceived harm associated with drug use (Mihić, 2011)	1st	155	3.34	0.34	.67	0.42
	2nd	155	3.48	0.33		
Scale of attitudes towards drug use (Mihić, 2011)	1st	155	1.77	0.41	.65	0.23
	2nd	155	1.67	0.47		
AVERAGE EFFECT SIZE:						**0.32**

PROGRAM (18) Parenting program VI

SCALES	TIME POINT	N	M	SD	Alpha	d
The parental stress scale (Berry and Jones, 1995)	1st	8	36.12	5.77	.62	0.78
	2nd	8	29.37	10.76		
Parental behaviour questionnaire (Keresteš, 1999) - Acceptance of children	1st	8	3.59	0.19	.59	1.72
	2nd	8	3.87	0.13		
Parental behaviour questionnaire (Keresteš, 1999) - Autonomy	1st	8	3.72	0.34	.59	0.63
	2nd	8	3.91	0.26		
Parental behaviour questionnaire (Keresteš, 1999) - Psychological control	1st	8	2.16	0.56	.59	0.20
	2nd	8	2.03	0.72		
Parental behaviour questionnaire (Keresteš, 1999) - Supervision	1st	8	3.44	0.26	.59	0.78
	2nd	8	3.66	0.30		
Parental behaviour questionnaire (Keresteš, 1999) - Positive discipline	1st	8	3.17	0.34	.59	0.72
	2nd	8	3.40	0.30		
Parental behaviour questionnaire (Keresteš, 1999) - Negative discipline	1st	8	2.20	0.51	.59	0.45
	2nd	8	1.97	0.51		
Parental behaviour questionnaire (Keresteš, 1999) - Compliance	1st	8	2.67	0.47	.59	0.51
	2nd	8	2.29	0.93		
Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	1st	8	63.25	5.75	.79	0.88
	2nd	8	70.37	9.90		
Scale of family members' communication quality (adapted according to Brajša , Žižak, Mejovšek, Bašić, 1990)	1st	8	72	7.05	.81	0.83
	2nd	8	78.50	8.52		
AVERAGE EFFECT SIZE:						**0.75**

PROGRAM (19) Underage drinking prevention

SCALES	TIME POINT	N	M	SD	Alpha	d
Coopersmith's self-esteem questionnaire (Lacković-Grgin et al., 2002)	1st	29	20.62	2.69	.58	-0.07
	2nd	20	20.20	7.24		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) – Cooperation	1st	28	11.07	4.22	.94	0.61
	2nd	20	13.60	3.90		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Assertion	1st	28	11.62	3.75	.94	0.34
	2nd	20	12.88	3.65		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) -Empathy	1st	28	12.15	5.10	.94	0.49
	2nd	20	15.18	7.39		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) – Self-Control	1st	28	10.47	4.56	.94	0.44
	2nd	20	12.45	4.36		
Scale of attitudes about problem solving strategies (Mihić, 2011)	1st	28	3.36	0.56	.90	-0.32
	2nd	20	3.20	0.38		
Scale of attitudes towards alcohol consumption (Mihić, 2011)	1st	28	2.22	0.48	.84	0.06
	2nd	20	2.19	0.56		
Scale of reactions to alcohol offers (Mihić, 2011)	1st	28	2.19	0.50	.81	0.31
	2nd	20	2.35	0.53		
Scale of perception of advertisements (Mihić, 2011)	1st	28	3.66	0.70	.80	-0.42
	2nd	20	3.35	0.79		
AVERAGE EFFECT SIZE:						**0.16**

PROGRAM (20) MH promotion through volunteerism

SCALES	TIME POINT	N	M	SD	Alpha	d
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Cooperation	1st	26	13.52	2.79	.81	0.92
	2nd	23	16.11	2.81		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Assertion	1st	26	11.61	2.51	.81	0.39
	2nd	23	12.79	3.46		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Empathy	1st	26	14.12	2.50	.81	0.53
	2nd	23	15.60	3.07		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) – Self-Control	1st	26	11.72	3.03	.81	0.95
	2nd	23	14.53	2.84		
Scale of altruism (Raboteg-Šarić, 2002)	1st	26	52.80	8.67	.77	0.13
	2nd	23	54.13	11.85		
AVERAGE EFFECT SIZE:						**0.58**

PROGRAM (21) MH promotion through dance

SCALES	TIME POINT	N	M	SD	Alpha	d
Coopersmith's self-esteem questionnaire (Lacković-Grgin et al., 2002)	1st	118	18.72	4.17	.70	0.01
	2nd	116	18.76	5.37		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Cooperation	1st	118	14.89	3.15	.86	0.03
	2nd	116	14.98	3.06		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) - Assertion	1st	118	12.05	3.42	.86	0.14
	2nd	116	12.52	3.25		
Social skills questionnaire - grades 7-12 (Gresham and Elliott, 1990) – Self-Control	1st	118	12.46	4.23	.86	0.21
	2nd	116	13.29	3.48		
Scale of social and emotional loneliness (Ćubela Adorić, 2004)	1st	117	29.67	12.32	.50	0.12
	2nd	112	28.14	12.68		
AVERAGE EFFECT SIZE:						**0.10**

PROGRAM (22) Creative free time program II

SCALES	TIME POINT	N	M	SD	Alpha	d
Scale of attitudes towards leisure time activities (Mihić, 2011)	1st	31	4.04	0.48	.68	0.15
	2nd	28	4.12	0.60		
Scale of leisure time activities (Mihić, 2011)	1st	31	3.39	0.52	.65	-0.05
	2nd	28	3.36	0.58		
AVERAGE EFFECT SIZE:						**0.05**

PROGRAM (23) Parenting program VII

SCALES	TIME POINT	N	M	SD	Alpha	d
The parental stress scale (Berry and Jones, 1995)	1st	12	35.25	4.39	.73	0.90
	2nd	12	31.67	3.55		
Parental behaviour questionnaire (Keresteš, 1999) - Acceptance of children	1st	12	3.79	0.26	.81	-0.38
	2nd	12	3.69	0.26		
Parental behaviour questionnaire (Keresteš, 1999) - Autonomy	1st	12	3.42	0.43	.81	0.59
	2nd	12	3.65	0.35		
Parental behaviour questionnaire (Keresteš, 1999) - Psychological control	1st	12	2.65	0.49	.81	0.81
	2nd	12	2.23	0.54		
Parental behaviour questionnaire (Keresteš, 1999) - Supervision	1st	12	2.99	0.35	.81	0.33
	2nd	12	3.12	0.44		
Parental behaviour questionnaire (Keresteš, 1999) - Positive discipline	1st	12	3.35	0.44	.81	-0.42
	2nd	12	3.17	0.43		
Parental behaviour questionnaire (Keresteš, 1999) - Negative discipline	1st	12	2.08	0.54	.81	0.70
	2nd	12	1.75	0.39		
Parental behaviour questionnaire (Keresteš, 1999) - Compliance	1st	12	2.83	0.41	.81	0.32
	2nd	12	2.67	0.57		
Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	1st	12	65.58	7.63	.79	0.76
	2nd	12	71.32	7.53		
Scale of family members' communication quality (adapted according to Brajša, Žižak, Mejovšek, Bašić, 1990)	1st	12	72.71	7.88	.67	0.32
	2nd	12	75.22	7.58		
AVERAGE EFFECT SIZE:						**0.39**

PROGRAM (24) Parenting program VIII

SCALES	TIME POINT	N	M	SD	Alpha	d
Parental behaviour questionnaire (Keresteš, 1999) - Acceptance of children	1st	12	3.50	0.33	.77	-0.10
	2nd	8	3.46	0.42		
Parental behaviour questionnaire (Keresteš, 1999) - Autonomy	1st	12	3.33	0.39	.77	0.51
	2nd	8	3.54	0.42		
Parental behaviour questionnaire (Keresteš, 1999) - Psychological control	1st	12	2.67	0.39	.77	0.41
	2nd	8	2.44	0.65		
Parental behaviour questionnaire (Keresteš, 1999) - Supervision	1st	12	3.19	0.49	.77	0.10
	2nd	8	3.24	0.49		
Parental behaviour questionnaire (Keresteš, 1999) - Positive discipline	1st	12	2.95	0.35	.77	0.72
	2nd	8	3.27	0.49		
Parental behaviour questionnaire (Keresteš, 1999) - Negative discipline	1st	12	2.27	0.57	.77	-0.11
	2nd	8	2.34	0.68		
Parental behaviour questionnaire (Keresteš, 1999) - Compliance	1st	12	2.58	0.50	.77	0.28
	2nd	8	2.42	0.60		
Scale of parents' perception of their competence for parental role (Gustović-Ercegovac, 1992)	1st	12	62.37	5.75	.75	0.03
	2nd	8	62.60	9.38		
Attachment to a partner questionnaire (Kamenov and Jelić, 2003) - ANXIETY	1st	12	25.24	10.71	.77	0.38
	2nd	8	21.65	8.33		
Attachment to a partner questionnaire (Kamenov and Jelić, 2003) - AVOIDANCE	1st	12	19.25	7.65	.77	0.28
	2nd	8	17.17	7.27		
AVERAGE EFFECT SIZE:						**0.25**

APPENDIX G

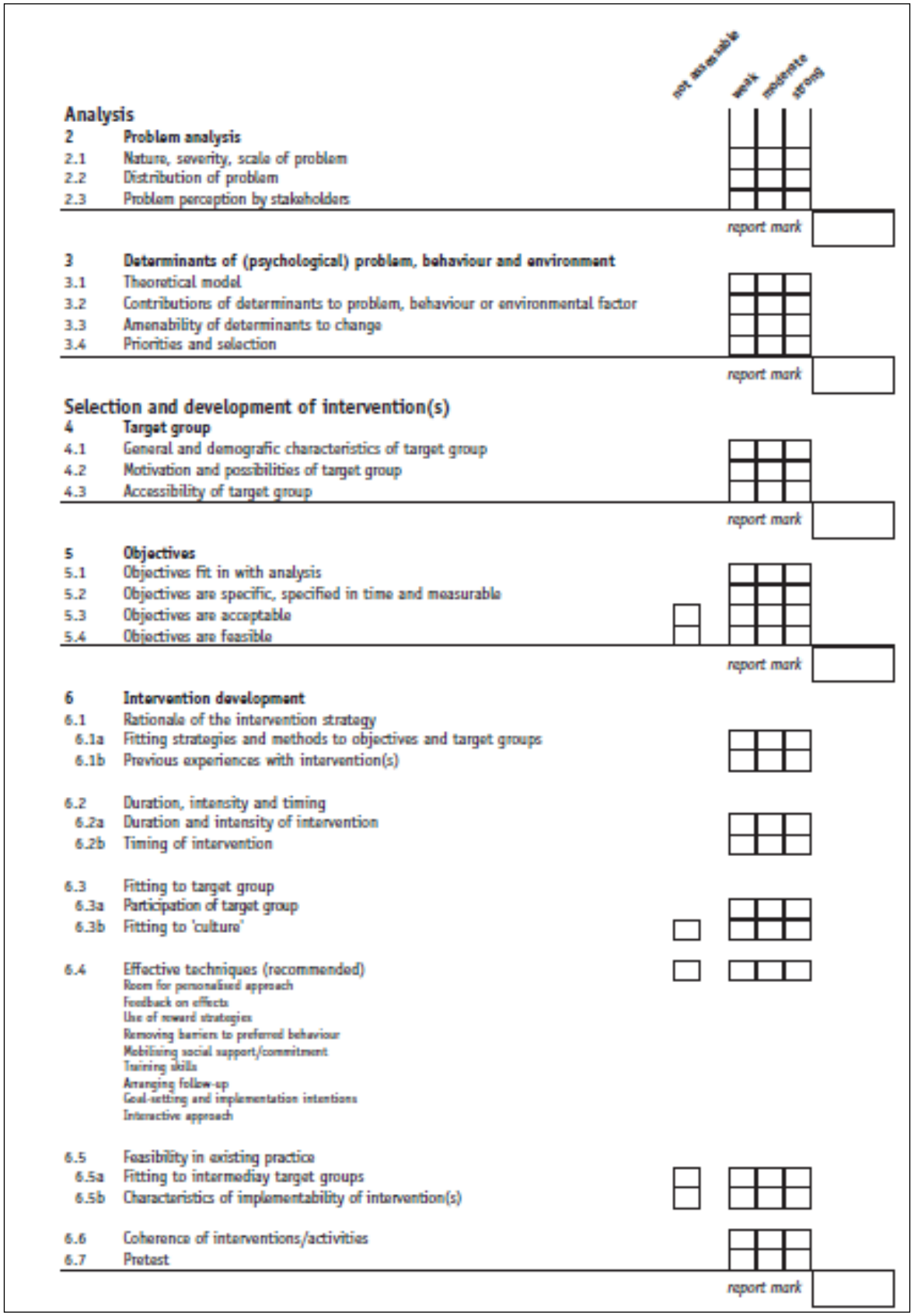
Results of the Kolmogorov-Smirnov Test of the Normality

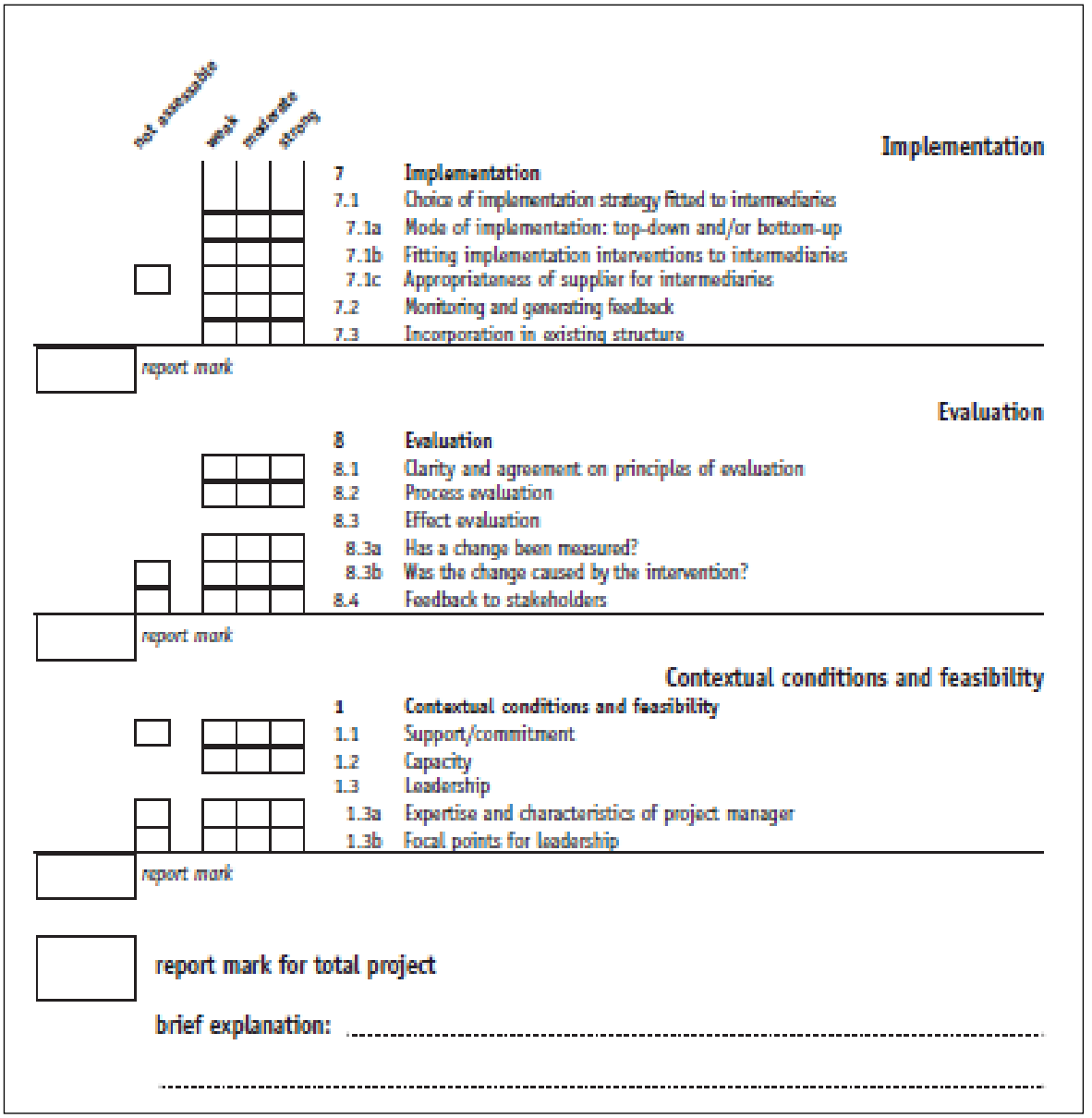
of Preffi 2.0 Scores' Distribution Assessed in Two Measurements

		BASELINE	POST-TEST
1ST PREFFI CLUSTER "Contextual conditions and feasibility"	K-S TEST	.770	.862
	p	.549	.447
	N	24	21
2ND PREFFI CLUSTER "Problem analysis"	K-S TEST	1.048	.870
	p	.222	.435
	N	24	21
3RD PREFFI CLUSTER "Determinants of behaviour and environment"	K-S TEST	.886	.767
	p	.412	.599
	N	24	21
4TH PREFFI CLUSTER "Target group"	K-S TEST	.969	.791
	p	.305	.558
	N	24	21
5TH PREFFI CLUSTER "Objectives"	K-S TEST	1.031	1.096
	p	.239	.181
	N	24	21
6TH PREFFI CLUSTER "Intervention development"	K-S TEST	.491	.835
	p	.969	.488
	N	24	21
7TH PREFFI CLUSTER "Implementation"	K-S TEST	.795	.748
	p	.553	.630
	N	24	21
8TH PREFFI CLUSTER "Evaluation"	K-S TEST	1.097	1.058
	p	.180	.213
	N	24	21
TOTAL PREFFI RESULT	K-S TEST	.446	.617
	p	.989	.841
	N	24	21

APPENDIX H

Overall Scheme of Preffi 2.0 (Molleman, 2005)




not assessable
weak
moderate
strong
Implementation
7 Implementation
7.1 Choice of implementation strategy fitted to intermediaries
7.1a Mode of implementation: top-down and/or bottom-up
7.1b Fitting implementation interventions to intermediaries
7.1c Appropriateness of supplier for intermediaries
7.2 Monitoring and generating feedback
7.3 Incorporation in existing structure
report mark
Evaluation
8 Evaluation
8.1 Clarity and agreement on principles of evaluation
8.2 Process evaluation
8.3 Effect evaluation
8.3a Has a change been measured?
8.3b Was the change caused by the intervention?
8.4 Feedback to stakeholders
report mark
Contextual conditions and feasibility
1 Contextual conditions and feasibility
1.1 Support/commitment
1.2 Capacity
1.3 Leadership
1.3a Expertise and characteristics of project manager
1.3b Focal points for leadership
report mark
report mark for total project
brief explanation:

CURICULUM VITAE OF THE DISSERTATION AUTHOR

Josipa Mihić was born on 10 March 1983 in Zagreb, where she finished her elementary and secondary education. In 2007 she graduated from the Department of Behavioural Disorders at the University of Zagreb, Faculty of Education and Rehabilitation Sciences with a degree in Social Pedagogy. The title of her thesis was "Characteristics of violent behaviour towards caregivers in institutions for children and youths with behavioural disorders." During her studies she received the award for the best student in academic year 2005/2006 and the Rector's Award of the University of Zagreb for the successful organization of the II European Congress of Students of Social Pedagogy, held in Zagreb.

Since 2006 she has worked as an external associate of the Centre for social welfare Peščenica and as head of the project activities in the NGO "Parents for Children". Since February 2008 she has been employed as a research assistant at the Department of Social Pedagogy at the University of Zagreb, Faculty of Education and Rehabilitation Sciences on the project „Communities That Care“, headed by professor Josipa Bašić. In the same year she enrolled in the postgraduate doctoral program “Prevention Science and Disability Study, study program - Prevention of Mental and Behaviour Disorders and Mental Health Promotion” at the Faculty of Education and Rehabilitation Sciences. She published six scientific papers and she was co-author of several book chapters. She actively participated at more than fifteen domestic and international conferences and she participated in the design and implementation of three international research projects and two local ones. She is continuously involved in the teaching of courses Theories of prevention II, Research in prevention, Prevention of externalizing and internalizing behavioural problems, Prevention media campaigns and Volunteer work. She co-mentored four graduate theses in the field of prevention of behavioural problems. She has participated in various training programs including the training for the implementation of meta-analysis at the University of Nijmegen, Netherlands. She has participated in three international volunteer camps in Croatia and Italy.

She is a member of American and European Societies for Prevention Research, European Network for Social and Emotional Competence and Croatian Society for Social Pedagogy.

Published papers and book chapters:

Bašić, J., Grozić-Živolić, S. (Eds.) (2011). *Communities That Care, Prevention Model of Behaviour Disorders in Children and Adolescents: Development, Implementation and Evaluation of Prevention in the Community*. Pula and Zagreb: University of Zagreb, Faculty of Education and Rehabilitation Sciences and Istria County; co-author of chapters (4.5., 5.4., 6.5. and 6.6.)

Bašić, J., Mihić, J., Novak, M. (2011). Risk Analysis in the Period of Growing-up of Children and Youth: Starting Point for Effective Prevention. *Journal of Public Health, 19, (1),* 3-11.

Mihić, J., Novak, M., Bašić, J. (2010). Communities That Care: CTC Needs Assessment Questionnaire for Planning Preventive Interventions in Children and Young people. *Annual of Social Work 17 (3),* 391 – 412.

Mihić, J., Bašić, J. (2008). Prevention Strategies - Externalizing Behavioural Problems and Risky Behaviour Among Children. *Annual of Social Work, 15, (3),* 445-471.

Mihić, J. (2008). Prevention of Underage Drinking - Policy Formulation at the Local Level from the Experience of Istria. *Criminology and Social Integration, 16, (2),* 61 - 72.

Ferić Šlehan, M., Mihić, J., Ricijaš, N. (2008). Youths' Risk Behaviour and Their Perception of Anticipated Parental Consequences for Risk Behaviour. *Criminology and Social Integration, 16, (2),* 47-59.