Utrecht University

Science and Business Management

Developing a Dialogue-Based Assessment Method for Depressive Disorder in Adolescents: A Qualitative Research on the Opportunities and Challenges for Integration into Practice

---

Author
Kim Kroes

Supervisor
Dr. Julian Frommel
Dr. ing. Sander Bakkes

June 15, 2023

Abstract

Interest in digital interventions for depressive disorders has grown, yet limited usage exists in practice. This study aimed to explore the opportunities, challenges, strategies, considerations, implementation improvements and general limitations associated with integrating a dialogue-based assessment method for depressive disorders into practice. Based on a thematic qualitative analysis with semi-structured interviews (N=6), the findings of this research revealed a range of insights. Opportunities included potential stimulation of honest responses, more reliable measurements, enhanced engagement and extra diagnostic support. Challenges involved high workloads, resistance from practitioners and finding funding. The strategies and considerations for successful implementation encompassed creating clarity on usage, linkage with the current system and writing implementation plans. Implementation improvements suggested changes to the storyline and characters but also further questioning symptoms, while also changing the target group and the questionnaire the dialogue is based on. Finally, the remaining limitations revealed difficulties with the heterogeneous and comorbid nature of depressive disorders and challenges related to using questionnaires as a base for the assessment method. Overall, this study contributes to understanding the integration of a dialogue-based assessment method for depressive disorders into practice, providing insights for future research to better bridge the gap between digital interventions and clinical practice.

Keywords: *Digital interventions, serious games, depressive disorders, dialogue-based assessment methods, integration into practice*

Contents

1 Introduction

Mental health disorders are occurring more frequently. They are defined as any mental, behavioural or emotional disorder [107]. The World Health Organisation (WHO) estimated that 1 in every 8 people lives with a mental health disorder, accounting for 970 million people worldwide [106]. Mental health disorders are estimated to become 1 of the 3 major causes of morbidity and mortality by 2030 [107]. They come with high individual suffering and reduced quality of life, leading to physical and social impairments, such as concentration problems and poor relationships with family and friends. Moreover, mental health disorders also impose an economic burden on individuals, employers, healthcare systems and governments [48]. To emphasise the global burden of mental health disorders, the Global Burden of Disease study announced that these disorders account for more than 15% of the disease burden in established market economies. This is more than the disease burden caused by all cancers [108]. Numerous disorders are identified as mental health disorders, such as depressive disorder, anxiety and post-traumatic stress disorder (PTSD) [109]. However, depressive disorder, a disorder with a constant feeling of sadness and/or an inability to experience pleasure, is one of the most common types where approximately 350 million people worldwide are affected, accounting for almost 4.5% of the entire population [95]. 50% of mental health disorders occur before the age of 14 during adolescence, which also applies to depressive disorders [93].

In the past decade, the prevalence of mental health disorders worldwide has increased. At the same time, the number of people with a mental health disorder who eventually seek help is decreasing, and the level of satisfaction with the help received, has remained unchanged [19]. According to Christensen et al. [19], only 35% of the people with a mental health disorder receive care. The number of people who can access professional help is even lower for adolescents, where less than two-thirds receive help [84]. In comparison, around 80% of all people with physical disorders do receive the care they need [7]. Moreover, many cases are undetected and untreated, especially in adolescents, due to misattributing the symptoms in adolescents to everyday stress or part of the transient stage of adolescence [93]. Besides this, other factors also contribute to the high number of undetected or untreated adolescents. This problem mainly arises from the stigma associated with mental health disorders and the reduced access to health services due to high costs and a shortage of professionals [47]. In the UK alone, people sometimes have to wait up to 18 weeks before they receive a referral to a mental health service [24]. The difference between the number of people having a depressive disorder and the actual, lower number of people who receive care for their depressive disorder is also referred to as the treatment gap [56]. As a result, individuals are reluctant to seek help or do not receive help, which could eventually lead to sub-optimal treatment and poor treatment outcomes [19]. Moreover, untreated mental health disorders, especially in adolescents, lead to poor health, academic and social outcomes but also higher levels of drug abuse and even suicidal behaviour [84]. There are also issues concerning the currently used methods for diagnosis, including interview techniques or self-

assessment questionnaires, for example, which bring along issues such as self-report bias [17]. Moreover, the growing number of people with a depressive disorder, which is increasing every year due to, amongst others, the COVID-19 pandemic, increases the prevalence of issues like limited access to help [17].

These negative outcomes emphasize the importance of early detection of mental health disorders and access to treatment. As a result, researchers have been studying the use of digital interventions like serious games for mental health to minimise this treatment gap and increase access to professional help. Digital interventions within the field of mental health care are becoming frequently studied topics as a potential solution to the current problems with diagnosis and treatment [32]. Digital interventions for mental health are defined as interventions delivered via the web or applications to increase access to mental health care [62]. Interest towards game-based digital interventions, including serious games, which are games that do not have a primary purpose of entertainment [1], has especially grown [62]. The research regarding the use of these serious games for therapeutic and preventive purposes of mental health care is becoming more mature [1]. Moreover, studies related to their effectiveness have shown that serious games for the treatment of a depressive disorder, have shown to be as effective as traditional non-digital treatment methods [76]. Where the research on the use of serious games for the treatment and prevention of depressive disorders is very extensive, research on the use of these games for assessment or diagnosis of depressive disorders is still very limited [17]. However, early diagnosis of a depressive disorder is as important as the treatment and is often also seen as the first step towards treatment. This is because early diagnosis can improve the effectiveness of the treatment, and it can limit severe harmful effects [17].

My research aims to develop a dialogue-based digital intervention for the assessment of symptoms of a depressive disorder in adolescents, aged 12 to 18, as this is a group with a high prevalence of depressive disorders [93]. This intervention could potentially be used as a serious game component or as a stand-alone tool for assessment. For this research, an existing method for the assessment of a depressive disorder is incorporated to maintain a close distance to practice. This method is the use of questionnaires. A frequently used questionnaire within mental health care and research is the self-assessment Patient Health Questionnaire for depressive disorder (PHQ-8) [61]. The PHQ-8 is used for the diagnosis of a depressive disorder but also for addressing the severity of the existing symptoms. The PHQ-8 will be included within a dialogue. The questionnaire is translated into a dialogue in a way where the non-player character (NPC) asks questions based on the items of the questionnaire. The adolescent can answer these questions based on the scale items of the questionnaire [11]. This has been previously studied for different questionnaires [11, 34].

After the development of the dialogue, it will be evaluated by conducting semi-structured interviews with mental health care professionals specialising in depressive disorders or digital interventions for mental health. The goal of this evaluation is to determine the opportunities and challenges for integrating the dialogue into practice. This is important to study as the acceptance and inclu-

sion of serious games into practice for the assessment of a depressive disorder remains unclear [28]. Moreover, even though serious games have potential within mental health care, they are not widely being used in practice [28]. With the current number of people with a depressive disorder increasing every year and the number of people who seek help decreasing, there is a need for new interventions like serious games. Similarly to the development of serious games for depressive disorders, the evaluation of the implementation into practice is also often limited to therapeutic usage. Therefore, the main research question that is designed for this research is the following question:

> "*What are the opportunities and challenges for integrating the dialogue-based assessment method for adolescents into practice?*"

Besides this main research question, the following sub-research questions are also designed for this research:

- *What are the implementation improvements for the dialogue-based assessment?*

- *What are the strategies and considerations for implementing the dialogue-based assessment into practice?*

- *What are the limitations of questionnaire-based assessments in general?*

First, an extensive literature review provides information on depressive disorders, depressive disorder development, assessment of the disorder and the Patient Health Questionnaire. Furthermore, the related work focuses on serious games for mental health, including the in-game assessment of mental health disorders. The related work section also covers the use of NPC dialogues within games which this research is partially based on. The next section covers the methodology for the research and explains how the PHQ-8 is integrated into the dialogue component, what the design of the dialogue is, how it will be evaluated using semi-structured interviews and how the thematic analysis is performed. Afterwards, the results will be presented, including the identified themes from the thematic analysis. Finally, the discussion and conclusion can be found.

2 Background

This section includes background information about depressive disorders, including the classification used to establish a diagnosis, the symptoms, the development of a depressive disorder and the self-assessment questionnaire, the PHQ.

2.1 Depressive disorder and symptoms

The Diagnostic and Statistical Manual of Mental Disorders (DSM), developed in 1950 by the American Psychiatric Association (APA), is a classification system

developed to classify different mental disorders [9]. The DSM is developed to arrange human behaviours, experiences, perceptions or characteristics. A group of certain symptoms together are referred to as a syndrome or a disorder, for example, a depressive disorder [9]. The latest version of the DSM is the DSM-5 (2014), where each disorder has certain diagnostic criteria. These diagnostic criteria are characteristic symptoms of these disorders that need to be present in a person in order to establish a certain diagnosis and often also include the duration of the symptoms and their development. In total, the DSM-5 classifies twenty different main categories of disorders that each differentiate from each other. Each of these categories is divided into more specific disorders. Examples of these categories are bipolar-mood disorders, anxiety disorders, eating disorders and depressive-mood disorders [9].

Depressive disorder is a mood disorder and can be characterised by a constant feeling of sadness and/or an inability to experience pleasure. These feelings often interfere with daily functioning [75]. A depressive disorder is also known as Major Depressive Disorder (MDD) or clinical depression. MDD, or clinical depression, refers to individuals suffering from the symptoms at a higher level of intensity [30]. Besides that, sub-clinical depression is widely used to label those individuals who are in the early stage of depressive disorder where the depressive symptoms are not severe enough to meet the criteria of an MDD or clinical depression [49]. The term depressive disorder thus ranges in seriousness of the disorder from very mild symptoms to severe and persistent depressive disorder. The term that will be used throughout this research is depressive disorder, including mild, moderate and severe symptoms.

The DSM-5 can be used to determine which symptoms must be present in order to establish a diagnosis of depressive disorder [81]. It requires five or more symptoms to be present within a two-week time period. At least one of the symptoms should be a depressed mood or anhedonia, which is the loss of interest or pleasure. These are considered the primary symptoms of depressive disorder [80]. The secondary symptoms are [101]:

- Weight loss or weight gain without changing the diet, or a sudden change in appetite

- Sleep difficulties, including insomnia or hypersomnia

- Feeling of agitation (state of anxiety) or retardation (hampering of mental and physical activities), which can also be observed by others.

- Fatigue or a loss of energy

- Feelings of worthlessness or excessive guilt

- Diminished abilities to concentrate or think

- Re-occurring thoughts of death, suicidal thoughts or suicidal plans

In addition to these criteria, it is crucial that the person has not had any manic or hypomanic episodes in the past. Manic episodes refer to a person having an

irritable or exaggerated mood daily within a one-week period. The exaggerated mood can be defined by an increase in targeted activities or an increased level of energy [81]. A hypomanic episode is characterised by the same symptoms as a manic episode. However, the symptoms are present for a minimum of four days, and it does not necessarily interfere with daily functioning [81]. When a person has experienced either manic or hypomanic episodes in the past, the disorder will be referred to as a bipolar disorder rather than a depressive disorder [81].

2.2   Development of depressive disorders

Multiple studies have shown that there are differences in the development of depressive disorder between biological sex (female/male). However, it is crucial to notice that these studies have addressed biological sex as gender, while it is important to recognize that biological sex and gender identity should be distinguished. This is because gender identity encompasses men, women and non-binary individuals. The relationship between the development of a depressive disorder and biological sex and gender identity is complicated, as some effects might be explained by biological sex while others might come from gender identity. In this section, the terminology as is used in the studies will be adopted.

Essau et al. [29] has shown that twice as many females as males meet the criteria for depressive disorder at some time in their lives. This difference in gender usually emerges around puberty, between the age of 15 and 18. The differences might be explained by gender socialisation, social and hormonal mechanisms and stressful events that occur around puberty. These are factors that can relate to either biological sex or gender identity. The results of studies differ when it comes to the first occurrence of depressive disorder and re-occurrence. The study by Kovacs [57] showed that females are more likely to experience a depressive order for the first time than males. However, studies researching the re-occurrence of depressive disorders are inconsistent. The study by Amenson et al. [5] showed that there were no gender differences in the new incidence of depressive disorder for people without a history of depressive disorder. In comparison, results of the study by Lewinsohn et al. [67] did show that the probability of re-occurrence is greater for females than for males. In general, for people who experienced a depressive disorder, the chance for re-occurrence is 50%. After a second episode occurred, the chance for a third one happening is around 70-80% [41]. When a person experienced multiple depressive episodes, then the diagnosis of recurrent depressive disorder is given. In order to determine whether a person has a recurrent depressive disorder, there must be at least a two-month time period where the symptoms of the depressive disorder were absent [81].

2.3   Patient Health Questionnaire

For the assessment of mental health disorders, trained experts typically use interview procedures or questionnaires that follow standardised manuals. These

interview procedures that are used arise from the DSM-5 (section 2.1). These interview methods are also called the Structured Clinical Interview for DSM-5 Axis I Disorders (SCID-I). Trained clinicians use the SCID-I, and it allows the professional to go into more depth on the answers that a patient gives [81].

On the other hand, questionnaires are also often used as a first screening method. These screening methods often include self-assessment questionnaires. Self-assessment questionnaires can be beneficial because they provide a starting point for professionals [103]. An example of such a questionnaire is The Primary Care Evaluation of Mental Disorders (PRIME-MD). The PRIME-MD was developed in the 1990s and has the purpose of diagnosing five of the most common mental health disorders [96]. These disorders are depressive disorder, anxiety, somatoform (physical complaints without somatic cause), alcohol and eating disorders [61]. As the PRIME-MD took too long to complete during clinical visits, there was a need for a self-administered version. As a result, the Patient Health Questionnaire (PHQ) was designed and validated. There is a general PHQ, but also one specific for depressive disorders called the PHQ-9 [61]. This questionnaire is used in both clinical and research settings [59].

The PHQ-9 can be used for the screening, the diagnosis, the assessment of the severity of the symptoms and the monitoring of symptoms of depressive disorder. It consists of 9 items. Based on the answers to the PHQ-9, a score can be produced, addressing the severity of the depressive disorder. There are cutpoints of 5, 10, 15 and 20, which represent mild, moderate, moderately severe and severe levels of depressive symptoms. A person is diagnosed with a depressive disorder when 5 out of the 9 symptoms are present more than half of the days, when the 9th symptom (suicide) is present several days, and one of the first two symptoms (depressed mood or loss of interest) is present. The symptoms are derived from the DSM-IV (version before the DSM-5) and are defined as [60]:

1. Little interest or pleasure in doing things

2. Feeling down, depressed, or hopeless

3. Trouble falling or staying asleep, or sleeping too much

4. Feeling tired of having little energy

5. Poor appetite or overeating

6. Feeling bad about yourself, or that you are a failure or have let yourself or your family down.

7. Trouble concentrating on things, such as reading the newspaper or watching television

8. Moving or speaking so slowly that other people could have noticed. Or the opposite, being so fidgety or restless that you have been moving around a lot more than usual

9. Thoughts that you would be better off dead, or thoughts of hurting yourself

Besides the PHQ-9, there is also the PHQ-8, which eliminates item 9 about self-harm and suicide. This item is sometimes removed when assessing the presence and severity of depressive symptoms because of a lack of clarity about the 9th item of the PHQ-9. This is because this item might identify people who had passive thoughts about death or self-harm but never actually considered it [85]. Research has shown that the removal of the 9th item only has minimal effect on the overall scoring on the questionnaire [85]. In this research, the PHQ-8 will be used, and the full questionnaire can be found in Appendix A.

It is crucial to notice that screening questionnaires are often recommended to clinicians but that they are often used to pre-identify symptoms or monitor them. The questionnaire alone does not serve as a diagnostic tool. Evaluation of those who have positive screening results is still needed, but the screening questionnaire could aid the clinician in offering appropriate treatment. Moreover, the use of screening questionnaires could also allow a quicker diagnosis [66]. These methods have been shown to be effective when assessing depressive disorder.

However, the PHQ has also been shown to have limitations. One of these limitations is that this questionnaire is less accurate in populations with high comorbidity [27]. This term refers to the existence of multiple conditions at the same time. This can be problematic as anxiety and depressive disorder tend to be comorbid. This limitation is not unique to this questionnaire alone, but it is frequently seen for other self-assessment questionnaires as well [27].

3   Related work

This section will cover earlier work on serious games for mental health, the in-game assessment of mental health disorders, including serious games developed for treatment, and the use of NPC dialogues within games. My research will build upon the existing work as a dialogue-based assessment method will be developed to identify symptoms of depressive disorder in adolescents and evaluated, which could potentially be used as a component in a serious game in future work.

3.1   Serious games for mental health

As of 2020, 63.2% of the population has shown to have access to the Internet, and a majority of these people also have access to either a computer, a mobile phone or both. As a result of the increasing access to the Internet and computers, the number of people playing video games also increased to a number of 3.2 billion by 2021 [17]. Subsequently, researchers' interest towards the application of digital interventions for health, including mental health, has grown.

A topic that has been studied frequently is the concept of applied games for (mental) health, including serious games [32]. Serious games often have the goal of educating people or persuading them in an educational, health or other

manner. Serious games make use of gaming as a central medium [32]. Serious games were introduced more than a decade ago, and since then, their use and development have significantly grown [63]. Serious games are being applied in a variety of application areas, e.g. military, governmental, educational and healthcare areas [100]. The application of serious games within health care has significantly increased over the years. In 2016, the healthcare serious game market size was over 16 billion USD. In 2020, this market was already valued at over 25.3 billion USD, showing the extensive growth of this market [44]. An example of a serious game within health care is the game *Re-Mission* [8]. This serious game was developed with the purpose of educating young people with cancer about their treatment and involving them in the process [8]. In the past few years, researchers have also been exploring the potential use of serious games within mental health care due to the increasing prevalence and the low number of people who are diagnosed and treated. It is still in its infancy, and they are not widely adopted, but studies have shown that they bring along promising benefits for different aspects of mental health [68].

3.1.1   Benefits and applications of serious games

Serious games can be beneficial within mental health care because of their three potentials. First of all, they have an *appealing potential*. The use of serious games for a mental health intervention might increase its reach to people who do not want to seek help or who cannot access help. This is because games are popular and considered more fun than traditional treatments or diagnostic methods for mental health [32]. Moreover, serious games have an *engaging potential*. This is because games can be competitive, and people want to experience the storyline of a game. This might reduce attrition rates for mental health interventions due to the dynamics of a game [32, 39] . At last, they have an *effectiveness potential* because these serious games can include conventional and non-traditional processes for behaviour change and learning. Serious games offer more sensory environments that support engagement or learning and allow users to learn new skills in a reactive but safe environment [32]. Besides that, serious games also allow rehearsal of new behaviours or interventions [32]. Moreover, game-based mental health interventions also increase accessibility because these interventions can be delivered digitally. This could allow interventions to be more accessible to those who are in need of them [71]. This is especially important since many cases of people with depressive disorders are not diagnosed and are undetected [93]. Another advantage of serious games for mental health is that they might save costs and time in comparison to face-to-face sessions [6].

An advantage of digital interventions in general, specific to the assessment of mental health disorders with questionnaires, is that important symptoms are immediately highlighted for the practitioner. This means that when someone performs an online assessment, the practitioner can easily see whether this person has an elevated score for suicide, for example, [6]. Moreover, online completion of questionnaires also automatically generates scores, which saves time and reduces administrative costs. In line with this is that practitioners can also

monitor patient progress in real-time and proactively support the patient before any harmful consequences occur [6].

Furthermore, serious games are used for different aspects of mental health disorders. These aspects are [74]:

- *Awareness*; the aim of these applications is to remove the stigma around many mental health disorders and to educate people in recognising certain mental health conditions.

- *Prevention*; the aim of these applications is to improve the physiological well-being of people and to avoid the occurrence of mental health disorders.

- *Detection*; the aim of these applications is to identify mental health disorders by looking through game data and patterns.

- *Therapy*; the aim of these applications is the use of existing or new techniques for therapy to improve the mental health and well-being of people already diagnosed with a disorder.

The research by Martinez et al. [74] analysed these aspects of serious games for mental health, including depressive disorder and anxiety and discovered that the most common application of a serious game was prevention (44%), followed by therapy (32%). Awareness and detection are two applications that are infrequently used for the development of serious games, and only two games each were found (6%) in this study [74]. The study also identified the mental disorder these serious games focused on. The most common disorder was anxiety, with 26%, 18% focused on depressive disorders, and 56% addressed both disorders. However, this study did not identify any serious games for the detection of a depressive disorder. Even though the study by Martinez et al. [74] did not identify any serious games for the detection of depressive disorders, the literature review of my research does show that there are some existing games for the diagnosis or detection of depressive disorders. These will be further evaluated in Section 3.2.

Previous research has shown that there is a need for more digital interventions, including serious games for mental health care diagnosis or detection. First of all, this is because early diagnosis of the occurrence of a mental health disorder is as crucial as the treatment of it. This is because early diagnosis could prevent the formation of harmful effects [74]. Other research has shown that the use of these serious games for diagnosis also makes the process more standardised, stimulating and entertaining and contributes to better monitoring of patients in the long term [43]. However, as shown by Martinez et al. [74], research on the use of serious games for the diagnosis or detection of mental health disorders is very limited. Therefore, the potential benefits of serious games for diagnosis or detection emphasise that there is a need for more diagnosis or detection games for mental health disorders, including depressive disorders [74].

3.1.2   Serious games for mental health in the clinical setting

The potential of the use of serious games within the field of mental health has been explored, and studies have shown that there are benefits (see Section 3.1.1). However, little is still known about its impact and implementation within the clinical setting [63], and there are only a few validated serious games for mental health, despite the benefits [89]. Moreover, the acceptance and the potential inclusion of serious games among therapists and patients still remain unclear. The study by Eichenberg et al. [28], aimed to address this research gap by gaining an understanding of the opinions, preconditions and expectations of therapists and patients of using digital mental health applications.

First of all, the results showed that therapists had addressed challenges with digital mental health applications. One of these challenges is their potential cost or their training implications. Other challenges include legal, ethical and procedural considerations. Moreover, these internet-based applications for mental health are also relatively new, resulting in a lack of guidelines, standards and policies [28]. Studies have also shown that practitioners have less confidence in and are more sceptic towards digital mental health interventions. This reduced confidence is especially high amongst older practitioners who are less familiar with the Internet [90]. This is also considered a challenge for implementing these interventions into practice as practitioners are reluctant to refer their patients to digital interventions [6].

Practitioners addressed another challenge of using digital interventions for mental health, which is that there is no formation of an interpersonal relationship. This relationship is only formed during face-to-face assessment and therapy [28]. Moreover, another research addressed that practitioners prefer physical treatment as there is a lack of personal contact online [90]. A lack of personal contact during the diagnosis and treatment process could result in a lack of nonverbal signals and missing important disease aspects, according to the research by Schuster et al. [92]. When there is personal contact during treatment, interventions can also be adjusted based on the patient, which results in more personalized treatment [6].

Other challenges of using digital mental health interventions relate to the difficulty of checking the accuracy of responses and obtaining additional information [6]. A limitation specific to the online assessment of mental health disorders is that it is difficult to establish a diagnosis when the symptoms are complex. This all might result in a lower diagnostic accuracy [6].

Another important limitation of digital interventions is that there is a high prevalence of comorbidity for disorders like depressive disorder. Most of these interventions target only one specific disorder. Face-to-face treatment, on the other hand, allows practitioners to create tailored treatments where comorbidity can be targeted [6].

The research by Eichenberg et al. [28] also analysed the precautions for implementing digital mental health applications. Therapists valued easy accessibility and clinical evidence concerning the effectiveness of the serious game the most. The therapists also addressed anxiety disorders, affective disorders

(such as depressive disorder) and adjustment disorders (feelings of sadness) as the most suited psychological disorders for a serious game. Moreover, most of the therapists believed that the serious game would be the most effective for mild forms of a disorder. Additionally, the age group that therapists found the most fitting for a serious game was young adults, followed up by adolescents and adults. Regarding the possible application areas of serious games, both patients and therapists found that the serious game was best suitable as an addition during therapy, as a supportive measure after therapy, as a preventive method or as part of a consultation. Using serious games for therapeutic usage instead of therapy or using therapy through serious games before a consultation were both not favourable. Patients, on the other hand, valued the entertaining factor of the game the most [28].

The research by Graham et al. [38] similarly aimed to determine strategies for digital mental health intervention implementation. This research identified that several factors influence the adoption of digital interventions, including alignment between treatments, the likelihood of meeting patients' demands, the integration with the workflow of the practitioner and the costs for implementation. A strategy for meeting patients' demands is conducting an assessment of their needs with, for example, focus groups, surveys and discussions. The information derived from these assessments could result in the creation of more valuable interventions for the target group [38]. Graham et al. [38] also emphasize the importance of familiarity as it will lead to greater acceptance of the intervention [105]. This includes clearly explaining what the intervention is and why it is different from the traditional one. Besides these factors, practitioners also expressed the importance of security. When there are debates regarding privacy guidelines such as sharing identifiable health information of participants, practitioners are often reluctant [38].

The researchers state that a clearer understanding of the opinions of practitioners is still required [28]. The proposed studies in this section all focused on the evaluation of games for therapeutic usage. The literature review of my research, to the best of my knowledge, has not identified any evaluations on the use of serious games, or other digital interventions, for the diagnosis or detection of mental health disorders.

3.2 In-game assessment of mental health disorders

As mentioned in Section 3.1.1, most of the digital interventions or serious games that have been developed, are limited to therapeutic usage. However, the literature study of this research has identified several digital interventions or serious games aimed at the assessment of different mental health disorders. These in-game assessments will be further discussed in this section.

As explained in Section 2.3, the assessment of mental health disorders is typically performed with interview procedures such as the SCID-I, and sometimes standardized questionnaires are used additionally [81]. Researchers have been studying the use of games to assess mental health disorders as a replacement for or in addition to using these interview procedures or questionnaires because

of their previously addressed limitations, including self-report biases [50]. Reporting biases, for example, include the social desirability bias. This can be defined as the tendency of people to portray themselves based on perceived cultural and social norms [2]. Additional steps to collect information about the behaviours and symptoms of the patient, such as reports, are very time and resource intensive. Therefore, there has been a specific interest in behavioural correlates that are predictive of certain mental health disorders as a replacement for the currently used methods. If these behavioural correlates are known for mental health disorders, researchers can develop games to monitor responses and performances indicative of that disorder [72].

The research by Dechant et al. [50] has studied the use of these behavioural correlates to predict the presence of the mental health disorder social anxiety. In this research, digital biomarkers were used to assess behaviours in social situations which are behavioural responses measured through a digital device. These biomarkers are already being used frequently through the use of smartphones or social media [50]. Within the game, players had to walk past a non-player character (NPC) in a room. Players were also able to customise their avatars in the game, resulting in a higher prediction of social anxiety for that player. This is because customized characters lead to a higher identification with them. This can create a situation that the avatar is more socially present in an environment, and the person can get the illusion that the other characters are also real human beings. Due to this social presence, the players might feel that the situation is closer to reality, which could lead to social anxiety. Players could also see their avatar from the third-person perspective and could see their representation throughout the entire game, also during interaction around the NPC. This led to the fact that the players had a stronger perception of their performance in front of the observing NPC [50]. Besides behavioural measures for social anxiety, a subjective assessment of the disorder was also performed by making use of the Liebowitz Social Anxiety Scale [69]. Results of the study showed that socially anxious people stayed further away from the NPC and tried to avoid any interaction [50].

In the context of my research, social anxiety is often a behavioural predictor of worse outcomes of a depressive disorder. This is because social anxiety disorder and depressive disorder are often comorbid disorders, meaning that these disorders are often coexistent. Together they can result in greater functional impairment than one condition alone [26]. Individuals who are suffering from social anxiety disorder are much more likely to develop a depressive disorder than those not suffering from social anxiety disorder [26]. Social anxiety disorder almost always starts first, after which a depressive disorder usually develops a few years after. Therefore, social anxiety disorders are very relevant to study in association with a depressive disorder as they pose a possible risk factor for the development of the disorder [98].

A similar behavioural-based in-game assessment of a different mental health disorder is the game *The Delivery*, researched by Tsionas et al. [102]. *The Delivery* is a story-based video game developed to assess and diagnose depressive disorder by making use of Artificial Intelligence (AI) and Machine Learning

(ML) methods. Within the game, players were able to alter certain scenarios with their actions and conversations with NPCs and had to deal with unexpected turns of events [102]. The game gathers data regarding the cognitive behaviour of the players during the gameplay. The player could also interact with the surroundings and objects in the environment. The choices and actions that are given to the player are based on the criteria from the DSM-5. The conversation with the NPCs is implemented in such a way that different answers signal different probabilities of the player meeting specific criteria of depressive disorder [102]. The final score of the game is also important because it indicates the severity of the depressive disorder. This score changes when players interact with NPCs, where every answer has a different weight relative to DSM-5 criteria. Interaction with the environment also increases the final score. Moreover, showing sensitivity to characters and the situation alters the final score in favour of a diagnosis of no depressive disorder [102]. After gameplay, participants were labelled as depressive disorder-positive or depressive disorder-negative. The results showed promising results for the distinction between depressive disorder positives and negatives. However, the authors emphasize the importance of further research regarding other, new and more efficient techniques for the diagnosis of depressive disorder. This originates from the lack of work on the subject of depressive disorder diagnosis within the field of gameplay [102].

Even though the research by Dechant et al. [50] and Tsionas et al. [102] showed promising results regarding the use of in-game behavioural assessment methods for mental health disorders, there are concerns. One of the concerns is the consideration of the privacy of the patient and ensuring a safe space to play the game in. Game developers also need to be aware of potential consequences for players after reporting mental health predictions. Moreover, an important factor that needs to be taken into account with in-game assessment is that players might behave differently within a gaming environment than they would in the physical world. This could potentially bias the results of the assessment [50].

A completely different approach was used in the research by Khamwang et al. [55]. This research did not approach the assessment of a mental health disorder from a behavioural approach but proposed a simulation game to help patients determine the severity of their depressive disorder. In this simulation game, the player was immersed in a situation with a certain dialogue. After each conversation, the game asked the player if they had experienced a similar situation. The player could recall this experience and explain it or share another experience or story if the player has not had a similar experience. This had the purpose of analysing the mindset of the player. After the dialogues, the players completed the PHQ-9 questionnaire to determine the severity of the depressive disorder. The goal of this was to give the patients advice for further treatment of their disorder. After the game was designed, an evaluation of the game was performed with both a psychiatrist and sample users. The psychiatrist expressed that the game is a creative tool and that it is suitable to use symbolically to represent depressive disorder. However, the psychiatrist also expressed that the game might be difficult to use due to a lack of motivation for people who are

already at a further stage of their depressive disorder. The users who evaluated the game explained that it was important to add more complex stories, clear instructions and more interaction with the game [55].

The research by Dechant et al. [50], Tsionas et al. [102] and Khamwang et al. [55], relate to my research in a way in which each of these studies aimed to develop a method for in-game assessment of a mental health disorder. Moreover, the study by Tsionas et al. [102] distinguished depressive disorder positives from negatives by making use of dialogues to change the gameplay and to measure the final score. This is similarly done in the research by Khamwang et al. [55] where dialogues are also used but then to let the player recall whether it has had the same experience before. In my research, dialogues will also be included to identify symptoms of a depressive disorder. Moreover, similar to the research by Khamwang et al.[55], the player was also asked to address whether they have ever experienced feelings of a depressive disorder. The difference between the research by Tsionas et al. [102] and my research is that the dialogue options will not directly be based on the DSM-V but rather on the PHQ items. Moreover, the goal of my research is to identify symptoms of a depressive disorder rather than distinguish the positive ones from the negative ones and establish a diagnosis. The difference between the research by Khamwang et al. [55] and my research is that the existing research had a clear distinction between playing a game and the actual assessment of the depressive disorder by making use of the PHQ-9 after playing the game. The only purpose of playing the game and reading the stories was to analyse the mindset of the player. This will be different in the design of the dialogue in my research. My research will include the PHQ within the dialogue. This will potentially identify symptoms of a depressive disorder similar to what the research of Khamwang et al. [55] aimed to do. Furthermore, what is important to note is that none of the games covered in this section, despite being evaluated or validated, is being used in (clinical) practice.

3.2.1 A serious game for mental health disorders used in practice

As mentioned in the previous Section 3.2, different games for assessing mental health disorders have been developed and evaluated. However, none of these games are actually being used in practice. This section will cover a serious game that has been used in (clinical) practice.

One serious game that is being used in practice is the game SPARX. This game is developed to use at home with family, at schools or in healthcare settings. SPARX is an interactive fantasy game aimed at adolescents to help them with their depressive disorder [76]. SPARX stands for Smart, Positive, Active, Realistic, X-factor thoughts and is a computerised cognitive behavioural therapy (CBT). CBT is a widely used and recommended treatment for people suffering from mild and moderate symptoms of depressive disorder. This treatment specifically focuses on the relationship between thoughts, emotions and behaviour [31]. Computerised CBT interventions have shown to be effective, but adherence to the treatment has been a problem in the past [18]. Therefore,

17

the researchers of SPARX have tried to focus on this specific aspect of the intervention. They wanted to ensure that SPARX was appealing for young people and that it would also be enjoyable, to ensure high adherence rates. Within the game, the player had to undertake challenges to restore the balance in the fantasy world that was dominated by Gloomy Negative Automatic Thoughts (GNATs). During the game, there was also a guide to put the game in context, provide education, gauge moods and monitor real-life challenges. The results of the program showed that using SPARX resulted in a significant reduction in depressive symptoms. It has shown to be at least as effective as traditional face-to-face treatment [76]. The game SPARX was especially a success in New Zealand [76].

A potential reason why the serious game SPARX is being used in practice while other validated games are not is that SPARX has been widely studied and evaluated. This has resulted in a substantial body of evidence of its effectiveness which could have contributed to the successful implementation into practice.

3.3   The use of NPC dialogues within games

The collection of data is important for the development of a game. This information entails data about, e.g. player behaviour and interactions. In order to collect this data, self-report questionnaires are often used. However, the problem with these questionnaires is that they tend to interrupt gameplay when included during the game [11]. When a person is disturbed during gameplay, this influences the execution of tasks and their motivation to complete the game. Game developers prefer to include these questionnaires during gameplay because afterwards, the data is reported retrospectively. Another disadvantage of including questionnaires during a game is that they bring along other issues, such as guessing behaviour. In the specific context of serious games, that have the goal of educating players or to persuade them in different manners, this process is disturbed when interruption takes place due to the use of questionnaires [35]. As a result of the found disadvantages of questionnaires during gameplay, researchers have been looking at alternatives for assessment within games.

The research by Frommel et al. [34] analysed the use of in-game dialogues in order to administer self-reports. In this research, dialogue-based self-reports (DBSR) were used to detect the emotions of the players. A DBSR is a self-report mapped to a dialogue with an NPC. Here, the NPC asked specific questions based on the self-reports and possible responses to this self-report are displayed as dialogue options. The results of this study showed that DBSR provided similar results as using the original questionnaire. This implies that this method could be used to gather self-report data from the player as a replacement for questionnaires. However, the results did not show that the DBSR led to a better player experience than a traditional questionnaire. A potential reason for this could be due to the repetitive dialogues. Therefore, it is concluded that dialogues to assess self-report data can be beneficial to an extent where it is not applied too often [34]. Another important factor to take into account is that the participants in this study might have recognized the original questionnaire

in the DBSR as this method was still close to the questionnaire itself [34].

The research by Frommel et al. [34] concluded that dialogues with NPCs could be used to measure different player states, such as their emotion during gameplay. The research by Bowey et al. [11] assessed whether these NPC dialogues could also be used for more stable constructs like players' beliefs. In this research, a validated questionnaire used to measure sexual beliefs was integrated into a dialogue within a game. This questionnaire was the Male Role Attitudes Scale (MRAS). The MRAS is an eight-item measure that measures attitudes regarding male gender roles [82]. Similarly to depressive disorders, sexist beliefs were modelled because this belief is challenging to assess within game dialogues due to social desirability. Social desirability plays a role because people tend to respond in dialogue responses in such a way that they are not predictive of their actual beliefs. This especially occurs when the beliefs of this person are unpopular such as racist beliefs [11]. Another aspect that is studied in the research by Bowey et al. [11] is the fact that players enter a magic circle when playing a game. This would mean that players might respond to a dialogue as how their character would respond, instead of how the player itself would respond. However, previous research has shown that players might unconsciously align their character responses to their own. This is because, for player identification with the character to be achieved, alignment between the payer's beliefs and the beliefs displayed by a game's character is needed [10]. This would mean that players will respond to the dialogues in a way which they reflect their actual traits. This is in line with the *Media Equation* [86] that explains that people interact with computers and other virtual entities in a way in which they would also interact with other people. However, research also emphasises that dialogue choices in games might affect the experience. This is because when there is an obvious choice in dialogue, people might not respond according to how they would actually do [64]. The study by Bowey et al. [11] aims to analyse different styles of dialogue framing and their effects on response behaviour and player experience.

For the narrative of the game, Bowey et al. [11] based their story on that of previous research [10]. In this story, a Kingdom's Royal Heir was captured by a wizard and had to be rescued. The player encounters obstacles, opponents and dilemmas in order to rescue the Royal Heir. In the research by Bowey et al. [11], the player is the Knight who has a conversation with the King, who is interviewing the player about rescuing the heir. The questions that the King asks are based on the MRAS, and the Knight could agree or disagree with things the King is saying [11].

Two different methods were used to convert the existing questionnaire into a dialogue, *direct* and *indirect* approaches. For *direct* prompting, you provide the player with the actual scale item, and you give the option to agree or disagree with it. *Indirect* prompting means giving the player a related statement where you can respond with the original item, and the other response is the opposite of the original item [11]. The choices that players made in the dialogue were eventually used to predict their beliefs.

Prior to the experiment with the dialogue, participants were also asked to fill

in the original MRAS. Results showed that the dialogue model accurately and significantly predicted the real-world beliefs of the players after comparing the results of the dialogue with the original MRAS [11]. This predictability did not differ between the direct and indirect groups, showing that there is no evidence that one of the approaches worked better than the other within the context of this research [11]. The results also showed that most of the participants responded to the dialogue as themselves and not as their imaginary characters, which is in line with previous research [10], [86].

My research is partially based on the results and method from the research by Bowey et al. [11] and the results from the study by Frommel et al. [34]. In my research, the PHQ-8 will be included within a dialogue to identify depressive disorder symptoms and their severity in adolescents, as a replacement for the traditional on-paper questionnaire. This is done similarly in the research by Bowey et al. [11] where the MRAS is included within a dialogue. However, a different approach to convert the items of the PHQ-8 will be used than for the MRAS. Bowey et al. [11] used direct and indirect approaches. However, in my research, identification with the NPC's emotions, feelings and experiences will be used to identify the symptoms of a depressive disorder in the players.

4 Dialogue-based assessment method

The goal of this research was to develop a dialogue-based assessment method for identifying symptoms of a depressive disorder in adolescents. This method could potentially be used as a serious game component. The dialogue-based assessment method was based on the PHQ-8, adjusted for adolescents, for the assessment of a depressive disorder. The PHQ-8 items were translated into a dialogue where the player could potentially be able to identify itself with the NPC and address symptoms of a depressive disorder.

4.1 Target group

The dialogue-based assessment method was specifically developed as an additional diagnostic tool for adolescents, ranging from the age of 12 to 18 years old [79]. This target group was chosen as the PHQ-8 is designed for adolescents aged 12 to 18 years old. There is also a version developed for adults. Moreover, as defined in Section 2.2, the chance for the development of a depressive disorder is the highest during adolescence. According to Hankin et al., [40], 28% of girls and 14% of boys expect a depressive episode before the end of adolescence. Moreover, Horgan and Sweeney [42] discovered that 30.8% of adolescents use the Internet to search for mental health information, and 68% indicated that they would use the Internet if they needed to. Therefore, digital mental health interventions are considered to have the potential for supporting mental health among the group of adolescents [65]. However, a majority of young people also still prefer face-to-face support if available. Research has also shown that young children and adolescents with mental health difficulties are significantly stigma-

tized, forming an important barrier to help-seeking [54]. Besides that, games are the most frequent source of entertainment for adolescents. In the United States alone, 91% of all children between the age of 2 and 17 play games at least one hour a day. This shows that games could be a suitable way to reach adolescents that are in need of mental health support [23]. The study by Eichenberg et al. [28] also identified that psychologists believe that the use of serious games for mental health disorders is the most suitable for children and adolescents (see Section 3.1.2).

4.2   Theme

It is crucial to engage the players of the dialogue by creating a compelling theme, which is considered a primary decision in designing a game [46]. Many different themes are already being used in the development of serious games for different purposes. The research by Bowey et al. [10] developed a theme where players had to rescue the Kingdom's Royal Heir from a dark wizard. This theme was reused in different research where sexist beliefs were predicted by making use of NPC dialogues [11]. The successful game SPARX also included a similar adventurous theme as the previously mentioned research. Here, the player had to undergo challenges in a fantasy world [76] (see Section 3.2.1). According to Reuter et al. [87], adventure themes like the one in the research by [10, 11, 76], offer a way to present learning content in a structured and guided way.

In my research, an adventurous theme was also chosen for the dialogue, similar to the themes in previous research. In the dialogue, the player could choose to explore different parts of a town called *Roaring River*. While exploring this town, the player was guided by a guide that told the player about the town. Throughout the dialogue, the player was able to make decisions to help the guide or to change the storyline of the dialogue based on the player's preferences, which could impact the outcome of the dialogue.

I have developed the dialogue-based assessment method by making use of Unity and its Fungus library. All the backgrounds in the dialogue were created by making use of the Artificial Intelligence (AI) text-to-image generator from Canva. I have drawn the characters in the dialogue-based assessment method myself. Examples of scenes in the dialogue are seen in Figure 1.

4.3   Narrative

Besides the use of a compelling theme, a strong narrative could contribute to the players becoming immersed in the game. The narrative is an important component when it comes to providing meaning to actions and building players' commitment to the game [52]. The narrative for the dialogue was that the player was visiting the town of *Roaring River*. Before exploring what the town had to bring, the player was able to choose which parts they wanted to explore. They had the option to explore the trails around the town, visit the popular markets or go to the river to go fishing. Different options were provided to the players so that the narrative would fit different interests, and it would allow

(a) Example of PHQ item
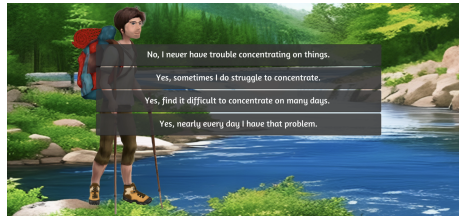
(b) Example of scale item

Figure 1: *Examples of a PHQ item and scale in the dialogue*

the players to play the dialogue again, where they explore different parts of the town. After choosing which part the player wanted to explore, they were able to choose a guide. The player could choose between three different guides based on their preference: *Scout*, *Hawk* and *Pixie* (see Figure 2). These guides are all humanoid game characters with different gender identities (men, women). After choosing a guide, the player could get to know their guide on their journey.

The aim of this research was to primarily assess the initial opinions and impressions of the dialogue-based assessment method. Therefore, only one storyline was developed and utilized during the evaluation process. This storyline was the narrative where the player could choose to catch some rare fish species at the river. For this specific storyline, only two guides have been developed as well which are *Scout* (man) and *Pixie* (woman). In this storyline, the player and the guide went to the market to buy food for their trip and hike to the river. During the hike to the river, the guide told the player about the environment and they also met another guide to have lunch and talk. Finally, the player and guide reached the river where they went fishing. The complete dialogue script can be found in Appendix C.

4.4 Patient Health Questionnaire-8

In this research, the PHQ-8 is translated into a dialogue-based assessment method. Here, the aim is to identify symptoms of a depressive disorder in adolescents based on the identification with the feelings of the character, in order to let the player identify their own feelings. The guide, as introduced in

(a) Scout   (b) Pixie

Figure 2: *Guides in the dialogue*

Section 4.2, mentioned the items of the PHQ-8 within the dialogue and mentioned that it was having difficulties with these specific symptoms. The guide expressed its feelings to the player. The guide would ask the player whether they were ever feeling the same. An example of a question asked by the guide is: "*Lately, I have been feeling very fidgety and restless. I just cannot stand still or focus on what I am doing. Are you ever feeling restless or fidgety?*" from the original PHQ-8 item "*… Or the opposite, being so fidgety or restless that you have been moving around a lot more than usual.*" (see Appendix A). This original item has been split into two items to translate the PHQ-8 item into the dialogue. An example of how the player could respond to this is:

- No, I never feel restless or fidgety.

- Yes, sometimes I do.

- Yes, I am experiencing that on many days.

- Yes, nearly every day.

In the original PHQ-8, people could indicate whether they experience the symptoms "*not at all*", "*several days*", "*more than half of the days*", or "*nearly every day*". Based on the answers of the player, a set of symptoms of the depressive disorder could be retrieved, including their severity based on the player addressing how often they experience these symptoms. The PHQ-8 itself can be found in Appendix A. The translation of how the PHQ-8 items are translated into dialogue items can be found in Appendix B.

4.5   Use of the dialogue-based assessment method

The goal of the dialogue-based assessment method is to identify symptoms of a depressive disorder in adolescents. As also described in Section 2.3, it is

important to notice that also this method, similar to the traditional questionnaires, does not serve as a diagnostic tool [66]. Also, for the dialogue, additional evaluation of the symptoms and severity is needed to establish a diagnosis of a depressive disorder. However, the dialogue could be used by practitioners in face-to-face sessions to focus on the pre-identified symptoms. This could potentially make the diagnosis process more standardised, stimulating and entertaining.

5 Evaluation

This dialogue-based assessment method was evaluated to determine the opportunities and challenges, strategies and considerations, implementation improvements and general limitations of integrating this method into practice. The method was evaluated by conducting interviews with mental health care professionals. This included psychologists, researchers and university lecturers specialized in mental health and psychology. As described in section 3.1.2, mental health care professionals have already addressed challenges about internet-based mental health applications. This includes legal, ethical and procedural considerations and other limitations. However, current research focuses on the evaluation of serious games for therapeutic usage. Therefore, this research has been conducted to develop a better understanding of how a dialogue-based assessment method could be integrated into practice and what the specific challenges for integration are for such an assessment method. This collaboration with professionals is crucial to increase the number of serious game or internet-based interventions in practice, especially for the diagnosis of a depressive disorder.

5.1 Data collection and procedure

Data for the research was collected through in-depth 45-60 minutes, online semi-structured interviews with mental health care professionals from both clinical and research backgrounds. Participant eligibility was restricted to those involved in the diagnosis or treatment of depressive disorders or those who have researched or developed digital interventions for mental health disorders, including depressive disorders.

The procedure of conducting the interviews consisted of several steps. First, participants received an email with an information sheet about the research (Appendix D). Here, participants were made aware that the interviews were going to be audio-recorded. The information sheet also included information about the background and the goal of the research, the researchers, how the research was conducted, what happened to participants' data and the rights of the participants. This email also included the informed consent form with a survey (Appendix E). In this form, participants were, amongst others, asked to consent to the interviews being audio-recorded. Additionally, a short survey was sent to the participants to gather information about their gender identity, age, educational background, profession and years of experience.

After participants completed the informed consent form and gave consent to participate in this research, they were invited for an interview that was audio-recorded. In this interview, several topics were discussed with the participants. These topics can be found in Table 1. The interview protocol, including the main questions that were asked during each interview, can be found in Appendix G.

Topic	Sub-topics
Diagnosis	Experience with diagnosis, steps taken in the process, use of standardized questionnaires and challenges for diagnosis
Digital therapies and diagnostics	Experience with digital therapies and diagnostics, integration into the treatment plan and barriers and challenges digital therapies
Evaluation of the dialogue	general opinion about the dialogue, addition to the current diagnosis and mitigation of existing challenges in the diagnosis
Challenges and integration	Challenges for integration, challenges for identification of symptoms, precautions for implementation and benefits of the dialogue
Suggestions and improvements	Suggestions and improvements of the dialogue, suggestions for integration

Table 1: *Interview guide*

After discussing the current methods of diagnosis and participants' experience with digital therapies, the dialogue-based assessment method was evaluated with the participants during the interview. Here, the dialogue was partially shown to the participants. Due to time constraints, not the entire dialogue could be shown. The beginning of the dialogue, where a short introduction about the town was given, was demonstrated. Moreover, the different storylines and characters that could be chosen were shown to the participants. Finally, I showed an example of how an item of the PHQ-8 was translated into the dialogue. The specific example that I showed was the original item "*Trouble concentrating on things such as school work, reading or watching television.*". Here, the scene was demonstrated where the NPC mentions "*I cannot concentrate on what to do now. Maybe you can just give it a try... Maybe it is just me who cannot concentrate anymore. Do you ever have trouble concentrating on things?*". However, if participants still had questions about the dialogue, other scenes where the NPC asks questions to the players were also shown.

Due to time constraints and the fact that only one researcher was involved in this research, data collection stopped after six interviews and not the moment when no new codes or themes emerged from the data. However, in order to achieve data saturation, the interview questions for each interview were structured in a way in which each participant was asked similar questions about similar topics (see Table 1) [36]. Moreover, as recommended by previous studies, a more diverse set of participants were asked to contribute to this research to

prevent someone with specialized information on the topic from overshadowing the data [36].

5.2 Thematic analysis

After all the interviews were performed, the data from the interviews were analysed by using a thematic analysis. In the first step of the analysis, before starting the code generation and theme generation process of the qualitative analysis, the interviews were transcribed in the language of the interview itself (Dutch) and were prepared for the analysis. Here, the audio recordings from the interviews were used to create transcripts. This process also included the anonymization of the data where personally identifiable information was removed, such as the names of people or specific names of organizations where the interviewees worked. Afterwards, the step of data familiarization was performed as a preliminary analysis. This involved reading the transcripts multiple times and making initial notes about the data. This was considered an important step in understanding the general context of the interviews.

After the anonymization and data familiarization, inductive thematic analysis was performed by making use of NVivo 12. The thematic analysis was performed on the original Dutch transcriptions. However, the codes generated through the process were in English. A thematic analysis was used to identify, analyse and interpret certain patterns of meaning, which are known as themes, within qualitative data [21]. The thematic analysis was inductive, which indicates that the analysis was data-driven rather than theory-driven [21]. This means that the process of coding the data did not include categorizing the codes based on a preexisting coding frame, but it was rather driven by the content of the data itself [12]. The codes and the eventual themes were thus derived from the content of the data. However, it is important to note that the coding was not completely inductive but also consisted of a certain level of deductive coding. This was done to ensure that all the potentially relevant data was coded and captured. In a deductive coding approach, a set of ideas, concepts and topics are used to code and interpret the data [12]. However, the inductive approach was predominant in this research. For the review of the data and the coding generation process, only one researcher was involved.

The set of codes was manually developed following a systematic and iterative approach. The coding process involved multiple stages of analysis. Each of the interviews were first individually coded by making use of an in vivo, line-by-line coding approach. This coding technique ensures that you can place an emphasis on the exact spoken words of the participant and allows for a more sophisticated analysis and also helps to familiarize with the data in an inductive nature [73]. During this coding process, a codebook was developed and maintained to ensure that the analysis was consistent throughout all the separate interviews. This codebook consisted of the in vivo code name and a clear definition of this code. Examples of these initial in vivo codes are "*You cannot tailor questionnaires to each client*", "*There is really just a need for a conversation*" and "*That I can just embed it with a code*". The initial round of in vivo coding resulted in 930

individual codes.

After the first phase of the coding process, the initial in vivo codes were categorized into higher-level categories such as "*Advantages of the dialogue*" and "*Disadvantages of the dialogue*". This ensured a clearer overview of the existing codes and made the process of organizing the codes into lower-level categories more efficient. After the initial in vivo codes were categorized into these higher-level categories, the process of code merging and parent code generation started. Codes that were similar to each other and that had the same meaning were merged together. This ensured a clearer representation of the data. Codes that could not be merged but captured the same subject were categorized under parent codes. An example is "*Include suicidality in the dialogue*". This parent code included all the codes as to why suicidality should be added to the dialogue or how suicidality could be approached in the dialogue. Examples of these child codes are "*To leave out suicidality, you miss a very important diagnosis part*" and "*Suicidality is to be spoken for*". After these two rounds of categorization, the number of codes (excluding child codes) resulted in 51 codes. After the process of merging codes and categorizing codes into parent codes, a mind map was created, including all 51 lower-level codes. The relationship and overarching topics of these codes were examined to develop different themes. Lower-level codes related to the advantages of the dialogue were, for example, grouped together, while specific suggestions for the dialogue were also grouped together. In the end, the inductive thematic analysis resulted in the generation of 5 final themes, including the 65 (parent) codes.

5.3   Participants

Participants for this research were recruited through the personal network and Universities. For this research, a total of six participants were interviewed, of which five identified as women and one identified as men, ranging from 24-64 years old (M=39.17, SD= 14.34). Overall, the participants have been practising their profession ranging from 1-40 years (M=12.64, SD=13.97). In total, 1 participant worked with or had experience with working with adults in mental health care. 2 participants worked with both adults and older adults and finally, 3 participants had experience with adolescents and children in mental health. An overview of each participant, their educational background, profession and experience with digital interventions can be found in Table 2. In Table 2, two terms require extra clarification. These terms are GZ psychologist and GGZ. A GZ psychologist is a person who is specialized in the diagnostics and treatment of multiple psychological disorders such as depressive disorder, anxiety but also addiction [104]. Moreover, the GGZ is a Dutch organization for mental health and addiction care.

Participant number	Educational Background	Profession	Experience with digital interventions
1	Pedagogical Sciences, Educational Sciences and Behavioural Science	University lecturer and Assistant Professor	Builds and tests interactive experiences, including games, for children's mental health
2	Neuro- and rehabilitation psychology	GZ psychologist	Worked with eHealth modules and VR for mental health disorders and phobias
3	Psychology and Sport- and performance psychology	Project leader in the GGZ and Entrepreneur of a sports psychology agency	Builds e-learning platforms and digital interventions for mental health
4	Remedial education	Senior Researcher and Clinical psychologist	Researched computerised and blended therapy for anxiety and depressive disorder, not much experience in practice
5	Psychology	University lecturer and GZ psychologist	Researched computerised and blended therapy for depressive disorder and used digital therapies in practice
6	Psychology	Psychologist	No experience with digital interventions

Table 2: *Participants' background, profession and experience*

5.4 Ethics and Privacy

The Ethics and Privacy Quick Scan of the Utrecht University Research Insititute of Information and Computing Sciences was conducted, and this study was classified as low-risk. No fuller ethics review or privacy assessment was required. The full results of the Ethics and Privacy Quick Scan can be found in Appendix F.

6 Results

In this research, where the opportunities and challenges for integrating the dialogue-based assessment method into practice are evaluated, the following five themes have been identified:

1. Advantages of the dialogue-based assessment over traditional paper questionnaires

2. Challenges for integrating the dialogue-based assessment into practice

3. Strategies and considerations for implementing the dialogue-based assessment into practice

4. Implementation improvements for the dialogue-based assessment

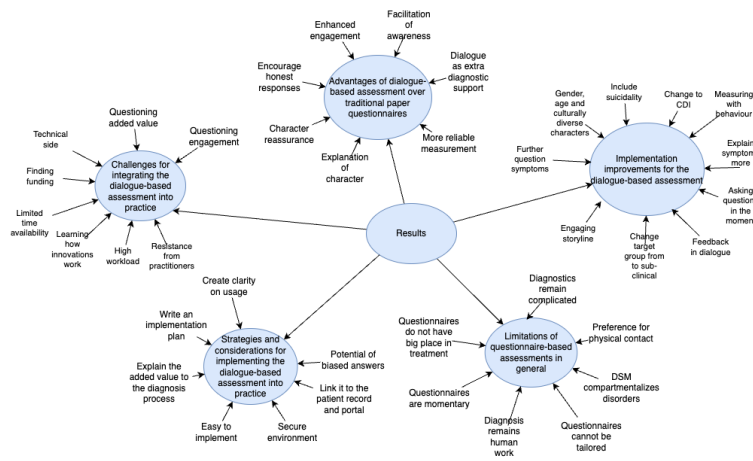5. Limitations of questionnaire-based assessment in general

Figure 3: *Mindmap of the results of the thematic analysis*

These themes are not mutually exclusive and exhaustive (MECE) as key elements, e.g. related to comorbidity and the heterogeneous nature of depressive disorders, could be coded under "*Challenges for integrating the dialogue-based*

assessment into practice" and "*Limitations of questionnaire-based assessment in general*". The themes will be further explained below and will be supported by quotes. Quotes will be labelled by participant number and profession (e.g. P1; University lecturer and Assistant Professor). Information about the participants and professions can be found in Table 2 (Section 5.3). The participants' quotes are translated from the original language of the interviews, which was Dutch, to English. All the themes and their key elements are presented in Figure 3.

6.1 Advantages of the dialogue-based assessment over traditional paper questionnaires

The first theme focuses on the advantages of the dialogue-based assessment method over traditional on-paper questionnaires. It highlights the benefits of this method and explores how the dialogue mitigates certain disadvantages or limitations of on-paper questionnaires.

Participants have addressed several challenges regarding traditional on-paper questionnaires. First of all, traditional questionnaires are often perceived as difficult by people filling them in on various aspects. One of these aspects is the difficulty of understanding the questions. Participants have mentioned that the questions are often unclear or formulated in a difficult way; "*...what is often the problem with questionnaires is that people have little clarity with what is meant by the questions.*" (P3; Project leader at GGZ). People find it especially difficult to understand what the symptoms mean and how these manifest in their daily lives; "*...what do you mean with concentration? What does it all have to do with?*" (P2; GZ Psychologist). This makes it difficult for these people to fill in the right answer to the questions. The consequence of this perceived difficulty is that wrong conclusions could be drawn from the results;

> "*I often find with the clinical lists, that if you just don't understand something and fill in something and therefore get the wrong conclusion, I also think it's, well not harmful, that's too big of a word.*" (P3; Project leader at GGZ)

Furthermore, participants have commented that people often misinterpret the questions in on-paper questionnaires; "*I think the disadvantage of those self-tests, ..., is that they are clinical. I can also get very restless or misinterpret a question.*" (P3; Project leader at GGZ). Another participant addressed that people also often misinterpret the scales within the questionnaires;

> "*...if it only says "I never have this, "sometimes", "regularly" or "often", then everyone has a completely different interpretation of "what is regular?"*" (P6, Psychologist)

Together, this can lead to unreliable results.

Moreover, participants have mentioned that people often also have difficulty estimating the severity of their symptoms when filling out questionnaires. Comparing symptoms to those of others is perceived as necessary in order to determine the severity of symptoms; "*...you always have to put it next to another*

and you have to compare it..."(P5; University lecturer and GZ Psychologist). Consequently, participants have addressed that determining the actual severity of the symptoms through questionnaires becomes challenging. The severity of symptoms is something that can only be estimated during a conversation;

> "*Also the severity of not feeling like something, ..., how extensive? And on what levels? In what areas of life does this all play out? You often don't get that from answering a questionnaire but you do in a conversation.*" (P2, GZ Psychologist)

Finally, another limitation of on-paper questionnaires that participants discussed is "*response bias*" which is the tendency of people to respond falsely to questions. Participants mentioned that people tend to downplay their symptoms; "*..also downplay it a bit like it's not too bad*" (P6, Psychologist). People downplay their symptoms either out of shame or due to social desirability; "*...young people can report under, so that is indeed a bit of a disguise out of shame..*" (P5; University lecturer and GZ Psychologist) and "*The fact that they answered the questionnaire in a socially desirable manner can, of course, play a major role in this...*" (P6, Psychologist). Downplaying symptoms especially occur for the item about suicidality, as addressed by a participant; "*...on that suicide item, for example. Young people often do not fill it in correctly because they think, oh that goes to my parents.*" (P4; Senior Researcher and Clinical Psychologist). In addition, participants acknowledged that people also sometimes fill in the questionnaires for fun or not seriously;

> "*We also screen young people a lot at schools, ..., and there you sometimes see over-reporting. So that they fill it in for fun and very much everything on very extreme.*" (P4; Senior Researcher and Clinical Psychologist)

Participants mentioned that the dialogue-based assessment method can potentially mitigate these limitations of traditional on-paper questionnaires. First of all, one of the advantages is the "*explanation that the character gives*" about the symptoms and the story told by the character, which can offer a clearer picture of symptoms and improve comprehension; "*..., I think it will give a much clearer picture because of the explanation.*" (P3; Project leader at GGZ). Another participant mentioned that the dialogue and the explanation can help address issues related to misinterpretation of questions; "*...those questions may sometimes not be well received. And with that explanation, it's probably better.*" (P4; Senior Researcher and Clinical Psychologist).

Furthermore, the dialogue-based assessment method can potentially address the reporting biases that occur with traditional on-paper questionnaires, by creating a safe atmosphere to encourage honest responses due to the reassurance that the character can give to the player; "*...especially to give the atmosphere of it's okay, it can happen to anyone, but be honest about it because then you also get better help.*" (P2, GZ Psychologist). Moreover, honest responses are also encouraged due to the format of the dialogue being less clinical than the traditional questionnaire;

> "…you interact more honestly in this format I think because there is a difference with whether you are going to fill out a game or a paper with the clinical list that says "the Patient Health Questionnaire-8" above it" (P3; Project leader at GGZ)

Moreover, participants have also referred to the dialogue as a "*concept of awareness*"; "*…a first step to get young people to think about what actually belongs to a depression and what is that like for me?*" (P2; GZ Psychologist) and "*…makes many people a little bit aware of things at the same time.*" (P3; Project leader at GGZ). One of the reasons why the dialogue is a concept of awareness is because you have to think for someone else; "*…that it makes people think more anyway because they have to think for other people.*" (P3; Project leader at GGZ). A more elaborate explanation of this concept is that people might be more open or think along more with the character in the dialogue;

> "*You get an example, you are in a conversation and people are very empathetic, even when you watch a film you are automatically inclined to think along or to be open or if someone says something violent in the film, you can think actually I also have that in a certain way.*". (P3; Project leader at GGZ)

On the whole, participants have addressed that this dialogue-based assessment method provides a more reliable measurement of the symptoms of depressive disorder. This is because, within the dialogue, there would be less social desirability bias; "*…that social desirability is somewhat less.*" (P6, Psychologist). Moreover, another factor that contributes to a more reliable measurement is the fact that players might be less concerned about the desired answer to a question; "*…you are also less concerned with what is a desired answer to this question.*" (P6, Psychologist). Another participant addressed that with the dialogue you get a more reliable measurement because the players might not fill in the dialogue randomly or quickly, but are really forced to think about their complaints;

> "*I also think that you get a more reliable measurement. So, what I just said, you don't get an "I don't feel like it so I just fill in something", you are really forced to think about it for a while.*" (P5; University lecturer and GZ Psychologist)

The "*concept of awareness*" also ensures that players are able to provide a better picture of their actual complaints, also resulting in more reliable measurements; "*…then you also give the young people more opportunity to think about this a bit more and thus give a better picture of their real complaints.*" (P2, GZ Psychologist). Besides this "*concept of awareness*", the "*explanation of the character*" can also contribute to a more reliable measurement. This is because the character provides a story around the symptoms and not only asks about the symptoms of the player;

> "*…if you see a character like that who probably also talks about it a little bit and not just asking those questions but also telling about it in*

*the story. Yes, then you also give the young people more opportunity
to think about this a bit more and thus give a better picture of their
real complaints"* (P2, GZ Psychologist)

Besides the advantages mentioned above, participants have also provided
arguments for the fact that the dialogue-based assessment method can ensure
"*enhanced engagement*". This is because there are certain issues with traditional
on-paper questionnaires such as that they are often perceived as boring by
adolescents; "*...the basic questionnaire, which is very boring and which is not
fun.*" (P1, University lecturer and Assistant professor) and "*...I just find the
current self-tests very boring.*" (P3; Project leader at GGZ). Another issue is
that patients are currently being overexposed to questionnaires, e.g. "*...because
they already get so many questionnaires.*" (P6, Psychologist). Participants have
commented that the dialogue could enhance engagement due to a few factors.
First of all, the participants perceive the dialogue as a "*fun approach to asking
questions*";

> "*And something that people find boring or think "oh here they come
> again with all those questions", that it is asked in a fun way is, of
> course, great.*" (P5; University lecturer and GZ Psychologist)

Furthermore, another participant similarly addressed that using the dialogue is
refreshing in comparison to using the on-paper questionnaires so often; "*...be-
cause they already receive so many questionnaires that it is actually quite re-
freshing to map it out in such a way.*" (P6, Psychologist). Another factor that
contributes to enhanced engagement is the "*interactive nature*" of the dialogue;
"*I like it because it is very interactive.*" (P5; University lecturer and GZ Psy-
chologist) and "*...you take the best things from different things, ..., and you
make it interactive.*" (P3; Project leader at GGZ). Another significant differ-
ence between the traditional questionnaires and the dialogue that could increase
engagement is "*attractiveness*". Here, participants have commented that the di-
alogue is more attractive to do than the traditional questionnaire; "*...this could
be much more attractive for them to do.*" (P4; Senior Researcher and Clini-
cal Psychologist). Besides the dialogue being potentially more attractive than
the traditional questionnaires, participants have also addressed that the dia-
logue itself is well-designed; "*...this is just really really nice and well designed*"
(P5; University lecturer and GZ Psychologist). This participant specifically ad-
dressed that this is not always the case for digital interventions for adolescents;

> "*...I have often seen something that..., such a game where you really
> think this is like it was developed 20 years ago you know. Then I
> think, yes, young people are really not waiting for that anymore*"
> (P5; University lecturer and GZ Psychologist)

Additionally, other participants mentioned that the dialogue could be an en-
richment to the currently used questionnaires; "*...that it can also be an en-
richment...*" (P4; Senior Researcher and Clinical Psychologist). Moreover, the

dialogue could be beneficial when varying between the dialogue and the questionnaires; "...*I think mostly for variety. This is because they already receive so many questionnaires..*" (P6, Psychologist). This can also lead to enhanced engagement. Other factors that contribute to this engagement are "*liveliness*" where participants have addressed that this could result in more people using the dialogue compared to the traditional questionnaires, e.g. "...*it makes something livelier so that it will be used by more people...*" (P3; Project leader at GGZ). Another factor is that the dialogue would be a nicer way to fill out a questionnaire than what you get now; "*It's like a nice replacement I guess for the questionnaires.*" (P6, Psychologist). This is also because participants have addressed that the dialogue is perceived as less boring than the traditional questionnaire, e.g. "*It's not boring...*" (P5; University lecturer and GZ Psychologist). This could potentially result in more adolescents filling in the dialogue compared to the questionnaire, e.g. "...*the chance that these young people will fill this in is greater.*" (P4; Senior Researcher and Clinical Psychologist). This is especially an advantage as the traditional questionnaire is often not completed by adolescents or they do not want to repeat them over and over again;

> "*We are also concerned that they often simply do not complete it and that at a certain point, they no longer want to repeat it and do not want to complete it again.*" (P4; Senior Researcher and Clinical Psychologist)

The final factor contributing to enhanced engagement is that the dialogue is designed in a way that appeals to adolescents and matches their preferences; "...*it is in a way that appeals to young people...*" (P5; University lecturer and GZ Psychologist).

The final advantage of the dialogue-based assessment method that has been addressed by the participants is that this method could be used as "*extra diagnostic support*". One participant commented that it could be an extra support as it gives a first indication of which symptoms are present and it could provide the adolescents with more insights into their symptoms;

> "...*it is an extra help in diagnosing anyway. It already gives a good start for what symptoms are there and also how does the young person view it, can he already say a little more about it when you speak to him, and provide some more insight?*" (P2, GZ Psychologist)

As a result of the dialogue providing an indication of which symptoms are present in the adolescent, the physical diagnostics could be performed more extensively; "...*diagnosing in the room can ultimately be done more extensively and more thoroughly.*" (P2, GZ Psychologist). Another reason why the dialogue could provide extra support is that with the dialogue it could be more accessible to see which symptoms are present in the adolescent; "...*with such a dialogue, with such a video game, it is certainly a bit more accessible to be able to see what is going on among the young people.*" (P2, GZ Psychologist).

In conclusion, these findings indicate that the dialogue-based assessment method offers several advantages over traditional on-paper questionnaires. This

method addresses the perceived difficulty of questionnaires, the misinterpretation of questions and scales and response bias. The dialogue can create clarity around questions, create a safe atmosphere to be open and honest, facilitate the awareness of complaints, ensure reliable measurements of symptoms, enhance engagement and support current diagnostics.

6.2 Challenges for integrating the dialogue-based assessment into practice

The second theme relates to the specific challenges for the integration of the dialogue-based assessment into practice. Here, barriers and obstacles for integration are included but also disadvantages regarding the method itself that could potentially result in challenges for integration.

The first challenge for integration addressed by participants is that practitioners need to learn how innovations and new technologies work. Practitioners often have to get to know these innovations by themselves and in their own time; "*...then you have to figure that out in your own time and then still know the day after what you want to give to your client.*" (P3; Project leader at GGZ). However, learning how these innovations and technologies work is a challenge due to the "*high workload*". This is because participants have addressed that practitioners work in a busy environment; "*...practitioners who already have a very busy agenda and where the workload is really very high...*" (P1, University lecturer and Assistant professor). Due to this high workload, there is limited time available to learn how new innovations work but also to try out different innovations; "*That GP doesn't have time and that psychologist doesn't have time to figure that out himself.*" (P3; Project leader at GGZ) and "*As practitioners, we just have very little time to actually try out new things like that...*" (P6, Psychologist).

Furthermore, the participants especially addressed that there is often resistance from practitioners when learning how these innovations work. Participants referred to this as; "*...unknown makes unloved.*" (P4; Senior Researcher and Clinical Psychologist) and "*...new is scary.*" (P3; Project leader at GGZ). Especially in the GGZ, the Dutch mental healthcare organization, resistance towards new technologies occurs frequently;

> "*I think that in the GGZ, we are a bit inclined to keep doing things as we have been doing them for ten thousand million years.*" (P6, Psychologist)

The main cause for resistance is thus introducing something new and forcing behavioural changes; "*...people who hold on to certain thought patterns.*" (P3; Project leader at GGZ) and "*And it's also behaviour change...*" (P4; Senior Researcher and Clinical Psychologist). In addition to resistance due to introducing something new, there is also resistance from practitioners with an anti-games mindset; "*...just resistance from many practitioners who actually just have a bit of an anti-games mindset.*" (P1, University lecturer and Assistant professor).

This anti-games mindset arises due to a few factors. First of all, this mindset starts during education where there is no focus on potentially using technologies within mental healthcare; "...it already starts with the education. And there is really very little in that." (P1, University lecturer and Assistant professor). Moreover, a lot of practitioners see games as a problem in adolescents and cannot make the switch to the positive aspects of games; "...especially in youth care, they also see too many young people who game too much, which means that they see it mainly as a problem..." (P1, University lecturer and Assistant professor). Additionally, in the past, many innovations have been introduced into organizations like the GGZ without any training or clarification which increased resistance and contributed to the anti-games mindset within mental healthcare;

> "...there's been a lot of these kind of innovations in the past, they've been thrown in, door open, innovation in, door closed again and sort it out yourself." (P1, University lecturer and Assistant professor)

Finally, resistance toward innovations and new technologies arises as it is not yet often integrated into the policies of organizations; "People have not yet really integrated eHealth in their vision, policy and approach..." (P3; Project leader at GGZ).

Besides these factors that have contributed to an anti-games mindset, participants have also highlighted specific challenges related to the dialogue-based assessment method. First of all, a challenge could be whether practitioners have enough trust in this method as the dialogue is currently not tested. Participants have addressed the importance of evidence-based methods;

> "...whether practitioners have enough trust. So you see that they are very much evidence-based, so if this has not been tested, I think that a lot of people say "We are not going to use it at all because I do not trust that"." (P1, University lecturer and Assistant professor)

In addition, some participants questioned whether practitioners could see the added value of of the dialogue. First of all, this is because participants addressed that the intake is the most important method to establishing a diagnosis;

> "...but how much-added value does it have? Because in the end, at least for us, the most important way to make a diagnosis is not a questionnaire, but the conversation." (P6, Psychologist)

Moreover, another participant mentioned that the dialogue does not have much-added value for clinical depression as there the focus is more on suicidality rather than establishing a diagnosis; "With us, they come in really, really depressed and suicidal. So then suicidality is actually the first thing you should pay attention to..." (P5; University lecturer and GZ Psychologist) and "When I think about our practice, I think, yes, I'm sorry, but it's nice, but it wouldn't be of much use to me." (P5; University lecturer and GZ Psychologist).

Another challenge addressed by participants is whether adolescents would find the dialogue engaging enough; "...whether they find this engaging enough."

(P1, University lecturer and Assistant professor). First of all, this is because the dialogue takes more time to complete than the on-paper questionnaire; "...whether that outweighs, this probably takes a bit longer than filling out a questionnaire, is it worth it? Or do they think I'll just fill out that questionnaire?" (P1, University lecturer and Assistant professor) and "...it takes more time than the PHQ-8..." (P4; Senior Researcher and Clinical Psychologist). Moreover, the fact that the dialogue is still very close to filling out a questionnaire could also result in both practitioners and patients not seeing the added value of it;

> "...because it is still very close to completing a questionnaire actually, because you still get the textual questions on the screen, only packaged in a slightly different jacket so that people do not necessarily see the added value." (P1, University lecturer and Assistant professor)

Another challenge relates to technical difficulties with the dialogue-based assessment method. This is because the dialogue must be integrated into the system with patient records and portals. This system is currently being used for logging all the patient records and data; "...we already have a system with a patient file..." (P4; Senior Researcher and Clinical Psychologist). The experience of the participants is that it is difficult to build something like the dialogue inside of this system; "...everything that has to be built in there is always complicated..." (P4; Senior Researcher and Clinical Psychologist).

Finally, participants have addressed that the main challenge for integrating any innovation or technology into the mental healthcare practice is finding funding. This is because participants have addressed that there is often funding available for research and development but not for implementation; "...we don't find the money, so to speak, to go from a research version to an implementable version." (P1, University lecturer and Assistant professor). This is partially because youth care in the Netherlands, compared to adult care, is very fragmented across municipalities which makes it very difficult to find funding in municipalities to implement the technology; "...youth care in the Netherlands is very fragmented across municipalities. So there are little bags of money all over all the municipalities..." (P1, University lecturer and Assistant professor). Moreover, innovations like the dialogue need to be approved by health insurers in order to implement them into practice; "...these kinds of things have to be approved by health insurers in order to be used in clinical practice." (P1, University lecturer and Assistant professor). However, this process is perceived as very complicated due to all the bureaucratic paperwork; "And it's a really horrible process, the bureaucratic paperwork." (P1, University lecturer and Assistant professor).

In summary, several challenges to the integration of the dialogue-based assessment method into practice have been identified by the results. These challenges include learning how innovations work, high workload and limited time availability. Moreover, there can be resistance from practitioners due to offering something new, behaviour change and an anti-game mindset. Other challenges

relate to a lack of evidence for the dialogue, uncertainty about the added value, lack of engagement in the dialogue, technical difficulties and finding funding.

6.3 Strategies and considerations for implementing the dialogue-based assessment into practice

The third theme includes which strategies and considerations need to be taken into account when implementing the dialogue-based assessment into practice.

First of all, participants have addressed that it is important to create clarity on the usage of the dialogue-based assessment method; ”*...make the ease of use clear...*” (P2, GZ Psychologist). This is an important consideration as resistance may arise when the usage is unclear or when the dialogue is very difficult to use; ”*...if you were to offer something and it would be difficult, I already see the resistance of all those people...*” (P3; Project leader at GGZ). Creating clarity on usage includes different aspects such as explaining when and how the dialogue could be integrated into the treatment relationship; ”*...when do you deploy this?... So there is a need for some clarity for that..*” (P2, GZ Psychologist) and ”*I think that is something good to think about for an integration. How can you really use it in treatment?*” (P2, GZ Psychologist). Another aspect is the importance of explaining to the practitioners what the dialogue is; ”*...that people know what it is...*” (P4; Senior Researcher and Clinical Psychologist). Finally, it is important to explain what the added value of the dialogue is to the diagnosis process; ”*...explain well how it can complement the next steps in the diagnosis process.*” (P2, GZ Psychologist). What is considered important here is that it does not increase the workload. Instead, it should be made clear how the dialogue can help practitioners save time; ”*...to be able to show what the supplement can be and not that it is something on top..*” (P2, GZ Psychologist). Moreover, it should also be made clear whether it could reduce the workload; ”*...the advantage of any innovation or technological application should in principle be that you reduce the workload of practitioners...*” (P1, University lecturer and Assistant professor). Finally, the added value can be made clear by showing the practitioners that they no longer have to burden patients with questionnaires; ”*...that the care provider also thinks “oh this is fun this and this can also be fun” so that he no longer has the feeling I burden the young people with a long questionnaire...*” (P4; Senior Researcher and Clinical Psychologist). For the patients, it is important to show here that the dialogue does not replace any physical therapy or sessions;

> ”*...that it will not replace live contact with a practitioner, with the therapist. That it is an addition, an extra help. But I think it is very important for those young people in particular that one thing does not exclude the other*” (P2, GZ Psychologist)

Additionally, participants have suggested several methods that could ensure clarity of usage. This includes the use of a demonstration video in which you include how the dialogue works for both practitioners and patients;

> "...such a demonstration video, films and videos do well, so how does it work? How does it work for the practitioners? But how does it work, what do you encounter for the patient?" (P2, GZ Psychologist)

Moreover, a step-by-step manual could be used to clearly show the practitioners how to use the dialogue. This manual could for example include information about when to deploy the dialogue, for which people to use it, where to find the data and how to interpret it;

> "So when do you deploy this? At which moment? For which young people do you have to do this and for which young people not? Where can you find the data? How should you interpret that data? Really, say, a step-by-step guide so that it is made really easy to use this"
> (P1, University lecturer and Assistant professor)

Furthermore, the technical side of the dialogue-based assessment method also needs to be considered. Despite the linkage between the system with patient records and portals and the dialogue being a challenge, participants have addressed that this is important for integration; "So you also want it to be linked to the patient file in some way." (P6, Psychologist). This is because all the data about the patient is saved in this system, including results from questionnaires; "...the results are also included." (P4; Senior Researcher and Clinical Psychologist). Moreover, the dialogue should also be made easy to implement outside of the patient records and portals; "...that it is easy to implement..." (P3; Project leader at GGZ). The participants addressed that in order for the dialogue to be easily implemented, it is important that it is a standalone function that can for example be embedded with a code; "...a standalone function,...That I can just embed it with a code..." (P3; Project leader at GGZ). Finally, the dialogue must be played within a secure environment where privacy is an important aspect to take into consideration; "...within a secure environment, you should be able to play the game..." (P6, Psychologist) and "So that is something that is an important rule for us, privacy." (P6, Psychologist).

Additionally, an aspect that needs to be considered when potentially implementing the dialogue into practice is the potential of biased answers. Participants have mentioned that the environment or the story may prime someone a bit which could lead to different results; "...that story already primes someone a bit..." (P4; Senior Researcher and Clinical Psychologist) and "Because the environment, that primes someone to answer that question with a bit heavier or lighter answer." (P4; Senior Researcher and Clinical Psychologist). Participants have addressed that before implementing the dialogue into practice, this needs to be investigated and studied "...something you should really look into." (P4; Senior Researcher and Clinical Psychologist).

Finally, a strategy that is often used when innovations for mental healthcare are being developed, is the process of writing an implementation plan with external organizations. This implementation includes information about the target group and the stage of symptoms of the mental health disorder that is

focused on. This implementation plan is important to create because the target group and stage of symptoms determine what type of innovation should be developed, for example, for lowering stigma or activating people with mental health disorders;

> "...write down very specifically who we focus on and especially do you know what the target group exactly is? But also at what stage of, yes, symptoms are we? Because of course what you develop depends on that." (P1, University lecturer and Assistant professor)

In line with this strategy is the aspect of taking along the users well when developing innovations for mental health. Participants have addressed that it is important to sit together with your target group from the beginning to retrieve information about their preferences and to create an engaging storyline; "...to sit with the target audience from the start. Because you could have gotten a lot of information from that..." (P1, University lecturer and Assistant professor).

These findings indicate that there are several strategies and considerations that need to be taken into account when implementing the dialogue-based assessment method into practice. The results include the importance of creating clarity on the usage, considering the technical side of the dialogue, studying the potential of biased answers, writing an implementation plan and taking along the users well in the development process.

6.4 Implementation improvements for the dialogue-based assessment

The fourth theme focuses on the improvements that should be made to the dialogue-based assessment. This includes specific improvements to the dialogue method itself but also alternatives to this method.

First of all, participants commented that the character in the dialogue could explain its symptoms more. This is because, as mentioned in Section 6.1, people often find it difficult to understand what symptoms mean and how these manifest in their lives. The character could prove and explain their symptoms more by providing more details about the symptom;

> "...really just explain more about what it is so "I have a hard time concentrating, I wanted to do this and this yesterday but I keep getting distracted and I can't do anything about it". Just prove it with something, try to use that reader a little more and immerse them into the story so that you understand the question even better." (P3; Project leader at GGZ)

Moreover, another improvement suggestion is further questioning symptoms in the dialogue. Participants have mentioned that one method for this could be the use of a flow diagram that when someone has certain symptoms, you can further question this by asking more specific questions; "Well, suppose someone overeats, that you then add some more specific questions..." (P4; Senior

Researcher and Clinical Psychologist). This flow diagram, where symptoms are further questioned, allows the practitioner to get a better picture of the severity of symptoms; "...*that you can present certain things to people with a kind of flow diagram so that you, as a clinician, immediately get a much better idea of what is going on.*" (P4; Senior Researcher and Clinical Psychologist).

Furthermore, participants have addressed that it would be better to change the PHQ in the dialogue to the CDI (Children's Depression Inventory) questionnaire. First of all, this is because the PHQ is used often only in research; "*So it is used a lot for research, but not much in clinical practice.*" (P4; Senior Researcher and Clinical Psychologist). On the other hand, the CDI is used a lot more in clinical practice as it is recommended in the Dutch multidisciplinary guidelines for depressive disorder. This guideline includes all the techniques and recommendations that a GZ psychologist should use; "*...the CDI is currently one of the better questionnaires to use and is also recommended.*" (P5; University lecturer and GZ psychologist). When the CDI is used for the dialogue, implementation into practice could be made easier because this questionnaire is widely being used at schools, the GGZ and the Municipal Health Services (GGD); "*...both in the GGZ and in schools, at the GGD...if you can join it. I think that makes it, even more, easier to implement.*" (P5; University lecturer and GZ psychologist).

Another improvement to the dialogue would be to specifically focus on subclinical depressive disorder instead of clinical depressive disorder. Thus, focus on the individuals that do not have severe enough symptoms to meet all the criteria; "*So I would change the target group anyway from clinical depressive disorder to sub-clinical depressive disorder...*" (P5; University lecturer and GZ psychologist). At a sub-clinical level, participants have addressed that the dialogue could be used more at schools, as a preventive measure; "*...more at schools, preventive.*" (P5; University lecturer and GZ psychologist). This is because, as mentioned in Section 6.2, with clinical depressive disorder, you are less concerned with establishing a diagnosis but the focus is on treating suicidality.

Additionally, participants have provided suggestions for the character in the dialogue. These suggestions included the creation of gender-diverse and culturally diverse characters for the sake of recognisability with the character; "*I would also do a non-binary one.*" (P5; University lecturer and GZ Psychologist) and "*...a girl with a headscarf...*" (P4; Senior Researcher and Clinical Psychologist). Moreover, participants also commented that the character could also be adjusted more age appropriately where the figure looks different based on the age group playing the dialogue;

> "*So then that in terms of age, it is really an adolescent up to 16 years old and then more of a young adult up to 18 years old. And then yes, the age structure and so on.*" (P4; Senior Researcher and Clinical Psychologist)

This would allow the dialogue to be more appropriate and fitting for various ages.

Moreover, participants have also suggested changes to the storyline. First of all, participants addressed the importance of an engaging storyline in the dialogue in order to create an advantage compared to the traditional on-paper questionnaires;

> "...what a storyline like that has to do is sort of grab people and carry them along, and if it's able to do that and so get those people very involved with what's happening, then you have an advantage over the basic questionnaire, which is very boring and which is not fun." (P1, University lecturer and Assistant professor)

When creating this engaging storyline, participants have mentioned that it is important to consider a storyline that makes the players feel for the character which results in more openness. This could be achieved by telling more about the background story of the character; "...I think the story surrounding it of the character in that can also kind of jump-start it and give it a little bit of openness..." (P2, GZ Psychologist). However, an important consideration for this storyline is to create a logical structure for symptom reporting to prevent only having a summary of complaints because that does not make the storyline engaging or stimulate openness; "...you want some sort of logical build-up to why that person is telling you what they are telling you..." (P1, University lecturer and Assistant professor) and "And they shouldn't be a kind of enumeration...because then there is no reason why it affects you so much.". (P1, University lecturer and Assistant professor).

Moreover, there are two contentious subjects for improvement where participants differ in their opinion. First of all, there is a debate about including feedback in the dialogue. Two participants have commented that feedback for the player should not be included in the dialogue; "Not for the patient himself.." (P2, GZ Psychologist). This is because these participants mentioned that scores of the dialogue could give the player a distorted picture or it could scare the player when they see an elevated score; "If you really start working with scores and the client sees that, that means nothing to the patient at all or that gives a distorted picture..." (P2, GZ Psychologist). Two other participants commented that it is important to give some feedback to the patient such as advice to go talk to someone when there is an elevated score;

> "So I would give some kind of feedback. Maybe not very intensely like "Hey you have depression" or whatever, but "Gosh we notice that you also have some difficulties, have you ever talked to someone about it or could you talk to someone about it?" " (P5; University lecturer and GZ Psychologist)

This is because it also should not be a surprise for adolescents when they are invited to further discuss the symptoms, for example at schools, in the case of an elevated score; "...that it is, therefore, no surprise that those young people are singled out or something, that you do mention something about it." (P5; University lecturer and GZ Psychologist). Another subject where there was

some disagreement between participants is the inclusion of the item of suicidality in the dialogue. One participant commented that the suicidality item should only be included for clinical depressive disorder and not for screening or prevention purposes in sub-clinical populations because of the certain heavy load of this question; "...*such a question immediately has a certain load...*" (P3; Project leader at GGZ). Moreover, if suicidality is added in screenings, for a sub-clinical population, there is a duty of care; "*The disadvantage is if we as an organisation were to use that and then they ask if it is in it, then you also have to do something with it and I often find that more difficult.*" (P3; Project leader at GGZ). The majority of the participants, however, highlighted that the inclusion of suicidality in the dialogue is of significant importance. This is because when suicidality is excluded, an important part of the diagnosis will be missed; "...*to leave that out, yes you are also missing a very important diagnostic part.*" (P2, GZ Psychologist). Moreover, participants have addressed that there is a stigma around suicidality and leaving out suicidality can contribute to even more stigma around it; "*The more we cover it up and oh we shouldn't talk about it and it's dangerous. The more such a young person thinks yes, I am not going to say it...*" (P5; University lecturer and GZ Psychologist). Potential ways to ask about suicidality in the dialogue are also suggested by participants such as; "...*already in the cautious phase of "do you ever think about it?"*" (P5; University lecturer and GZ Psychologist) or "...*saying at the end of the dialogue like "gosh you noticed, I also sometimes have problems with concentrating...and then "I sometimes wonder what it would be like to be dead" for example.*" (P5; University lecturer and GZ Psychologist).

Finally, participants have also highlighted alternative methods to the dialogue-based assessment. First of all, one of these alternatives is asking questions in the moment. This means completing multiple short questionnaires a day to ask about the symptoms of a person in the moment. Participants have addressed that this might be a better method than the dialogue-based assessment because the dialogue, based on the PHQ-8, asks about symptoms over a two-week time period;

> "...*ask in the moment. So that you receive a short questionnaire several times a day about "I am very tired now", "I am very hungry" or "I don't feel like eating at all", "I feel very happy", "I feel not happy at all"...*" (P1, University lecturer and Assistant professor)

Based on these short daily questionnaires, you can develop personal fluctuating symptom networks where symptoms are linked together. This method ensures that more personal symptom profiles of depressive disorder can be developed, allowing practitioners to better help patients;

> "...*based on that you can create certain symptom networks over time that are specific to each person...there are all these networks of symptoms that are linked to each other..where we also come to that individual symptom profile much more and that we can help people much better with it.*" (P1, University lecturer and Assistant professor)

Another alternative for the dialogue-based assessment method, suggested by the participants, is measuring depressive disorder through behaviour. This is because behaviour is perceived as a better predictor of a depressive disorder than answers that people give to questionnaires. This is because everyone interprets questions differently;

> "*So when you ask us a question, we interpret that question, so everyone can interpret a question in a different way. And we kind of give a cognitively constructed answer to that...those behavioural degrees are often a better predictor of behaviour than the answers people give.*" (P1, University lecturer and Assistant professor)

Similarly to asking questions in the moment, measuring based on behaviour allows the creation of behavioural profiles that are also very specific for each person; "*...create a very large set of behavioural profiles...*" (P1, University lecturer and Assistant professor). This is considered a better method than questionnaires because questionnaires assume that there is only one profile for depressive disorder, while in fact there are many more; "*...one profile of depressive disorder while research has been done and there are over 100 different symptom profiles of people with depressive disorder.*" (P1, University lecturer and Assistant professor).

The results reveal several implementation improvements for the dialogue-based assessment method. These improvements include explaining the symptoms more, further questioning the symptoms with a flow diagram, changing the questionnaire to the CDI and focusing on sub-clinical depressive disorder instead of also clinical depressive disorder. Moreover, the character can be made more gender and culturally diverse and age appropriately. The importance of creating an engaging and logical storyline is also highlighted. Furthermore, opinions differed about the inclusion of suicidality in the dialogue and the type of feedback that should be included. Finally, participants addressed a preference for other assessment methods including asking questions in the moment and measuring depressive disorder based on behaviour.

6.5 Limitations of questionnaire-based assessment in general

The fifth theme covers limitations of traditional on-paper questionnaires that are not being mitigated by the dialogue-based assessment or any other questionnaire-based assessment in general. Moreover, this theme also covers challenges for diagnosis that remain, even when implementing the dialogue-based assessment method into practice.

First of all, as the dialogue-based assessment is based on the PHQ, there are a few limitations of questionnaires in general that also limit this assessment method. One of these limitations is that participants highlighted that questionnaires are considered a momentary reflection of the symptoms; "*It is a momentary reflection*" (P5; University lecturer and GZ Psychologist). Therefore, people completing these questionnaires often find it difficult to determine

the right answer to a question; "*What is the correct answer then? They find it complicated to fill that in.*" (P6, Psychologist). This is because how you have been feeling at that moment, your thoughts, and who you have been with, all affect how you answer the questions. Moreover, how you feel today can fluctuate from how you feel tomorrow;

> "*...how you feel at that moment and how your day has been, the people you've been with, all have an effect on how you fill out that questionnaire at that moment. And that can therefore be very different today than tomorrow.*" (P1, University lecturer and Assistant professor)

Finally, another limitation that remains is that questionnaires cannot be tailored for each person, while every person has other symptoms. Questionnaires are thus very general; "*...they are not specific to the specific complaints. Nor can you tailor them individually for each client...it is very generic.*" (P2, GZ Psychologist). Because of the fact that the dialogue is based on a questionnaire with fixed symptoms, the dialogue cannot adapt to people potentially having different symptoms that are not included in the original questionnaire.

Furthermore, there are also challenges regarding the diagnosis process that are not mitigated by the dialogue-based assessment method nor can be by any questionnaire-based method. First of all, questionnaires currently do not have a big place in the treatment and often, practitioners spend very little time discussing the results of questionnaires; "*...you notice that it doesn't really have a big place in the treatment to talk about the results of the questionnaires.*" (P2, GZ Psychologist). The conversation and the intake are considered more important by participants; "*Because in the end, at least for us, the most important way to make a diagnosis is not a questionnaire, but the conversation.*" (P6, Psychologist). This conversation is considered important because it often reveals symptoms that questionnaires do not cover; "*...you often see other symptoms that the questionnaires don't address...*" (P2, GZ Psychologist). This relates back to the questionnaires being generic and the fact that also the dialogue cannot be tailor-made to all the symptom profiles of patients. Therefore, participants have also mentioned that diagnosis remains human work;

> "*...I think it really remains just human work to really make that assessment of "is it a depression or is it, not a depression?""* (P5; University lecturer and GZ Psychologist)

Moreover, there is also a significant preference for physical contact during diagnosis from both the practitioner and patient side; "*Diagnosing is really something of, yes, seeing someone, assessing someone, often also talking to a loved one.*" (P2, GZ Psychologist), "*Just the human aspect in it, that's what I notice, people just find that really important.*" (P6, Psychologist) and "*I notice that young people prefer to just sit in the room.*" (P5; University lecturer and GZ Psychologist). From the practitioners' side, there is a preference for physical contact during diagnosis because they can estimate suicidality better and are

able to read someone better physically; "...*the risk of suicidality...when you see someone in real life, you have the feeling that you can assess it better.*" (P4; Senior Researcher and Clinical Psychologist) and "...*you can read someone a little better when someone is sitting with you compared to sitting online.*" (P5; University lecturer and GZ Psychologist). Despite the preference for having physical conversations over using questionnaires, one participant did address that using questionnaires is an efficient method to monitor symptoms and measure treatment effects. Therefore, this participant argued that questionnaires should have a bigger place in treatment;

> "...*the power of those questionnaires is actually somewhat, yes, nullified because it is a good method...So basically it should have a bigger place.*" (P2, GZ Psychologist)

Finally, participants have addressed that any assessment method based on the DSM, including the assessment method used during conversations, has a limitation. This limitation is that the DSM compartmentalizes all mental health disorders; "*I think one of the complexities is that the DSM compartmentalizes it all.*" (P6, Psychologist). This means that patients with certain complaints are always categorized in one of the boxes of the DSM, while in practice, patients often cannot be categorized in only one box. This is because participants highlighted that in practice, mental health disorders, including depressive disorder, are very heterogeneous; "...*depressive disorder is actually quite heterogeneous.*" (P4; Senior Researcher and Clinical Psychologist). There are many correlational connections for the development of a depressive disorder such as biology, genes, and hormonal development;

> "*You see it also has to do with biology, but with the genes, with family burden, with the hormonal development in young girls where things change.*" (P4; Senior Researcher and Clinical Psychologist)

These correlational connections make the diagnostics very complicated as practitioners need to determine the cause of the depressive disorder; "*And with that diagnosis, it is actually very difficult to determine exactly what could be the cause...*" (P4; Senior Researcher and Clinical Psychologist). Besides these connections, comorbidity is also considered a complexity in diagnostics. This is because a depressive disorder is often accompanied by fears or anxiety; "*And the comorbidity is also a problem, so it's often not just depressive disorder but also real anxiety.*" (P4; Senior Researcher and Clinical Psychologist). Overall, participants have addressed that diagnostics remain very complicated, despite effective interventions; "*I think the diagnostics still remain complicated...*" (P4; Senior Researcher and Clinical Psychologist).

In conclusion, several remaining limitations have been found including that questionnaires are a momentary reflection and they cannot be tailor-made. Furthermore, questionnaires currently do not have a big place in the treatment as the intake and conversation are considered more important and there is a preference for physical contact during the diagnosis process. Moreover, another

limitation is that the DSM compartmentalizes all mental health disorders while they are actually very heterogeneous with different correlational connections and are comorbid.

7 Discussion

Previous research has shown increasing interest in mental health games [32, 71]. This research has indicated that serious games have significant benefits for mental health regarding awareness, prevention, detection and treatment of disorders [74]. Several games have been developed to assess and treat mental health disorders, aiming to increase access to help, improve assessment efficiency, reduce stress, and mitigate bias issues [50, 55, 76, 102].

However, few games developed for mental health disorders are used in practice [89]. Therefore, further understanding is needed on the integration of games into practice and practitioners' opinions, as there is limited research available on the integration of digital interventions into practice [28].

Similarly, in my research, in the research by Bowey et al. [11], a dialogue was used to predict sexist beliefs. My research followed Bowey et al.'s [11] approach of translating on-paper questionnaires into dialogue items. Bowey et al.'s [11] research tested a dialogue method to predict real-world beliefs. However, this approach lacks evaluation with expert opinions and has not been studied in the context of mental health disorders.

Therefore, my study aims to address this gap by developing and evaluating a dialogue-based assessment method. In total, five themes are identified by the results of the thematic analysis, that provide insights into the specific opportunities and challenges for integrating a dialogue-based assessment method for depressive disorder into practice. Moreover, my study also identifies themes related to strategies and considerations for implementation, implementation improvements for this assessment method and remaining limitations of questionnaire-based assessment methods. In this section, these themes will be discussed in relation to previous research and limitations and future work will be considered.

7.1 Opportunities for integration

Regarding the opportunities for integrating the dialogue-based assessment method into practice, the advantages proposed by the participants are considered. The findings support the idea that games for mental health can be beneficial because of their *appealing potential* [32, 39]. Results suggest enhanced engagement with the dialogue due to enjoyment compared to traditional questionnaires, according to the experts' opinions.

Moreover, traditional on-paper questionnaires are subject to issues such as guessing behaviour, social desirability bias and, in the specific context of mental health disorders, also often shame and stigma [17, 35]. These issues are all considered response biases that influence the responses of people in a way in which

47

they provide inaccurate or untruthful responses [16, 35]. The participants who contributed to my study indicate that the dialogue-based assessment method could potentially partially prevent reporting biases by encouraging more honest responses. This is because participants believed that people might be less concerned with giving a desired answer to the questions in the dialogue. This might indicate that the dialogue might lower the chances of a reporting bias. This result is in line with the research by Bowey et al. [11], which indicated that when transforming on-paper questionnaires into dialogues, this might be more predictive of people's actual thoughts and beliefs. Moreover, according to the experts' opinions, my findings also suggest that when the dialogue-based assessment method is used, people might respond more honestly because a different method is used for assessment and because the character in the dialogue reassures the player. However, in future research, an evaluation must be performed of whether this is actually the case by conducting an experiment with people completing the dialogue.

While the literature review did not encompass all the findings identified in this study, certain results can still be addressed. For example, the results indicate that adolescents often have difficulty understanding on-paper questionnaires. Moreover, they often misinterpret questions and find it difficult to estimate the severity of the symptoms themselves. Results from a study by Clarke et al. [20] reveal that certain items in mental health questionnaires are suspect to misunderstanding and misinterpretation, where my results provide confirmation for this. This limitation of on-paper questionnaires could potentially be mitigated by the explanation that the character in the dialogue gives about the symptoms, according to the participants. However, to maximise this potential of the dialogue-based assessment method, an improvement for future development of the dialogue would be to explain the symptoms even more. This would mean, amongst others, giving specific examples of how the symptom manifests in daily life.

Similarly, the experts indicate that the dialogue could facilitate awareness of symptoms, where there is the potential that the dialogue allows players to be more open and share their own symptoms. This is in line with previously identified research that showed that games could be used to create awareness of mental health disorders and their aspects [14, 74].

Furthermore, the participants revealed that the dialogue-based assessment method could be used as extra diagnostic support to, for example, give an indication of which symptoms are present. Afterwards, the diagnosis could be performed more extensively. Moreover, participants have also addressed that the dialogue could be used as a preventive measure, amongst others, at schools. This is in line with previous research that showed that practitioners believe the most suitable application of serious games was as an addition to the therapy or as a preventive method [28].

7.2 Challenges for integration

Previous research has shown that practitioners are often reluctant towards internet-based health interventions for therapeutic usage, partially because they are relatively new in practice [28]. The findings indicate that resistance from practitioners might also occur for the dialogue-based assessment method. This is especially because a new method is introduced into practice and, as addressed by Eichenberg et al. [28], there is a lack of policies for digital mental health interventions. The latter was also confirmed by the results that indicate that these digital interventions are not yet often integrated into practitioners' policies. However, the findings also suggest that this resistance might be caused by previous introductions of innovations into practice where no support was provided to practitioners. This could have resulted in a decrease in trust towards these innovations. Another factor contributing to resistance is practitioners being unfamiliar with the Internet [90]. In line with this, the analysis suggests that for the dialogue-based assessment method, resistance could arise due to an anti-games mindset, partially as a result of a lack of knowledge about these innovations.

Other challenges identified in the results relate to finding funding to implement the dialogue-based assessment into practice. This challenge was not previously identified during the literature review in my study. However, several studies indicate that there is limited public health funding as it is difficult to establish a clear image of the cost-benefit outcomes of interventions [45]. Integration is often performed through reimbursement and health insurance coverage, as also supported by the results of my study. Important factors of this reimbursement and funding process are that the method is evidence-based [33]. Despite the benefits shown for several digital mental health interventions, there is an absence of reimbursements. This is considered the primary obstacle to the adoption of digital interventions into practice, which is also the case for the dialogue and could thus limit integration [77].

Moreover, despite the time that internet-based interventions might save practitioners [6], the results of my study reveal that limited time availability still remains a challenge. This is because practitioners experience high workloads with little to no time to learn how innovations work. My results confirm findings from studies that have shown that introducing new technologies into practice actually leads to an increased workload. This increased workload arises from technical and organisational obstacles [53]. However, research also suggests that once these obstacles are overcome, digital interventions could potentially bring flexibility to practitioners and help them save time [37].

Furthermore, the results reveal that the technical side of the dialogue-based assessment method, which needs to be built into the patient records and portals, is complicated. Previous research is in line with this that shows that these patient health record systems lack interoperability, meaning that it is difficult to connect different systems and applications to the patient health record system [3, 22]. Despite the difficulties with integrating an intervention like the dialogue-based assessment method into the patient record system, this is still considered

a crucial aspect. A further evaluation of how the dialogue can be integrated into this system is needed for successful implementation.

7.3 Strategies and considerations

The findings regarding strategies and considerations for implementation of the dialogue-based assessment method reveal that creating clarity of usage, explaining the added value, and easy implementation are crucial. This confirms previous research suggesting that the adoption of digital mental health interventions is influenced by, amongst others, creating familiarity with interventions and explaining why and how the intervention is different from the traditional treatment or assessment method [38, 105].

Furthermore, similar to the research by Graham et al. [38], for the dialogue-based assessment method, participants also addressed the importance of playing the dialogue in a secure environment where privacy is considered a crucial factor. The findings also indicate that the target group of the dialogue should be taken along well. This target group should be well established while writing an implementation plan. This is because this could lead to an intervention that better fits the needs and preferences of the target group, as previously identified in my study [38].

7.4 Implementation improvements

The results indicate that several improvements to the dialogue-based assessment method could be made. One limitation of internet-based assessment methods is that it is difficult to obtain additional information [6]. The results indicate that this addressed limitation could partially be mitigated by further questioning symptoms in digital assessment methods in future research, according to the participants. This would allow practitioners to get a more complete image of the symptoms of the depressive disorder.

Furthermore, previous research has also shown that serious games or other digital mental health interventions might be more suitable for people with mild symptoms of a disorder [28]. In the case of depressive disorder, when a person has mild symptoms and does not meet all the criteria, it is referred to as sub-clinical. The findings of my study that indicate that practitioners would suggest using the dialogue-based assessment method for a sub-clinical rather than a clinical population provides evidence for the results from previous research [28]. This is also because, for clinical populations, you are less concerned with establishing a diagnosis.

Additionally, several studies indicate that the PHQ is frequently used in clinical practice [58, 59, 70]. A result contradicting this literature is the recommendation for changing the PHQ as a base for the dialogue-based assessment to the CDI (Children's Depression Inventory). However, this result could be explained by the fact that the PHQ is frequently used in the United Kingdom. As my study is performed with mental health care professionals based in the

Netherlands, they recommend the CDI as it is in the multidisciplinary guidelines for depressive disorder for youth [94]. Moreover, according to the GGZ, the PHQ in the Netherlands is often used for assessing adults, while the CDI is recommended for assessing children or adolescents [97]. This could also explain the recommendation to change the questionnaire to the CDI. However, future research could analyse the actual prevalence of the use of the PHQ within practice.

Other findings from my research reveal several suggestions regarding the character in the dialogue. The results indicate the importance of adding, for example, non-binary characters besides men and women and the inclusion of culturally diverse characters. Within the field of games for mental health, characters are perceived as one of the most important characters [78]. This is because characters allow self-representation and the expression of emotions. The effectiveness and usability of a mental health intervention are influenced by this degree of user representation [78]. This does not only apply to gender-diverse characters but also to culturally diverse ones, including, for example, women with a headscarf or men/women with darker skin [83]. In conclusion, several studies have shown that people express the importance of the inclusivity of genders and cultural diversity within games, confirming the findings of my study [78, 83, 99].

Furthermore, the results reveal several areas for improvement regarding the storyline of the game. Here, participants addressed the importance of a storyline that grabs people to encourage engagement. The engaging potential of games is also considered one of the games' potentials [32]. Moreover, the surrounding story of the character is also considered crucial for stimulating openness in the player. For example, the storyline of The Last of Us, a post-apocalyptic game where a father who lost his child escorts a teenage girl through the world, shows the success and importance of a story that makes the players understand the characters and feel for them [51, 91]. Another research also emphasizes that players value an involving story that is able to elicit emotions and feelings, confirming the results of my study [25].

7.5 Limitations in general

The results show that there are several factors that could challenge the integration of any questionnaire-based method into practice. First of all, the results of my study confirm previous research that showed that practitioners value personal contact during the treatment process [28]. The research by Schroder et al. [90] specifically addressed that this is because, with online diagnosis or treatment, specific symptoms, aspects, or non-verbal signals could be missed. The results from my analysis show that the participants also raised this issue with questionnaire-based assessments as they emphasized the importance of seeing someone physically and including the human aspect in the process.

Moreover, the results show that the conversation is considered more important than the questionnaires during the treatment process, according to practitioners and that questionnaires do not have a big place in treatment. The

results of my study reveal that this conversation is considered more important as you often see other symptoms that questionnaires do not address. However, this contradicts several studies that indicate a high prevalence of self-assessment questionnaires being used in practice and that it is considered a valuable tool [15, 58, 59, 70]. This discrepancy highlights the importance of future evaluation of the actual prevalence of different self-assessment questionnaires for mental health disorders in practice.

Moreover, previous research has shown that digital diagnosis might be complex due to various factors, including complex symptoms and comorbidity [6]. The results of my study confirm these challenges when using questionnaire-based assessment methods. The analysis showed that people refer to a depressive disorder as very heterogeneous and comorbid, which makes the network of symptoms with correlational connections and comorbid disorders very complex. Previous research has already shown that self-assessment questionnaires are considered less accurate when there is a high prevalence of comorbidity [27]. The findings of my study also reveal that these self-assessment questionnaires assume only one profile of depression and do not take into account comorbidity or these correlational connections.

Furthermore, as previously identified, face-to-face treatments allow the creation of tailor-made interventions. For traditional on-paper questionnaires, this is not possible, limiting the flexibility to adapt to the patient during the process [6]. This aligns with the results of my study that identified that a limitation of self-assessment questionnaires, and also the dialogue-based assessment method, is that they cannot be tailor-made.

7.6 Limitations and future work

In this study, several participants were interviewed, and almost all had a background in psychology and either were primarily working as a psychologist or combined this with another profession in this field. Only one participant was not working as a psychologist but had experience with developing interactive experiences for children's mental health. Since the sample of my research is primarily composed of experts working in the same field, this might limit the diversity of perspectives and generalizability. For future work, a broader pool of participants could be considered with experts from different professional backgrounds, such as general practitioners or psychiatrists. Moreover, as participants have addressed that the dialogue-based assessment method could be used as a preventive measure at schools, teachers or school directors could also be interviewed in future research. A more diverse range of participants can provide more insights from different perspectives, which might reveal additional opportunities and challenges [4].

Another limitation of this research is that due to time constraints during the interviews, not the entire dialogue could be shown to the participants. During the interviews, only small parts of the dialogue, including the beginning and how certain symptoms are addressed, were shown for evaluation. As a result, participants were not able to comment on the entire dialogue and could not

review the entire context and flow of the dialogue. Consequently, valuable insights regarding the opportunities and challenges for integration could remain unrevealed. This highlights the potential for further research and exploration in future studies where an evaluation of the complete dialogue is performed.

Additionally, the thematic analysis performed in this study is also subject to limitations as one researcher performed this study. This could be considered a limitation in terms of objectivity and reliability. This is because there might be researcher subjectivity which is considered a bias [13]. This bias is often managed by using multiple coders and measuring their agreement between coders. Multiple coders prevent the unintentional involvement of personal preconditions, beliefs or expectations, which can influence the analysis [13]. Despite the limitation of only one researcher, a codebook was used to try to minimise this bias. This helped ensure consistency in the coding process while also making the research transparent and reproducible [13].

Furthermore, for future work, it is recommended that the dialogue-based assessment method is also tested regarding its effectiveness and accuracy in predicting the symptoms of depressive disorder in adolescents. Moreover, participants also addressed the importance of investigating the potential of biased answers due to the environment of the dialogue. Participants have highlighted that testing is an important consideration when implementing new methods into practice, as practitioners are very much evidence-based. Previous research on the serious game SPARX [76] showed that a large body of evidence could significantly contribute to the chance of integration into practice. Besides that, the results of my study reveal concerns regarding the *engaging potential* of the dialogue [32]. This is because the dialogue takes longer to complete, and it is still close to filling out a questionnaire. This was also similarly found in the research by Frommel et al. [34], where a potential limitation of their research was that participants might have recognised the original questionnaire as the method was close to the questionnaire itself. Testing the dialogue with adolescents could ensure a better understanding of whether they find the dialogue engaging enough. This could reveal further improvements for implementing the dialogue into practice. The engaging potential is especially important as research has shown that patients value the entertaining part of interventions [28]. For testing the dialogue-based assessment method, a similar approach could be used as the research by Bowey et al. [11] where scores of the on-paper MRAS were compared with the responses from the dialogue. To prevent response bias, in future research, the dialogue for depressive disorders could be tested with adolescents who are already in the treatment process to ensure that responses to the on-paper questionnaire are indicative of their state.

Moreover, another recommendation can be made regarding the creation of an immersive and engaging storyline to ensure that the dialogue achieves the *appealing* and *engaging* potential of serious games [32]. Here, participants addressed the importance of an immersive storyline that makes players feel for the characters in the story. This could be achieved by involving adolescents in the early stages of further development to understand their preferences and evaluate the dialogue. Moreover, game researchers who have experience with interactive

storytelling could be consulted. This collaborative approach could ensure that the *appealing*, *engaging* and *effectiveness* potentials of serious games for mental health are maximised [32].

In addition, the results of this study revealed several contradictions regarding the inclusion of feedback in the dialogue and the item of suicidality. For future work, it is recommended to evaluate the inclusion of these aspects further. This could involve studying the effect of the inclusion of feedback in the dialogue by involving both patients and practitioners while also further analysing the benefits and challenges of addressing suicidality in a dialogue-based assessment method. This could provide further insights into how these factors could effectively be handled.

Finally, research by Dechant et al. [50] aimed to assess the presence of social anxiety in games by making use of behavioural correlates. This meant that this mental health disorder was measured based on the behaviour of the player while playing the game. A similar approach was used in the game *The Delivery* where depressive disorders were assessed based on the cognitive behaviour of players [102]. The results from my study reveal that assessing mental health disorders and measuring symptoms might be more effective when using behaviour as a predictor, according to participants. Besides measuring based on behaviour, previous studies also aimed to analyse the use of Ecological Momentary Assessment (EMA) in mental health. EMA aims to address the limitations of self-assessment questionnaires, such as recall bias and a lack of generalizability of the results to real-life situations [88]. With EMA, the feelings of people are measured multiple times a day, which can reveal more symptoms and characteristics of the disorder than with a momentary questionnaire like the PHQ-8 [88]. The findings of my study also reveal that practitioners often refer to the traditional questionnaires as a momentary reflection which is a limitation that is not addressed by the dialogue-based assessment method, which confirms the results from the research on EMA [88]. However, as the results from the study by Bowey et al. [11] show that dialogue-based in-game assessment significantly predicts real-world beliefs, future research could study whether there is indeed a difference in prediction between behavioural-based or EMA-based and dialogue-based methods. This could ensure a better understanding of which methods would be the most effective for assessing mental health disorders.

8   Conclusion

Despite the extensive research on the use of serious games or other digital interventions for mental health, very little work has focused on both the use of these interventions for the assessment of mental health disorders and the evaluation of the integration into practice. The results from my study suggest that there are several opportunities and challenges for integrating a dialogue-based assessment method for depressive disorder into practice. The opportunities for integrating this method into practice relate to the potential mitigation of issues with traditional on-paper questionnaires, such as misinterpretation and misunderstanding

of questions in the questionnaire and response biases. Furthermore, the dialogue could be used as a method to create clarity around questions, encourage honesty in patients, facilitate awareness about the complaints and the disorder, ensure more reliable measurements, enhance engagement and help practitioners pre-identify symptoms of the depressive disorder, according to the participants. However, this study also reveals several challenges that could make the integration into practice more difficult. These challenges are specifically related to the high workload of practitioners, where they also need to learn how new innovations work for which there is limited time available. Moreover, introducing new innovations into practice often brings along resistance from practitioners due to issues related to behaviour change or anti-game mindsets. Other specific challenges related to a lack of a solid body of evidence for the dialogue-based assessment method, uncertainty about whether it adds value to the current process, and whether the dialogue is engaging enough. Furthermore, the dialogue also brings along technical difficulties, including the integration with the system of patient records and portals. Finally, another major challenge for integration is the transient phase from a research version to an implementable version, where a significant amount of funding is needed.

Besides the main research question that identifies opportunities and challenges, participants also addressed several strategies and considerations that need to be considered for implementation into practice. These include creating clarity around usage, considering integration with the system of patient records and portals, testing the dialogue and writing an implementation plan where users are taken along well to ensure that the intervention meets patients' demands. However, before the dialogue-based assessment method is integrated into practice, several implementation improvements need to be performed. These refer to explaining and further questioning symptoms within the dialogue, including the CDI as a questionnaire to fit the Dutch guidelines and focus on sub-clinical depression as a target group. Moreover, several adjustments need to be made to the characters, including making them more gender-culturally diverse and age-appropriate. Furthermore, the inclusion of feedback and suicidality in the dialogue need to be considered and studied in future work to determine whether these items need to be included. Another aspect that needs to be evaluated is the potential of behavioural or EMA-based assessment methods for depressive disorder, as a replacement to the current dialogue-based assessment method.

Finally, this study revealed that certain challenges are difficult to mitigate with any (questionnaire-based) assessment method, whether digital or not. These challenges include the fact that questionnaires are momentary and not tailor-made to specific patients. Moreover, physical contact is very much valued by practitioners and patients, where the conversation is considered more important than questionnaires. This results in questionnaires not being used frequently during face-to-face sessions. Other challenges relate to the DSM placing mental health disorders in boxes while correlational connections and comorbidity are often not taken into account when establishing a diagnosis, making the process very complex for practitioners.

Together these results suggest that there are several opportunities that could help integrate more digital assessment methods into practice. However, there are remaining obstacles that could limit the chances for integration. The work of this study contributes to a better understanding of these obstacles, which could be used in future research when developing other assessment methods. Taking these challenges into consideration could ensure the development of more effective and efficient interventions that are more likely to be used in practice. However, it could also potentially aid researchers who already developed interventions in determining strategies for implementation where the results of my study are considered. This could eventually potentially increase access to help, improve the received care and make the process more engaging, stimulating and less stressful for patients.

References

[1] Alaa Abd-Alrazaq, Eiman Al-Jafar, Mohannad Alajlani, Carla Toro, Dari Alhuwail, Arfan Ahmed, Shuja Mohd Reagu, Najeeb Al-Shorbaji, Mowafa Househ, et al. The effectiveness of serious games for alleviating depression: systematic review and meta-analysis. *JMIR Serious Games*, 10(1):e32331, 2022.

[2] Swann Arp Adams, Charles E Matthews, Cara B Ebbeling, Charity G Moore, Joan E Cunningham, Jeanette Fulton, and James R Hebert. The effect of social desirability and social approval on self-reports of physical activity. *American journal of epidemiology*, 161(4):389–398, 2005.

[3] Manal Alghamdi, Courtney Stanley, William K Willis, and Alberto Coustasse. Personal health record interoperability. *BHAA Annual Conference*, 2019.

[4] Peter Allmark. Should research samples reflect the diversity of the population? *Journal of medical ethics*, 30(2):185–189, 2004.

[5] Christopher S Amenson and Peter M Lewinsohn. An investigation into the observed sex difference in prevalence of unipolar depression. *Journal of Abnormal Psychology*, 90(1):1, 1981.

[6] Gerhard Andersson and Nickolai Titov. Advantages and limitations of internet-based interventions for common mental disorders. *World Psychiatry*, 13(1):4–11, 2014.

[7] Gavin Andrews and Gregory L Carter. What people say about their general practitioners' treatment of anxiety and depression. *Medical journal of Australia*, 175:S48–S51, 2001.

[8] Ivan L Beale, Pamela M Kato, Veronica M Marin-Bowling, Nicole Guthrie, and Steve W Cole. Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *Journal of Adolescent Health*, 41(3):263–270, 2007.

[9] Theo Bouman. Classificatie. In *Klinische psychologie: Theorieën en psychopathologie*, pages 175–187. Noordhoff Uitgevers, 2015.

[10] Jason T Bowey, Ansgar E Depping, and Regan L Mandryk. Don't talk dirty to me: How sexist beliefs affect experience in sexist games. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1530–1543, 2017.

[11] Jason T Bowey, Julian Frommel, Brandon Piller, and Regan L Mandryk. Predicting beliefs from npc dialogues. In *2021 IEEE Conference on Games (CoG)*, pages 1–8. IEEE, 2021.

[12] Virginia Braun and Victoria Clarke. *Thematic analysis.* American Psychological Association, 2012.

[13] Virginia Braun and Victoria Clarke. Can i use ta? should i use ta? should i not use ta? comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, 21(1):37–47, 2021.

[14] Adolfo J Cangas, Noelia Navarro, José MA Parra, Juan J Ojeda, Diego Cangas, Jose A Piedra, and Jose Gallego. Stigma-stop: a serious game against the stigma toward mental health in educational settings. *Frontiers in psychology*, 8:1385, 2017.

[15] Mauro Giovanni Carta, Maria Carolina Hardoy, Mariangela Cadeddu, Andrea Murru, Andrea Campus, Pier Luigi Morosini, Alex Gamma, and Jules Angst. The accuracy of the italian version of the hypomania checklist (hcl-32) for the screening of bipolar disorders and comparison with the mood disorder questionnaire (mdq) in a clinical sample. *Clinical practice and epidemiology in mental health*, 2(1):1–5, 2006.

[16] Jennifer LK Charles and Patrick V. Dattalo. Minimizing social desirability bias in measuring sensitive topics: The use of forgiving language in item development. *Journal of Social Service Research*, 44(4):587–599, 2018.

[17] Vibhav Chitale, Nilufar Baghaei, Daniel Playne, Hai-Ning Liang, Yinshu Zhao, Aysu Erensoy, and Yvette Ahmad. The use of videogames and virtual reality for the assessment of anxiety and depression: A scoping review. *Games for Health Journal*, 2022.

[18] Helen Christensen, Kathleen M Griffiths, Louise Farrer, et al. Adherence in internet interventions for anxiety and depression: systematic review. *Journal of medical Internet research*, 11(2):e1194, 2009.

[19] Helen Christensen and Ian B Hickie. Using e-health applications to deliver new mental health services. *Medical Journal of Australia*, 192:S53–S56, 2010.

[20] Aileen Clarke, Tim Friede, Rebecca Putz, Jacquie Ashdown, Steven Martin, Amy Blake, Yaser Adi, Jane Parkinson, Pamela Flynn, Stephen Platt, et al. Warwick-edinburgh mental well-being scale (wemwbs): validated for teenage school students in england and scotland. a mixed methods assessment. *BMC public health*, 11:1–9, 2011.

[21] Victoria Clarke, Virginia Braun, and Nikki Hayfield. Thematic analysis. *Qualitative psychology: A practical guide to research methods*, 3:222–248, 2015.

[22] Pascal Coorevits, Mats Sundgren, Gunnar O Klein, Anne Bahr, Brecht Claerhout, Christel Daniel, Martin Dugas, Danielle Dupont, Andreas

Schmidt, Peter Singleton, et al. Electronic health records: new opportunities for clinical research. *Journal of internal medicine*, 274(6):547–560, 2013.

[23] Oana Alexandra David, Cristina Costescu, Roxana Cardos, and Cristina Mogoaşe. How effective are serious games for promoting mental health and health behavioral change in children and adolescents? a systematic review and meta-analysis. In *Child & Youth Care Forum*, volume 49, pages 817–838. Springer, 2020.

[24] NHS Digital. How to access mental health services., 2018.

[25] Danilo Barros dos Santos, Cristiano Maciel, Vinicius Carvalho Pereira, and Eunice Pereira dos Santos Nunes. Digital empathic games and their relation with mortality: Analysis of discussion forums. In *HCI in Games: First International Conference, HCI-Games 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pages 307–319. Springer, 2019.

[26] M Taylor Dryman and Richard G Heimberg. Emotion regulation in social anxiety and depression: A systematic review of expressive suppression and cognitive reappraisal. *Clinical psychology review*, 65:17–42, 2018.

[27] Shaun M Eack, Catherine G Greeno, and Bong-Jae Lee. Limitations of the patient health questionnaire in identifying anxiety and depression in community mental health: many cases are undetected. *Research on social work practice*, 16(6):625–631, 2006.

[28] Christiane Eichenberg, Gloria Grabmayer, and Nikos Green. Acceptance of serious games in psychotherapy: an inquiry into the stance of therapists and patients. *Telemedicine and e-Health*, 22(11):945–951, 2016.

[29] Cecilia A Essau, Peter M Lewinsohn, John R Seeley, and Satoko Sasagawa. Gender differences in the developmental course of depression. *Journal of affective disorders*, 127(1-3):185–190, 2010.

[30] Maurizio Fava and Kenneth S Kendler. Major depressive disorder. *Neuron*, 28(2):335–341, 2000.

[31] Kristina Fenn and Majella Byrne. The key principles of cognitive behavioural therapy. *InnovAiT*, 6(9):579–585, 2013.

[32] Theresa M Fleming, Lynda Bavin, Karolina Stasiak, Eve Hermansson-Webb, Sally N Merry, Colleen Cheek, Mathijs Lucassen, Ho Ming Lau, Britta Pollmuller, and Sarah Hetrick. Serious games and gamification for mental health: current status and promising directions. *Frontiers in psychiatry*, 7:215, 2017.

[33] Anna Paldam Folker, Kim Mathiasen, Sigurd Mørk Lauridsen, Ellen Stenderup, Els Dozeman, and Marie Paldam Folker. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: A comparative case study of implementation challenges perceived by therapists and managers in five european internet services. *Internet Interventions*, 11:60–70, 2018.

[34] Julian Frommel, Fabian Fischbach, Katja Rogers, and Michael Weber. Emotion-based dynamic difficulty adjustment using parameterized difficulty and self-reports of emotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, pages 163–171, 2018.

[35] Julian Frommel, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinau, Ramona Schabenberger, Valentin Riemer, Claudia Schrader, and Michael Weber. Integrated questionnaires: Maintaining presence in game environments for self-reported data acquisition. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pages 359–368, 2015.

[36] Patricia I Fusch Ph D and Lawrence R Ness. Are we there yet? data saturation in qualitative research. *Walden faculty and Staff Publications*, 2015.

[37] Jordan S Gardner, Brittany E Plaven, Peter Yellowlees, and Jay H Shore. Remote telepsychiatry workforce: a solution to psychiatry's workforce issues. *Current psychiatry reports*, 22:1–9, 2020.

[38] Andrea K Graham, Emily G Lattie, Byron J Powell, Aaron R Lyon, Justin D Smith, Stephen M Schueller, Nicole A Stadnick, C Hendricks Brown, and David C Mohr. Implementation strategies for digital mental health interventions in health care settings. *American Psychologist*, 75(8):1080, 2020.

[39] Juho Hamari, David J Shernoff, Elizabeth Rowe, Brianno Coller, Jodi Asbell-Clarke, and Teon Edwards. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in human behavior*, 54:170–179, 2016.

[40] Benjamin L Hankin, Lyn Y Abramson, Terrie E Moffitt, Phil A Silva, Rob McGee, and Kathryn E Angell. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. *Journal of abnormal psychology*, 107(1):128, 1998.

[41] Florian Hardeveld, Jan Spijker, Sophie A Vreeburg, Ron De Graaf, Sanne M Hendriks, Carmilla MM Licht, Willem A Nolen, Brenda WJH Penninx, and Aartjan TF Beekman. Increased cortisol awakening response was associated with time to recurrence of major depressive disorder. *Psychoneuroendocrinology*, 50:62–71, 2014.

[42] Aine Horgan and John Sweeney. Young students' use of the internet for mental health information and support. *Journal of psychiatric and mental health nursing*, 17(2):117–123, 2010.

[43] Rosa Redolat Iborra, Mercedes Fernandez Rios, Nuria Martinez, Ana Moron, and Soledad Corachan. Scientific evidence for the use of" serious games" or therapeutic games in people with alzheimer's disease and other dementias. *Technium Soc. Sci. J.*, 12:173, 2020.

[44] Aneeqa Ijaz, Muhammad Yasir Khan, Syed Mustafa Ali, Junaid Qadir, and Maged N Kamel Boulos. Serious games for healthcare professional training: A systematic review. *European Journal of Biomedical Informatics*, 15(1), 2019.

[45] Ihoghosa Iyamu, Oralia Gómez-Ramírez, Alice XT Xu, Hsiu-Ju Chang, Sarah Watt, Geoff Mckee, and Mark Gilbert. Challenges in the development of digital public health interventions and mapped solutions: Findings from a scoping review. *Digital Health*, 8:20552076221102255, 2022.

[46] Alice Jaffray, Conor Finn, and Jason RC Nurse. Sherlocked: A detective-themed serious game for cyber security education. In *International Symposium on Human Aspects of Information Security and Assurance*, pages 35–45. Springer, 2021.

[47] Afzal Javed, Cheng Lee, Hazli Zakaria, Robert D Buenaventura, Marcelo Cetkovich-Bakmas, Kalil Duailibi, Bernardo Ng, Hisham Ramy, Gautam Saha, Shams Arifeen, et al. Reducing the stigma of mental health disorders with a focus on low-and middle-income countries. *Asian journal of psychiatry*, 58:102601, 2021.

[48] Rachel Jenkins, Florence Baingana, Raheelah Ahmad, David McDaid, and Rifat Atun. Social, economic, human rights and political challenges to global mental health. *Mental Health in Family Medicine*, 8(2):87, 2011.

[49] Jianlin Ji. Distinguishing subclinical (subthreshold) depression from the residual symptoms of major depression. *Shanghai Archives of Psychiatry*, 24(5):288, 2012.

[50] Martin Johannes Dechant, Julian Frommel, and Regan Mandryk. Assessing social anxiety through digital biomarkers embedded in a gaming task. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[51] Stephen Michael Johnson. "go. just take him.": Ptsd and the player-character relationship in the last of us part ii. *Games and Culture*, page 15554120221139216, 2022.

[52] W Lewis Johnson, Hannes Högni Vilhjálmsson, and Stacy Marsella. Serious games for language learning: How much game, how much ai? In *AIED*, volume 125, pages 306–313, 2005.

[53] Hélène Kane, Jade Gourret Baumgart, Wissam El-Hage, Jocelyn De-loyer, Christine Maes, Marie-Clotilde Lebas, Donatella Marazziti, Jo-hannes Thome, Laurence Fond-Harmant, Frédéric Denis, et al. Oppor-tunities and challenges for professionals in psychiatry and mental health care using digital technologies during the covid-19 pandemic: systematic review. *JMIR Human Factors*, 9(1):e30359, 2022.

[54] Anya Kaushik, Evgenia Kostaki, and Marinos Kyriakopoulos. The stigma of mental illness in children and adolescents: A systematic review. *Psy-chiatry Research*, 243:469–494, 2016.

[55] Kanyakon Khamwang, Charoenchai Wongwatkit, and Chitphon Yachu-lawetkunakorn. A simulation game to diagnosing major depressive disor-der based on phq-9 for psychiatrist program in thailand: An evaluation of game interface design and prototype. *International Conference on Com-puters in Education*, page 541–549, Dec 2018.

[56] Robert Kohn, Shekhar Saxena, Itzhak Levav, and Benedetto Saraceno. The treatment gap in mental health care. *Bulletin of the World health Organization*, 82(11):858–866, 2004.

[57] Maria Kovacs. Gender and the course of major depressive disorder through adolescence in clinically referred youngsters. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(9):1079–1085, 2001.

[58] Kurt Kroenke. Phq-9: global uptake of a depression scale. *World Psychi-atry*, 20(1):135, 2021.

[59] Kurt Kroenke and Robert L Spitzer. The phq-9: a new depression diag-nostic and severity measure, 2002.

[60] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. The phq-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9):606–613, 2001.

[61] Kurt Kroenke, Robert L Spitzer, Janet BW Williams, and Bernd Löwe. The patient health questionnaire somatic, anxiety, and depressive symp-tom scales: a systematic review. *General hospital psychiatry*, 32(4):345–359, 2010.

[62] Emily G Lattie, Elizabeth C Adkins, Nathan Winquist, Colleen Stiles-Shields, Q Eileen Wafford, and Andrea K Graham. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. *Journal of medical Internet research*, 21(7):e12869, 2019.

[63] Ho Ming Lau, Johannes H Smit, Theresa M Fleming, and Heleen Riper. Serious games for mental health: are they accessible, feasible, and effec-tive? a systematic review and meta-analysis. *Frontiers in psychiatry*, 7:209, 2017.

[64] Paul J Lavrakas. *Encyclopedia of survey research methods.* Sage publications, 2008.

[65] Susanna Lehtimaki, Jana Martic, Brian Wahl, Katherine T Foster, Nina Schwalbe, et al. Evidence on digital mental health interventions for adolescents and young people: systematic overview. *JMIR mental health*, 8(4):e25847, 2021.

[66] Brooke Levis, Andrea Benedetti, and Brett D Thombs. Accuracy of patient health questionnaire-9 (phq-9) for screening to detect major depression: individual participant data meta-analysis. *bmj*, 365, 2019.

[67] Peter M Lewinsohn, Antonette M Zeiss, and Edward M Duncan. Probability of relapse after recovery from an episode of depression. *Journal of abnormal psychology*, 98(2):107, 1989.

[68] Tim MH Li, Michael Chau, Paul WC Wong, Eliza SY Lai, and Paul SF Yip. Evaluation of a web-based social network electronic game in enhancing mental health literacy for young people. *Journal of medical Internet research*, 15(5):e2316, 2013.

[69] MR Liebowitz. Liebowitz social anxiety scale. *Modern Problems of Pharmapsychiatry*, 22:141–173, 1987.

[70] Alice Malpass, Chris Dowrick, Simon Gilbody, Jude Robinson, Nicola Wiles, Larisa Duffy, and Glyn Lewis. Usefulness of phq-9 in primary care to determine meaningful symptoms of low mood: a qualitative study. *British Journal of General Practice*, 66(643):e78–e84, 2016.

[71] Regan Lee Mandryk and Max Valentin Birk. Toward game-based digital mental health interventions: player habits and preferences. *Journal of medical Internet research*, 19(4):e6906, 2017.

[72] Regan Lee Mandryk and Max Valentin Birk. The potential of game-based digital biomarkers for modeling mental health. *JMIR mental health*, 6(4):e13485, 2019.

[73] Jimmie Manning. In vivo coding. *The international encyclopedia of communication research methods*, 24:1–2, 2017.

[74] Kim Martinez, Maria Isabel Menéndez-Menéndez, Andres Bustillo, et al. Awareness, prevention, detection, and therapy applications for depression and anxiety in serious games for children and adolescents: Systematic review. *JMIR serious games*, 9(4):e30482, 2021.

[75] Robert M. McCarron, Bryan Shapiro, Jody Rawles, and John Luo. Depression. *Annals of Internal Medicine*, 2021.

[76] Sally N Merry, Karolina Stasiak, Matthew Shepherd, Chris Frampton, Theresa Fleming, and Mathijs FG Lucassen. The effectiveness of sparx, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. *Bmj*, 344, 2012.

[77] David C Mohr, Francisca Azocar, Andrew Bertagnolli, Tanzeem Choudhury, Paul Chrisp, Richard Frank, Henry Harbin, Trina Histon, Debra Kaysen, Camille Nebeker, et al. Banbury forum consensus statement on the path forward for digital mental health treatment. *Psychiatric Services*, 72(6):677–683, 2021.

[78] Helen Morgan, Amanda O'donovan, Renita Almeida, Ashleigh Lin, and Yael Perry. The role of the avatar in gaming for trans and gender diverse young people. *International journal of environmental research and public health*, 17(22):8617, 2020.

[79] Danielle N Moyer, Kara J Connelly, and Amy L Holley. Using the phq-9 and gad-7 to screen for acute distress in transgender youth: findings from a pediatric endocrinology clinic. *Journal of Pediatric Endocrinology and Metabolism*, 32(1):71–74, 2019.

[80] Lawrence T Park and Carlos A Zarate Jr. Depression in the primary care setting. *New England Journal of Medicine*, 380(6):559–568, 2019.

[81] Frank Peeters, Marcus JH Huibers, J van der Bout, and S Perreijn. Depressieve-en bipolaire-stemmingsstoornissen. In *Klinische psychologie, theorieën en psychopathologie*. Bohn Stafleu van Loghum, 2015.

[82] Joseph H Pleck, Freya L Sonenstein, and Leighton C Ku. Attitudes toward male roles among adolescent males: A discriminant validity analysis. *Sex roles*, 30(7):481–501, 1994.

[83] Aung Pyae. Understanding the role of culture and cultural attributes in digital game localization. *Entertainment computing*, 26:105–116, 2018.

[84] Jerica Radez, Tessa Reardon, Cathy Creswell, Peter J Lawrence, Georgina Evdoka-Burton, and Polly Waite. Why do children and adolescents (not) seek and access professional help for their mental health problems? a systematic review of quantitative and qualitative studies. *European child & adolescent psychiatry*, 30:183–211, 2021.

[85] Ilya Razykov, Roy C Ziegelstein, Mary A Whooley, and Brett D Thombs. The phq-9 versus the phq-8—is item 9 useful for assessing suicide risk in coronary artery disease patients? data from the heart and soul study. *Journal of psychosomatic research*, 73(3):163–168, 2012.

[86] Byron Reeves and Clifford Nass. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK*, 10:236605, 1996.

[87] Christian Reuter, Viktor Wendel, Stefan Göbel, and Ralf Steinmetz. Multiplayer adventures for collaborative learning with serious games. In *6th European Conference on Games Based Learning*, pages 416–423, 2012.

[88] Jeroen Ruwaard, Lisa Kooistra, and Melissa Thong. Ecological momentary assessment in mental health research: A practical introduction, with examples in r (–build 2018-11-26). *Amsterdam: APH Mental Health*, 2018.

[89] Elke A Schoneveld, Monique Malmberg, Anna Lichtwarck-Aschoff, Geert P Verheijen, Rutger CME Engels, and Isabela Granic. A neurofeedback video game (mindlight) to prevent anxiety in children: A randomized controlled trial. *Computers in Human Behavior*, 63:321–333, 2016.

[90] Johanna Schröder, Thomas Berger, Björn Meyer, Wolfgang Lutz, Martin Hautzinger, Christina Späth, Christiane Eichenberg, Jan Philipp Klein, and Steffen Moritz. Attitudes towards internet interventions among psychotherapists and individuals with mild to moderate depression symptoms. *Cognitive therapy and research*, 41:745–756, 2017.

[91] Stefan Schubert. Playing as/against violent women: Imagining gender in the postapocalyptic landscape of the last of us part ii. In *Gender Forum*, volume 80, pages 30–54, 2021.

[92] Raphael Schuster, Raffaela Pokorny, Thomas Berger, Naira Topooco, and Anton-Rupert Laireiter. Advantages and disadvantages of online and blended therapy: Attitudes towards both interventions amongst licensed psychotherapists in austria. *Journal of medical internet research*, 20(12):e11007, 2018.

[93] Shefaly Shorey, Esperanza Debby Ng, and Celine HJ Wong. Global prevalence of depression and elevated depressive symptoms among adolescents: A systematic review and meta-analysis. *British Journal of Clinical Psychology*, 61(2):287–305, 2022.

[94] Henny Sinnema, Trimbos-instituut, Matthijs Oud, and Daniëlle van Duin. *Handleiding voor de implementatie van de multidisciplinaire richtlijn depressie bij jeugd*. Trimbos-instituut, 2013.

[95] Kerri Smith and IBC De Torres. A world of depression. *Nature*, 515(181):10–1038, 2014.

[96] Robert L Spitzer, Janet BW Williams, Kurt Kroenke, Mark Linzer, Frank Verloin deGruy, Steven R Hahn, David Brody, and Jeffrey G Johnson. Utility of a new procedure for diagnosing mental disorders in primary care: the prime-md 1000 study. *Jama*, 272(22):1749–1756, 1994.

[97] GGZ Standaarden. Depressive stoornissen, 2020.

[98] Murray B Stein, Martina Fuetsch, Nina Müller, Michael Höfler, Roselind Lieb, and Hans-Ulrich Wittchen. Social anxiety disorder and the risk of depression: a prospective community study of adolescents and young adults. *Archives of general psychiatry*, 58(3):251–256, 2001.

[99] Penelope Strauss, Helen Morgan, Dani Wright Toussaint, Ashleigh Lin, Sam Winter, and Yael Perry. Trans and gender diverse young people's attitudes towards game-based digital mental health interventions: A qualitative investigation. *Internet Interventions*, 18:100280, 2019.

[100] Tarja Susi, Mikael Johannesson, and Per Backlund. Serious games: An overview. *Developing Industrial Strategies Through Innovative Cluster and Technologies*, 2007.

[101] Julio C Tolentino and Sergio L Schmidt. Dsm-5 criteria and depression severity: implications for clinical practice. *Frontiers in psychiatry*, 9:450, 2018.

[102] Athanasios Tsionas, Aristotelis Lazaridis, and Ioannis Vlahavas. Serious game development for the diagnosis of major depressive disorder cases using machine learning methods, 2020.

[103] SG van Gunst and VG Pigmans. *NHG-standaarden voor de praktijkassistente 2014*. Bohn Stafleu van Loghum, 2014.

[104] Marinus Johannes Petrus Maria Verbraak, Sako Visser, PEHM Muris, and K Hoogduin. *Handboek voor gz-psychologen*. Boom, 2020.

[105] Christiaan Vis, Mayke Mol, Annet Kleiboer, Leah Bührmann, Tracy Finch, Jan Smit, Heleen Riper, et al. Improving implementation of emental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. *JMIR mental health*, 5(1):e9769, 2018.

[106] WHO. Mental disorders, 2022.

[107] Leonie Wolf. Depression and digital games. an investigation of existing uses of therapy games. *Violence — Perception — Video Games*, page 151–170, 2019.

[108] Robynn Zender and Ellen Olshansky. Women's mental health: depression and anxiety. *Nursing Clinics*, 44(3):355–364, 2009.

[109] Alina Zuberi, Ahmed Waqas, Sadiq Naveed, Md Mahbub Hossain, Atif Rahman, Khalid Saeed, and Daniela C Fuhr. Prevalence of mental disorders in the who eastern mediterranean region: A systematic review and meta-analysis. *Frontiers in Psychiatry*, page 1035, 2021.

A   Appendix: PHQ-8

The PHQ-8 can be found in Table 3.

Over the last 2 weeks, how often have you been bothered by any of the following problems?	Not at all	Several days	More than half of the days	Nearly every day
Little interest or pleasure in doing things				
Feeling down, depressed, irritable or hopeless				
Trouble falling asleep or staying asleep, or sleeping too much				
Feeling tired or having little energy				
Poor appetite or overeating				
Feeling bad about yourself, or that you are a failure or have let yourself or your family down				
Trouble concentrating on things such as school work, reading or watching television				
Moving or speaking so slowly that other people could have noticed. Or the opposite, being so fidgety or restless that you have been moving around a lot more than usual				

Table 3: *PHQ-8*

B Appendix: Translation of PHQ-8 into dialogue

The translation of the PHQ-8 items into the dialogue can be found in Table 4.

Original item	Prompt
Little interest or pleasure in doing things	I need to find something new that brings me pleasure. Lately, I just have little pleasure or interest in doing things. Do you ever feel like that no matter what you do, it brings you no pleasure?
Feeling down, depressed, irritable or hopeless	I really think that some time for myself will help me. Making some time for self-care could maybe allow me to have a good time again. I don't want to feel this down, depressed and hopeless anymore.
Trouble falling asleep or staying asleep	I really have trouble falling asleep lately. Once I go to bed, I just stare at the ceiling all night.
Sleeping too much	I actually never have trouble falling asleep. You know what my problem is, sleeping way too much! Last week, I missed my morning alarm 4 times!
Feeling tired or having little energy	Pixie, remember how I used to run up and down the hill to go to the river? Well, now I really do not have the energy for that anymore.
Overeating	I am wondering if I am the only one who feels like overeating. Do you ever experience that?
Poor appetite	And what about losing your appetite?
Feeling bad about yourself, or that you are a failure or have let yourself or your family down	I don't get it, all the others are catching fish and I am not. Ugh, I always feel like such a failure and that I have let myself down. I don't know how to handle these situations. Do you ever feel like this?
Trouble concentrating on things such as school work, reading or watching television	Maybe it is just me who cannot concentrate anymore. Do you ever have trouble concentrating on things?
Moving or speaking so slowly that other people could have noticed	Well, some people also experience the opposite of what I feel. They just move or speak very slowly that other people can even notice that. Is that something that you experience?

Or the opposite, being so fidgety or restless that you have been moving around a lot more than usual	Lately, I have been feeling very fidgety and restless. I just cannot stand still or focus on what I am doing. Are you ever feeling restless or fidgety?

Table 4: *Dialogue for the assessment method, derived from the PHQ-8*

C Appendix: Dialogue script

Scene of the town Roaring River:

Game narrator: "Welcome to Roaring River [NAME]! This small town is home to dozens of fishermen and merchants. Even though the town is small, it is bustling!

Scene of a map where Roaring River is shown:

Game narrator: "The town of Roaring River was once a quiet town. It grew in power as it was visited by many fishermen and merchants when they heard about the rare and imperiled species of fish in the river."

Back to scene of the town Roaring River:

Game narrator: "Roaring River is becoming more popular than ever! Adventure awaits! Let's explore everything that this town has to bring."

New scene of Roaring River:

Game narrator: "As many people visit the small town of Roaring River, there are lots of different guides that will explore the town and surroundings with you!"

Game narrator: "What would you like to do today?"

Player options:

1. Let's explore the trails around Roaring River!

2. Show me what all these merchants offer in this town!

3. I want to catch some rare fish species in the river!

Game narrator: "Good choice! Let's pick a guide"

Screen where player sees the guides:

Player options:

1. Guide 1: Scout

2. Guide 2: Hawk

3. Guide 3: Pixie

Zoomed in photo of Scout in Roaring River:

Game narrator: "Great! Enjoy your tour, you can find Scout at the market. Your adventure will start there."

Player option: Go to the market

Scene somewhere in town:

Scout: "Good to see you [NAME]! I am Scout, born and raised in Roaring River."

Scout: "I believe I spend more time near the river than I do at my own house, so you can ask me anything on this trip."

Scout: "Alright, follow me!"

Player option: Follow Scout

Market scene 2:

Scout: "Our first stop is the market. This is where all the merchants from Roaring River and nearby towns sell their fish, fruits, vegetables and more!"

Scout: "We need to bring some food for on the way."

Scout: "Lately I have been eating very strange. One day I am overeating and the other day, I completely lose my appetite."

Scout: "Maybe it's for the best if you choose something for us on this trip."

Player options:

1. Bring sandwiches from the bakery

2. Get some fresh fruit from the greengrocer

3. Pick some sweets from the candy shop

Fresh fruit shop scene:

Scout: "Good choice. I am wondering if I am the only one who feels like overeating. Do you ever experience that?"

Player options:

1. No, I am not experiencing this myself.

2. I also feel like overeating sometimes.

3. Yes, I am overeating on many days.

4. A common problem for me, I overeat nearly every day.

Scout: "And what about losing your appetite."

Player options:

1. No, I am not experiencing this myself.

2. I sometimes lose my appetite.

3. Yes, I lose my appetite on many days.

4. I know this problem, I lose my appetite nearly every day.

TWO OPTIONS FOR ANSWERS:

Scout option 1 (IF answer 1 for overeating appetite): "Well, good to hear that you do not have this problem! Let's get us some fruit."

Scout option 2 (IF answer 2/2/3): "Thank you for sharing that with me. Let's get us some fruit for on the way."

Hike 1, leaving town:

Scout: "[NAME], this is the first part of the hike to the river. Look around you, maybe we can spot some peace lily."

Scout: "The original citizens of Roaring River believed that this plant brings good luck. Now, Roaring River is well known for this plant and it is sold by many merchants!"

Scout: "If we take a back path of the trail, we can go look for some. Follow me!

Searching for peace lilies:

Scout: "Uhh, they are not here. Come quick, let's find another place."

Searching for peace lilies new place:

Scout: "Well, I used to always find them here. Wait let me just search again."

Searching for peace lilies new place "Scout far in the background":

Scout: "I cannot calm my mind anymore. I need to keep searching. . . ."

Scout back on the trail for the hike:

Scout: "Lately, I have been feeling very fidgety and restless. I just cannot stand still or focus on what I am doing. Are you ever feeling restless or fidgety?"

Player options:

1. No, I never feel restless or fidgety.

2. Yes, sometimes I do.

3. Yes, I am experiencing that on many days.

4. Yes, nearly every day.

Scout: "Well, some people also experience the opposite from what I feel. They just move or speak very slowly that other people can even notice that."

Scout: "Is that something that you experience?"

Player options:

1. No, I never speak or move very slowly.

2. Yes, sometimes I do.

3. Yes, I am doing that on many days.

4. Yes, nearly every day.

TWO OPTIONS FOR ANSWERS:

Scout option 1 (IF answer 1): "That is good to hear. Let's just continue our journey, maybe I will then feel less restless."

Scout option 1 (IF answer 2/3/4): "Thank you again for sharing it with me. It always feels good to talk about this with others. Let's continue!"

Scout and player meeting another guide:

Scout: "Pixie!"

Pixie: "Scout! Good to see you here."

Pixie: "I was just about to have a short break to eat something, do you want to join me?"

Scout: "Yes sure, I am here with [NAME]. We are on our way to the river but also need to eat the food we bought on the market."

Scout, pixie and player around campfire:

Scout: "Pfoe, it is good to finally sit down!"

Scout: "[NAME], I used to do this hike to the river every day. And I have been a guide for 10 years! So that would mean......let me think"

Scout: "3650 times! I think...."

Scout: "Pixie, remember how I used to run up and down the hill to go to the river? Well, now I really do not have the energy for that anymore."

Pixie: "Why is that Scout?"

Scout: "Maybe it is because I really have trouble falling asleep lately. Once I go to bed, I just stare to the ceiling all night."

Pixie: "I sometimes also feel tired or feel like I have less energy like you do Scout. What about you [NAME]?

Player options:

1. No, I never feel tired or have little energy.

2. Yes, sometimes I do.

3. Yes, I am tired and less energized on many days.

4. Yes, nearly every day.

TWO OPTIONS FOR ANSWERS:

Scout option 1 (IF answer 1): "I am happy to hear that. But what about having trouble falling asleep?"

Scout option 1 (IF answer 2/3/4): "I am glad to know that you both feel the same sometimes. Do you also sometimes have trouble falling asleep [NAME]?"

Player options:

1. No, I do not have trouble falling asleep.

2. Yes, sometimes I do.

3. Yes, on many days I find it difficult to fall asleep.

4. Yes, nearly every day.

Pixie: "Scout and [NAME], I actually never have trouble falling asleep. You know what my problem is, sleeping way too much! Last week, I missed my morning alarm 4 times..!"

Scout: "That is though Pixie! Some people also experience both. They never fall asleep but also oversleep at the same time. Do you feel like ever sleeping too much [NAME]?
Player options:

1. No, I never oversleep.

2. Yes, sometimes I do.

3. Yes, on many days I oversleep.

4. Yes, nearly every day.

TWO OPTIONS FOR ANSWERS:
Scout option 1 (IF answer 1): "Thanks for sharing that with us. It feels good to talk about it with you"
Scout option 1 (IF answer 2/3/4): "Wow, I really feel relieved to talk about this with you. It already feels like we have known each other for so long!"

At the river:
Scout: "Finally, [NAME] we made it! Have you ever tried fishing before?"
Player options:

1. Yes, I am basically an expert!

2. Yes, but not that many times.

3. No, never in my life.

4. No, but I really want to do it once.

TWO OPTIONS FOR ANSWERS:
Scout option 1 (IF answer 1/2): "Cool! Then you can go catch some fish first!"
Scout option 1 (IF answer 3/4): "No worries, I will teach you today."
Scout: "By the way, I am a three times champion Roaring River champion in fishing. So don't worry [NAME] we will catch some fish!"

At the river fishing:
Scout: "Uhh, usually we would have caught something by now."
Scout: "I don't get it, all the others are catching fish and I am not..."
Scout: "Ugh, I always feel like such a failure and that I have let myself down. I don't know how to handle these situations. Do you ever feel like this [NAME]?
Player options:

1. No, I never feel like I let myself down or that I am a failure.

2. Yes, sometimes I do feel like that.

3. Yes, I feel like that on many days.

4. Yes, nearly every day I feel like that.

TWO OPTIONS FOR ANSWERS:

Scout option 1 (IF answer 1): "I am happy to hear that you are not feeling this way. OKAY, let's focus again!!"

Scout option 1 (IF answer 3/4): "I am sorry to hear that. Let's try to make this a positive day and focus again!!"

New fishing scene:
Scout: "Okay uhm let me think, what can we change?"
Scout: "Okay focus Scout, what do you need...."
Scout: "I cannot concentrate on what to do now. Maybe you can just give it a try."

Screen where player catches fish:
Scout: "Wow, AWESOME [NAME]. You did it!

Back to "new fishing scene":
Scout: "Maybe it is just me who cannot concentrate anymore."
Scout: "Do you ever have trouble concentrating on things?"
Player options:

1. No, I never have trouble concentrating on things.

2. Yes, sometimes I do struggle to concentrate.

3. Yes, find it difficult to concentrate on many days.

4. Yes, nearly every day I have that problem.

TWO OPTIONS FOR ANSWERS:

Scout option 1 (IF answer 1): "That is good to hear. Congratulations on catching your first fish!"

Scout option 1 (IF answer 3/4): "I am sorry that you are also going through this. But you did just catch your first fish, that is so cool!"

Storytelling about rare fish:
Scout: "Did you know that we call this fish the Whaliathan? The original citizens of Roaring River gave this fish this name and it means "fortune fish".

It was all about luck back in the day."

Back to "new fishing scene":
Scout: "[NAME], I hope you enjoyed this day. I think this is going to be my last trip for now. I need some time off."

Scout: "I need to find something new that brings me pleasure. Lately I just have little pleasure or interest in doing things."

Scout: "Do you ever feel like that no matter what you do, it brings you no pleasure?"

1. No, I never feel like that.

2. Yes, sometimes I do feel like that.

3. Yes, on many days I feel like nothing gives me pleasure.

4. Yes, nearly every day I feel like that.

TWO OPTIONS FOR ANSWERS:
Scout option 1 (IF answer 1): "I am happy for you. I feel like we are already becoming friends!"

Scout option 1 (IF answer 3/4): "Thank you for sharing that with me. I do feel like better days are coming."

Saying goodbye:
Scout: "I really think that some time for myself will help me. Making some time for self-care could maybe allow me to have a good time again. I don't want to feel this down, depressed and hopeless anymore."

Scout: "Is that something that you experience?

1. No, I never feel like that.

2. Yes, sometimes I do feel down, depressed or hopeless.

3. Yes, on many days I feel like that.

4. Yes, nearly every day I feel like that.

TWO OPTIONS FOR ANSWERS:
Scout option 1 (IF answer 1): "Thank you for sharing so much with me today. Even though I am not feeling the best, you made my day. I would love to see you again in Roaring River!"

Scout option 1 (IF answer 3/4): "Know that it is always okay to ask for help. I will always be there for you. Even though I am not feeling the best, you made my day. I would love to see you again in Roaring River!"

D   Appendix: Information sheet

Introduction

I would like to invite you to participate in this research study. Before you decide to take part, you need to understand why this research is being performed and what it would mean for you as a participant. Please take the time to read the following information carefully. You are free to ask questions about anything that you need if something is not clear or if you would like more information. Take the time to decide whether or not you would like to participate.

What is the background and purpose of this study

In this research, a dialogue will be developed that adolescents can play on their computer. In this dialogue, players will talk with a character. This character is expressing that it has symptoms of depression. The character asks the player where it is experiencing the same or not. The symptoms and formulation of the questions is based on an existing questionnaire for the assessment of depression symptoms, the Patient Health Questionnaire-8. The goal of this dialogue is to assist the diagnosis of depression by creating a stimulating and distracting environment for adolescents in which they can express their feelings. Important to note is that this dialogue is not a replacement of face-to-face diagnosis, but rather an additional tool for diagnosis that can take place before, during and after diagnosis. This research aims to develop an understanding of the opportunities and challenges regarding the implementation of such dialogue into practice. This understanding is obtained by conducting interviews with professionals who have experience with diagnosing adults/adolescents/children with depression.

Who will carry out the study?

This study is carried out by Kim Kroes, k.m.d.kroes@students.uu.nl, as part of my master's thesis under the supervision of Dr. Julian Frommel, j.frommel@uu.nl. This research is performed as partial fulfilment of the requirements for the degree of Master of Science in the Master's program of Science and Business Management at Utrecht University.

How will the study be carried out?

In this study, you will be interviewed. In this interview, questions will be asked about your experience with diagnosing depression and/or identifying symptoms of depression. You will also be asked to address any experience with using computerized therapies or diagnostic tools for depression. After these questions, the dialogue that is created will be shown to you. Afterwards, your general attitude

towards the dialogue will be evaluated and you will be asked to identify possible benefits and disadvantages. Moreover, your opinion about the acceptability and usability of the dialogue will also be evaluated. Lastly, your thoughts about the implementation of the dialogue in practice will be discussed. The interview can take place either online or remotely, based on your preferences and time. The entire interview will take approximately 50-60 minutes.

What will we do with your data?

If you consent to this, an audio recording will be made. The recording will be transcribed so that participants' opinions about the dialogue are captured into text. The audio will be securely deleted after transcription (within 2 months of the study). The transcribed text will be anonymized so that you will not be identifiable. The transcript will become part of my thesis. My thesis will not include your name or any other individual information by which you could be identified.

What are your rights?

Participation is voluntary. We are only allowed to collect your data for our study if you consent to this. If you decide not to participate, you do not have to take any further action. You do not need to sign anything. Nor are you required to explain why you do not want to participate. If you decide to participate, you can always change your mind and stop participating at any time, including during the study. You will even be able to withdraw your consent after you have participated. However, if you choose to do so, we will not be required to undo the processing of your data that has taken place up until that time. The personal data we have obtained from you up until the time when you withdraw your consent will be erased (where personal data is any data that can be linked to you, so this excludes any already anonymised data).

Approval of this study

This study has been allowed to proceed by the Research Institute of Information and Computing Sciences on the basis of an Ethics and Privacy Quick Scan. If you have a complaint about the way this study is carried out, please send an email to: ics-ethics@uu.nl. If you have any complaints or questions about the processing of personal data, please send an email to the Faculty of Sciences Privacy Officer: privacy-beta@uu.nl. The Privacy Officer will also be able to assist you in exercising the rights you have under the GDPR. For details of our legal basis for using personal data and the rights you have over your data, please see the University's privacy information at www.uu.nl/en/organisation/privacy.

More information about this study?

If you have any questions or concerns about this research, please contact Kim Kroes at k.m.d.kroes@students.uu.nl or my supervisor Dr. Julian Frommel at j.frommel@uu.nl. Questions can be asked and additional information can be obtained before, during and after the study.

E Appendix: Consent form and survey

Consent form

Please complete the form below by ticking the relevant boxes and signing on the line below. A copy of the completed form will be given to you for your own record.

- I confirm that the research project "*Using non-player character dialogues to identify symptoms of depression in adolescence*" has been explained to me. I have had the opportunity to ask questions about the project and have had these answered satisfactorily. I had enough time to consider whether to participate.

- I consent to the material I contribute being used to generate insights for the research project "*Using non-player character dialogues to identify symptoms of depression in adolescence*".

- I consent to audio recordings being used in this study as explained in the information sheet. I understand that I can request to stop recordings at any time.

- I understand that if I give permission, the audio recordings will be held confidentially so that only Kim Kroes and Dr. Julian Frommel have access to the recordings. The recordings will be held in a secure and password protected storage for up to 10 years after which period they will be securely destroyed. In accordance with the General Data Protection Regulation (GDPR) I can have access to my recordings and can request them to be deleted at any time during this period.

- I understand that my participation in this research is voluntary and that I may withdraw from the study at any time without providing a reason, and that if I withdraw any personal data already collected from me will be erased.

- I consent to allow the fully anonymized data to be used in future publications and other scholarly means of disseminating the findings from the research project.

- I understand that the data acquired will be securely stored by researchers, but that appropriately anonymized data may in future be made available to others for research purposes. I understand that the University may publish appropriately anonymized data in appropriate data repositories for verification purposes and to make it accessible to researchers and other research users.

- I agree to take part in the above research project on "*Using non-player character dialogues to identify symptoms of depression in adolescence*".

Survey

Below you will find a number of questions that provide insight into your background characteristics. We collect this to assess demographic data. You are not required to complete every question.

1. How do you identify?

    (a) Man

    (b) Woman

    (c) Non-binary

    (d) Prefer not to say

    (e) Prefer to self-describe

2. What is your age?

3. What did you study?

4. What is your profession?

5. How long have you been practicing this profession?

F Appendix: Ethics and Privacy Scan output

Section 1. Research projects involving human participants

- P1. Does your project involve human participants? This includes for example use of observation, (online) surveys, interviews, tests, focus groups, and workshops where human participants provide information or data to inform the research. If you are only using existing data sets or publicly available data (e.g. from Twitter, Reddit) without directly recruiting participants, please answer no.

    – Yes

- P2. Does your project involve participants younger than 18 years of age?

    – No

- P3. Does your project involve participants with learning or communication difficulties of a severity that may impact their ability to provide informed consent?

    – No

- P4. Is your project likely to involve participants engaging in illegal activities?

    – No

- P5. Does your project involve patients?

    – No

- P6. Does your project involve participants belonging to a vulnerable group, other than those listed above?

    – No

- P8. Does your project involve participants with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues, or clients?

    – No

Informed consent

- PC1. Do you have set procedures that you will use for obtaining informed consent from all participants, including (where appropriate) parental consent for children or consent from legally authorized representatives? (See suggestions for information sheets and consent forms on the website.)

– Yes

- PC2. Will you tell participants that their participation is voluntary?

    – Yes

- PC3. Will you obtain explicit consent for participation?

    – Yes

- PC4. Will you obtain explicit consent for any sensor readings, eye track-ing, photos, audio, and/or video recordings?

    – Yes

- PC5. Will you tell participants that they may withdraw from the research at any time and for any reason?

    – Yes

- PC6. Will you give potential participants time to consider participation?

    – Yes

- PC7. Will you provide participants with an opportunity to ask questions about the research before consenting to take part (e.g. by providing your contact details)?

    – Yes

- PC8. Does your project involve concealment or deliberate misleading of participants?

    – No

Section 2. Data protection, handling, and storage

The General Data Protection Regulation imposes several obligations for the use of personal data (defined as any information relating to an identified or identifiable living person) or including the use of personal data in research.
- D1. Are you gathering or using personal data (defined as any information relating to an identified or identifiable living person)?

    – No

Section 3. Research that may cause harm

Research may cause harm to participants, researchers, the university, or society. This includes when technology has dual-use, and you investigate an innocent use, but your results could be used by others in a harmful way. If you are unsure regarding possible harm to the university or society, please discuss your concerns with the Research Support Office.

- H1. Does your project give rise to a realistic risk to the national security of any country?

    – No

- H2. Does your project give rise to a realistic risk of aiding human rights abuses in any country?

    – No

- H3. Does your project (and its data) give rise to a realistic risk of damaging the University's reputation? (E.g., bad press coverage, public protest.)

    – No

- H4. Does your project (and in particular its data) give rise to an increased risk of attack (cyber- or otherwise) against the University? (E.g., from pressure groups.)

    – No

- H5. Is the data likely to contain material that is indecent, offensive, defamatory, threatening, discriminatory, or extremist?

    – No

- H6. Does your project give rise to a realistic risk of harm to the researchers?

    – No

- H7. Is there a realistic risk of any participant experiencing physical or psychological harm or discomfort?

    – No

- H8. Is there a realistic risk of any participant experiencing a detriment to their interests as a result of participation?

    – No

- H9. Is there a realistic risk of other types of negative externalities?

    – No

Section 4. Conflicts of interest

- C1. Is there any potential conflict of interest (e.g. between research funder and researchers or participants and researchers) that may potentially affect the research outcome or the dissemination of research findings?

    – No

- C2. Is there a direct hierarchical relationship between researchers and participants?

    – No

Scoring

- Privacy: 0

- Ethics: 0

G Appendix: Interview protocol

This interview protocol gives an overview of the questions that were used during the interviews. It is considered a guide used during the interviews. Sometimes, additional questions were asked for clarification or if participants addressed a certain new topic. These questions are not included in the interview protocol.

Diagnosis

1. Can you introduce yourself and share your experiences with diagnosing and treating depression in young people?

2. What steps do you usually take when diagnosing depression in young people?

3. Do you typically use standardized questionnaires, such as the CDI-2 used in your research, in the diagnostic process? If yes, which questionnaires and why? If not, why not?

4. What do you consider the biggest challenges or limitations in diagnosing depression in young people?

Digital therapies and diagnostics

1. Do you also use digital therapies yourself, and if so, how do you integrate them into your treatment plan?

2. What types of digital therapies do you use personally?

3. Could you briefly explain the challenges you encounter when using these digital therapies? OR Could you summarize the challenges you see in using these digital therapies?

4. How do you address these challenges?

5. What would be necessary to increase the practical use of these digital therapies, such as games?

6. In my own research, I have seen few computerized diagnostic interventions. What do you think could be the reason for this?

Evaluation

For the evaluation, the dialogue was partially shown to the participants. I explained the following to the participants: The current development of the dialogue can be seen as the initial phase. Since I don't have a background in psychology, it is important to incorporate your vision and opinion in further developing this dialogue, with the ultimate step being validation.

I would like to learn more about your thoughts on the dialogue. The goal of the research is also to explore how it could be further developed. In my research, I am already focusing on potential challenges and barriers to integrating this dialogue into practice, which can also be considered in further development.

1. What is your general initial impression of the developed dialogue?

2. Do you think it would be a useful addition to the current diagnosis? If yes, in what way? If no, why not?

3. How do you think the dialogue could potentially help with the current limitations you experience in diagnosing depression?

Challenges and integration/Suggestions and improvements

1. What challenges do you see regarding the integration of this dialogue into practice?

2. What potential limitations of this dialogue do you see for identifying depression in young people?

3. What precautions should be taken to ensure the effectiveness and safety of this dialogue in practice?

4. What is your perspective on the potential benefits of the dialogue for identifying symptoms of depression in young people?

5. What is your perspective on reducing social desirability bias during the completion of standardized questionnaires through this dialogue?

6. Do you have any suggestions for the dialogue to improve it?

7. Are there topics that we have not yet addressed but are important to address?