The Role of Implementation: Happy Class in The Hague

Name	Anna Maria Verhagen
Student number	3349861
Date of application	23 June, 2014
Faculty	Faculty of Social Sciences
Master	Education, Socialization and Youth Policies (MOV), University of Utrecht
Thesis Supervisor	Rianne van der Zanden, Trimbos-institute
Thesis Professor	Monique van Londen-Barentsen, University of Utrecht

Abstract

Objective. Happy Class is an online universal depression prevention given in class for students with VMBO education school track in The Hague. This study aimed to understand whether differences in implementation aspects between two Happy Class implementations (HC1 and HC2) influenced the quality of implementation (outcome scores).

Method. A pre-test post-test design was used with a sample of 740 Dutch students from four schools in The Hague. HC1 had programme developer involvement, HC2 did not. Online questionnaires which are part of the Happy Class programme were analysed, which allowed us to collect complete data from the real world setting. Responsiveness could only be measured as satisfaction.

Results. HC1 students were more satisfied and its sub-clinical depression group decreased significantly in depression. In HC2, the clinical group increased significantly in well-being as the only group in both HC1 and HC2. Implementation differed in gender-ratio, schools, programme staff and dosage, and these factors predicted well-being, depression and satisfaction scores.

Conclusion. Factors impacted the implementation quality, but should be further explored. We recommend planning implementation-evaluation beforehand, a randomised controlled trial to assess effectiveness, clarification of the mechanisms of change within Happy Class and its target group, and exploring the role of responsiveness in Happy Class.

Keywords: Happyles, e-mental health, depression, prevention programme, implementation evaluation, responsiveness, satisfaction, positive psychology

In 2012-2013, the Trimbos-institute worked together with the city of The Hague and the Dutch youth health care organization (Jeugd-GGZ; De Jutters) to implement the online prevention programme Happyles at three schools in The Hague. Happyles, from now on referred to as Happy Class, is a universal stepped care online prevention programme given in class at school for VMBO students with slight to mild depression symptoms (sub-clinical depression symptoms). Happy Class reduces sub-clinical depression symptoms by enforcing well-being with exercises based on positive psychology, but it addresses problem-solving and cognitive behavioural techniques as well. The Trimbos-institute's involvement ended in 2013, yet The Hague and the youth health care organization continued the programme in 2013-2014 with the same number of student participants (n = 380).

In this study, the main question is whether Happy Class results remain the same or improve with the next implementation and why this happens. What makes this implementation study unique is that with the integrated online questionnaires within Happy Class, we collected complete data from both the implementation period where researchers were involved as well as the period where they had no involvement. With this data, two implementation periods are compared to explore how implementation differences relate to outcomes in depression and well-being. We study the role of student responsiveness as well, since it is a part of the implementation process and may influence outcomes. The Happy Class programme will benefit from this study in the form of new insights about the role of its implementation and responsiveness. These insights can reveal barriers and facilitators during implementation and stimulate implementation improvement of the Happy Class programme.

Depression in Youth

Depression is a mood disorder defined by persistent feelings of sadness, guilt, irritable mood, worthlessness and a loss of pleasure in activities; it is the experience of a pervasive unhappy mood (Gray, 2007; Wicks-Nelson & Israel, 2009). According to a study by Wijga, Scholtens, Van Oeffelen and Beckers (2010), depression can accompany other disorders and diseases as well (comorbidity), indicating that one increases the risk of the other being present or following. They related depression to chronic fatigue, chronic pain or obesity, and claimed symptoms at childhood are associated with the presence of anxiety and other mood disorders years later. The World Health Organization (WHO) studied 245 404 participants who were eighteen years old, from 60 countries in all regions of the world (Moussavi et al., 2007). By comparing mean scores

for health progression, they concluded depression to cause the greatest impairment in health, compared with chronic diseases such as angina, arthritis, asthma and diabetes. Combined with one or more chronic diseases, depression had the worst health scores of all the disease states.

Aside from clinical depression, people can also suffer sub-clinical depression; depression which does not match the DSM-criteria for a depression disorder yet still has consequences for those involved (Cuijpers & Smit, 2008b). Cuijpers and Smit revealed even sub-clinical depression symptoms impair social life and mental well-being. With data from the Netherlands Mental Health Survey and Incidence Study (Dutch: NEMESIS-onderzoek), they calculated the economical consequences in the Netherlands as well. This revealed expenses such as a total cost of 1.5 billion euro each year for sub-clinical depression, and 1.8 billion euro each year for clinical depression. These costs even grew with the number of depression symptoms. The authors argue a sub-clinical depression can develop into a clinical depression if it is left untreated, marking sub-clinical depression as an equally important problem.

High prevalence and incidence rates of depression form another area of concern (Waraich, Goldner, Somers & Hsu, 2004). Depression can be experienced by both youth and adults, with a higher prevalence for females; sometimes even occurring thrice as often for females than males between the age of 13 to 17 (De Hollander, Hoeymans, Melse, Van Oers & Polder, 2006). A study by the WHO (Moussavi et al., 2007) revealed that worldwide, the 1-year prevalence for ICD-10 (International Classification of Diseases tenth revision) depressive episode was 3.2% (N = 245 404, 95% CI 3.0-3.5). As for the Netherlands, comparable rates for depression were found for youth aged between 13-18 (3.8%) and a much higher rate for those with sub-clinical depression (21.4%; Verhulst, Van der Ende, Ferdinand & Kasius, 1997; Smit, Bohlmeijer & Cuijpers, 2003). A more recent Dutch study among 18-64 year olds by De Graaf, Ten Have and Van Dorsselaer (NEMESIS-2, 2010), revealed that of the 18-24 year old participants (n = 480), 13% had experienced depression disorder at a moment in their lives and 6.7% had experienced depression disorder in the past year. Age in general seems to be related to depression, as the start and severity of depression is at a maximum between the age of 13 to 18 (Lewinsohn, Clarke, Seeley & Rohde, 1994; Masi, Favilla, Mucci & Millepiedi, 2000; Kim-Cohen et al., 2003). Between the age of 12 to 18, stressful experiences and sub-clinical depression symptoms are considered to be two of the most important risk factors (Thapar, Collishaw, Pine & Thapar, 2012).

Online Prevention for Youth

A prevention programme in a school setting seems to be a viable and effective option for both youth's emotional and behavioural problems when interventions are properly integrated into the school environment (Greenberg et al., 2003). With this kind of prevention, risk factors during school years can be compensated for by protective factors to create a balance and resilience against difficulties in life (Zimmerman & Arunkumar, 1994).

According to a study by Cuijpers, Van Straten, Smit, Mihalopoulos and Beekman (2008c), preventive interventions are at least capable of significantly reducing the incidence rate of depressive disorders by 22% compared with treatment-as-usual control groups, independent of the target group, and in a more recent study this increased to reducing the incidence rate by 38% (Cuijpers, Muñoz, Clarke & Lewinsohn, 2009). However, there are various forms of prevention; universal (everyone), selective (group at risk) and indicative (group with depressive symptoms; Kroes et al., 2007; Muñoz, Cuijpers, Smit, Barrera & Leykin, 2010). Universal programmes seem to have a lower impact than selective or indicated programmes (Cuijpers et al., 2008c; Horowitz & Garber, 2006). A study by Van Oorsprong (2011) explains sub-groups in universal preventions can impact overall effectiveness; she found Happy Class to be effective for the sub-clinical depression group, but not for the normal or clinical group.

Muñoz and colleagues (2010) discuss several options for prevention in their review study, namely cognitive behavioural preventions, interpersonal psychotherapy, problem-solving therapy and prevention via the internet (online prevention). The first two of these have some evidence for their effectiveness in reducing the incidence rate, but less is known about online depression prevention for youth (Riper, Van Ballegooijen, Kooistra, De Wit & Donker, 2013). For example, Riper and colleagues (2013) mention in their review study how little research is actually available yet about specific online depression prevention for youth, and how the little which is available often shows insignificant results or methodological problems. They listed the online preventions which could be found and were studied for effectiveness: MoodGym, EverythingUnderControl (Dutch: Allesondercontrole), MasterYourMood and Catch-It. Of these four, they only found MasterYourMood to produce significant results compared to the control group.

The online indicated prevention MasterYourMood (MYM) is the parent programme on which Happy Class is based. MYM seemed to mainly reach and work for a group of youth with

higher education (e.g. VWO), which left a significant portion of the target group neglected (Van der Linden & Van der Zanden, 2011, p. 127). Van der Linden and Van der Zanden point out Happy Class forms a useful addition to MYM, since it is a depression prevention given at school and meant for students with a MBO or lower education track. They explain Happy Class improves mental well-being to decrease depression, with the help of exercises based on positive psychology. Positive psychology focuses on people's strengths rather than their problems (Lubyomirsky, 2007). Some existing studies support the idea of well-being forming a protective factor against depression and anxiety (Ruini et al., 2009; Patton et al., 2011). Yet the Happy Class exercises also address cognitive behavioural techniques (e.g. the relationship between thoughts, feelings and actions) and problem solving skills (e.g. different types of problems and what to do with them).

Implementation

The term Implementation is often used when the researchers explore the execution of a programme and its possible determinants (Fleuren, Wiefferink & Paulussen, 2004; Koppelaar, Knibbe, Miedema & Burdorf, 2009; Berkel, Mauricio, Schoenfelder & Sandler, 2011). A study by Salmivalli, Kaukiainen and Voeten (2005) refers to school-programme implementation as "the extent to which the components of the programme are put into practice" (p. 469). Durlak and DuPre (2008) broaden this concept, defining implementation as "what a program consists of when it is delivered in a particular setting" (p. 329). Rossi, Lipsey and Freeman (2004) claim an implementation evaluation can focus on programmes in circumstances where the integrity of programme operations, service delivery and other executional aspects are unclear (p.175).

For the majority of prevention programmes, evaluations of the programme by the developers only occur in an experimental setting and neglect implementation. Durlak (1997) listed 1200 educational, mental and physical health prevention studies published just before 1996 of which less than 5% mentioned programme implementation. In a more recent study, Perepletchikova, Treat and Kazdin (2007), demonstrated that programme integrity was only examined for 3.5% of the 202 evaluated psychosocial interventions. Yet according to a study by Tibbits, Bumbarger, Kyler and Perkins (2010), differences may appear for implementations over time. They measured post-funding sustainability of school prevention programmes for crime and delinquency, which had previously received funding at least a year ago, with the question "To

what extent is your program still operating?". Overall, of 67 interventions assessed after funding had ended (minimum one year, maximum five years), 21% were no longer implemented, 40% had reduced quality and 39% were functioning at the same level or a higher level of quality than before. For the interventions after five years of post-funding (n = 9), 33% of the interventions were inactive, 22% had decreased in quality and 45% remained at the same level of quality or reached a higher level of quality than before. The researchers also found that post-funding sustainability could be predicted when programme staff and school support formed no barriers in its implementation.

Furthermore, Durlak and DuPre (2008) reviewed five meta-analyses with a total of 542 studies related to implementation evaluation, and found that nearly all studies they analysed had outcomes affected by implementation and that the way a programme is implemented is an important determinant in outcomes both statistical and in benefits for participants. They list eight aspects of importance to implementation which influence outcomes: (1) integrity – the extent to which the programme is implemented as intended, (2) dosage – how much of the original program is delivered, (3) quality – how well (clearly and correctly) programme components are conducted, (4) participant responsiveness – the degree to which the program interests or holds the attention of the participants, (5) differentiation – the degree to which programme theory and practices are unique compared to other programmes, (6) monitoring of control groups – the nature and amount of services received, (7) reach – the rate of involvement and representativeness of programme participants, and (8) modification – the changes made to the original programme during implementation. These were assessed by either observational data, self-reports, attendance figures or interviews.

For preventions in specific, research by TNO (Fleuren, Paulussen, Dommelen & Buuren, 2012) provided an instrument called the Measuring Instrument for Determinants of Innovations (MIDI), which enables assessment of four main categories with 29 determinants considered essential for implementation quality. These categories are: programme, user, organization, and social political context. Determinants are measured by one item each, although they can involve multiple sub-items depending on the implementation. The MIDI and the eight aspects of implementation do not contradict each other but add to each other. Where the MIDI goes into detail, the aspects are not measured in one specific way, and where the eight aspects offer a broader view, the MIDI is more limited to a few dimensions and only takes the shape of a self-report of programme staff.
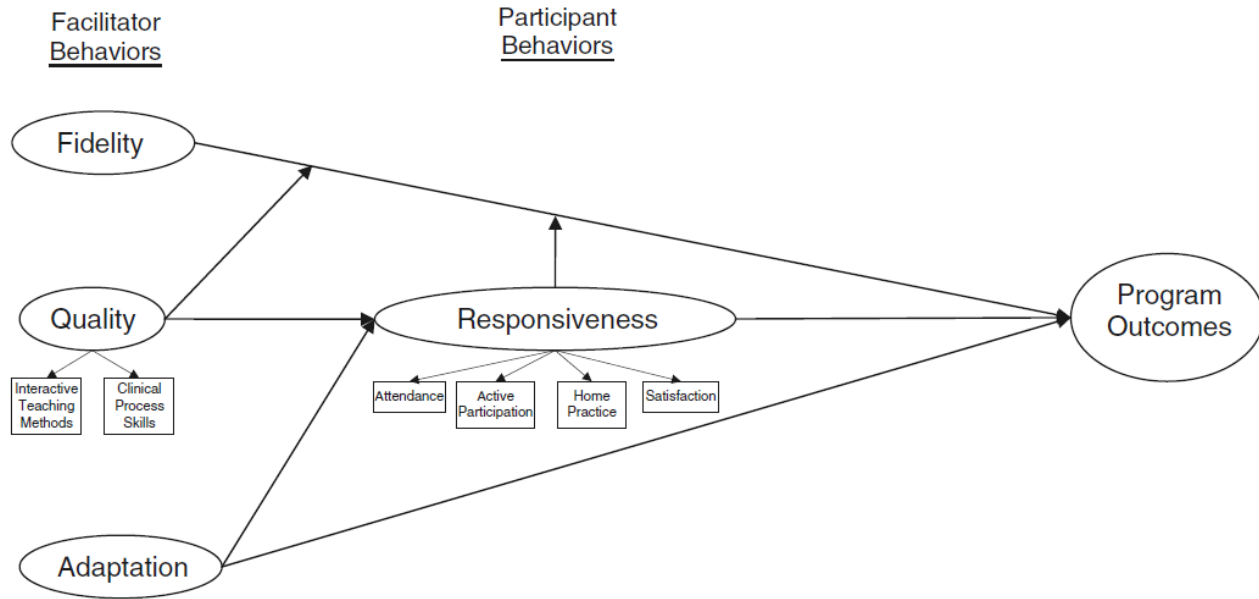
Figure 1. An integrated theoretical model of program implementation (from Berkel et al., 2011, p. 25).

Responsiveness

Berkel and colleagues (2011) argue that determinants of implementation do not function isolated and should be studied in relation to each other. For this, Berkel and colleagues provide a theoretical framework of the relationship between facilitator and participant behaviours which in turn influence programme outcomes positively (figure 1). In this model, they define participant responsiveness as "participants' level of enthusiasm for and participation in an intervention" (p.23). Responsiveness is then operationalised with the following indicators: sessions attended, active participation, satisfaction and home practice completion (attempts at practising assigned skills at home). Fidelity is operationalised as either assessment of which core components (the driving elements of a programme) were delivered or the amount of time dedicated to each of the core components. Quality of delivery is operationalised as interactive teaching methods and clinical process skills (enthusiasm and clarity of delivery, and fostering cohesion in a group). Finally, Adaptation is operationalised as modifications or additional elements added to the programme in order to make it more relevant for a given population.

Adaptation and fidelity share an interesting coexistence here as they appear to be antonyms at first glance, even though both are considered essential for implementation. This makes more sense when considering Berkel and colleagues' operationalisation of adaptation next

to two characteristics of programmes consistently related to implementation: adaptability and compatibility (Durlak & DuPre, 2008, pp. 336-338). These characteristics represent modifications made to the programme to fit the user's needs, current practices and goals. According to this, programme implementation benefits from integrity, but would have a higher chance of effectiveness when it is adjusted to match needs, practices and goals. Furthermore, relationships are expected between adaptation and responsiveness, quality of delivery and responsiveness, fidelity and quality of delivery, and fidelity and responsiveness. It becomes clear that responsiveness plays a large role in this model, as it either moderates or mediates all of the effects from implementation execution on outcomes.
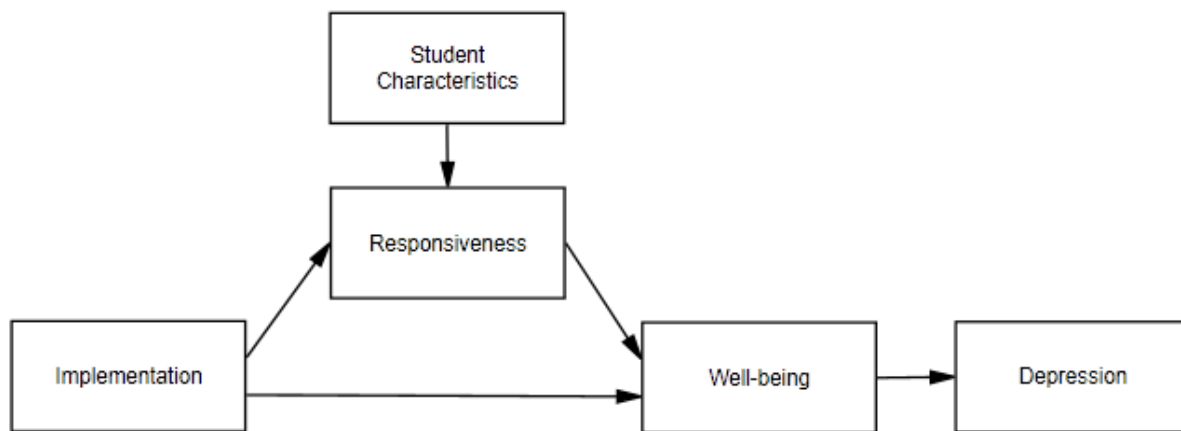
Statement of the Problem

Figure 2. The theoretical model of programme implementation for Happy Class.

The literature review clarifies depression is a problem best addressed before it grows more serious, as adolescence is often the onset-age for depression and sub-clinical depression predicts clinical depression at a later age (Thapar et al., 2012). Although several solutions are provided, not all are effective and limited knowledge exists about the effectiveness of online depression prevention (Riper et al., 2013). So far, only the intervention MYM is known to show effectiveness in studies. Since Happy Class is based on this online preventive intervention and taught as an online class at school, it may provide a useful solution for those youth MYM does not reach. However, implementation of Happy Class may variate in effectiveness, since previous research on implementation processes has found both improvement and decline in quality and many determinants which can influence implementation quality (Durlak & DuPre, 2008). Out of

these determinants, especially responsiveness seems of importance, since it is related to many factors in the implementation process (Berkel et al., 2011).

In this study, we compared two Happy Class implementations to gain insights into the differences in quality of implementation. The 2012-2013 implementation of Happy Class will be referred to as Happy Class 1 (HC1), and the 2013-2014 implementation will be referred to as Happy Class 2 (HC2). Those who implement Happy Class will be referred to as the Happy Class staff or programme staff.

In addition, both Happy Class's own theory of well-being's causal relationship with depression will be tested as well as this study's theory that responsiveness plays a role within the implementation process (see figure 2). We expect that implementation differences and responsiveness will influence well-being and depression, and that in turn student characteristics and programme characteristics will influence responsiveness.

In other words, we aim to discover which factors influence the effectiveness of Happy Class. To do this, the following research questions will be answered: (1) How do HC1 and HC2 differ in programme and student characteristics, student well-being, student depression and student responsiveness? (2) How do the implementation factors which differed (question one) relate to student well-being, depression and responsiveness? (3) What is the relationship between student well-being and student depression? (4) How does responsiveness relate to student well-being and student depression, and is responsiveness influenced by student characteristics?

Method

Procedure and Participants

Two implementation cohorts were compared as independent groups (HC1 & HC2). The first implementation (HC1) was conducted under guidance of the programme developers which were studying Happy Class's effectiveness and execution as it was implemented. The second implementation (HC2) was conducted in real world circumstances where the developers had no involvement and research was announced after the programme. Comparing HC1 with HC2 will reveal whether the programme's effectiveness or execution have changed from one implementation to the other, especially when programme components are adjusted or removed entirely, and whether changes and effectiveness are related.

When the cohorts are taken together, a total of 740 VMBO students (n_{HC1} =373; n_{HC2} = 367) participated in this study during either spring or fall in their mentor classes. Students came from four schools in total, with two of the three schools remaining constant per implementation. participated in the programme when their school had agreed to participate. However, the school informed parents of Happy Class, the data collection and the Trimbos-institute before the start of the program, so parents could object to this and exclude their child if they wished.

Thanks to the online method of data collection Happy Class uses, data from its questionnaires is automatically archived whenever students complete the Happy Class online questionnaires. Because of this, we could use complete data-sets for both HC1 and HC2; a pre-test and post-test questionnaire with questions about students' personal situation, measures for mental well-being and depression and an evaluation questionnaire on their personal situation and thoughts about Happy Class (further information about this data will follow). Only the students, Happy Class staff and Trimbos-institute staff can view this data.

Happy Class Programme

The city of The Hague and the Happy Class staff (youth health care) selected a sample from schools which were interested in and would benefit from the programme according to all parties involved. Each year, the city of The Hague aimed to reach roughly 380 VMBO-students from three schools in The Hague. These students were mainly from the first two grades (age 12-14) and received Happy Class as part of the standard school curriculum. While Happy Class also offers students to participate in a chat-group course after the main intervention, this component

will not be discussed as it comes after the post-test and evaluation. To implement Happy Class at the selected schools, The Hague, the Happy Class staff and the Trimbos-institute developed a programme plan together (figure 3).

In the first week (a week before the start of Happy Class), promotional material is spread throughout the schools such as posters, cards and brochures. In addition, all students attend a theatre show about feeling sad and depressed and an exhibition (held at school) which focuses on reducing stigma around psychological problems.

In week two, all classes have the Happy Class group discussion in their mentor-class, in which they talk about themes such as happiness. This discussion is organised by a (Jeugd-GGZ, Jutters) prevention worker, who can be seen as one of the programme staff. This person answers all the questions students may have, whether practical, content-specific or personal.

In week three, students register for Happy Class behind a personal computer in a classroom and begin the programme with a pre-test. The pre-test also contains questions about their personal lives (e.g. situation at home), a depression-scale (CES-D) and well-being-scale (WEMWBS). Afterwards, they receive a Happiness score (an inverted CES-D score) where a high score represents a low CES-D score. When the pre-test is finished, students start with the first e-learning lesson, in which they watch short movie-clips and complete several positive psychology, cognitive behavioural and problem-solving exercises on the computer (e.g. ways to be happy, relation between feelings, behaviour and thoughts, and various types of problems and problem solving skills). Individually, students need to either type out answers or interact with the images or options on the screen (e.g. dragging pictures, clicking options) to complete the exercises. They receive personal earphones or headphones so they can view the movie-clips by themselves as well.

Week four continues where e-lesson one left off, providing the final exercises the students have to complete. In addition, students can check their personal file and the chill-out during the e-lessons. The personal file presents the student's completed assignments so far and the chill-out offers them a few more tips and tricks to use for when they are feeling sad.

Week five forms the conclusion in which the students complete the Happy Class Test (only CES-D and WEMWBS) once more to check their final Happy Score. After the post-test, students also complete an evaluation form which focuses on their satisfaction with the programme, the enjoyment they got from it and how interesting they thought Happy Class was.
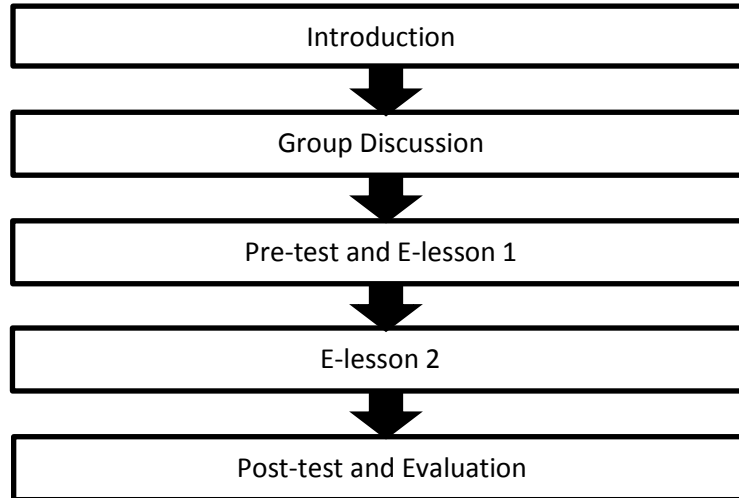
```
┌─────────────────────────────────────────────┐
│                 Introduction                  │
└─────────────────────────────────────────────┘
                      ▼
┌─────────────────────────────────────────────┐
│               Group Discussion                │
└─────────────────────────────────────────────┘
                      ▼
┌─────────────────────────────────────────────┐
│             Pre-test and E-lesson 1           │
└─────────────────────────────────────────────┘
                      ▼
┌─────────────────────────────────────────────┐
│                  E-lesson 2                   │
└─────────────────────────────────────────────┘
                      ▼
┌─────────────────────────────────────────────┐
│            Post-test and Evaluation           │
└─────────────────────────────────────────────┘
```

Figure 3. Happy Class Programme Design for The Hague.

Motive for This Study

The original motive for this study came from interviews and e-mail contact with the head of the programme staff about the two implementations. In table 1, differences found this way are presented: (1) HC1 was implemented with involvement and assistance from the intervention's developers (The Trimbos-institute) who were doing an implementation study on HC1 at the same time. The programme staff was aware of the implementation study being conducted at the time and were observed during the implementation. Developer involvement was entirely removed in HC2, so programme staff only communicated with the Happy Class website providers (a third party). (2) The group discussion component of Happy Class was improved in HC2 with the feedback from the implementation evaluation of HC1 (Van der Zanden & Van der Linden, 2013) and the programme staff's own ideas. (3) The introduction originally contained a theatre show and exhibition, but it was only included in HC1. The programme staff originally had practical issues for a one-time exclusion in HC2, but finally excluded them entirely since the involved schools thought they were redundant components which did not think it fit the students very well. (4) For HC2, new programme staff assisted, who were recruited and trained by the original programme staff who had received training from the Trimbos-institute. (5) For HC1, one school was going through many changes (teachers were fired, reorganisation, classes delayed, mentors were sometimes absent from Happy Class).

Table 1

Implementation characteristics differences according to the head of the programme staff

Difference Domain	Happy Class 1	Happy Class 2
Developer Involvement	Involvement, observation and pre-announced research	None
Group Discussion	Implemented as planned	Improved with HC1-feedback from the programme staff and the Trimbos-institute's study
Introduction	Implemented as planned	Theatre show and exhibition excluded
Programme Staff	Original staff trained by the Trimbos-institute	Original staff and new staff trained by the original staff
School	School A experienced reorganisation;	School A was stable in organisation;

Measures

Student Depression

Student depression is evaluated with the Dutch version of the Center Epidemiologic Studies-Depression (CES-D; Bouma, Ranchor, Sanderman & Van Sonderen, 2012). It contains twenty items with a 4-point answer-scale from 0 = rarely or never to 3 = often or always.

The total scores are computed by the sum of all scores and range from 0 to 60. According to Bouma and colleagues (2012), depression begins at a CES-D score of 16, and clinical depression begins at a score of 22 or higher (Cuijpers, Boluijt, Van Straten, 2008a). Thus this study defines three subgroups of depression: normal/no depression (CES-D 0-16), sub-clinical depression (CES-D 16-22) and clinical depression (CES-D > 21). Sometimes, when numbers in subgroups are too small for analysis, only the entire sample is addressed. The relevant group is always mentioned along with analysis.

The internal consistency of the scale was checked, which proved to be decent (α = .73, n = 740; α = .82, n = 604). Content validity of the scale is also supported, as a factor analysis with scree plots revealed both pre-test and post-test have one underlying explaining factor which accounted for 32.1% of the variance on its own.

Student Well-being

Student well-being is measured with the Warwick-Edinburgh Mental Well-Being Scale, which is suited for people aged 16 and older (WEMWBS; Tennant et al., 2007) and found suitable for school students aged 13 to 16 as well (Clarke et al., 2011). It contains fourteen items with statements about positive feelings using a 5-point scale where 1 = none of the time and 5 = all of the time. The scale ranges from 14 to 70, where a high score represents high well-being, with a median score for males at 52 (95% CI: 51-52), for females 51 (95% CI: 50-52) and for the population at 51 (95% CI: 51-52; Tennant et al., 2007). Furthermore, the user guide by Stewart-Brown and Janmohamed (2008) indicates the scale has good scaling properties which allow numbers to be compared easily; a score of 22 is twice as bad as a score of 44.

According to Tennant and colleagues (2007), the internal consistency of their scale is high for students and the entire population ($\alpha = .89$, n = 82; $\alpha = .91$, N = 1749), and this study found high internal consistencies for both pre-test and post-test ($\alpha = .83$, n = 740; $\alpha = .89$, n = 604). The scale's internal validity was tested with a factor analysis; one main factor was found which accounted for 32.4% of variance on its own in the pre-test and 42.3% in the post-test. Both scree plots supported the one-factor hypothesis, cutting off at the second component.

Student Satisfaction

Berkel and colleagues (2011) identified the following components of responsiveness: satisfaction, sessions attended, active participation and home practice completion (attempts at practising assigned skills at home). However, because of the nature of this study, only satisfaction could be measured reasonably and proved to have a decent internal consistency. To avoid confusion, henceforth we will speak of student satisfaction instead of student responsiveness during analyses.

Satisfaction in HC1 and HC2 has been measured with six items from the student evaluation questionnaire. For this study, only the items using a 10-point scale are included for quantitative analysis. This scales ranges from 1 = worthless to 10 = excellent. Student satisfaction is measured as the mean of these six item scores. The internal consistency of the six satisfaction items was high ($\alpha = .88$, n = 659) and the factor analysis revealed one clear factor explaining 64.1% of the variance and clearly cutting off at the second component on the scree plot.

Timing, Dosage and Programme Staff

Timing of Happy Class, dosage and programme staff were coded into quantitative measures so we could explore their influence and control for it. Timing of the prevention was measured per student by the date on which they registered for Happy Class. Two timing moments were observed in each implementation and coded as 0 (spring) and 1 (fall). Dosage indicates whether the exhibition and theatre show were shown to the student or not, where 0 = neither was shown, 1 = only one was shown and 2 = both were shown. This was measured with an item in the online Happy Class evaluation questionnaire about whether the student had seen the theatre show or not and completed with information from the programme staff about which groups had seen the exhibition and which had not. Finally, programme staff was measured in values from 0 (original trained staff) to 1 (original trained staff and new recruits).

Analysis

The main analyses are independent t-tests for differences, paired-samples t-tests for changes in well-being and depression over time, and multiple linear regression analyses (with stepwise method) to identify predictor variables. Since the study uses a pre-test and post-test model, the sample size varies per analysis. Analyses only addressing the pre-test measures (such as depression and well-being at pre-test) have samples such as 740 for the entire group, 536 for the normal group, 100 for the sub-clinical group and 104 for the clinical group. Of these 740 at the pre-test, 373 are from HC1 and 367 from HC2. For analyses looking at both pre-test and post-test (e.g. when looking at change over time in depression) this sample size will decrease since some students did not complete the post-test. The sample size will always be mentioned along with the analysis.

Random Missing Students and Data

Students who completed the post-test had also completed the pre-test, so in this sample all students have pre-test scores even if there is no post-test score. A number of students had not completed the post-test (n = 136) and some of these students had not completed the evaluation questionnaire either (n = 81). To make sure these missing cases (students) were random, frequencies were calculated for missing post-test cases per implementation period. The number of missing post-test for HC1 and HC2 are very similar to each other (n = 69; n = 67), which

seems to indicate this number is normal for the setting. Frequencies were checked for each class over both periods as well (HC1 + HC2), which revealed a maximum of nine missing post-test cases per class. Classes had a minimum number of twelve students who completed the pre-test, so no class was entirely excluded from the study at post-test.

Since the missing students appear to be random, missing answers (data) on the questionnaires were checked next. By observation and with support from Little's Missing Completely at Random Test (MCAR), we can conclude these missing data are not random; MCAR showed significant results ($p < .001$) for well-being, depression and satisfaction. For this reason, missing data were excluded from analysis and analyses will be done with the data available (students will be excluded pairwise).

Results

Implementation Differences (research question 1)

Programme and Student Characteristics

Pre-test Frequencies for programme and student characteristics were observed and compared with a chi-square test for HC1 and HC2. Significant differences in the frequencies are discussed and complete results are presented in table 2.

Differences in frequencies were found for programme staff and dosage (theatre and exhibition), which are both missing an entire subgroup in one of the implementation periods. Further significant differences were found for student ethnicity $\chi2$ $(1, n = 740) = 7.647, p < .05$, student gender $\chi2$ $(1, n = 740) = 8.620, p < .01$, and student age $\chi2$ $(1, n = 740) = 23.003, p < .001$. HC2 had fewer foreign students, a more balanced number of females and males and more students between the ages 12-14 than HC1. Among programme characteristics, significant differences were found for timing $\chi2$ $(1, n = 740) = 210.604, p < .001$. A large group of students in HC2 (83.7%) started Happy Class in fall, while HC1 had most students starting in spring (69.2%).

Table 2

Number of students with certain student and programme characteristics for HC1 and HC2 at the pre-tes; chi-square analysis

Characteristics	Happy Class 1 n = 373		Happy Class 2 n = 367	
	n	*%*	*n*	*%*
Programme Characteristics				
Computer Problems				
No	270	83.9	295	87.5
Yes	52	16.1	42	12.5
Theatre and Exhibition				
Seen Both	36	9.7	0	0
Seen One	217	58.2	8	2.2
Seen Neither	120	32,2	359	97.8
Programme Staff				
Original Trained Staff	373	100	169	46
Trained Staff and New Staff	0	0	198	54
Timing***				
Fall	115	30.8	307	83.7
Spring	258	69.2	60	16.3
Student Characteristics				
Age***				
12 to 14 years	318	85.3	351	95.6
15 to 17 years	55	14.7	16	4.4
Ethnicity*				
Dutch	297	79.6	320	87.2
Foreign	76	20.4	47	12.8
Gender**				
Female	144	38.6	181	49.3
Male	229	61.4	186	50.7
Missed Something				
No	300	94.3	315	94.9
Yes	18	5.7	17	5.1
Pre-test and Post-test				
No	69	18.5	67	18.3
Yes	304	81.5	300	81.7
School				
A	72	19.3	60	16.3
B	203	54.4	109	29.7
C	98	26.3	0	0
D	0	0	198	54
Subgroup				
Clinical	52	13.9	52	14.2
Sub-clinical	48	12.9	52	14.2
Normal	273	73.2	263	71.7

Note.$*p < .05. **p < .01. ***p < .001.$

Change in Student Depression and Well-being

First, change in well-being and depression from pre-test to post-test was analysed separately for HC1 and HC2 and the depression subgroups with paired-samples t-tests. Significant results are discussed and complete results are presented in table 3.

For the normal group in HC1, a significant increase in depression was found from pre-test (M = 7.70, SD = 4.03) to post-test (M = 10.13, SD = 8.40); t (221) = 4.79, *p* < .001. This was not a clinically significant increase, as students remained within the normal group (CES-D < 16).

In contrast, the sub-clinical group in HC1 significantly decreased in depression from pre-test (M = 17.98, SD = 1.55) to post-test (M = 14.58, SD = 6.12); t (42) = -3.78, *p* < .001. Since the mean depression score went from 18 to 15, this change is clinically significant as well; the subgroup went from sub-clinical depression (CES-D 16-21) to normal (CES-D < 16).

The clinical group in HC1 had a significant decline in depression as well from pre-test (M = 30.46, SD = 7.98) to post-test (M = 26.90, SD = 11.48), but it was not clinically significant (still > 21); t (38) = -2.35, *p* < .05.

For the normal group in HC2, a significant increase was found as well in depression from pre-test (M = 7.75, SD = 4.17) to post-test (M = 9.04, SD = 8.17); t (219) = -3.78, *p* = .01. As was the case for the normal group in HC1, this group's increase in depression remains within the score-range of the normal group.

Finally, the clinical group in HC2 significantly decreased in depression from the pre-test (M = 29.63, SD = 6.06) to the post-test (M = 22.21, SD = 10.96); t (42) = -4.72, *p* < .001. This change was nearly clinically significant as well (< 22). Moreover, this is the only group in both implementations where well-being improved significantly from pre-test (M = 42.19, SD = 6.86) to post-test (M = 45.00, SD = 10.67); t (42) = 2.02, *p* = .05.

Table 3

Change in student well-being and depression for HC1 and HC2 from pre-test to post-test;
paired samples t-test

Measure per Group	Happy Class 1 n = 304					Happy Class 2 n = 300				
	M_0 (SD)	M_1 (SD)	t	df	p	M_0 (SD)	M_1 (SD)	t	df	p
Entire Group										
Depression (CES-D)	12.08 (9.09)	12.91 (10.22)	1.81	303	.835	12.20 (9.01)	11.96 (9.81)	- .49	299	.623
Well-being (WEMWBS)	51.78 (8.19)	51.88 (9.38)	.21	303	.072	52.45 (8.04)	53.10 (9.34)	1.51	299	.131
Normal Group										
Depression (CES-D)	7.70 (9.09)	10.13 (10.22)	4.79	221	**.000**	7.75 (4.17)	9.04 (8.17)	2.57	219	**.011**
Well-being (WEMWBS)	53.88 (7.15)	53.93 (8.32)	.10	221	.918	54.98 (6.62)	55.08 (8.24)	.22	219	.827
Sub-clinical Group										
Depression (CES-D)	17.98 (1.55)	14.58 (6.12)	- 3.78	42	**.000**	18.35 (1.64)	17.41 (6.75)	- .92	36	.366
Well-being (WEMWBS)	48.35 (6.63)	48.42 (7.54)	.07	42	.948	49.35 (6.37)	50.73 (8.37)	1.00	36	.325
Clinical Group										
Depression (CES-D)	30.46 (7.98)	26.90 (11.48)	- 2.35	38	**.024**	29.63 (6.06)	22.21 (10.96)	- 4.72	42	**.000**
Well-being (WEMWBS)	43.64 (9.04)	44.00 (11.62)	.23	38	.822	42.19 (6.86)	45.00 (10.67)	2.02	42	**.050**

Note. CES-D range = 0-60 (20 items). WEMWBS range = 14-70 (14 items). High scores on these scales relate to high depression/well-being. M_0 = Mean at pre-test. M_1 = Mean at post-test. Change statistics (e.g. *t*-statistics) are reported as post-test – pre-test. Bold = significant difference between pre-test and post-test mean.

Outcome Measures per Group

Student depression, well-being and satisfaction scores were compared per depression subgroup between the implementations; independent t-tests were conducted per subgroup with implementation as grouping variable. Complete results can be viewed in Appendix A.

Analysis for all students revealed a significant difference between HC1 (M = 6.43, SD = 2.02) and HC2 (M = 6.13, SD = 2.07) on satisfaction scores (t (657) = 1.93, *p* = .05); the entire group of students were more satisfied in HC1 than in HC2.

Students in the normal group significantly differed on satisfaction as well between HC1 (M = 6.60, SD = 2.05) and HC2 (M = 6.09, SD = 2.09); t (476) = 2.66, *p* < .05. On average, normal group students scored higher on satisfaction in HC1.

For students in the sub-clinical group, a significant difference was found in depression scores at the post-test between HC1 (M = 14.58, SD = 6.12) and HC2 (M = 17.41, SD = 6.75); t (78) = -1.96, p = .05. On average, sub-clinical group students were more depressed at the post-test in HC2 compared to the students in HC1. This score is clinically significantly higher from those in HC1 as well; those in HC1 scored within the range of the normal group, while those in HC2 remained within the range of the sub-clinical group.

The Role of Implementation (research question 2)

For the entire group (both HC1 and HC2), the relationship was analysed between perceived implementation differences (research question 1) and satisfaction, depression and well-being pre-test, post-test and difference scores. The variables we found notable or significant differences in were: dosage, programme staff, timing, age, ethnicity, gender, school, satisfaction and depression at the post-test (subclinical group only). Because some of the subgroups were too small to test these variables (n < 30), only the entire group is observed. Including implementation period (HC1 or HC2) as variable did not alter results or reveal it to be a relevant factor on its own after controlling for the other variables (e.g. dosage). Table 4 presents an overview of all significant models and predictors.

A Multiple regression analysis using the stepwise method was conducted for the dependent variables well-being and depression at pre-test, post-test and difference scores and for satisfaction. The independent variables were gender, programme staff, timing, ethnicity, age (raw data from 12 to 17), school, dosage and interaction effects between programme staff and school (their standardized scores were multiplied).

A significant model emerged for well-being at pre-test (F(2, 737) = 12.89, p < .001, R Square = .034); the model explains 3.4% of the variance in well-being at pre-test. In this model, gender was a significant predictor for well-being at pre-test (B = -2.25, t = -3.78, p < .001); for females scored lower on well-being at pre-test than males. Programme staff was also a significant predictor (B = 2.44, t = 3.66, p < .001); students with original and new staff scored higher on well-being at pre-test than students with only the original staff.

To analyse well-being at post-test, well-being scores at pre-test were added to the regression so it could be controlled for. We found a significant model (F(2, 601) = 194.22, p < .001, R Square = .39); the model explains 39% of the variance in well-being at post-test. In this

model, well-being pre-test scores were a significant predictor for well-being scores at post-test (B = .71, t = 19.39, p < .001); students with a high well-being score at pre-test also scored relatively higher on well-being at post-test. The other significant predictor was school B (B = -1.65, t = -2.73, p < .05); students in school B scored lower on well-being than those in other schools.

For well-being difference-scores a significant model was found as well (F(2, 601) = 33.95, p < .001, R Square = .102); the model explains 10.2% of the variance in well-being difference scores. In this model well-being pre-test scores were a significant predictor for well-being difference scores (B = -.29, t = -7.81, p < .001); students with a high well-being score at pre-test increased relatively less in well-being. Dosage was a significant predictor as well (B = -1.48, t = -2.89, p < .01); for each unit extra in dosage (seen either exhibition/theatre show or both), students increased less in well-being than those who had seen neither exhibition or theatre show.

Next, a significant model was found explaining depression pre-test scores (F(3, 736) = 157.19, p < .001, R Square = .391); the model explains 39.1% of the variance in depression at pre-test. In this model, well-being pre-test scores were a significant predictor for depression pre-test scores (B = -.67, t = -21.01, p < .001); students with a high well-being score at pre-test had lower scores on depression at pre-test. Gender was another significant predictor for depression at pre-test (B = 1.46, t = 2.81, p < .05); girls scored higher on depression at pre-test than boys. Finally, school A predicted depression pre-test scores as well (B = -1.48, t = -2.20, p < .05), where students in school A had lower scores on depression at pre-test than other students.

A significant model also emerged for depression at post-test (F(2, 601) = 80.93, p < .001, R Square = .212); this model explains 21.2% of the variance in depression at post-test. In this model, well-being pre-test scores were a significant predictor for depression post-test scores (B = -.55, t = -12.24, p < .001); students with a high well-being score at pre-test scored lower on depression at post-test. Dosage was the other significant predictor for depression at post-test (B = 1.92, t = 3.08, p < .01); for each unit extra in dosage, students scored higher on depression scores at post-test than those who had seen neither exhibition or theatre show.

Furthermore, a significant model emerged for depression difference scores (F(3, 600) = 11.16, p < .001, R Square = .053); the model explained 5.3% of the variance in depression difference scores. In this model, gender was the largest predictor for depression difference scores (B = -2.33, t = -3.52, p < .001); girls had a larger decrease in depression than boys. Dosage was

the second predictor (B = 1.68, t = 2.99, p < .01); students who had seen both had a larger increase in depression than those who had seen neither exhibition or theatre show. Finally, well-being pre-test scores were a significant predictor for depression difference scores (B = .11, t = 2.81, p < .001); students with a high well-being score at pre-test decreased less in depression than others.

Lastly, we found a significant model to explain student satisfaction (F(4, 654) = 14.34, p < .001, R Square = .081); the model explained 8.1% of the variance in satisfaction. In this model, school D (B = -1.17, t = -6.49, p < .001) and C (B = -.73, t = -3.04, p < .01) were significant predictors; students in school D and C scored lower on satisfaction than students from other schools. Gender was also a significant predictor (B = .46, t = 2.83, p = .005); girls scored higher on satisfaction than boys. Finally, well-being pre-test scores were a significant predictor for satisfaction scores (B = .027, t = 2.81, p = .005); students with a high well-being score at pre-test scored higher on satisfaction.

The Role of Student Well-Being in Student Depression (research question 3)

To understand the relationship between well-being and depression, we conducted a multiple linear regression analysis using stepwise method for the entire group (both HC1 and HC2). Gender, programme staff, school and dosage were also added to the regression (these variables were significant in question 2). Since well-being pre-test scores have already been analysed in question 2, we will only analyse well-being difference scores.

A significant model emerged for depression difference scores (F(3, 600) = 17.08, p < .001, R Square = .074); the model explains 7.4% of the variance in depression difference scores. Of this variance, 4% is explained by well-being difference scores alone (partial R = -.20, so R Square = .04). In the model, well-being difference scores (B = -.21, t = -4.99, p < .001), gender (B = -2.32, t = -3.56, p < .001) and dosage (B = 1.35, t = 2.42, p < .05) were significant predictors of depression difference scores. An increase in well-being difference scores (so, a higher well-being at post-test compared to pre-test), is related with a decrease in depression. Gender is also related to a decrease in depression, where girls show a higher decrease than boys. Finally, dosage is related to depression in the sense that students who had seen one or both increased in depression.

Table 4

Significant predictors and models for depression, well-being and satisfaction scores; linear regression (stepwise)

| | Depression | | | | | | | | |
| | Pre-test | | | Post-test | | | Difference | | |
Predictors	B	β	$R^2_{partial}$	B	β	$R^2_{partial}$	B	β	$R^2_{partial}$
Dosage				1.92	.11	.016*	1.68	.12	.015*
Gender	1.46	.08	.011*				-2.33	-.14	.020**
School A	-1.48	-.06	.007*						
Well-being pre-test	-.67	-.61	.375**	-.55	-.44	.200**	.11	.11	.013*
R^2 (Adjusted)	.391 (.388)**			.212 (.210)**			.053 (.048)**		

| | Well-being | | | | | | | | |
| | Pre-test | | | Post-test | | | Difference | | |
Predictors	B	β	$R^2_{partial}$	B	β	$R^2_{partial}$	B	β	$R^2_{partial}$
Dosage							-1.48	-.11	.014*
Gender	-2.25	-.14	.019**						
Programme Staff	2.44	.13	.018**						
School B				-1.65	-.09	.012*			
Well-being pre-test				.71	.62	.384**	-.29	-.30	.092**
R^2 (Adjusted)	.034 (.031)**			.393 (.391)*			.102 (.099)*		

| | Satisfaction | | |
Predictors	B	β	$R^2_{partial}$
Gender	.46	.11	.012*
School C	-.73	-.12	.014*
School D	-1.17	-.25	.061**
Well-being pre-test	.03	.11	.012*
R^2 (Adjusted)	.081 (.075)*		

Note. Difference = post-test – pre-test. $R^2_{partial}$ = variance explained by one predictor alone. * = $p \leq$ .05. ** = $p \leq$ .001.

The Role of Student Satisfaction (research question 4)

First, satisfaction was related to well-being and depression pre-test, post-test and difference scores in the entire group (both HC1 and HC2). Multiple linear regressions with stepwise method

were conducted while controlling for gender, programme staff, school, the interaction effect between school and staff, and dosage. Significant results are discussed and full results for satisfaction as predictor can be found in table 5.

A significant model emerged for well-being pre-test scores ($F(3, 655) = 10.25$, $p < .001$, R Square = .045); the model explains 4.5% of the variance in well-being pre-test scores. Of this variance, 1.1% is explained by satisfaction alone (partial R = .107, so R Square = .011). In the model, gender (B = -2.47, t = -3.91, $p < .001$), programme staff (B = 2.86, t = .72, $p < .001$) and satisfaction (B = .43, t = 2.75, $p < .05$) were significant predictors of well-being pre-test scores. Girls scored lower on well-being at pre-test than boys, students with original and new staff scored higher on well-being at pre-test than students with only original staff, and those who were more satisfied scored higher on well-being at pre-test.

Well-being post-test scores could be explained by a significant model as well ($F(3, 600)$ = 137.90, $p < .001$, R Square = .408); the model explains 40.8% of the variance in well-being post-test scores. Of this variance, 2.6% is explained by satisfaction alone (partial R = .160, so R Square = .026). In the model, well-being pre-test scores (B = .70, t = 19.31, $p < .001$), satisfaction (B = .58, t = 3.97, $p < .001$) and school B (B = -2.09, t = -3.44, $p = .001$) were significant predictors of well-being post-test scores. Students with higher well-being pre-test scores also scored higher on well-being at post-test. Students with higher satisfaction scored higher on well-being at post-test, and students from school B scored lower on well-being at post-test than students from other schools.

In line with this, a significant model emerged for well-being difference scores ($F(3, 600)$ = 25.62, $p < .001$, R Square = .114); the model explains 11.4% of the variance in well-being difference scores. Of this variance, 1.3% is explained by satisfaction alone (partial R = .116, so R Square = .013). In the model, well-being pre-test scores were a significant predictor (B = -.29, t = -8.02, $p < .001$); students with a higher well-being score at pre-test increased less on well-being than others. Dosage was the second significant predictor (B = -1.58, t = -3.10, $p < .01$); students who had seen one or both, scored lower on well-being. Satisfaction was the other significant predictor (B = .41, t = 2.86, $p < .01$), where students with a higher satisfaction showed a larger increase in well-being.

Next, the relationship with depression was assessed. A significant model emerged for depression post-test scores ($F(3, 600) = 56.75$, $p < .001$, R Square = .221). Of this 22.1% explained variance, satisfaction explained 1.1% variance. In the model, well-being pre-test

scores were a significant predictor (B = -.54, t = -12.11, $p < .001$); students with a higher well-being score at pre-test scored lower on depression at post-test. Next, dosage was a significant predictor (B = 2.04, t = 3.27, p = .001), so students who had seen one or both scored higher on depression at post-test. Satisfaction was the final significant predictor (B = -.46, t = -2.61, $p <$ .05), where students with a higher satisfaction showed lower scores in depression at post-test.

Finally, we tested the impact of student characteristics on satisfaction. A multiple regression analysis with a stepwise method was conducted with the dependent variable student satisfaction and independent variables gender, age, school, programme staff, ethnicity, dosage, completion of both pre-test and post-test, whether they were able to participate in all Happy Class components, and well-being scores at pre-test. A significant model emerged for satisfaction ($F(4, 645) = 14.15$, $p < .001$, R Square = .081), explaining 8.1% of the variance in satisfaction. In this model, programme staff (B = -1.17, t = -6.44, $p < .001$), School C (B = -.73, t = -3.01, $p < .01$), gender (B = .46, t = 2.81, $p < .05$) and well-being pre-test scores (B = .027, t = 2.79, $p < .05$), were significant predictors of satisfaction. Students with mixed programme staff scored lower on satisfaction than those with original programme staff, students in school C scored lower on satisfaction than students in the other schools, girls scored higher on satisfaction than boys and students with high well-being pre-test scores scored higher on satisfaction than others.

Table 5

Satisfaction as predictor for depression and well-being scores; linear regression (stepwise)

Dependent Variables	Satisfaction		
	B	β	$R^2_{partial}$
Depression			
Pre-test		-.08	.006*[1]
Post-test	-.46	-.10	.011*
Difference		-.04	.001[1]
Well-being			
Pre-test	.43	.11	.011*
Post-test	.41	.11	.013*
Difference	.58	.13	.026**

Note. Difference = post-test – pre-test. $R^2_{partial}$ = variance explained by one predictor from a model.
[1] = satisfaction was excluded during stepwise regression. * = $p \leq .05$. ** = $p \leq .001$.

Discussion

This study aimed to understand which factors in the Happy Class implementation in The Hague could influence its effectiveness. Four topics were studied: (1) implementation differences between HC1 and HC2, (2) the role of implementation, (3) the role of student well-being in student depression, and (4) the role of student responsiveness.

(1) Implementation Differences

First of all, there were significant differences between HC1 and HC2 in both student and programme characteristics as well-being, satisfaction and depression scores. When looking at frequencies, seven aspects in total were significantly different. This first result meant the effectiveness of HC1 and HC2 could vary because of at least seven factors, if all seven were significant predictors of depression, well-being or satisfaction. These factors were: (1) dosage – no-one had seen both theatre show and exhibition in HC2, and only eight students had seen one of them in HC2. (2) Programme staff – HC1 only used original staff trained by the Trimbos-institute, while HC2 had both groups with original staff and groups with newly recruited staff working together with the original staff. (3) Timing – in HC2, a large number of students (83.7%) had Happy Class in fall, as opposed to HC1 (30.8%). (4) Age – in HC2, more students were between the age of 12-14 (95.6%) compared to HC1 (30.8%). (5) Ethnicity – in HC2, more students were Dutch (87.2%) compared to HC1 (79.6%). (6) Gender – in HC2, more students were female (49.3%) than in HC1 (38.6%). (7) School – in HC2, school C was replaced by school D and vice versa for HC1.

Secondly, this study would make no sense if no differences were found between implementation effectiveness. HC1 and HC2 were compared on change in well-being and depression, and on satisfaction, well-being and depression pre-test, post-test and difference scores. When comparing significant improvement for both depression and well-being, HC1 and HC2 revealed varying results for the sub-clinical group and the clinical group. In HC1, the sub-clinical group had a significant decrease in depression, while this was not the case for HC2. In contrast, the clinical group in HC2 revealed a significant increase in well-being, while those in HC1 did not. When comparing mean scores for depression, well-being and satisfaction,

differences were found for the entire group, the normal group and the subclinical group. In HC1, the entire group was more satisfied (6.4) than the entire group of HC2 (6.1). In HC1, students in the normal group scored higher on satisfaction compared to HC2. In HC1, students in the sub-clinical group scored lower on depression at post-test (normal range) compared to HC2 (sub-clinical range).

In other words, something happened in HC1 to make students significantly more satisfied and to let the sub-clinical group in HC1 improve both significantly and clinically better than the sub-clinical group in HC2. On the other hand, HC2 is the only implementation which has a group (the clinical group) with a significant increase in well-being along with a significant decrease in depression. This is odd, as Happy Class is supposed to decrease depression by improving well-being, yet in all other subgroups where depression decreased significantly, it did this without a significant increase in well-being. A possible reason for this could be that well-being and depression have a more correlational than causal relationship. Happy Class, with exercises about problem solving and cognitive therapy, could be decreasing depression in other ways than through well-being.

(2) The Role of Implementation

We analysed the role of implementation; the factors on which HC1 and HC2 differed (topic 1) were analysed to understand if they had a significant relationship with outcome scores. Only the entire group was analysed, because the number of students within some factors were too small per subgroup to include in analysis (n < 30). Five factors emerged as relevant for outcome scores: (1) Gender – females scored lower on well-being and higher on depression at pre-test, yet their difference scores for depression showed a larger decrease in depression compared to boys (this was the strongest predictor for depression difference scores). They were also more satisfied than boys. (2) Programme staff – the combination of new and original staff working together was related to higher well-being scores at pre-test only. (3) Dosage – the students who saw either one of or both exhibition and theatre show had difference scores which revealed a smaller increase in well-being and a smaller decrease in depression. They also scored higher on depression at the post-test. (4) School – School A scored lower on depression at pre-test, school B scored lower on well-being at post-test, and both school C and D scored lower on satisfaction than the other schools. (5) High well-being pre-test scores predicted high well-being post-test scores, a smaller

increase in well-being, lower depression pre-test scores, lower depression post-test scores and a smaller decrease in depression. They also predicted higher satisfaction.

It appears girls may begin worse off, but eventually strive and surpass their male counterparts in both depression and well-being scores. Perhaps their higher satisfaction reflects this improvement. While gender was a much stronger predictor for satisfaction than well-being pre-test scores, this is slightly contradicting as higher well-being scores at pre-test are related to higher satisfaction scores. Perhaps it has something to do with the factor programme staff, which was related to higher well-being pre-test scores. Since programme staff only mattered for the pre-test scores, the mixed programme staff group either succeeded more in the first e-lesson and discussion group, or they were paired with students with higher well-being by coincidence. Furthermore, it appears school is an important factor as well; each of the schools is in some way related to either satisfaction, depression or well-being. This could be related to personal situations surrounding the school setting; for example, the mixed programme staff group only went to school D. And last, dosage revealed a negative influence on both well-being difference scores, depression difference scores and depression post-test scores. Either the students became more depressed after seeing the exhibition and theatre show, for example because students became more aware of their situation or feelings, or there is another factor responsible for the cause.

(3) The Role of Student Well-being in Student Depression

Well-being is a central mechanism by which Happy Class aims to decrease depression and has been studied before in the implementation evaluation of HC1 (Van der Zanden & Van der Linden, 2013). Since no real proof of the relationship between well-being and depression was found in that study, we decided to explore it as well. As the previous topic included well-being pre-test scores so it could be controlled for, it is already clear well-being pre-test scores can predict lower depression pre-test scores and post-test scores. However, high well-being pre-test scores also predicted a smaller decrease in depression, and a significant increase in well-being is often absent when there is a significant decrease in depression. We did find a relationship between well-being difference scores and depression difference scores, but well-being difference scores explained only 4% of the variance in depression difference scores – what about the other 96%? All of this slightly supports the connection between well-being and depression, but mostly

at a correlational level; it cannot be said with certainty that Happy Class's main mechanism is well-being, or that it greatly impacts well-being at all.

(4) The Role of Student Responsiveness

In accordance with the model of Berkel and colleagues (2011) about programme implementation, we explored the role of a component of responsiveness – student satisfaction – as a predictor. We also studied whether student characteristics would influence satisfaction. Higher scores on satisfaction predicted lower scores on depression at post-test, higher scores on well-being pre-test, post-test and a larger increase in well-being. In addition, mixed programme staff (instead of original staff only) and school C predicted lower scores on satisfaction, while gender (girls) and students with higher well-being pre-test scores predicted higher scores on satisfaction. These results support that student satisfaction plays a role in well-being and depression, even when accounting for well-being pre-test effects. Satisfaction is also influenced by well-being pre-test scores, gender, school and programme staff. Since school C and D predicted lower satisfaction scores in topic 2 and mixed programme staff is only connected to school D, this seems to be a confirmation that mixed programme staff is tied to lower student satisfaction scores. However, because mixed programme staff so closely overlaps with school D, it may as well be a factor unique to school D which explains the variance.

While it is true we found satisfaction to predict well-being and depression scores alone and in combination with other implementation factors, the question is which role it takes. The problem with assessing satisfaction here, is that no definite statements can be made about its causal relationship with well-being; satisfaction was not influenced solely to test its effect on well-being or depression. It could simply be that students who have a higher well-being at the end of Happy Class, become more satisfied about the programme or that another aspect of the programme (e.g. staff or dosage) influences both in comparable ways. This would mean satisfaction is an outcome score such as depression itself. The other aspects of responsiveness (sessions attended, active participation, home practice completion) are more likely to influence outcomes, but in this study we were unable to measure these aspects in a reliable way. However, there are some things satisfaction scores can be used for, even with confusion about its role. At the very worst, a positive correlational relationship was found between well-being and satisfaction; when one decreases or increases, the other does so as well. In other words, it seems

at though student satisfaction can give an idea of whether things went well or not for students in the case of well-being.

Strengths and Limitations

With this study, it became clear how difficult it is to conduct a thorough implementation study and make clear statements about causal relationships or relationships at all. While the online and class aspects of Happy Class ensured there was comparable data to work with, there was no control group to see whether improvement was truly due to Happy Class. Another problem was that some documents (such as the division of programme staff over schools) and details of the implementation were often incorrectly archived or not archived at all and had to be traced down and corrected personally. This lack of documentation is further problematised when there is a variation in several major variables and the sample, and when other potential factors are missing (e.g. other aspects of responsiveness, the school situation, problem solving skills, cognitive skills). Some variables were also constructed in a way which made it hard to study them separately, because they overlapped their own categories or those of other variables (e.g. programme staff and schools). This made the impact of developer involvement difficult to detect as well, since it overlapped with several other changes. In the future, it may be wise to decide if and how annual implementation quality studies will be conducted – before implementing the actual prevention or intervention programme on a large scale.

Future Research

In the literature review it was stated that Happy Class is a child programme of the effective-proven MasterYourMood programme. While it is good to know an effective programme was used as the basis for this prevention programme, it does not ensure that Happy Class is truly effective for treating student depression. This study and previous ones before it (Van Oorsprong, 2011) provide signs that Happy Class likely is effective, but the programme would profit from real evidence from a randomised controlled trial. In this case, it's especially important to include a placebo-group as control group instead of a care-as-usual group, since the mere initiative of undertaking a depression prevention programme (of any kind) may result in improvement.

Secondly, the predictors we found and the mechanisms of Happy Class should be further explored; are the factors we found truly able to impact the effectiveness of Happy Class or are

the relationships of a different nature? Why are most of the students improving in depression but not in well-being, when this decrease in depression is supposed to go through well-being? Why are only the positive psychology exercises considered a Happy Class mechanism, but not the equally present problem-solving and cognitive behavioural exercises? Why have normal group students in both HC1 and HC2 significantly (although not clinically) increased on depression? These questions relate to core aspects of the prevention programme which should be clarified to understand whether the program is working as it should or whether it needs to be improved upon.

Third, what is the exact target group of Happy Class? It aims to reach more students with slight to mild depression symptoms (sub-clinical depression). Yet at the same time, the prevention appears to work for the clinical group as well; sometimes even though it does not work for the sub-clinical group (see table 3, HC2). Is this common or unique to the Happy Class implementation in The Hague? If these effects can be proven to be common, Happy Class can be seen as – and perhaps improved to be – a programme which does not only aim to prevent clinical depression, but decreases it as well.

Conclusion

In conclusion, we found implementation effectiveness does indeed differ between HC1 and HC2. In HC1, Student satisfaction is higher and the sub-clinical group improved significantly and clinically, while it did not improve at all in HC2. On the other hand, the only group which improved significantly on well-being is the clinical group from HC2; this group nearly improved clinically in depression as well (from clinical to sub-clinical). The decision to exclude the exhibition and theatre show (i.e. dosage items) from HC2 seems to have been a positive one, as inclusion was related to decreases in well-being and increases in depression. Yet the question then remains why HC2 is still off worse than HC1. Furthermore, well-being and satisfaction do seem to play a part in the implementation proces, specifically with each other, but it remains unclear what their exact role is; well-being may not affect depression at all and satisfaction could either be an end product such as depression, or actually influence outcomes.

While the exact roles are still uncertain, we did narrow the domain of Happy Class implementation factors down to a few: well-being pre-test scores, school, gender, programme staff and dosage seem to be relevant factors which influence implementation quality. In addition, satisfaction still proved to be related to various factors in the implementation process and

outcome scores even if those relationships were small and it is only a part of responsiveness. If the entire concept of responsiveness could be measured for Happy Class, new mechanisms of change to improve the quality of implementation may reveal themselves.

References

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Society for Prevention Research, 12*, 23-33. doi:10.1007/s11121-010-0186-1

Bouma, J., Ranchor, A. V., Sanderman, R., & Van Sonderen, E. (2012). *Het meten van symptomen van depressie met de CES-D*. Rijksuniversiteit Groningen: NCG.

Clarke, A., Friede, T., Putz, R., Ashdown, J., Martin, S., Blake, A., ... Stewart-Brown, S. (2011). Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Validated for teenage school students in England and Scotland. A mixed methods assessment. *BMC Public Health,* 11:487. doi:10.1186/1471-2458-11-487

Cuijpers, P. Boluijt, P., & Van Straten, A. (2008a). Screening of depression in adolescents through the internet: Sensitivity and specificity of two screening questionnaires. *European Child and Adolescent Psychiatry, 17*, 32-38. doi:10.1007/s00787-007-0631-2

Cuijpers, P., & Smit, F. (2008b). Subklinische depressie: een klinisch relevante conditie? *Tijdschrift voor Psychiatrie, 50*, 519-528.

Cuijpers, P., Muñoz, R. F., Clarke, G. N., & Lewinsohn, P. M. (2009). Psychoeducational treatment and prevention of depression: the "Coping with Depression" course thirty years later. *Clinical Psychology Review, 29*, 449-458. doi:10.1016/j.cpr.2009.04.005.

Cuijpers, P., Van Straten, A., Smit, F., Mihalopoulos, C., & Beekman, A. (2008c). Preventing the onset of depressive disorders: A meta-analytic review of psychological interventions. *American Journal of Psychiatry, 165*, 1272-1280. doi:10.1176/appi.ajp.2008.07091422

De Graaf, R., Ten Have, M., & Van Dorsselaer, S. (2010). *De psychische gezondheid van de Nederlandse bevolking. NEMESIS-2: Opzet en eerste resultaten*. (NEMESIS-report). Utrecht: Trimbos-institute.

De Hollander, A. E. M., Hoeymans, N., Melse, J. M., Van Oers, J. A. M., & Polder, J. J. (2006). *Zorg voor gezondheid. Volksgezondheid Toekomst Verkenning 2006. RIVM-rapport.* Houten: Bohn Stafleu Van Loghum.

Durlak, J. A., & 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*, 327-350. doi:10.1007/s10464-008-9165-0

Fleuren, M. A. H., Paulussen, T. G. W. M., Van Dommelen, P., & Van Buuren, S. (2012). *Ontwikkeling meetinstrument voor determinanten van innovaties (MIDI)*. (TNO report R10625). Leiden: Behavioural and Societal Scienes.

Fleuren, M., Wiefferink, K., & Paulussen, T. (2004). Determinants of innovation within health care organizations: Literature review and Delphi-study. *International Journal for Quality in Health Care, 16*, 107-123. doi:10.1093/intqhc/mzh030

Gray, P. (2007). *Psychology* (5th ed.). United States, NY: Worth Publishers.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*, 466-474. doi:10.1037/0003-066X.58.6-7.466

Horowitz, J. L., & 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. doi:10.1037/0022-006X.74.3.401

Kim-Cohen, J., Caspi, A., Moffit, T. E., Harrington, H., Milne, B. J., & Poulton, R. (2003). Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective- longitudinal cohort. *Archives of General Psychiatry, 60*, 709-717. doi:10.1001/archpsyc.60.7.709

Koppelaar, E., Knibbe, J. J., Miedema, H. S., & Burdorf, A. (2009). Determinants of implementation of primary preventive interventions on patient handling in healthcare: a systematic review. *Occupational & Environmental Medicine, 55*, 353-360. doi:10.1136/oem.2008.042481

Kroes, M. E., Mastenbroek, C. G., Couwenbergh, B. T. L. E., Van Eijndhoven, M. J. A, Festen, C. C. S., & Rikken, F. (2007). Van preventie verzekerd. Diemen: College voor zorgverzekeringen.

Lewinsohn, P. M., Clarke, G. N., Seeley, J. R., & Rohde, P. (1994). Major depression in community adolescents: Age at onset, episode duration, and time of recurrence. *Journal of the American Academy of Child and Adolescent Psychiatry, 33*, 809–818. doi:10.1097/00004583-199407000-00006

Lubyomirsky, S. (2007). De maakbaarheid van het geluk. Amsterdam: Uitgeverij Archipel.

Masi, G., Favilla, L., Mucci, M., & Millepiedi, S. (2000). Depressive comorbidity in children and adolescents with generalized anxiety disorder. *Child Psychiatry and Human Development, 30*, 205-215. doi:10.1023/A:1021351722364

Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., & Ustun, B. (2007). Depression, chronic diseases, and decrements in health: Results from the World Health Surveys. *Lancet, 370*, 851-58. doi:10.1016/S0140-6736(07)61415-9

Patton, G. C., Tollit, M. M., Romaniuk, H., Spence, S. H., Sheffield, J., & Sawyer, M. G. (2011). A prospective study of the effects of optimism on adolescent health risks. *Pediatrics, 127*, 308 – 316. doi:10.1542/peds.2010-0748

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*, 829-841. doi:10.1037/0022-006X.75.6.829

Riper, H., Van Ballegooijen, W., Kooistra, L., De Wit, J., & Donker, T. (2013). *Preventie & eMental-health: Onderzoek dat leidt, technologie die verleidt, preventie die bereikt en beklijft*. (Kennissynthese 2013 commissioned by ZonMw). Den Haag: Preventieprogramma IV.

Ruini, C., Ottolini, F., Tomba, E., Belaise, C., Albieri, E., Visani, D., … Fava, G. A. (2009). School intervention for promoting psychological well-being in adolescence. *Journal of Behavior Therapy and Experimental Psychiatry, 40*, 522 – 532. doi:10.1016/j.jbtep.2009.07.002

Rossi, P. H., Lipsey M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). United States, NY: Worth Publishers. Thousand Oaks, CA: SAGE.

Salmivalli, C., Kaukiainen, A., & Voeten, M. (2005). Anti-bullying intervention: Implementation and outcome. *British Journal of Educational Psychology, 75*, 465-487. doi:10.1348/000709905X26011

Smit, F., Bohlmeijer, E., & Cuijpers, P. (2003). *Wetenschappelijke onderbouwing depressiepreventie. Epidemiologie, aangrijpingspunten, huidige praktijk en nieuwe richtingen*. Utrecht: Trimbos instituut.

Stewart-Brown, S., & Janmohamed, K. (2008). *Warwick-Edinburgh Mental Well-being Scale (WEMWBS): User guide: Version 1*. University of Warwick: NHS Health Scotland.

Tennant, R., Hiller, L., Fishwick, R. Platt, S. Joseph, S. Weich, S., … Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. Health and Quality of Life Outcomes, 5:63. doi:10.1186/1477-7525-5-63

Thapar, A., Collishaw, S., Pine, D. S., & Thapar, A. K. (2012). Depression in adolescence. *Lancet, 379*, 1056-1067. doi:10.1016/S0140-6736(11)60871-4

Tibbits, M. K., Bumbarger, B. K., Kyler, S. J., & Perkins, D. F. (2010). Sustaining evidence-based interventions under real-world conditions: Results from a large-scale diffusion project. *Prevention Science, 11*, 252-262. doi:10.1007/s11121-010-0170-9

Van der Linden, D., & Van der Zanden, R. (2011). Happyles: Online stepped care aanbod voor studenten in het middelbaar beroepsonderwijs. *Tijdschrift voor Orthopedagogiek, 50*, 126-130.

Van der Zanden, R., & Van der Linden, D. (2013). *Evaluatieonderzoek Happyles Den Haag: Implementatie van Happyles in het VMBO en de jeugdzorgketen ter bevordering van de mentale veerkracht van jongeren*. (Report AF1219). Utrecht: Trimbos-instituut.

Van Oorsprong, S. (2011). *Happyles als preventieve interventie voor depressieve klachten.* (Master-thesis). University of Utrecht: Faculty of Social Sciences.

Verhulst, F. C., Van der Ende, J., Ferdinand, R. F., & Kasius, M. C. (1997). De prevalentie van psychische stoornissen bij Nederlandse adolescenten. *Nederlands Tijdschrift voor de Geneeskunde, 141*, 777-781.

Waraich, P., Goldner, E. M., Somers, J. M., & Hsu, L. (2004). Prevalence and incidence studies of mood disorders: A systematic review of the literature. *Canadian Journal of Psychiatry, 49,* 124–138. PMID:15065747

Wicks-Nelson, R., & Israel, A. C. (2009). *Abnormal Child and Adolescent Psychology* (7th ed.). Upper Saddle River, NJ: Pearson Education.

Wijga, A. H., Scholtens, S., Van Oeffelen, A. A. M., & Beckers, M. (2010). Klachten en kwalen bij kinderen in Nederland. Omvang en gevolgen geïnventariseerd. RIVM-rapport.

Zimmerman, M. A., & Arunkumar, R. (1994). Resiliency research: Implications for schools and policy. *Society for Research in Child Development, 8*, 1-20. doi:10.1.1.132.8277

Appendix A: Differences between HC1 and HC2

Mean score differences in student depression, well-being and satisfaction per student group for HC1 and HC2; independent samples t-test

Measure per Student Group	M_{HC1} (SD)	M_{HC2} (SD)	t	df	p
Entire Group					
Depression (CES-D)					
Pre-test	12.08 (9.09)	12.20 (9.01)	- .16	602	.870
Post-test	12.91 (10.22)	11.96 (9.81)	1.16	602	.245
Difference	.83 (8.03)	- .24 (8.33)	1.61	602	.109
Well-being (WEMWBS)					
Pre-test	51.78 (8.19)	52.45 (8.04)	- 1.01	602	.313
Post-test	51.88 (9.38)	53.10 (9.34)	- 1.60	602	.110
Difference	.10 (7.99)	.65 (7.40)	- .88	602	.380
Satisfaction	6.43 (2.02)	6.13 (2.07)	1.93	657	**.054**
Normal Group					
Depression (CES-D)					
Pre-test	7.70 (4.03)	7.75 (4.17)	- .13	440	.894
Post-test	10.13 (8.40)	9.04 (8.17)	1.37	440	.169
Difference	2.42 (7.54)	1.27 (7.44)	1.60	440	.111
Well-being (WEMWBS)					
Pre-test	53.88 (7.15)	54.98 (6.62)	- 1.68	440	.094
Post-test	53.93 (8.32)	55.08 (8.24)	- 1.45	440	.147
Difference	.05 (7.83)	.10 (6.77)	- .07	440	.947
Satisfaction	6.60 (2.05)	6.09 (2.09)	2.66	476	**.008**
Sub-clinical Group					
Depression (CES-D)					
Pre-test	17.98 (1.55)	18.35 (1.64)	- 1.05	78	.297
Post-test	14.58 (6.12)	17.41 (6.75)	- 1.96	78	**.053**
Difference	- 3.40 (5.89)	- .95 (6.28)	- 1.80	78	.076
Well-being (WEMWBS)					
Pre-test	48.35 (6.63)	49.35 (6.37)	- .69	78	.494
Post-test	48.42 (7.54)	50.73 (8.37)	- 1.30	78	.198
Difference	.07 (7.01)	1.38 (8.41)	- .76	78	.450
Satisfaction	5.98 (1.71)	6.63 (1.84)	- 1.74	89	.086

Clinical Group					
Depression (CES-D)					
Pre-test	30.46 (7.98)	29.63 (6.06)	.54	80	.594
Post-test	26.90 (11.48)	22.21 (10.96)	1.89	80	.062
Difference	- 3.56 (9.48)	- 7.42 (10.31)	1.76	80	.083
Well-being (WEMWBS)					
Pre-test	43.64 (9.04)	42.19 (6.86)	.82	71	.418
Post-test	44.00 (11.62)	45.00 (10.67)	- .41	80	.686
Difference	.36 (9.89)	2.81 (9.14)	- 1.17	80	.246
Satisfaction	6.02 (2.02)	5.81 (2.14)	.47	88	.639

Note. CES-D range = 0-60 (20 items). WEMWBS range = 14-70 (14 items). High scores on these scales relate to high depression/well-being. Satisfaction scores range from 1-10 (1 = worthless, 10 = excellent). For pre-test measures, only students who completed both pre-test and post-test were included. M_{HC1} = HC1 Mean. M_{HC2} = HC2 Mean. Bold = significant difference between HC1 and HC2.