From the Division of International Health (IHCAR),
Department of Public Health Sciences,
Karolinska Institutet, Stockholm, Sweden

Turning understanding into clinical practice
**An intervention in continuing professional development
based on readiness to change**

Mandana Shirazi

Abstract

BACKGROUND: Current continuing medical education (CME) programmes are often insufficient in changing doctors' performance, and there is a need to improve CME and shift toward a more comprehensive continuing professional development (CPD).The aim of this study was to develop and assess the effects of an educational intervention, based on a modified stages-of-change model, on general physicians' stages of readiness to change, knowledge, attitudes and performance regarding management of depressive disorders in Iran. **METHOD:** A randomized controlled trial with 192 general physicians in primary care (GPs), who were equally distributed to an intervention and control arm, following stratification related to stage of change, sex, age and work experience. The intervention comprised an interactive workshop for a small group at a higher stage of readiness-to-change ('intention') and an interactive large group meeting for those demonstrating a lower propensity to change ('attitudes') at the pre-assessment stage. **MEASURES**: All the measures were validated in the Iranian context. The GPs' stages-of-change were assessed based on the Modified Prochaska Questionnaire (MPQ), their knowledge and attitudes were assessed using written questionnaires and their performance was assessed by standardized patients (SPs), who filled in checklists regarding the encounter, and collected prescriptions if any. Five different scenarios for depression disorders were compiled by an expert group. The validity and reliability of checklists, SPs' portrayals and SPs' ways of completing checklists were documented. The pre-assessment of GPs' performance was done two months before and the post-assessment two months after the intervention. **RESULTS**: GPs in the intervention arm significantly shifted to a higher stage with an intervention effect of 47 percentage units. Their overall mean scores on the knowledge test also improved, with an intervention effect of 12 percentage units. Although their attitudes changed in the post-test in comparison with the pre-test, the difference between the intervention and control arms was not significant. The performance of the GPs in the intervention arm also improved for mean scores regarding diagnosis, with an intervention effect of 14 percentage units, and for appropriate management regarding treatment and referral, with an intervention effect of 20 percentage units. The largest changes appeared in the small intervention group with intervention effects of 28 and 38 percentage units, respectively. **CONCLUSIONS:** The model was successful in improving both knowledge and practice according to the theoretical assumptions. It can be used in educational interventions within a CPD context.

Key words: Continuing Professional Development, Continuing Medical Education, Stages of change, Depression, Primary care and Iran.

List of Publications

I. Shirazi M, Assadi SM, Sadeghi M, Zeinaloo AA, Kashani AS, Arbabi, M, Alaedini F, Lonka K and Wahlstrom R. Applying a modified Prochaska's model of readiness to change for general practitioners on depressive disorders in CME programmes: validation of tool. Journal of Evaluation in Clinical Practice. 2007;13(2):298-302.

II. Shirazi M, Zeinaloo AA, Parikh SV, Sadeghi M, Taghva A, Arbabi M, Kashani AS, Alaeddini F, Lonka K and Wahlström R. Effects on readiness to change of an educational intervention on depressive disorders for general physicians in primary care based on a modified Prochaska model: a randomized controlled study. Family Practice. 2008 Apr;25(2):98-104.

III. Shirazi M, Parikh SV, Alaeddini F, Lonka K, Zeinaloo AA, Sadeghi M, Arbabi M, Nejatisaf AA, Sharivar Z, Wahlstrom R. Effects on knowledge and attitudes of using stages of change to train general practitioners on management of depression: a randomized controlled trial (Submitted).

IV. Shirazi M, Lonka K, Parikh SV, Zeinaloo AA, Alaeddini F, Sadeghi M, Wahlstrom R. A learner-centred stages-of-change model in educational intervention improves doctors' performance in managing depression: a randomized controlled trial (Submitted).

The articles will be referred to by their Roman numerals.

Contents

List of abbreviations

CME	Continuous Medical Education
CONSORT	Consolidated Standards of Reporting Trials
CPD	Continuous Professional Development
DSM IV	Diagnostic and Statistical Manual of Mental Disorders
GP	General Physicians working in primary care
MPQ	Modified Prochaska Questionnaire
RCT	Randomized Controlled Trial
SP	Standardized Patient
TUMS	Tehran University of Medical Sciences
WHO	World Health Organization
WPA	World Psychiatry Association

PREFACE

My background is midwifery. I graduated with a Master degree from Shaheed Beheshti school of midwifery in 1993. From 1993 until 1998, I was working as a faculty member of Tehran University of Medical Sciences (TUMS) in School of Midwifery. Since 1998 I have worked as a lecturer at the Educational Development Centre (EDC) of TUMS. I was later promoted to the post of Executive Manger of the CME (Continuous Medical Education) office of TUMS. The focus of most of the research I have done in this field, has been on needs assessment. The results showed that doctors are not satisfied with the current CME programs and also that there is a gap between their desired needs and the needs based on the experts' opinion. Generally, I have found a real demand in this area for understanding the physicians' real needs and designing the programs based on these identified needs.

In 2003 I held a WHO scholarship in medical education for two months and I went to New South Wales University in Sidney, Australia. That was a good opportunity for me to be familiar with their CME organizations, including the needs assessment methods and different types of CME activities that they provide. The policies of the CME organization in New South Wales have some similarities with Iran because of the compulsory policy regarding the participation of doctors in CME programmes. At the same time I did a review of the literature in the CME field, and I found the Prochaska model to be an appropriate paradigm for implementation in the Iranian context. Based on the results of studies performed in Canada, this model was successful in their context for designing a CME programme based on the doctors' needs. The assessment showed that participants changed their stages of change after attending in a tailored programme, although the research design should be improved. Then I tried to write a brief proposal for a study and also find qualified experts in the field of Medical Education to help me to develop a robust design and implement it in the Iranian context. I found Karolinska Institutet as a high degree research institute and I applied to continue my education in this field in order to answer two questions in my mind: " How can we improve doctors' performance through continuing educational programmes? And how can we shift from CME toward Continuous Professional Development (CPD) in Iran?" Fortunately, I found knowledgeable supervisors, who kindly accepted to work with me during these years. We had a very competent research group. Our team has a multi-professional perspective, such as family practice, medical

education, continuing education, epidemiology, psychiatry and also one scientist with the experience of testing the stages of change model for the first time in the Canadian context. I hope that the findings of this study can help the Iranian health policy makers to focus more on physicians' performance in their continuing education in order to make a shift towards CPD and improve health care outcomes.

1 BACKGROUND

Quality improvement in health care and quality management are common concepts affecting all professionals in the field. Different factors have influence on the health care quality such as organization, policy and education. Quality has become a critical issue in all sections of education including the health sciences. Quality of education is frequently addressed when the health system faces dilemmas regarding insufficient competence of health professionals or when the educational institutions want to plan their future programmes.

One of the important indicators which reveal the quality of education of health professionals is the outcome of the health care system. The outcome could be assessed based on the care provided and patients' outcomes. To ensure the quality of health care, it is a worldwide concern to assess the effects of continuing medical education (CME) or continuing professional development (CPD) not only on physicians' knowledge and skills but also on their performance and clinical outcomes. One important challenge in the health system is that conventional CME programmes have limited effect on physicians' performance (Davis, Thomson et al. 1992; Davis, Thomson et al. 1995). Hence there is a real need for more in-depth studies of the effects on physicians' performance of practical interventions based on different theories.

This thesis is about the implementation and assessment of the effects of an intervention of an based on the readiness-to-change theory on physicians' stages of change, knowledge, attitudes and practice in the CPD context in comparison with current CME programmes in Iran.

1.1 Problem of transfer between knowledge and practice

There is a gap between what physicians must know based on the guidelines and what they actually do (Wahlstrom 1997; Olesen and Hjortdahl 1999; Grimshaw, Eccles et al. 2004; Timmermans and Mauck 2005; Liang 2007). Thus, the aim of CPD is to narrow and close the gap between knowledge and practice. Effective CPD interventions should be tailored based on an accurate assessment of needs and causal variables that support and hinder targeted health care outcomes.

In order to develop and assess interventions, theory is necessary in the field of knowledge translation research (Eccles, Foy et al. 2006). We need a theory to understand what is happening in the black box of the mind of health professionals when trying to bring about change in professional practice. We can also use the theory to target and design interventions and assess their effects (Eccles, Grimshaw et al. 2005). Oxman claimed that a useful theory should be less theorizing, clearer, andwith simple logic based on the evidence. When selecting a theory for implementation, two steps should be considered: *1) The usefulness of evidence in the context in which it is to be used. 2) A clear, understandable and assessable hypothesis* (Oxman, Fretheim et al. 2005).

In the current study, the theory of readiness to change (Prochaska and DiClemente 1992) was considered because it had been tested in the CME context previously (Parker and Parikh 2001; Buckley, Goering et al. 2003), and had also been suggested in the literature (Parker and Parikh 2001; Dalton and Gottlieb 2003; Miller and Spilker 2003). The findings were promising regarding changing physician behaviour, although the sample size and study design were improved in this study in comparison with previous studies.

In the following overview, different theories and models of learning and changing physicians' practice will be described and discussed in the context of CME and CPD in order to provide a theoretical framework to support an educational intervention design.

1.2 Current ideas of learning and change

Different models have been designed to describe how learning and change occur in professional practice. Some of these models are discussed below.

Social cognitive theory

Bandura's Social Cognitive Theory (Bandura 1997), views change in understanding as a dynamic and interactive process in certain settings (Mann 2002). Dall' Alba (2007) claimed that the procedure of becoming a physician is an active, dynamic and changing process over time. Learning and understanding are an integrated process of knowing, acting and being within a domain of practice, such as general physicians or midwives. Learning is sequentially reproduced and renewed, so that CPD and CME programmes must include a related responsiveness to the changing requirements (Dall'Alba and

Barnacle 2007). Dall'Alba also claimed that knowledge acquisition and skills in continuing education and professional education programmes are common, but argued that this is not enough as it does not automatically make doctors capable in clinical practice. Instead, developing skilful practice should be based on understanding through practice (Dall'Alba 2004).

Learning from experience

Michael Balint can be seen as a pioneer in learning from experience and developing a practice-based continuing education for general practitioners (Hodges, Inch et al. 2001). Already around1950, he initiated and facilitated regular group sessions over a longer period of time with a group of doctors focusing on their experiences and emotional reactions in managing patients with various types of difficult medical problems (Balint 1957). "Balint groups" have later become a component of CPD in many countries. The aim of such groups are to emphasize the crucial role of the doctor-patient relationship and to learn from peers (Kjeldmand D, Holmström I,etal . 2004).

Another pioneer in this field is Kolb, who illuminated the role of experience in the learning and understanding process (Kolb 1984). Kolb described the cycle of experience, reflection, abstraction and generalization, to apply to future experience. Donald Schoen described an important model, by focusing on the "reflective practitioner", as a means of understanding learning in practice. He contrasted the formal learning, which occurs in universities, with informal learning, which usually happens during the daily work when encountering problems and solving them. Schoen described five stages of experiential learning, which he termed "reflective practice". The first stage is "knowing-in-action". The knowledge and skills available to solve most clinical problems are called "zones of mastery". The second stage is "surprise" which occurs because of the complexity, ambiguity or uniqueness of events. The third stage is "reflection-in-action", which occurs during professional practice. The fourth stage is "experiment", which refers to professionals taking actions that will help to understand or resolve the dilemma arising from the surprise events. The fifth and final stage is "reflection-on-action". At this level, professionals focus on and critically analyze what they have encountered, the effectiveness of their decision-making or actions, and the implementation for continuing practice, until the next cycle happens.

Another model in this field is "reflective process" consisting of three phases: returning to the experience, attending to feelings, and re-evaluating the experience (Mann 2002).

Returning to the experience enables the learner to recall both factual and emotional aspects. Both positive and negative feelings are considered in the concepts of attending feelings, which could have an effect on future learning. Discussion of the experience, usually in a group of peers, gives new insights and allows for reorganization of both cognitive and emotional aspects and thus, the experience may be re-evaluated and acted upon, if needed.

In the field of experiential learning, most of the research in the educational literature is descriptive in nature and at the level of teachers and undergraduate students (Mann 2002). In Canada, Schoen's model has formed the conceptual framework for an approach to the maintenance of competence developed by the Royal College of Physicians and Surgeons of Canada. One method used is that the physicians record data using two recording systems; traditional (paper recording) and "PC-Diary", physicians record the questions brought up during their practice, the trigger for the questions and their learning actions in response (Mann 2002). By using different group work methods, such as using standardized patients and snowball groups (Box 1), we tried, in this study, to enable participants to put themselves in a real situation, make a decision and then receive feed back from their colleagues and tutor.

Box 1. Educational techniques

Snowball groups

The group is given a problem to solve: First, all group members work individually for five minutes and write down their ideas. Second, They share their ideas in pairs for 10 minutes. Then, pairs join up to form groups of four and then eight, and group members write categories and solutions on a large sheet of paper (Jaques 2003).

The jigsaw approach

In small groups of similar size, each student is given a different learning task to uncover only part of the information needed for the full lesson. Students from Group 1 then meet with students from other groups who have been given the same role. Within this "expert group," participants engage in research and discussion to gain full knowledge about the issue they have been assigned. After the "expert groups" have finished, each member returns to his or her original group and teaches what he or she has learned to allow the full lesson to be pieced together (Sharen 1994).

Double helix model

The "double helix model", described by Nowlen (1988), demonstrates a practical model of CME. Professional practice is the result of an interaction of two extremes of development; individual and cultural. The individual strand consists of past experience, growth and personal and professional life history, while the cultural strand comprise cultural meaning, expectation and norms of the society of which the individual is a part and also of the profession (Mann 2002).

This theory is linked to the theory 'community of practice', which describes that learning is a socialization process where the professional identity is constructed by integrating formal knowledge with informal tacit knowledge (Eraunt 2000). The theory explains how communities are shaped around practice by the members in any kind of organization. They are not just a group of people, or a net of organizations, but they are a group who share an overall view of the domain in which they practise and have a sense of belonging and common loyalty to this view (Bolander Laksov, Mann et al. 2008). The procedure demands the active participation of learners, and suggests that the acquisition of implicit knowledge is dependent on the context (Bolander Laksov, Mann et al. 2008).

CME and CPD providers must have a thorough understanding of clinical practice and identify factors, which are related to change in physicians' performance such as culture and society of practice. In the present study, we allowed time for group works to give learners the opportunity to learn from the experiences of their peers.

PRECEDE or Green's model

The Predisposing, Reinforcing, and Enabling (PRECEDE) model is a model of factors affecting clinical practice. It was originally introduced in the field of patient education, but has also been applied to explain the way physicians change their practice patterns (Tamblyn and Battista1993).

Systematic reviews of educational interventions in CME have found this model useful in both classifying and explaining the effectiveness of CME interventions (Davis, Thomson et al. 1995). Recently Davis et al. provided comprehensive implications of Greens' theory in the field of knowledge translation. The standpoint of the stakeholders (patients, practitioners, teams, policy- makers or population) is represented by Greens' model, which shows progress from awareness, agreement, adoption, to adherence with

evidence-based practice. They proposed that "interventions work in three ways: to predispose to change by increasing knowledge or skill; to enable the change by promoting conditions in the practice and elsewhere and; to reinforce the change, once it is made" (Davis, Evans et al. 2003).

The fundamental assumption of cognitive theory is that learning is an active process, which can be examined and understood. The cognitive theory of learning suggests a number of generalisable determinants of successful learning (Bosch, van der Weijden et al. 2007).

Expertise and mental model

Dreyfus and Dreyfus (1986) developed their model regarding changes in the physician's own level of expertise. They argued that skill acquisition in each new area typically proceeds through five levels: novice, advanced beginner, competent, proficient, and expert. However, not all physicians achieve expert status. They will need to promote their own capability to judge task locations and select appropriate modes of cognition. Based on the Dreyfus and Dreyfus model, it has been argued that the actions of experts are necessarily intuitive (Hamm 1991).

Educational research has shown that experts learn differently from novices, mainly because they have more pre-existing domain-specific knowledge (Patel, Glaser et al. 2000). Experts are better able to translate a problem into smaller problems, using categories, which are related to potential solutions. Between these levels, another phenomenon is known as the "intermediate effect". It occurs at various levels of expertise and varies from comprehension of clinical cases and explanation of clinical problems to problem solving. The pattern of development may be explained by the fact that intermediates have acquired an extensive body of knowledge but have not yet recognized this knowledge in a functional manner. Superior expert performance is mediated by highly structured and richly interconnected domain-specific knowledge. While a novice's knowledge base is sparse and an expert's knowledge base is intricately interconnected, an intermediate may have a lot of the pieces of knowledge in place but lack the extensive connectedness of an expert. The intermediate effect is not a one-time phenomena, rather it occurs repeatedly at strategic points in a physician's training that follow periods in which large bodies of new knowledge or complex skills are acquired. As mentioned, moving from novice to superior expert level is a procedural shift and needs a gradual change in the role of encapsulated knowledge

(active knowledge) as the level of expertise increases. Expertise is not only about acquiring knowledge, but it is about growing to be a member of a community of practice, sharing values, attitudes, roles and norms as well as building identities (Lonka and Ahola 1995; Patel, Glaser et al. 2000; Rikers, Schmidt et al. 2002; Lingard, Hodges et al. 2004; Rikers 2004; Bolander Laksov, Mann et al. 2008).

Stages- of- change or Prochaska theory

The aim of all types of learning is change - changing mental models and creating new ways of thinking. One of the models for change is based on the "stages of change", which is defined on the basis of people's propensity to change a specific behaviour and understanding. Firstly, individuals contemplate the problem and possible behaviours, then they plan and undertake actions, and finally they appraise the outcomes of these actions. The trans- theoretical model, or "stages- of- change" model was developed by James Prochaska and Carlo DiClemente in the early 1980s to explain the stages of change observed in persons striving to change addictive behaviour (Prochaska and DiClemente 1992). The model postulates that individuals are at different stages of readiness to change, that effective interventions must assess the individual's stage of readiness for change, and assist individuals to move from one stage to the next. Moreover, the interventions will differ for each of the five stages of change: pre-contemplation, contemplation, preparation, action and maintenance. Pre-contemplation refers to people who do not think about change; contemplation refers to people who are ambivalent about change; preparation refers to people who prepare to make a specific change; action refers to people who take a definitive action to change, and; maintenance refers to people who maintain new behaviour over time. Frequently, the change process is not effectively completed, and the cycle begins again. This theory has mostly been used in the field of addictive behaviours and health promotion such as smoking cessation (Tomson 2005). Recently, the stages-of-change theory has been used in research into the implementation of interventions;, such as in a study in the Netherlands where different stages- of- change theories were integrated and compiled into a ten-step model for changing professional behaviour (Grol and Wensing 2004).

However, as cited by Riemsma et al. (2002), West believed that the theory seemed to be inconsistent; he criticized the concept of "stage". Stages here were considered as a theoretical framework, which was regarded as inadequate to explain the more dynamic

process of behaviour change. Moreover, based on one of the systematic reviews of interventions using stages-of-changes, limited evidence for the effectiveness of interventions based on the model was found (Riemsma, Pattenden et al. 2002). In contrast, Sutton argued that the reviews which have been published regarding stages-of-change interventions up until 2005 have included studies that did not have appropriate implementation strategies for the stages-of-change model (Sutton 2005). Further, it has been suggested that the model provides a deeper understanding of participants' needs within the CPD context, and educational methods customized to the learners' stage of readiness to change have been proposed (Buckley, Goering et al. 2003; Davis, Barnes et al. 2003; MacDermid, Solomon et al. 2006). Studies on the effects of tailored interventions based on the Prochaska model have been suggested in the literature (Parker and Parikh 2001; Dalton and Gottlieb 2003; Miller and Spilker 2003).

In one study, a modified Prochaska model was used as an evaluation tool to assess the participants' stages of behaviour change before and after three months of a research transfer training programme. The model was modified to include only three stages: attitude, intention and action. In the *Attitude* stage, the original pre-contemplation and contemplation stages were combined, e.g., physicians who attend in CME programs, have no intention to change or if they are aware of a problem, no commitment to take action. The *Intention* stage was the same as the original preparation for action stage, describing people who are ready to change. The *action* stage refers to recent change in the original Prochaska model (Buckley, Goering et al. 2003). The findings showed positive change in the participants' behaviour between pre- and post-tests in the attitude and intention stages, but at the time of follow-up, most of them had not reached the action stage. It was revealed that participants changed their process of thinking and intended to do research in the future, but they did not change their practice patterns (Buckley and Georing 2003). These results may be related to external and internal barriers existing at the action stage. To identify these, further studies were recommended. The results demonstrated that the tools could be applied as a formative assessment to determine the learners' real needs and to tailor a course based on these needs.

In one review article (Rowe, Savigny et al. 2005), factors which affect health professionals' performance were assessed and strategies for improving it were suggested. Theories applied to change health-worker practice in low- and middle-income countries were categorized. One category specified the stages-of-change and

PRECEDE models. The researchers' assumption was to change individuals' performance through the stages, and those different interventions were needed at different stages. They suggested interventions based on a combination of these two theories consisting of: Predisposing strategies, to progress through pre-contemplation to contemplation (education programmes, conference); enabling strategies, to progress from contemplation to action (clinical guidelines); and reinforcing strategies, to progress from recent action to maintenance (Rowe, Savigny et al. 2005; Jamtvedt, Young et al. 2006). Their design is to some degree similar to the one presented in this thesis.

The "stages-of-change theory" was used in the current study for two important reasons. First, it had been tested in the CME context and the results of the intervention were successful (Buckley and Goering 2003; Parker and Parikh 2001). Second, bearing in mind that GPs are in different stages, it would be ideal for participants in a single large CME event to be assigned to different groups, matching their actual needs. These two important issues were also emphasized by Oxman when choosing appropriate theories in health system research (Oxman, Fretheim et al. 2005). Overall, this model helps to assess the learning needs and to tailor the interventions on the basis of readiness to change.

Interactive educational methods

Based on the literature, single educational interventions had slightly less effect (60%) than those using two methods (64%), while multifaceted interventions were more effective (79%) in creating change in physicians' practice (Oxman, Thomson et al. 1995; Mann 2002). Davis et al. conducted a review of 99 educational strategies, and concluded that effective strategies are reminders, patient-mediated interventions, outreach visits, opinion leaders and multifaceted activities (Davis, Thomson et al. 1995). However, in one of the review articles, the effects of multifaceted learning methods when trying to close the gap between clinical research and practice were contested (Shojania and Grimshaw 2005). They claimed that multifaceted interventions had median effect sizes on participants' practice that were not significantly greater than single ones. However, in one of the articles, which focused on the dissemination and implementation of knowledge in clinical practice, the authors stated that the use of multifaceted learning methods in comparison with one educational intervention could

nevertheless make a swifter progress in learners. The varying results require careful consideration of the influence of the context (Dopson, FitzGerald et al. 2002).

 The idea of activating instruction (Lonka and Ahola 1995) is to diagnose the "mental models" and activate participants' prior beliefs. In instruction, these internal models are supposed to be made overt to discussion and reflection. Therefore it is essential that instruction is interactive and supports this process. Getting feedback for the ideas is the third cornerstone of activating instruction and therefore peer discussions and reflections are an integral part of CPD activities. "Mental models" are not only mental, but are also cultural. It is important that when teaching people who are supposed to learn new ways of thinking, like how to diagnose depression, the instructors should take into account how the learners think and how they understand the phenomenon. This helps the learning process by providing opportunities to reflect on one's own mental models.

In the present study, interactive and multifaceted learning activities were used in the large and small intervention groups, such as case illustrations presented through standardized patients, role playing, jigsaw, buzz groups, programmed lectures and snowball techniques (Box 1). Moreover, printed materials regarding diagnosis and management of depression were distributed between participants.

1.3 CME, CPD and Translating Knowledge into practice

1.3.1 Definitions

Definitions of CME vary. Based on the definition provided by the American Association of Medical Colleges, continuing medical education is "a definable activity that supports the professional development of physicians and leads to improved patient outcomes. It encompasses all learning experiences the physicians engage in with the conscious intent of regularly and continually improving their performance"(Davis, Barnes et al. 2003). The demand for continuous learning as a part of a physician's profession is justified because doctors should be able to respond to the ever- changing needs of their patients. The best ways of introducing and developing this learning have been the subject of much controversy, and the quality of medical education at all levels has been questioned and debated in many countries (Holm 1998).

The CPD concept is more comprehensive than CME and includes a wide range of formal and informal learning activities in the professional lives of the health care

providers (Mann 2002). The definition of CPD as outlined by Stanton and Grant is more holistic than CME (Stanton and Grant 1999). It refers to the educational methods beyond didactic ones, and incorporates notions of self-directed learning and personal development, and organizational factors.

The differences between CME and CPD could be summarized in three components; setting, content and format. The CME setting is educational, but the CPD setting is practical. The content of CME is based on clinical problems, while the CPD content is community-based and holistic. The format of CME is lecture-based, whereas the CPD format is practice-based (O'Brien, Freemantle et al. 2001). Based on the theories mentioned above, CPD is founded on understanding, doing and acting. In addition, it refers to learning from experience and peers using interactive methods, while CME is based on acquiring knowledge and skills.

The translation of knowledge into practice is related to the purpose of CPD, which is outcome. Knowledge translation refers to the procedure of recomposing, interpreting and translating research findings into practice, or into an outcome (O'Brien, Freemantle et al. 2001). Based on the definition established by the Canadian Institute of Health Research, knowledge translation refers to the exchange, synthesis and use of knowledge – within a complex system of interaction among researchers and users - to hasten the uptake of the benefits of research for people through improved health, more effective services and products, and a strengthened health care system (Graham, Logan et al. 2006). In order to shift from CME towards CPD and make it more effective, we need to design an educational intervention based on the practical setting and models of change and learning. Changing performance based on the evidence is possible, although, it needs comprehensive approaches at different levels, which should be adapted to specific locations and target groups (Grol and Grimshaw 2003).

There are hardly any studies that demonstrate how CPD could actually change clinical practice. It is challenging to measure the outcomes of such interventions from the point of view of expertise and developing clinical skills. One possible method is using standardized patients (SPs). In the present study, we chose to use unannounced SPs because of the necessity for assessing physicians' real performance and to eliminate bias related to a prior relationship with the physician.

1.4 Measuring doctors performance using unannounced standardized patients

SPs were introduced for learning and assessment in medical education in the 1960s (Rethans, Gorter 2007). An SP is defined as an individual, who has been well trained to replicate the patient presentation in a clinical encounter consistently and realistically. SPs have been used as a valid and reliable tool for assessing health professionals' skills (Rethans, Gorter 2007). Unannounced SPs are preferably used for assessing physicians' performance under real conditions, which are completely different from structured examination conditions (Rethans, Gorter 2007). Unannounced SPs have been used in two randomized clinical trials (RCT) to assess the effectiveness of educational interventions on the skills of trainees in general practice regarding cancer control (Carney, Dietrich et al. 1995) and nutrition (Maiburg, Rethans et al. 2003). It was stated that the use of SPs is a valid tool for evaluating the effectiveness of an intervention programme (Carney, Dietrich et al. 1995; Maiburg, Rethans et al. 2003). An unannounced SP is not known by the participating doctors and they do not know when they will be visited by the SP. It has been suggested that the use of simulations, in CME and CPD should be increased, based on the clinical reasoning and skills of health professionals in continuous learning (Mann 2002).

1.5 Management of depression in primary health care

Depressive disorders are common and frequently unrecognized in general practice both in Iran and worldwide (Ormel, VonKorff et al. 1994; Murthy 2003), and a majority of patients do not receive adequate treatment (Lecrubier 2007). Based on the study which has been done in Iran, regarding mental health in the adult population, the prevalence of mental disorders was estimated 21.3% in rural areas and 20.9% in urban areas. Depression and anxiety symptoms were more prevalent than somatisation and social dysfunction. Prevalence increased with age and was higher in the married, widowed, divorced, unemployed and retired people. The most common symptoms of depression in Iran in addition to depressed mood are somatic symptoms. Feelings of guilt and delusion of guilt are uncommon. About a fifth of the people in the study (25.9% of the women and 14.9% of the men) were detected as likely cases (Noorbala, Bagheri Yazdi et al. 2004).

One review article, based on the literature published over more than five decades, concluded that there is a high demand for general physicians' mental health training (Hodges, Inch et al. 2001). Based on a study in Iran regarding psychiatric disorders in general practice, which has also been reported by WHO, it was shown that 35% of ambulatory clinic patients were assessed by general physicians as having psychiatric problems (WHO 2001). In another study carried out in a general medical ward in Iran, depression disorders represented the largest diagnostic category, amounting to 43% of all psychiatric cases (WHO 2001). Therefore, there is a need to increase doctors' knowledge, skills and understanding regarding diagnosis and treatment of depression. Improved treatment of depression by general physicians is recommended by many health authorities, like WHO, the World Psychiatric Association and the Iranian Ministry of Health (Moussaoui Lopez-Ibor et al.2003; Murthy 2003).

1.6 CME situation in Iran

The Islamic Republic of Iran is located in the south west of Asia. It is a middle- income country covering 1 648 000 km^2, of which less than a quarter is arable land. The population was 69 515 000 in 2005 (WHO 2006). Iran has 30 provinces. The health care system and medical education have been integrated in Iran since 1985, which is unique in the world (Marandi 1996). Medical sciences universities are responsible for providing health services to all parts of the country. Forty-one universities and some scientific organizations are in charge of directing CME programmes in Iran. The capital of Iran, Tehran, is a big city with 15 million inhabitants and is divided in 20 regions. Health and medical services are provided to each of these regions by one of the three medical sciences and health services universities in Tehran; the northern part of the city is served by the Shaheed Beheshty University, the western part is served by the Iran University and the southern part of Tehran, which consists of eight regions, is served by the Tehran University of Medical Sciences (TUMS). One of the largest groups of physicians in Iran is general physicians, who graduate from medical school after seven years of studies.

Two different components are considered for CME in all parts of the world; compulsory and voluntary. In some parts of the world such as the United States of America, Australia (New South Wales) and Iran, CME is compulsory for all health professions. All health care professionals must acquire CME credit points for a defined

period of time, otherwise they will not be able to continue their careers and their certificates will not be renewed. In Iran, every five years, physicians, dentists, pharmacists and laboratory specialists must gain 125 credit points from these kinds of activities.

In the TUMS region in 2004, there were 1 614 registered general physicians working in primary care (GPs). Most of them (80%) were engaged in private clinics, while (20%) of them were employed in public clinics and hospitals. The majority were men (74%), the mean age was 38 years (SD=9.3) and their mean work experience was 7 years (SD=8.3) (Shirazi M, Zeinaloo et al 2004). In other parts of the world, such as the United Kingdom and other countries in Europe, CME is not obligatory. Physicians participate in CME meetings based on their own interests or sometimes their organizations' requirements or recommendations.

Different kinds of CME activities are presented by the use of various types of educational formats; formal and informal. Formal educational activities include seminars, congresses, workshops etc. Informal educational activities consist of reading books, journals, discussing with colleagues, etc. In Iran, CME credit points can be gained from: congresses, conferences, workshops (small group meeting), planned programmes, practical training and self-directed learning programmes such as reading journals and books. However, 40% of doctors' credit points should be gained from participation in a planned programme, the curriculum of which is based on the topics that physicians must know as a doctor in different domains. The curriculum is compiled by the Ministry of Health. These kinds of programmes should be presented in medium to small groups of doctors (less than 50 persons). Interactive methods are used in some of these programmes, but the dominant CME programme in Iran is still the conference and the common educational method is still lecture-based, so for changing doctors' performance, a change of the programmes is inevitable (Shirazi, Zeinaloo et al. 2004).

2 STUDY SETTING, MATERIAL, METHODS

2.1 Hypothesis

An educational intervention based on modern learning theories, and using the stages-of-change model, will help physicians enhance their stages of change, knowledge, attitudes and performance in a CPD context in comparison with current CME programmes.

2.2 Aim

To assess the effects of a tailored intervention, based on the Modified Prochaska Model for general physicians in primary care (GPs) in Tehran, on their stages of readiness to change, knowledge, attitudes and performance in the field of diagnosis and management of depression disorders compared with a current CME programme.

2.3 Specific Objectives

1) To validate tools (the Modified Prochaska questionnaires, knowledge and attitudes questionnaire and observational checklists) for GPs' management of depression disorders (I)

2) To develop a tailored educational intervention for GPs based on the Modified Prochaska model in the field of diagnosis and management of depression disorders (II, III, IV)

3) To evaluate the effects of the educational intervention compared with a current CME programme on:

   a) GPs readiness to change (II)

   b) Knowledge and attitudes (III)

   c) Clinical performance (IV)

2.4 Study Design

2.4.1 Type of study

The study was a randomized controlled trial intervention with two study arms, i.e. the intervention arm was an educational intervention based on the modified Prochaska model and the control arm comprised current CME activities with 96 GPs in each arm (Figure 1).

The topic of this study was depression disorders due to the importance of depression as a common illness in Iran (Noorbala, Bagheri Yazdi et al. 2004). Improved treatment of depression by general physicians is recommended by many health authorities, like WHO, World Psychiatric Association and the Iranian Ministry of Health (Moussaoui, Lopez-Ibor et al. 2003; Murthy 2003).
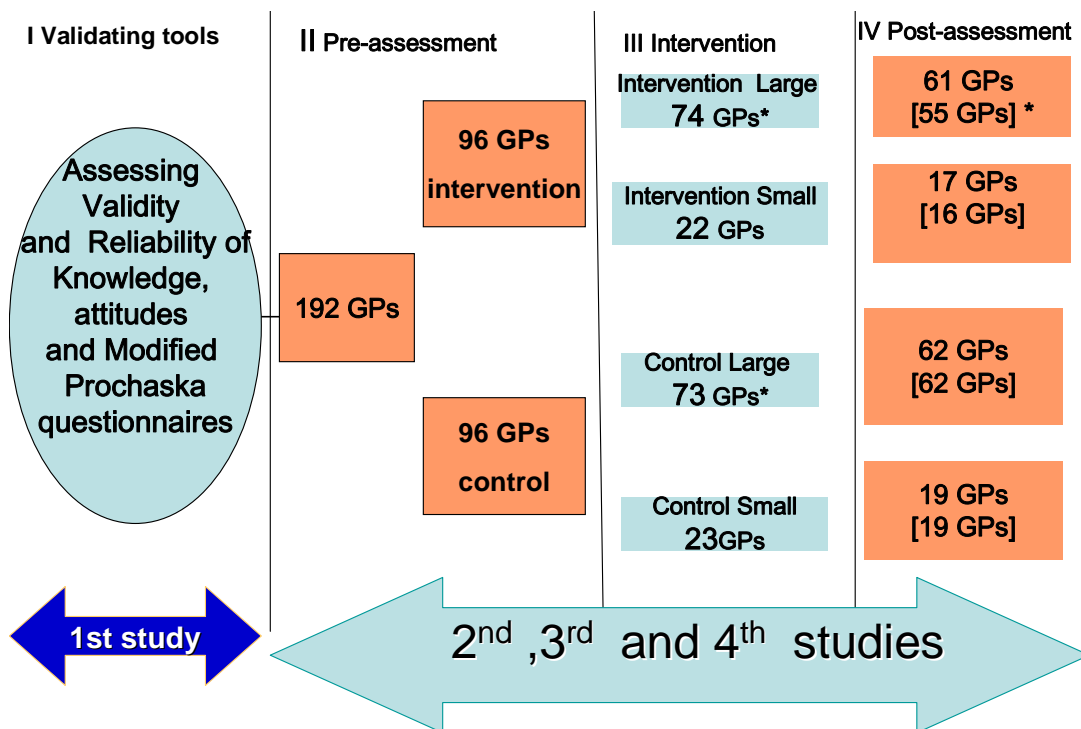
Figure 1: General study design

*GPs in fourth sub-study in brackets

The study was conducted in four phases (Table 1): Testing the assessment tools for validity and reliability in Iran [phase I]; Assessing the GPs' knowledge, attitudes and performance and their stages of change regarding management of depression disorders before (phase II), and after (phase IV) the intervention; Developing and implementing an educational intervention based on the modified Prochaska model (phase III); Comparing the outcomes of the education based on the modified Prochaska model in the intervention arm, with the current CME program in the control arm (phase IV).

Table 1: Summary of study design and methods

Study	Design/data collection method	Subjects	Data collection period
I	Cross-sectional: developing and validating of Modified Prochaska Questionnaire (MPQ), Knowledge and attitudes using written questionnaires	350 GPs	July and September 2005
	Interviews for assessing the concurrent validity of MPQ	39 GPs	September and October 2005
	Reliability (test-retest) of the MPQ, knowledge and attitudes questionnaires were assessed	59 GPs	September and October 2005
	Education intervention in Control arm	96 GPs in each arm	5th and 6th of July 2006
	Education intervention in Intervention arm		25 and 26 of July 2006
II	Randomized Control Trial (RCT): GPs' stages of change were assessed by the use of MPQ	192 GPs	Pre-assessment: April-May 2006 Post-assessment: August 2006
III	RCT: GPs' Knowledge and Attitudes were assessed by the use of a questionnaire	192 GPs	Pre-assessment: April-May 2006 Post-assessment: August 2006
IV	RCT: GPs' performance was assessed by the use of standardized patients	192 GPs	Pre-assessment: April-May 2006 Post-assessment: September-October 2006

2.4.2 Randomization procedure

Based on the results which had been derived from our pilot study in Iran, the percentage of GPs in the intention stage in pre-assessment was ascertained at 15 percentage units. The expected change in the intervention group in this stage in post-assessment was considered to be 40 percentage units. The expected change in the control group was considered to be 20 percentage units. The difference of expected change in post- assessment between interventional and control group in the second stage was thus 20 percentage units. The sample size was set at 164 for both groups and 82 for each group. This was determined by the use of the following formula for comparison of proportions (Smith and Morrow 1996). The power was considered 80% and α was equal to 5%.

$n= [(z_1+z_2)^2 2p^-(1-p^-)]/ (p_1-p_2)^2$

$z_1=1.96, z_2=0.84, p_1=20\%, p_2=40\%$

$n= [(1.96+0.84)^2 *2*.3(1-.3)]/ (20\%- 40\%)^2 =82$

Assuming that some of the samples would not want to participate, 20 percentage units more GPs were added to the first calculated sample size. Finally, 200 GPs, working in the TUMS district were selected using stratified randomized sampling. An extra 50 percentage units was added in order to allow for up to 33 percentage units initial attrition.

2.4.3 Participants

Inclusion criteria were: being physician registered as GP in the TUMS districts and; actively working in the southern part of Tehran with patient consultations, including depression disorders. Initially, 300 GPs, were randomly selected and invited to a meeting to receive information about the project, and 220 GPs attended the meeting. However, 98 of them did not have a clinic or worked outside the target area and were thus not eligible. Therefore, another 200 GPs were randomly sampled and invited to a meeting, which 185 of them attended. Again, those who did not have a clinic or worked outside the target area (n = 115) were excluded. The remaining 192 GPs were informed about the objectives of the study and methods used and that participation would exempt them from having to pay the usual participation fee for the education. After signing an informed consent, all of them filled in the Modified Prochaska Questionnaire (MPQ),

and the knowledge and attitudes questionnaires. The MPQ had previously been validated and tested for reliability in the Iranian context (Sub study I).

The 192 GPs were first grouped according to their stages of readiness to change and stratified based on gender, age (three levels), and length of work experience (three levels). Thereafter, they were randomly allocated within each sub-group to the intervention (n=96) or the control arm (n=96). The majority of the GPs were in the attitude stage (intervention group = 77%; control group = 76%), while the others were in the intention stage (intervention group = 23%; control group = 24%). None of them were in the action stage (Figure 2).

2.4.4   Definitions of GP distributions in different stages

Attitude stage: The physicians have no intention to change or, if they are aware of the problem, show no commitment to take action. The GPs in this stage are thus not ready to change their practice regarding management of depression disorders. Their knowledge about the importance of depression, diagnosis of depression disorders as well as consulting methods is not sufficient. *Intention* stage: GPs at this stage are ready to change. They have sufficient intention to change their practice in relation to the problem and their knowledge of diagnosis of depression disorders is also sufficient, although they want to know more about the new treatment and consulting methods. *Action* stage: The GPs in this stage have enough knowledge in diagnosis and they are also sufficiently informed regarding different methods of treatment. They have recently changed their method of treatment.

```
                    ┌─────────────────────────┐
                    │    Random Sampling      │
                    │     (n=300+200)         │
                    └────────────┬────────────┘
┌──────────────────┐            │
│  Doctors did not │            │
│  show up in the  ├────────────┤
│ meeting (n=95)   │            │
└──────────────────┘   ┌────────┴───────────────┐
                       │ Doctors attending the  │
                       │   meeting (n=405)      │
                       └────────┬───────────────┘       ┌────────────────────┐
                                │                       │ Doctors did not have│
                                ├───────────────────────┤ a clinic or were not│
                                │                       │ eligible (n=213)    │
                       ┌────────┴─────────────────┐     └────────────────────┘
                       │ Doctors formally assigned │
                       │ to the project and did    │
                       │ their pre-test (n=192)    │
                       └────────────┬──────────────┘
                                    │
                       ┌────────────┴─────────────┐
                       │     Randomization        │
                       └───┬──────────────────┬───┘
```
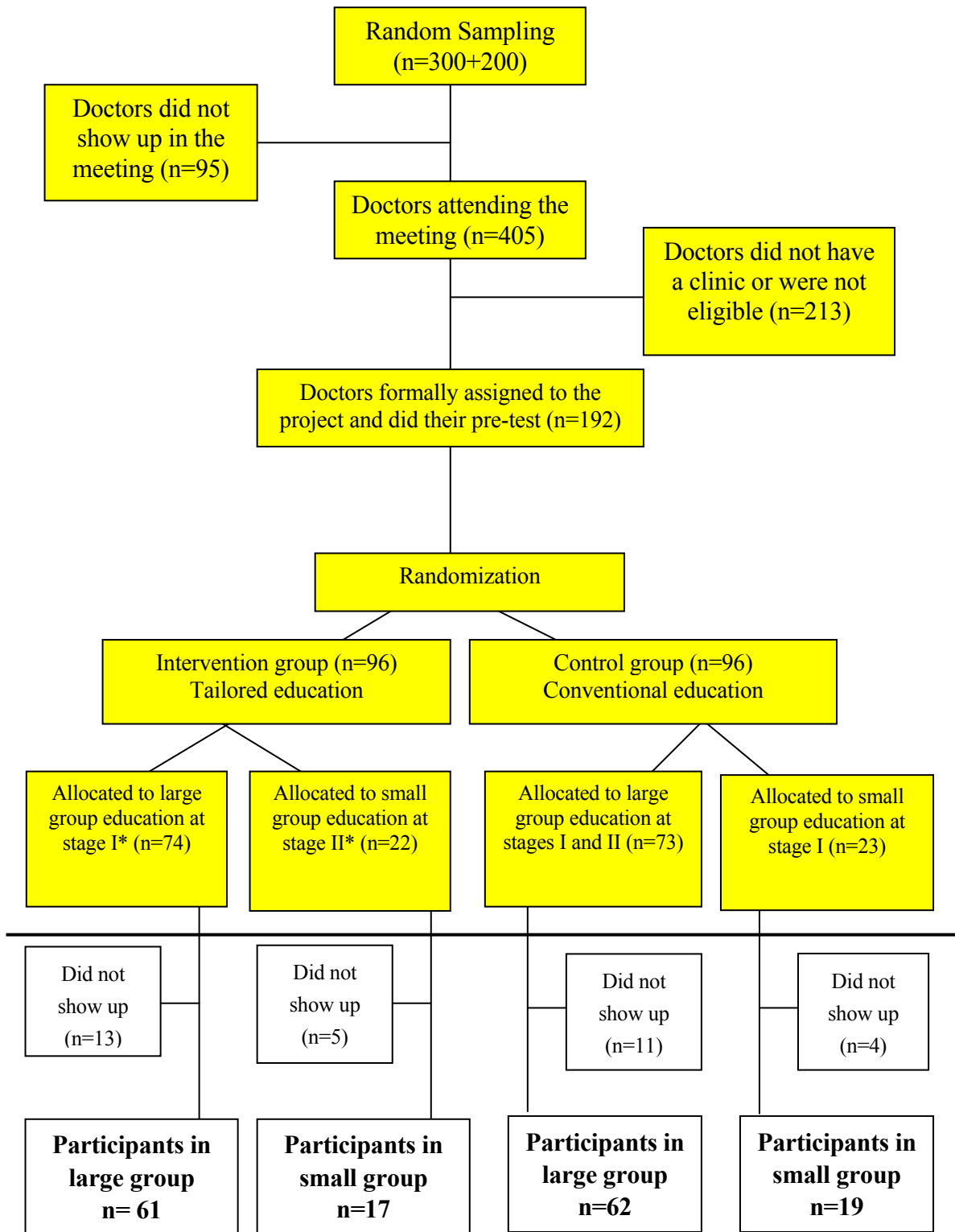
Figure 2: Flow of GPs through trial

*Stage I = Attitude; Stage II = Intention

Intervention group (n=96) — Tailored education

- Allocated to large group education at stage I* (n=74)
- Allocated to small group education at stage II* (n=22)

Control group (n=96) — Conventional education

- Allocated to large group education at stages I and II (n=73)
- Allocated to small group education at stage I (n=23)

Did not show up (n=13)
Did not show up (n=5)
Did not show up (n=11)
Did not show up (n=4)

Participants in large group n= 61

Participants in small group n=17

Participants in large group n=62

Participants in small group n=19

2.4.5 Intervention process

Eight CME teachers from the Department of Psychiatry were trained at a one-day workshop on the application of interactive educational methods. After training, they ran two-day courses for the GPs in the intervention arm. The courses were conducted for about half a day on two consecutive days. The two-day courses for the GPs in the control arm were conducted before the trainers participated in their own training.

The same educational background material was used in both groups: a book with comprehensive evidence-based guidelines for GPs regarding the diagnosis and treatment of depressive disorders. The book had been compiled by members of the research group, based on WHO documents (Ormel, VonKorff et al. 1994) and the WPA Bulletin on depressive disorders (Moussaoui, Lopez-Ibor et al. 2003) as a component of the CPD approach. The educational methods used in the intervention group were based on the participants' stage of readiness to change, while the control group followed a current CME programme.

2.4.5.1 Intervention arm

GPs in the attitude stage (n=74) were allocated to large group education (Figure 1), where the importance of management, epidemiologic aspects, diagnosis of depressive disorders and their relationship with patients were emphasized. Interactive educational methods relevant to a large group were used, such as programmed lectures, modified buzz groups and lectures followed by videos and discussion. GPs in the intention stage (n=22) were allocated to a small group in a workshop setting, mainly using case illustrations presented through standardized patients, role playing, jigsaw, buzz groups and snowball techniques (Box 1), all emphasizing novel treatments of depressive disorders, differential diagnosis and relationship with patients. The total education time in both groups in the intervention arm was twelve hours, but the contact time with the teachers was only eight hours.

2.4.5.2 Control group

The number of doctors in both the large and small group tallied with the intervention groups, but both diagnosis and treatment of depressive disorders were emphasized in both groups. The reasons for introducing a small group in the control arm was to partly control for the potential bias of a presumed better effect if the learning takes place in a small group setting. All GPs in the intention stage (n=23) were allocated to the large

group, which took part in a current CME programme on depressive disorders using conventional teacher-centred educational methods, such as lectures. A corresponding number (n=23) of the 52 GPs in the attitude stage were randomly assigned to the small group, participating in a workshop, where current CME methods for small groups were applied, such as mini-lectures followed by questions and answers. The total education time in both groups was eight hours, corresponding to the total contact time with the teachers.

Risk of contamination

Tehran is a big city with a large population of 15 million inhabitants, the majority living in the southern part of Tehran, where TUMS is responsible for providing health services. The GPs who are working in separate offices in this area have no formal meetings except for participation in different CME courses. Furthermore, the content of education in both intervention and control groups was the same, which makes the issue of contamination less important.

2.4.6 Educational material

2.4.6.1 Guidebook for managing depression

Members of the research group compiled a guidebook on the diagnosis and management of depression by GPs. The group consisted of three psychiatrists, two GPs, one epidemiologist and one medical education researcher. The evidence used for compiling the guideline was the WPA bulletin, and WHO guidelines regarding management of depression disorders (WHO 2001; Moussaoui, Lopez-Ibor et al. 2003). Most of the criteria of the AGREE instrument (The Appraisal of Guidelines for Research and Evaluation) (Oxman, Schunemann et al. 2006) were considered in developing the guidebook. The book is composed of ten chapters; each of them has the same construction with objectives, content, summary and assessment questions regarding the content of the chapter. The first aim of composing this guidebook was to have reliable educational material for our research purpose, while the second aim was to produce self–directed learning material, which can help GPs to learn about management of depression in the future. This book was published in 2007 and is accessible as one educational material in CME programmes for self-directed learning (Sadeghi, Shirazi et al. 2007).

2.4.6.2 Training videos

Two videos on the management of depression in primary care were bought from the Department of Psychiatry, Manchester University,UK. The videos were translated to Farsi by the main investigator and they were edited by an English native speaker and a psychiatrist. Then captions in Farsi were inserted below the videos. These videos were shown in the SP training course. The video on recognition of depression in general practice was shown to the large intervention group and the second video was shown to both the large and small intervention groups. The videos are in two parts;

Part 1 - Recognition in general practice. The tape draws attention to the skills doctors must have to detect and evaluate depressive illness, illustrated through role-played scenarios and a panel discussion. Running time was 40 minutes.

Part 2 - How to plan and assess treatment. The videos explore the skills that are necessary to assess depression as it presents itself in primary care, and also the further skills and strategies that are important in negotiating a treatment plan with the patient. Running time was 40 minutes.

2.5 Measures

2.5.1 Sub study I

Modified Prochaska Questionnaire

The MPQ, which had been developed and used in a Canadian study within a research transfer course (Buckley, Georing 2003), was used as a model for the Iranian tool. It was translated to Farsi, and further modified in order to adapt it to depression disorders and to the different cultural context. The final MPQ had eleven statements to be answered by a binary alternative, yes or no (Table3). The first three questions corresponded to the attitude stage, the next four questions to the intention stage, and the last four questions were related to the action stage.

Participants

The participants in the first sub-study were three groups of GPs. The first group comprised 350 GPs who were recruited for the development of the Modified Prochaska Questionnaire (MPQ) and attended in four other CME programmes between July and September 2005. The second group was recruited for validating the assessment tool and

was composed of 39 GPs, who came to the CME office at TUMS to register for other CME programmes between September and October 2005. The third group included 59 physicians, who were working in Tehran and were also participating in CME programmes at the TUMS. The respective questionnaires were completed by the three groups of GPs after signing forms of informed consent. Participants were recruited by offering one CME course free of charge in exchange for taking part in this study.

Definitions and procedure of validations

Validity can be defined as the degree to which a test measures what it was intended to measure in a particular setting. It should be noted that validity is a property of the interpretation given to the results; it is not a property of an instrument per se (Shea and Fortna 2002). An instrument may lead to valid interpretation when used in one setting but this doesn't necessarily extend to other settings. For several reasons, validity cannot be measured directly, but must be inferred from empirical observation.

The validity of the MPQ was assessed by means of content and criterion validity. Content validity asks how adequately the items match the concept being measured. This type of validity is most important when a scale or instrument is first being designed (Shea and Fortna 2002. The content validity of the MPQ was ensured through group discussions within the team of experts who received the results of the pilot groups and consequently made the required modifications. The first version of the Iranian MPQ was compiled in July 2005 and was piloted on a group of GPs. The process was repeated three times and a final version was compiled in September 2005. The team of experts consisted of three psychiatrists, two specialists in medical education, one epidemiologist, one specialist in public health and family medicine and the author of this thesis.

Criterion validity involves drawing inferences between scale scores and some other measure of the same behaviour construct. There are two types of criterion validity: concurrent and predictive. Concurrent validity amounts to asking if the observed scale scores are related as expected to other assessments of the same characteristic made at the same time (Shea and Fortna 2002).

To assess concurrent validity of the questionnaire, a semi-structured interview was performed with the respondents. Firstly, the recruited 39 GPs filled out the MPQ. Secondly, they were interviewed directly afterwards, to explore their level of readiness to change. All interviews were semi-structured and conducted by the principle

investigator. The average time for each interview was 20 minutes. The interview started with open-ended questions and if needed probed into certain areas to illuminate GPs' views regarding the importance of depressive disorders and their level of knowledge, their planning to improve their management of depressive disorders and their attendance at different CME programmes. All interviews were tape-recorded and transcribed.

There are two types of **reliability**: reproducibility (alternate forms, test-re test, inter-rater and intra–rater agreement) and internal consistency (also called homogeneity). When reliability is used as an expression of reproducibility, reliability addresses the extent to which scores obtained on two occasions (test-retest) or with two equivalent forms (alternate forms) or perhaps by two different raters or assessors (inter-rater) are similar. When used as an index of internal consistency, reliability refers to the degree to which the items measure a similar, unified concept or construct (Shea and Fortna 2002). The MPQ was examined in the group of 59 GPs through test–retest technique. The first test was run by sending the MPQ, knowledge and attitude questionnaires to the participants by trained carriers. The carriers were instructed to collect the test 20 minutes after delivery. The retest was gathered 3-7 days after the first, and it was again collected 20 minutes after delivery.

Analysis

The results of the interviews were analyzed qualitatively by the first author and then interpreted as binary alternatives (*yes* or *no)* responses to the MPQ statements without knowing the answers already given by the doctors. The correlations between the performances of the participants on the two measures were assessed by SPSS software version 13, which was used to determine the Kappa coefficient as a measure of agreement between interview results and MPQ and also between test and re-test (Cicchetti and Nelson 1994).

Knowledge and attitudes

Participants

The participants for developing and validating the knowledge and attitudes questionnaires were 56 physicians from a group of 59 GPs, working in Tehran and also participating in CME programmes at the TUMS.

The knowledge and attitudes questionnaires were combined and the development was based on published articles (Meredith, Rubenstein et al. 1999; Krupinski and Tiller 2001) and the guidelines for management of depression in primary care were compiled by the research group and experts' consensus.

Knowledge questionnaire

The knowledge questionnaire consisted of participants' demographic characteristics, seven Multiple Choice Questionnaires (MCQ) related to management of depression, eleven statements rated on a five-point Likert scale, ranging from 'Strongly disagree' to 'Strongly agree', assessing participants' knowledge about the treatment of depression disorders, and one open-ended question exploring the GPs' more active knowledge about symptoms of depression. The last part of the questionnaire related to three case vignettes describing different types of patients, asking about the diagnosis and suggested treatment for each. All MCQ were designed by experts within the group, but the Likert questions were based on published research (Meredith, Rubenstein et al. 1999). The vignettes were derived from one published article (Krupinski and Tiller 2001) and were then modified and adapted to the context of Iranian GPs (Appendix 2).

Different types of questions were designed in order to assess GPs' knowledge and skills, reflecting Bloom's taxonomy of three levels (Dave 1975). These three levels, including also attitudes, encompass recall, interpretation and problem solving, traversing superficial to deep understanding. A broad range of content was also captured by the questions. Most of the MCQ and Likert questions were related to treatment of depressed patients. The essay question related to the diagnostic criteria of depression, while the vignettes were related to the diagnosis and treatment of three different scenarios with depressed patients.

Validation of the knowledge questionnaire

The validity and reliability of the questionnaires were assessed in a pilot study. The validity of the knowledge and attitudes questionnaires were assessed by the use of content validity. These tools were designed based on the guidebook on depressive disorders compiled in the present study and other research questionnaires in the same field. Thus the experts' verification in the process was used. The questionnaires were then consecutively modified and developed based on the results of the pilot studies which focused on the dissemination of GPs between stages. The content validity of the instrument was ensured through group discussions within the team of experts.

In the present study, an item analysis for knowledge and attitudes questionnaires was done, which is described in Appendix 7.1. The accepted score based on consensus in the expert group for difficulty indices was considered 0.30 - 0.85 and for discrimination indices was 0.25- 0.80.

For assessing reliability during the pilot phase, the test–retest technique was used in a group of GPs attending CME programmes in TUMS. The procedure was the same as the MPQ reliability. The mean κ coefficient regarding all knowledge questions was 0.75.

Attitudes questionnaire

The attitudes questionnaire was initially based on 13 statements assessing GPs attitudes to depression. This questionnaire has been used in many studies (Kerr, Blizard et al. 1995). It was adapted to the Iranian context based on the results of the pilot study. The validation procedure was the same as for the knowledge questionnaire and the required modifications were consequently made within the team of experts. The final attitudes questionnaire version had ten items on a five-point Likert scale ranging from strongly agree to strongly disagree.

Standardized patients, scenarios and checklists

Participants

The GPs' performance was assessed by unannounced SPs. Twenty-five volunteer psychology and nursing students were interviewed based on certain criteria (well-being, availability, payment) and 15 were admitted as eligible, although only ten of them decided to participate regularly on a five-day training course, which was run by three psychiatrists and one senior officer in medical education. Nine of them passed the final tests (validity and reliability). Their age was between 21-43 years, four of them were men. All of them were undergraduate students and they had knowledge of psychiatric disorders.

Scenarios

Five different scenarios regarding depression disorders were compiled by three psychiatrists. The scenarios were related to depression combined with different problems such as: the risk of suicide, pathologic grief, other disorders (headache and hypothyroidism) and bipolar disorder (Appendix 4).

Checklists

Based on the scenarios, five checklists were compiled by the same expert group that compiled the guidebook on management of depression. The specified content of the checklists were based on DSM IV criteria (American Psychiatric Association 2000) for diagnosis and management of depression. The expert team consisted of three psychiatrists, two general physicians, one epidemiologist and a senior officer of CME.

Each checklist had two parts. The first part (Diagnosis) included 13-16 items based on the scenarios. In this part, the binary questions (yes or no) were related to the most important aspect to be considered by GPs when diagnosing depression disorders. It was completed by the SPs. The second part (Treatment) had 2-4 items related to the appropriate treatment procedure. This part was completed by three psychiatrists based on documents received by the SP (prescription, laboratory tests, and referral to specialists). The scores for each GP consisted of the number of correctly performed items by considering the weight of each item. Content validity of the checklist was assessed by a group of ten psychiatrists who were faculty members of TUMS. The reliability of the checklists was checked by assessing inter-rater reliability between groups of SPs, each SP assessing one's own role plays and two others', and was found to be acceptable ($\kappa=0.82$.

Procedure of SP training

SPs were trained in a small group setting and the duration of training was five educational group sessions, each for two hours. They were trained by three psychiatrists and one senior CME officer. In each educational meeting, SPs read handouts about depression disorders, watched videos regarding depressed patients, practised in the group and received feedback on role playing with the use of audio tape recorder.

Procedure of validation

Each of the ten SPs portrayed their roles' twice with the same doctor. All interactions were video-taped (total 20 videos). The validity and reliability of the SPs' portrayals and their filling-in of the checklists were documented. After these procedures, one potential SP was excluded because of an un satisfactory portrayal.

SP portrayals

Three psychiatrists individually watched each of the 20 videos and filled the observational rating scale compiled by the same group of psychiatrists regarding each

SP portrayal. The criterion validity of SPs' portrayals was assessed by comparison of each SP score with a gold standard, determined as 90% of maximum score (=2.7). The mean score was 2.95 (range 0-3). Test-retest approach was used for assessing the reliability of SP portrayals, which was highly acceptable (mean $\kappa=0.90$).

SP filling-in of checklists

Concurrent validity was assessed through the correlation between SPs' completed checklists and the checklists filled in by the three psychiatrists (mean $\kappa=0.55$). The reliability was assessed using the test-retest approach, where the SPs' initial checklist was compared with a checklist assessment made one week later after watching the video recording of their own consultation (mean $\kappa=0.72$).

Another important criterion for assessing the accuracy and consistency of the SP portrayal was regarding the reported detection rate by GPs. This depends on how well the SP played their role and how many of them were known by GPs. Only one of the GPs actually detected the SP at the second visit.

2.5.2 Sub study II

Participants

The total number of physicians participating in the education was 159 (83%), 78 in the intervention group and 81 in the control group

Measures

The final MPQ was used. GPs who gave positive answers to less than six questions (the first three questions plus two questions from the intention part) were allocated to the attitude stage. If they gave positive answers to six to nine questions (including at least all the attitude and three of the intention stage questions), they were allocated to the intention stage. If they gave ten or more positive answers, they were allocated to the action stage. The pre-intervention assessment was done 2-3 months before the intervention during the meetings held to introduce the project to the physicians. The post-intervention assessment was performed after one month by sending the MPQ to all participants by trained carriers, who were instructed to collect the test 20 minutes after delivery.

Data analysis

The data were entered using SPSS software version 13. Pearson's chi-square test for independence was used to compare the intervention and control groups regarding proportions of doctors reaching a higher stage of readiness to change after the intervention. Corresponding sub-group analyses for different initial stages and educational methods were done. All analyses were performed with an intention-to-treat approach.

2.5.3 Sub study III

Participants

The same participants and the same time for assessment as in sub-study II.

Measure

The knowledge and attitudes questionnaires were used (appendix 7.2 and 7.3). The answer to the essay question and the vignettes were corrected by the main researcher based on guidelines for correction, compiled by the psychiatrists in the research group. The kappa coefficient regarding the correlations between the scores, given by the main investigator and one of the psychiatrists (inter-rater reliability), proved a high degree of correlation between these two (κ=0.93).

Data analysis

The data were entered using SPSS software version 13. The paired t-test was used to compare the knowledge and attitudes scores before and after the interventions in both arms. Student's t-test and analysis of variance were also used to compare change of scores between the two study arms. Multiple regressions were used to compare between sub-groups, with significance set for P values less than 0.05. Bonferroni corrections were applied to control for multiple comparisons. We used a general linear model repeated measurement for assessing the effect of the education model between groups for showing the effect by consideration of time and repeated measure.

2.5.4 Sub study IV

Participants

Seven physicians of the 159 participating in the study (Fig.2) were not accessible in their offices for the SPs, and of the 33 physicians, eight were not accessible and twenty fifth of them who did not participate in the education were assessed. Thus, in total,, 177 GPs (92%) were assessed by SPs both before and after the intervention.

Measure

Each GP was visited by two different SPs', with the same scenarios, two months before and two months after the intervention. The SPs completed the specific checklists immediately after leaving the GPs' practice. The assessment of appropriateness of the treatment strategy was based on a protocol for treatment of depression, used in the biggest psychiatric hospital in Tehran. Each checklist was also given weight by the psychiatrists' views through a Delphi method. The normalized maximum score was 100 for the first part of the checklists related to diagnosis, and 10 for the second part related to treatment and referral.

Data analysis

The data was entered using SPSS software version 13. Student's t-test was used to compute the significance of the differences in the scores between the intervention and the control groups in pre- and post-assessments. To assess the differences between small and large groups in the intervention and control arms, we used ANOVA and Post Hoc test.

Multiple regressions were used to compare between sub-groups, with significance set for p values less than 0.05. We used an intention-to-treat analysis.
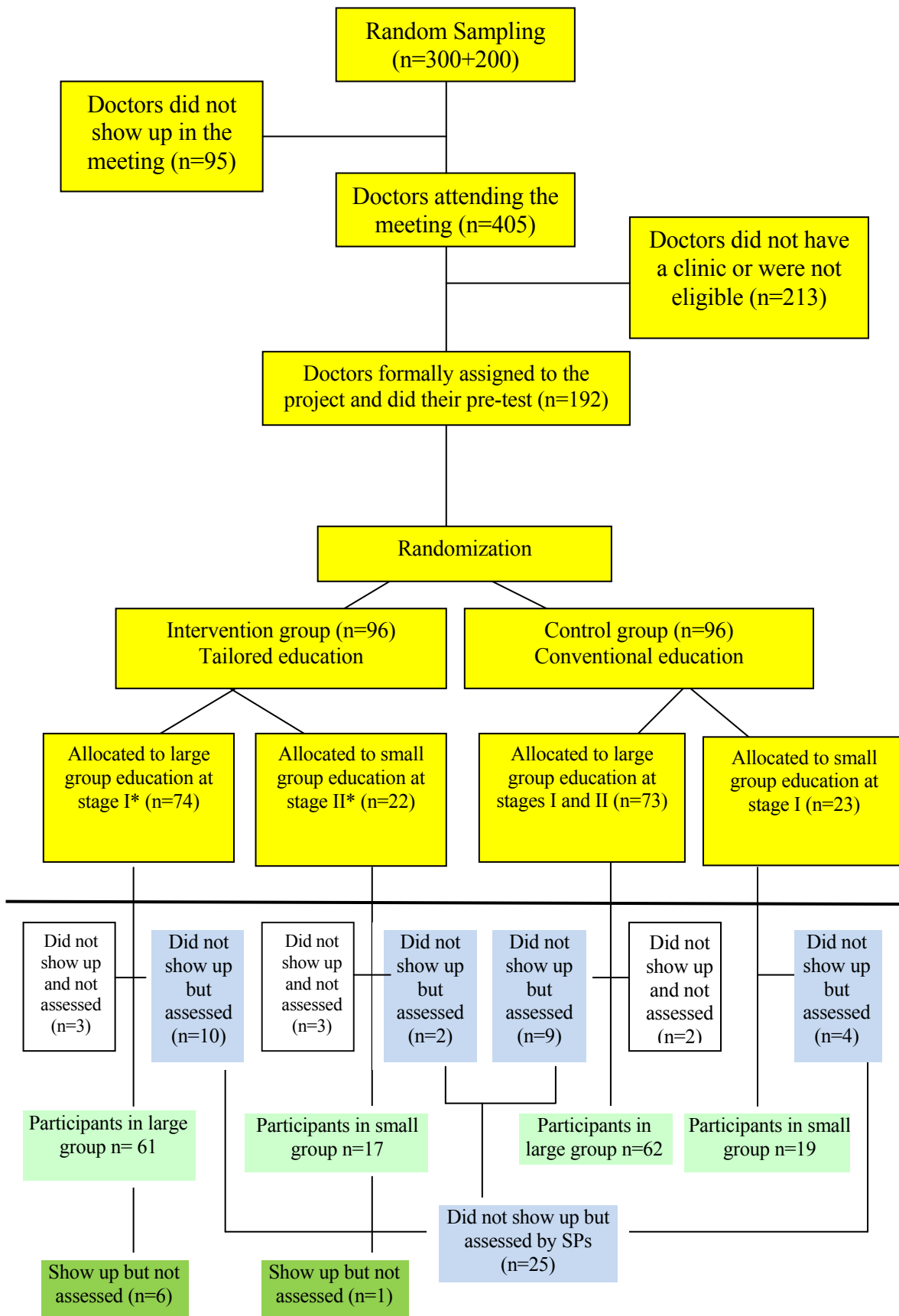
Figure 3: Flow of GPs in sub-study IV

3 FINDINGS

3.1 Sub-Study I

Demographic characteristics of participants

Group 1 for assessing content validity: 350 GPs (women 39%); age range 26–72 years (Mean=37 years, Standard Deviation=6 years); work experience 1–35 years (M=9 years, SD=6 years).

Group 2 for assessing concurrent validity: 39 GPs (women 32%); age range 26–71 years (M=37, SD=8); work experience 1–35 years (M=9 years, SD=6 years).

Group 3 for assessing reliability: 59 GPs (women 43%); age range 29–72 years (M=39 years, SD=7.5 years); work experience 2–44 years (M=9, SD=7).

Validity

Content validity of the MPQ was ensured based on experts' consensus. After each pilot study, the findings were analyzed by the team. Based on the participants' response rate, some questions were modified and revised by an expert panel within the research group and the new version of the MPQ was again tested. For example: if one question showed 100 percentage units 'yes' or 'no' responses, that question was not discriminating and should be revised or replaced by another question at the same stage. Gradually, the final version of the MPQ was developed to be used in the validation of the instrument (Table 3).

For concurrent validity, agreement between the direct responses to the questionnaire and the responses derived from the interviews with 39 GPs, was assessed. The κ coefficients for the different questions in the MPQ are shown in Table 3. The total κ coefficient for concurrent validity was 0.80 (0.77 for the attitude stage, 0.78 for the intention stage, and 0.93 for the action stage).

Table 2 shows how the participants were allocated to the three stages of readiness to change according to their direct responses or the interviews. According to both methods of measurement, four of five participants were in the attitude stage.

We were able to show that our measures were reasonably valid and reliable in the present context. The next studies will show how it works in a larger population and how physicians' stages could be changed based on this valid and reliable measure.

Table 2: Number of participants (N=39) at each stage according to Modified Prochaska Questionnaire (MPQ*) and interview responses

	MPQ*			
Interview	Attitude (%)	Intention (%)	Action (%)	Total (%)
Attitude (%)	30 (76.9)	2 (5.1)	0	32 (82.1)
Intention (%)	0	5 (12.8)	0	5 (12.8)
Action (%)	1 (2.6)	0	1 (2.6)	2 (5.1)
Total (%)	31 (79.5)	7 (17.9)	1 (2.6)	39 (100)

Table 3: Agreement between Modified Prochaska Questionnaire (MPQ) and responses based on interviews (n=39) and agreement between test-retest (n=59) for each question

Stages of Change	English translation of MPQ	κ questionnaire vs. interviews	κ test–retest	Percentage of identical responses
Attitude	1. I believe that depression is an important disorder in the professional of all GPs.	1	1	100
	2. I am consulted by depressed patients about every week regularly.	0.50	0.88	98
	3. I believe that my knowledge regarding depressive disorders is not adequate.	0.82	0.93	98
Total		0.77	0.93	99
Intention	4. I have planned to spend some time studying depressive disorders on a regular basis in the near future.	0.79	0.89	95
	5. Treating depressive patients has become a preoccupation of mine.	0.82	0.91	96
	6. I have planned to regularly attend CME meetings regarding depressive disorders.	0.84	0.74	88
	7. I have decided to do research in the field of depressive disorders on a regular basis.	0.68	0.85	93
Total		0.78	0.85	94
Action	8. I regularly consult my psychiatrist colleagues regarding diagnosis and treatment of depressive disorders and ask for their recommendations.	1	0.96	98
	9. In the last 6 months, I have attended CME programmes regarding depressive disorders.	1	0.88	98
	10. In the last 6 months, I have studied six articles from scientific sources, journal and books regarding the diagnosis and treatment of depressive disorders and I have compared my own diagnosis and treatment methods with the novel and up-to-date methods.	0.87	0.77	92
	11. During the last week, I have counseled more than five patients with depressive disorders.	1	0.93	97
Total		0.93	0.88	96

The findings of sub-study II, III and IV were reported in agreement with the CONSORT requirements (Moher, Schulz et al. 2001).

3.2 Sub-study II

The demographic characteristics of participants in sub-studies II, III and IV are shown in Table 4. The allocation of participants among the three stages of readiness to change according to their responses to the MPQ in pre- and post-test in both groups (post-test only for those participating in the education) are shown in Table 4.

The process of change to higher stages showed significant differences between the intervention and control groups for all comparisons (Table 5). Our hypotheses were strongly confirmed. In total, 57 GPs out of 96 (59%) moved to a higher stage in the intervention arm, compared with 12 out of 96 (13%) in the control arm (p<0.01), an intervention effect of 47 percentage points. In the large group setting, 46 out of 74 (62%) GPs in the intervention arm moved to a higher stage compared with 12 out of 73 (16%) GPs in the control arm (p<0.01), an intervention effect of 46 percentage points. The corresponding figures in the small group setting were 11 out of 22 GPs in the intervention arm and none of 23 in the control arm (p<0.01), an intervention effect of 50 percentage points.

GPs in stage I in the intervention arm moved to higher stages of change in significantly more cases than GPs in the control arm (46/74 = 62% vs 9/73 = 12%; p=<0.001). The same was true for GPs initially in stage II (11/22 = 50% vs 3/23 = 13%; p=0.001). It should be noted that these analyses were done according to the intention-to-treat model. If changes were to be considered only for those, who actually participated, the differences would be even more dramatic.

One potential limitation of this study was that the GPs in the intervention arm received twelve hours of training compared with eight hours in the control arm. However, teacher involvement was the same in all groups in both arms, as the methods in the intervention arm helped the participants to learn from peers and by themselves during their group work. A second limitation was the attrition in both study arms from randomization at the start of the intervention, although it was similar in the two arms. Notably, the final sample size of 159 GPs corresponds well with the originally calculated necessary sample size of 164.

The next study will present the level of GPs' knowledge and attitudes in both arms. The educational effects regarding the knowledge of GPs' in sub-groups will show whether the participants in small groups who progressed their stages in this study, have increased their level of knowledge and attitudes or not.

Table 4: Demographic characteristics of 96 participants in the intervention arm and 96 participants in the control arm

Demographic characteristics	Intervention	Control	*p*-value
Age, mean (SD)	40.2 (6.1)	40.5 (8.8)	0.81
Men (%)	72 (75)	68 (71)	0.52
Work experience in years, mean (SD)	11.7 (6,2)	12.1 (8.3)	0.77

Table 5: Allocation of participants among the three stages of readiness to change according to their responses to the MPQ* in pre- and post-test in both groups (post-test only for those participating in the intervention)

Stage of readiness to change	Pre-test		Post-test	
	n	%	n	%
Intervention				
Attitude	74	77	15	19
Intention	22	23	49	63
Action	0	0	14	18
Total	96	100	78	100
Control				
Attitude	73	76	53	65
Intention	23	24	25	31
Action	0	0	3	4
Total	**96**	**100**	**81**	**100**

* Modified Prochaska Questionnaire

Table 6: GPs' stages of change before and after the educational intervention and by participation in type of educational group, including non-respondents (nr). All differences between the intervention and control arm are significant (p≤0.001)

		Pre test		Post test		
		Stage I*	Stage II*	Stage I	Stage II	Stage III*
Intervention arm	**n=96**	**74 (77%)**	**22 (23%)**	**28 (nr:13)** (29%)	**54 (nr:5)** (56%)	**14 (15%)**
Large group – stage I	n=74	74	0	28 (nr: 13) (38%)	43 (58%)	3 (4%)
Small group – stage II	n=22	0	22	0	11 (nr:5) (50%)	11 (50%)
Control arm	**n=96**	**73 (76%)**	**23 (24%)**	**66 (nr:13)** (69%)	**27 (nr:2)** (28%)	**3 (3%)**
Large group – stage I	n=50	50	0	41 (nr:9) (82%)	9 (18%)	0
- stage II	n=23	0	23	2 (9%)	18 (nr:2) (78%)	3 (13%)
Small group – stage I	n=23	23	0	23 (nr:4) (100%)	0	0

* Stage I = Attitude, Stage II = Intention, Stage III = Action

** Non-respondents are assigned to their original stage in the post-test. All differences between the intervention and control arm are significant (p< 0.001).

3.3 Sub- study III

The cooperation rate among the 192 GPs, who consented to participate, was 81% (78 GPs) in the intervention group and 84% (81 GPs) in the control group (Figure 2). All these GPs took part in the educational programmes and answered the questionnaires both before and after. There were no significant differences in demographic characteristics between the GPs in the two study arms (Table 1).

Table 7 shows the results of GPs' knowledge regarding depression disorders, divided in two parts: the first including the MCQ and Likert scale questions, the second including

the three vignettes and the essay question. There was significant improvement in knowledge mean scores regarding MCQ and Likert (*p*= 0.002) as well as case vignettes and the essay question *(p= 0.011)* in the intervention arm in comparison with the control arm. The intervention effect regarding the first part was 6 percentage units and for the second part 12 percentage units. Table 8 shows the differences between the sub-groups. The Bonferoni correction test was also calculated. Regarding MCQ and Likert questions, a significant improvement was detected in the small intervention group in comparison with both control groups (*p*=0.02). The results regarding the case vignettes and essay question showed positive changes in the large and small intervention group versus other groups, close to being statistically significant (*p*=0.051). The intervention effect of the MCQ and Likert questions between the small groups was 14 percentage units and 4 percentage units between the large groups. The intervention effect regarding case vignettes and the essay question was 5 percentage units between the small groups and 13 percentage units between the large groups. Although there were favourable changes in the attitudes statements, no significant differences in mean scores on the attitudes questionnaire were detected between the two arms (Table 9).

The results showed modest effects of the educational intervention on GPs' knowledge in the intervention arm in comparison with the control arm. In the small group, this could be interpreted as being in the action stage (being expert), related to the expertise theory. Medical experts do not always remember essential facts better than novices but they are good at problem solving procedures. It still remains to be seen, how the actual performance is related to this interpretation of data. It may be not be regarded as particularly surprising that the GPs' attitudes did not change, because changing of attitudes is a more time-consuming process than what can be provided by a short two days course.

Table 7: Changes of GPs' knowledge scores with respect to different types of questions for both the intervention and control arms, including percentages of maximum score

Knowledge test	INTERVENTION		CONTROL		Difference between Intervention and control groups
	Pre-intervention Mean (SD[*])	Post-intervention Mean (SD)	Pre-intervention Mean (SD[*])	Post-intervention Mean (SD[*])	
MCQ and Likert	10.4 (2.1) 58%[¤]	14.4 (1.7) 80%[¤]	10.2 (1.9) 57%[¤]	13.1 (2.1) 73%[¤]	
Difference Pre-post	4.0 (22%[¤]) (t [77] = -12.7, P < 0.001)		2.9 (16%[¤]) (t = -9.5,P <0.001)		**1.1 (6%[†])** **p= 0.002**
Case vignettes and essay question	3.6 (2.0) 45%[¤]	4.9 (1.8) 61%[¤]	3.5 (1.9) 44%[¤]	3.8 (1.6) 48%[¤]	
Difference Pre-post	1.3 (16%[¤]) (t [77] = -5.2, p < 0.001)		0.3 (4%[¤]) (t = -1.3, p < 0.185)		**1.0 (12%[†])** **p = 0.011**

SD*= Standard Deviation

[†] = Intervention effect

[¤] = Percentage of maximum score

Table 8: Changes in GPs' knowledge scores in the large and small groups in the intervention and control arms

Knowledge questions		INTERVENTION		CONTROL		*p*
		Small group	Large group	Small group	Large group	
Pre-Intervention Mean (SD)*	MCQ and Likert	10.4 (2.1) 58%¤	10.4 (2.2) 58%¤	10.4 (2.2) 58%¤	10.2 (1.9) 57%¤	0.159
	Case vignettes and essay question	5.2 (1.2) 65%¤	3.2 (2.0) 40%¤	3.4 (2.0) 43%¤	3.5 (1.9) 44%¤	0.002
Post-Intervention Mean (SD)	MCQ and Likert	15.4 (1.5) 86%¤	14.1 (1.6) 78%¤	12.9 (2.3) 72%¤	13.2 (2.0) 73%¤	< 0.001
	Case vignettes and essay question	6.3 (1.6) 79%¤	4.5 (1.7) 56%¤	4.1 (1.6) 51%¤	3.7 (1.7) 46%¤	< 0.001
Difference pre-post Intervention Mean (SD)	MCQ and Likert	5.0 (2.7) 28%¤	3.7 (2.7) 21%¤	2.5 (2.7) 14%¤	3.0 (2.8) 17%¤	0.020
	Case vignettes and essay question	1.1 (1.6) 14%¤	1.3 (2.2) 16%	0.7 (2.3) 9%¤	0.2 (2.2) 3%¤	0.051

¤ = *Percentage of maximum score*

Intervention effect between small groups regarding MCQ and Likert questions is 14%

Intervention effect between large groups regarding MCQ and Likert questions is 4%

Intervention effect between small groups regarding Case vignettes and essay question is 5%

Intervention effect between large groups regarding Case vignettes and essay question is 13%

Table 9: Changes in GPs' attitude scores in the large and small groups in the intervention and control arms

Attitude statements	INTERVENTION		CONTROL		
	Small group	Large group	Small group	Large group	p
Pre-Intervention Mean (SD)*	5.47 (1.50)	5.30 (1.93)	4.84 (2.06)	5.56 (1.96)	0.53
Post-Intervention Mean (SD)	7.13 (1.26)	6.85 (1.81)	6.84 (1.07)	7.39 (1.43)	0.24
Difference pre and post Intervention Mean (SD)	1.69 (1.85)	1.45 (2.03)	2.00 (2.38)	1.80 (2.10)	0.74

SD= Standard Deviation*

3.4 Sub-study IV

GPs in the intervention group significantly improved their mean scores for performance regarding the questions asked about appropriate diagnosis (part 1 of the checklist) by 15 percentage units, with an intervention effect of 14 percentage units (Table 10). For appropriate management regarding treatment and referral (part 2 in the checklist), the significant improvement in the intervention group was 16 percentage units, with an intervention effect of 20 percentage units (Table 10).

The changes in the whole intervention arm (with both large and small groups) were also compared with the large control group, which was a mix of GPs in the attitude and intention stages. The effects on both diagnosis (part 1) and treatment and referral (part 2) were of the same magnitude, as when comparing with the whole control arm, with intervention effects of 13 percentage units (p=0.001) and 18 percentage units (p=0.004), respectively (Table 12) .

Table 11 shows the changes on sub-group level. Post-hoc tests revealed that the significant improvements in the intervention group for both parts of the checklist were due mainly to changes in the small group. For appropriate diagnosis (part 1), the

intervention effect was 32 percentage units compared with the small control group ($p<0.0001$), and 28 percentage units compared with the large control group ($p<0.0001$). For treatment and referral (part 2), the intervention effect was 38 percentage units compared with the small control group ($p=0.002$) and 31 percentage units compared with the large control group ($p=0.002$).

For the large intervention group, the significant improvements in the post-test (11 percentage units for part 1 and 12 percentage units for part 2), were not significantly different from the control groups.

In summary, the results showed significant and moderately high effect on the GPs' performance in the intervention arm in comparison with the control arm. Overall, the particularly higher effects (around 30 percentage units) regarding both diagnosis and treatment of depression in the small intervention group were shown. These findings are in line with our hypothesis that GPs' in the action stage, as shown in sub-study II, should change their performance and they actually did. They can be considered as experts.

Table 10: Changes in practice scores in the intervention and control arms with intention-to-treat analysis

Checklists		Intervention	Control	t-test	*p* -value
Pre-assessment Mean (SD)	diagnosis[1]	48 (22)	48 (25)	0.07	0.95
	treatment[2]	3.3 (2.8)	2.6 (3.1)	1.7	0.09
Post-assessment Mean (SD)	diagnosis	63 (20)	49 (24)	4.5	<0.0001
	treatment	4.9 (3.5)	2.2 (2.5)	6.2	<0.0001
Difference post - pre	diagnosis	15 (25)	1.08 (26)	3.8	<0.0001
	treatment	1.6 (4.1)	-0.4(4.0)	3.4	<0.001

[1] *Maximum score = 100*

[2] *Maximum score = 10*

Table 11: Changes in practice scores between sub-groups in the two study arms with intention-to-treat analysis

Part of tools*		Intervention		Control		*p*- value
		Small group	Large group	Small group	Large group	
Pre-intervention Mean (SD)	diagnosis[1]	45 (21)	49 (22)	47 (27)	48 (24)	0.91
	treatment[2]	2.9 (2.7)	3.5 (3.0)	3.2 (3.6)	2.4 (3.0)	0.20
Post-intervention Mean (SD)	diagnosis	75 (20)	60 (19)	45 (29)	50 (22)	<0.0001
	treatment	5.8 (3.5)	4.7 (3.5)	2.2 (2.5)	2.3 (2.5)	<0.0001
Diff post- pre Mean (SD)	diagnosis	30 (23)	11 (24)	-0.2 (26)	2.0 (27)	<0.0001
	treatment	2.9 (3.8)	1.2 (4.1)	-0.9 (4.7)	-0.2 (3.8)	0.002

* Part 1 = appropriate diagnosis, Part 2 = correct treatment and referral notes

[1] Maximum score = 100

[2] Maximum score = 10

Table 12: Changes in practice scores in the intervention and control arms with intention-to-treat analysis (between *intervention large and small groups* as an intervention arm and *Control large group* as a Control arm)

Checklists		Intervention	Control	t-test	*p* -value
Pre-assessment Mean (SD)	diagnosis[1]	48 (22)	48 (24)	-0.01	0.99
	treatment[2]	3.3 (2.8)	2.4 (3.1)	2.0	0.05
Post-assessment Mean (SD)	diagnosis	63 (20)	50 (22)	4.0	<0.0001
	treatment	4.9 (3.5)	2.2 (2.4)	5.6	<0.0001
Difference between post - pre	diagnosis	15 (25)	2.0 (26.6)	3.3	<0.001
	treatment	1.6 (4.1)	·0.19(3.8)	2.9	<0.004

[1] Maximum score = 100

[2] Maximum score = 10

4 DISCUSSION

The results of this thesis confirm that a tailored intervention based on the modified stages-of-change model was effective to improve Iranian GPs' stages of change, knowledge and practice regarding diagnosis and treatment of depression, although their attitudes did not change.

4.1 Effectiveness of the stages-of-change model and interactive educational methods

Criteria for selecting theory should be based on evidence of usefulness in the context in which it is to be used and a clear, understandable hypothesis (Oxman, Fretheim et al. 2005). Both of these criteria were considered in this study. The theoretical model had been tested in previous studies (Parker and Parikh 2001; Buckley, Goering et al. 2003) and results were promising. However, their study design was different and I tried to make it stronger through a rigorous design. The allocation of GPs to different stages of change in the study area was assessed in the pilot study.

Needs and objectives of tailored intervention

Learner needs can be assessed subjectively (desired needs) or objectively (objective evidence showing what they need to know). One way of assessing objective needs is by gathering information regarding actual knowledge, performance and attitudes of GPs'(Hodges, Inch et al. 2001). In the current study, this was done by using the MPQ, through which GPs' real needs were assumed to be assessed. Based on these findings, the objectives of the course were determined. Overall, MPQ is a comprehensive measure for assessing physicians' needs and tailoring the intervention based on their real needs and, it is not just a self-assessment questionnaire. It illuminates the extent to which physicians in different stages encounter with the specific disease and how important the topic is from their points of view. In addition, it assesses their commitment to acquiring knowledge about the topic and it also appraises their interpersonal relationship with other health professionals, something which has been emphasized as important by some researchers (Lingard, Hodges et al. 2004). GPs were also involved in designing the educational programme.

Educational methods

Important variables which affect on educational interventions in continuing education include the duration of the intervention, level of active participation, involving learners in tailoring the education and relevance to the clinical context (Hodges, Inch et al. 2001).

Educational interventions can be of different duration, although, the most frequent format is a lecture-based short course. One of the review articles mentioned that in order to change doctors' attitudes and performance, it is probably useful to implement longitudinal educational interventions. However, some improvement of doctors' knowledge and skills in the field of psychiatry have nevertheless been shown to occur after interventions of a duration from a few sessions to six months length (Hodges, Inch et al. 2001). In the present study, we implemented two-day courses in both arms. This could be considered as a limitation of the study and if it had been perfomed as an ongoing intervention, then the effects could have been even greater and changes in doctors' attitudes might have occurred.

The use of interactive methods made it easier for participants to rehearse knowledge and skills. These methods included the use of standardized patients with giving feedback on participants' performance. Clinical and contextual relevance was incorporated into the design of the current study by the psychiatrist teachers, who were encouraged to act in their real clinical roles by using standardized patients or other interactive methods and giving feedback to participants. All the psychiatrists in this study were familiar with the context of primary care.

Effects

The results showed that 75% of the GPs in the large group shifted from the attitude to intention stage and a few then even to the action stage. This is highly significant for confirming our hypothesis. Based on the stages-of-change assumptions, GPs who moved to the "action stage" in the intervention small group should improve their practice significantly. The results of the thesis confirm that the process of change is according to the model. In the small group, 65% of the participating GPs' shifted to the action stage and the results of doctors' knowledge assessment regarding MCQ and Likert questions, with the emphasis on treatment of depression, showed that performance improved and that the intervention effect between the small intervention

and control groups was 14 percentage units. Moreover, the intervention effects regarding actual performance between small intervention and control groups showed that 30 percentage units improved their treatment mean scores. Overall, our findings confirm that the results are in line with the hypothesis, according to which physicians in the action stage are the ones that could be expected to actually change performance and diagnosis of depression.

However, the effects may be regarded as moderate, although we should take into account that change in physicians' understanding of practice is difficult to accomplish and in comparison with other studies, the results are still quite impressive. Grol and Grimshaw did a systematic review of 235 assessments of guideline dissemination and implementation strategies. One conclusion was that it is important to tailor interventions based on identified barriers to change. Most intervention studies had some effect, although on average lower (about 10%) than in this study. None of the theories were appropriate to make a change in all settings. It was concluded that we probably need to continue to use different theories (Grol and Grimshaw, 2003). In some recent studies, with a randomized controlled trial design and assessing primary care physicians' knowledge, skills and attitudes, the results showed that doctors changed their knowledge and in some of the studies, also changed their attitudes slightly. However, their practice did not change significantly (Naismith, Hickie et al. 2001; Shuval, Berkovits et al. 2007; Vicente, Kohn et al. 2007) and it was concluded that change in this domain is not an easy task.

Hence, changing physician practice is a hard task, for achieving to this goal we combined the stages of change model with multifaceted interactive learning methods in order to reinforce our design. We should take it into account that the average effects of intervention for changing doctors performance in the previous studies were about 10% (Grol and Grimshaw 2003). In addition, regarding the effectiveness of multifaceted learning strategies on physicians performance, there is a controversy between scientists (Oxman, Thomson et al. 1995; Mann 2002; Shojania and Grimshaw 2005). Based on these evidences for changing doctors performance applying one approach for implementing research findings into clinical practice in different situation is not effective enough so we assume that combination of theory and interactive methods could be a sufficient way to changing physicians' understanding into clinical practice.

4.2 Methodological Considerations

Assessment intervals

Assessment of the GPs' stages-of-change, knowledge and attitudes was done 2-3 months before and one month after the intervention (Table 3). The physicians' performance was assessed two months before and two months after the intervention. The advantage of the short- term interval of performance assessment is that it keeps the GPs in the study. The disadvantage is that we do not have enough information about the long-term effects of the educational intervention in order to understand how change is maintained.

Overall, the research group made a decision to perform the assessment as soon as possible after the intervention because if the GPs were not accessible due to moving to other areas, then the number of participants would be reduced and not enough to show significant results with accurate statistical power. Performing post-assessment twice would have been too expensive for this project.

Educational intervention duration and methods

It may be seen as a limitation of the study that the GPs in the intervention arm actively participated in twelve hours of training compared with eight hours in the control arm. However, the time of teacher involvement was the same in all four groups in both arms. The extra time used in the intervention arm was intended to help the participants to learn from peers during their group work. It was considered that the extra peer group activities in the intervention groups were for elaboration only, typical for a workshop-type activity. The control group also had the possibility for individual study and self-elaboration by themselves. Delivering the information in eight hours and then adding the interactive methods and peer discussion can be expected to make the training more effective. However, it is not a question of mechanical things, such as only study hours, as commented by Handelsman: "Why do outstanding scientists who demand rigorous proof for scientific assertions in their research continue to use and, indeed, defend on the basis of the intuition alone, teaching methods that are not the most effective? Many scientists are still unaware of the data and analyses that demonstrate the effectiveness of active learning techniques" (Handelsman, Ebert-May et al. 2004).

Full control of other potentially influencing factors could not be achieved, as the educational methods were different. However, it should be emphasized that the

comparison was in fact done between these different methods. The assignment of participants in the intention stage in the intervention arm to the small group was a consequence of using the stages-of-change model, including appropriate learning methods for each stage.

Although the participants were randomly selected among all GPs working in southern Tehran, the external validity is partially reduced by the fact that one fifth of the initially selected GPs did not respond to the invitation. In addition, 8% of the participating doctors could not be assessed regarding their performance. However, the participating GPs match reasonably well with the total population of GPs regarding age, gender and length of work experience. Particular focus should be put on attitudinal topics such as the opportunity to discuss GPs' perceptions regarding inter-professional relationship and the health care system.

Using unannounced standardized patients

Arguments have been raised regarding the use of unannounced SPs. One of the drawbacks of unannounced SPs is the high expense in comparison with other assessment methods. Another one is that some people regard it is an unethical issue to send unannounced SPs' to doctors' offices. However, based on the literature, I argue that SPs are a valid tool for evaluating the effectiveness of an intervention programme in real clinical practice, which is completely different to a structured examination (Carney, Dietrich et al. 1995; Maiburg, Rethans et al. 2003). Moreover, in Iran the patient recording system for private clinics is not available so using unannounced SPs is the most valid and feasible method.

4.3 Strengths and practical implications

Applying the stages-of-change model as a needs-assessment tool was helpful in designing the educational events which aimed at changing GPs' understanding of management of depressive disorders and which led to change in their performance. A definite strength of the study is the rigorous design, including using intention-to-treat analysis.

Using unannounced SPs as a measuring instrument was adapted and extensively validated in Iran for the first time. Assessing physicians' performance by unannounced SPs in their real work setting is essential and it is regarded as the most appropriate

method (Hodges, Inch et al. 2001). Based on one systematic review article, the number of studies covering all parts of unannounced SP validation is very limited (Rethans Gorter et al. 2007). Furthermore, the total number of assessments of GPs' performance, makes the current study the largest study of this kind found in the literature (Rethans Gorter et al. 2007).

The documents that SPs collected from the doctors' offices (prescriptions, referral papers and laboratory examinations) were important from two perspectives: First, the documents showed how well the GPs' diagnosed and treated patients. Second, written documents illustrated the co-operation among health professionals and in appropriate situations, whether the doctor referred patients to further care or not. Experts understand these factors in implicit, dynamic and discipline-specific ways but novices are not capable of writing relevant prescriptions or referral letters because they lack a comprehensive understanding of influential factors (Lingard, Hodges et al. 2004). Using these written materials, psychiatrists decided whether the GPs had made correct or wrong judgments regarding different cases. It seems that it is possible to distinguish between a novice and an expert based on their treatment mean scores.

GPs' competence with respect to other theories

Our goal in a CPD context is training competent doctors and turning GPs' competence into possible expertise based on the needs assessment and current theories of learning. Based on one the definitions of competence and incompetence, it was recommended to avoid testing only pure knowledge, or merely general skills, or just using highly standardized measures or assessing doctors' competence just based on their capability of reflection (Hodges 2006). Based on the current theories of understanding of professional practice, many scientists underpin the importance of sociological, political, economic and cultural effects on changing physicians' practice (Bandura 1997; Lundborg, Wahlstrom et al. 1999; Dall'Alba 2004; Hodges 2006; Dall'Alba and Barnacle 2007). Having a variety of assessment methods enables us to look at the whole picture of the effects of an educational intervention.

In order to train competent physicians in a CPD context, the stages-of-change model was tailored in the present study and in some degree the design was supported by other theories, such as community of practice, learning from experience and expertise. In tailoring the intervention, moreover the acquisition of knowledge and skills, the opportunity to learn from peers was taken into the account when considering the

cultural and sociological perspectives of learning. The possibility of receiving feedback based on the theory developed by Schön was also considered in the intervention.

According to the expertise theory, those who shifted to the action stage could be considered as "experts". Their diagnosis and treatment pattern improved significantly. Their knowledge was also modestly enhanced also, but not as much as their problem-solving capability when treating patients. These findings are in line with the expertise theory that expert development is not always linear. For instance, medical experts do not always remember essential facts better than medical students but they are good at problem-solving procedures, which is in line with our assumption (Hobus, Schmidt et al. 1987; Lonka and Ahola 1995).

Educational material

Finally, a strength of the study from a CPD perspective was the compiling of a guidebook as a self- directed learning method for the management of depression. This guidebook can remind the physicians of the practical points about the treatment of depression after participating in the education.

4.4 Recommendations and suggestions for further research

We have demonstrated that a theoretical model of medical understanding and behavioural change can be used to develop educational formats that suit learners' different stages and that participation in such tailored education can help primary care doctors to change their practice behaviour regarding the management of depressive disorders. Our findings have practical implications for how CME shifts toward the broader context of CPD programmes in Iran and are possibly also applicable in other parts of the world. Using unannounced SPs seems to be a valuable tool in assessing performance in clinical practice and can be recommended.

Assessing the long-term effect of the intervention should be considered in the future. We recommend a follow up on the views of participants in the attitude and intention stages, who did not shift to the higher stage, in order to understand more about barriers to change. In addition, for those moving to the intention stage, the need to arrange continued meetings for them should be taken into account, so that they get time to reflect on their own practices and discuss with peers, which was emphasized in the

intervention small group in this study. The stages-of-change theory should be used to design and test new educational interventions in continuing professional development.

51

4.5 Conclusion

Through a tailored intervention based on the stages-of-change model and using interactive methods, we facilitated for doctors to improve their understanding and clinical practice, emphasizing a shift from CME towards CPD. Implications for continuing professional development arising from the study are the need for a clear and explicit focus on developing doctors' understanding of medical practice throughout CPD programmes.

5 ACKNOWLEDGEMENTS

6 REFERENCES

American Psychiatric Association, 2000. Diagnosis and statistical Manual of Mental Disorders DSM-IV-TR. American Psychiatric Association, Washington DC.

Balint, M.;1957. The Doctor, His Patient and the Illness. International Universities Press, New York.

Bandura, A., 1997. Self efficacy. The exercise of control. Freeman & co, New York.

Bolander Laksov, K., Mann, S., Dahlgren, L: O:., 2008. Developing a community practice around teaching: case study. Higher Education research and development 27(2), 121-132.

Bosch, M., van der Weijden, T., Wensing, M., Grol, R., 2007. Tailoring quality improvement interventions to identified barriers: a multiple case analysis. J Eval Clin Pract 13(2), 161-8.

Buckley, L. L., Goering, P., Parikh, S. V., Butterill, D., Foo, E. K., 2003. Applying a 'stages of change' model to enhance a traditional evaluation of a research transfer course. J Eval Clin Pract 9(4), 385-90.

Carney, P. A., Dietrich, A. J., Freeman, D. H.; Mott, L. A., 1995. A standardized-patient assessment of a continuing medical education program to improve physicians' cancer-control clinical skills. Acad Med 70(1), 52-8.

Cicchetti, D. V., Nelson, L. D., 1994. Re-examining threats to the reliability and validity of putative brain-behavior relationships: new guidelines for assessing the effect of patients lost to follow-up. J Clin Exp Neuropsychol 16(3), 339-43.

Dall'Alba, G., 2004. Understanding professional practice: investigation before and after an educational programme. Studies in Higher Education 29(6), 679-692.

Dall'Alba, G., Barnacle, R., 2007. An ontological turn for higher education. Studies in Higher Education 32(6), 679-91.

Dalton, C. C., Gottlieb, L. N., 2003. The concept of readiness to change. J Adv Nurs 42(2), 108-17.

Dave R: H:, 1975. Developing and Writing Behavioural Objectives. Educational Innovator Press.

Davis, D., Barnes, B.E., Fox, R.D. (2003). The Horizon of Continuing Professional Development. The continuing professional development of physicians from research to practice: United States of America, AMA**:** 394.

Davis, D. A., Thomson, M. A., Oxman, A. D., Haynes, R. B., 1992. Evidence for the effectiveness of CME. A review of 50 randomized controlled trials. JAMA 268(9), 1111-7.

Davis, D. A., Thomson, M. A., Oxman, A. D., Haynes, R. B., 1995. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA 274(9), 700-5.

Davis, D., Evans, M., Jadad, A., Perrier, L., Rath, D., Ryan, D., Sibbald, G., Straus, S., Rappolt, S., Wowk, M., Zwarenstein, M., 2003. The case for knowledge translation: shortening the journey from evidence to effect. BMJ 327(7405), 33-5.

Dopson, S., FitzGerald, L., Ferlie, E., Gabbay, J., Locock, L., 2002. No magic targets! Changing clinical practice to become more evidence based. Health Care Manage Rev 27(3), 35-47.

Eccles, M., Grimshaw, J., Walker, A., Johnston, M., Pitts, N., 2005. Changing the behaviour of healthcare professionals: the use of theory in promoting the uptake of research findings. J Clin Epidemiol 58(2), 107-12.

Eccles, M. P., Foy, R., Bamford, C. H., Hughes, J. C., Johnston, M., Whitty, P. M., Steen, N., Grimshaw, J. G., 2006. A trial platform to develop a tailored theory-based intervention to improve professional practice in the disclosure of a diagnosis of dementia: Study protocol [ISRCTN15871014]. Implement Sci 1, 7.

Eraunt, M. (2000). Non-formal learning, implicit learning and tacit knowledge in professional work. The necessity of informal learning. F.Coffield. Bristol, The policy press**:** 12-31.

Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W., Robinson, N., 2006. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 26(1), 13-24.

Grol, R., Grimshaw, J., 2003. From best evidence to best practice: effective implementation of change in patients' care. Lancet 362(9391), 1225-30.

Grol, R., Wensing, M., 2004. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 180(6 Suppl), S57-60.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., Gentile, J., Lauffer, S., Stewart, J., Tilghman, S. M.,Wood, W. B., 2004. Education. scientific teaching. Science 304(5670), 521-2.

Hobus, P. P., Schmidt, H. G., Boshuizen, H. P., Patel, V. L., 1987. Contextual factors in the activation of first diagnostic hypotheses: expert-novice differences. Med Educ 21(6), 471-6.

Hodges, B., 2006. Medical education and the maintenance of incompetence. Med Teach 28(8), 690-6.

Hodges, B., Inch, C., Silver, I., 2001. Improving the psychiatric knowledge, skills, and attitudes of primary care physicians, 1950-2000: a review. Am J Psychiatry 158(10), 1579-86.

Holm H.A., 1998. Continuing medical education: Quality issues in continuing medical education. BMJ 316, 621-624.

Jamtvedt, G., Young, J. M., Kristoffersen, D. T., O'Brien, M. A.,Oxman, A. D., 2006. Audit and feedback: effects on professional practice and health care outcomes. Cochrane Database Syst Rev (2), CD000259.

Jaques D., 2003. ABC of learning and teaching in medicine: Teaching small groups. BMJ 326(7387), 492-4.

Kerr, M., Blizard, R., Mann, A., 1995. General practitioners and psychiatrists: comparison of attitudes to depression using the depression attitude questionnaire. Br J Gen Pract 45(391), 89-92.

Kjeldmand D., Holmström I., Rosenqvist U,2004. Balint training makes GPs thrive better in their job. Patient Educ Couns, 55(2):230-5)

Kolb, D. A.,1984. Experience as the Source of Learning and Development. Experiential Learning. Englewood Cliffs, N.J, Prentice-Hall, Inc.

Krupinski, J., Tiller, J. W., 2001. The identification and treatment of depression by general practitioners. Aust N Z J Psychiatry 35(6), 827-32.

Lecrubier, Y., 2007. Widespread under recognition and under treatment of anxiety and mood disorders: results from 3 European studies. J Clin Psychiatry 68 Suppl 2, 36-41.

Lingard, L., Hodges, B., MacRae, H., Freeman, R., 2004. Expert and trainee determinations of rhetorical relevance in referral and consultation letters. Med Educ 38(2), 168-76.

Lonka, K. 2004. Foreword. How to motivate healthy learning,. Rochester, NY: MHH Publications.

Lonka, K., Ahola, K., 1995. Activating instruction: How to foster study and thinking skills in higher education. Euro J Psycho Edu, 10(4), 351-8.

Lord, F.M., 1952. The relationship of the Reliability of Multiple-Choice Test to the Distribution of Item Difficulties. Psychometrika, 18, 181-194.

Lundborg, C. S., Wahlstrom, R., Dall'Alba, G., 1999. Ways of experiencing asthma management. Variations among general practitioners in Sweden. Scand J Prim Health Care 17(4), 226-31.

MacDermid, J. C., Solomon, P., Law, M., Russell, D., Stratford, P., 2006. Defining the effect and mediators of two knowledge translation strategies designed to alter knowledge, intent and clinical utilization of rehabilitation outcome measures: a study protocol [NCT00298727]. Implement Sci 1, 14.

Maiburg, B. H., Rethans, J. J., Schuwirth, L. W., Mathus-Vliegen, L. M.,van Ree, J. W., 2003. Controlled trial of effect of computer-based nutrition course on knowledge and practice of general practitioner trainees. Am J Clin Nutr 77(4 Suppl), 1019S-1024S.

Mann K.V., 2002. Continuing Medical Education. The international Handbook of Research in Medical Education. Norman, GR, Vander Vluten Cees, PM and Newble, DL. Netherlands, KLUWER. **1:** 415-457.

Marandi, A., 1996. Integrating medical education and health services: the Iranian experience. Med Educ 30(1), 4-8.

Meredith, L. S., Rubenstein, L. V., Rost, K., Ford, D. E., Gordon, N., Nutting, P., Camp, P.,Wells, K. B., 1999. Treating depression in staff-model versus network-model managed care organizations. J Gen Intern Med 14(1), 39-48.

Miller, E. T., Spilker, J., 2003. Readiness to change and brief educational interventions: successful strategies to reduce stroke risk. J Neurosci Nurs 35(4), 215-22.

Moher, D., Schulz, K. F., Altman, D. G., 2001. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 357(9263), 1191-4.

Moussaoui, D., Lopez-Ibor J J, Okasha A, Maj M, 2003. Depression in general practice. WPA Bulletin on Depression 1, 1-2.

Murthy, J. M., 2003. Some problems and pitfalls in developing countries. Epilepsia 44 Suppl 1, 38-42.

Naismith, S. L., Hickie, I. B., Scott, E. M., Davenport, T. A., 2001. Effects of mental health training and clinical audit on general practitioners' management of common mental disorders. Med J Aust 175 Suppl, S42-7.

Noorbala, A. A., Bagheri Yazdi, S. A., Yasamy, M. T., Mohammad, K., 2004. Mental health survey of the adult population in Iran. Br J Psychiatry 184, 70-3.

Nowlen, P. M. (1988). The performance model.A new approach for business and the professions. New York, NY, Macmillan.

O'Brien, M.A., Freemantle, N., Oxman, A.D., Wolf, F., Davis, D.A., Herrin, J., 2001. Continuing education meetings and workshops: effects on professional practice and health care outcomes. Cochrane Database of Systematic Reviews, (1).

Ormel, J., VonKorff, M., Ustun, T. B., Pini, S., Korten, A., Oldehinkel, T., 1994. Common mental disorders and disability across cultures. Results from the WHO Collaborative Study on Psychological Problems in General Health Care. Jama 272(22), 1741-8.

Olesen, F. and Hjortdahl, P. (1999). CME in primary care: the way forward. Scand J Prim Health Care 17(3): 131-4.

Oxman, A. D., Fretheim, A., Flottorp, S., 2005. The OFF theory of research utilization. J Clin Epidemiol 58(2), 113-6; discussion 117-20.

Oxman, A. D., Schunemann, H. J., Fretheim, A., 2006. Improving the use of research evidence in guideline development: 16. Evaluation. Health Res Policy Syst 4, 28.

Oxman, A. D., Thomson, M. A., Davis, D. A., Haynes, R. B., 1995. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. CMAJ 153(10), 1423-31.

Parker, K., Parikh, S. V., 2001. Applying Prochaska's model of change to needs assessment, programme planning and outcome measurement. J Eval Clin Pract 7(4), 365-71.

Patel V, Glaser R, Arocha J, 2000. Cognition and expertise: acquisition of medical competence. Clinical Invest Med 23(4), 256-260.

Prochaska, J.O, DiClemente, C.C., Norcross, J.C., 1992. In search of how people change. Applications to addictive behaviours, 47, 1102-1114.

Rethans JJ, Gorter S, Bokken L, Morrison L, 2007. Unannounced standardised patients in real practice: a systematic literature review. Medical Education 41(6), 537-49.

Riemsma, R. P., Pattenden, J., Bridle, C., Sowden, A. J., Mather, L., Watt, I. S., Walker, A., 2002. A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behaviour change. Health Technol Assess 6(24), 1-231.

Rikers, R. M., Schmidt, H. G., Boshuizen, H. P., Linssen, G. C., Wesseling, G. and Paas, F. G. (2002). The robustness of medical expertise: clinical case processing by medical experts and subexperts. Am J Psychol 115(4): 609-29.

Rowe, A.K., Savigny, D.D., Lanata, C.F., Victora, C.G., 2005. How can we achieve and maintain high-quality performance of healthworkers in low-resource settings? Lancet 366(17), 1026-1034.

Sadeghi, M., Shirazi, M., Arbabi, M., Assadi, S.M., Zeinaloo, A., 2007. Guidebook for management and treatment of depression disorders by General Physicians. Tehran University of Medical Sciences, Tehran.

Sharen, S., 1994. Handbook of Cooperative Learning Methods. Westport: Greenwood Press.

Shea, J., A., Fortna, Gregory, S., (2002). Psychometric Methods. The international Handbook of Research in Medical Education. In: Norman GR, Vander Vluten Cees PM, Newble DL, editors. Netherlands, Kluwer. 1.

Shirazi, M., Zeinaloo, A A., Kashani,. A S., Alaeddini, F, 2004. Assessing the gap between current and desirable needs in TUMS CME Unit: participants' viewpoints. Iranian Journal of Medical Education, 5(1), 17-22.

Shojania, K. G., Grimshaw, J. M., 2005. Evidence-based quality improvement: the state of the science. Health Aff (Millwood) 24(1), 138-50.

Shuval, K., Berkovits, E., Netzer, D., Hekselman, I., Linn, S., Brezis, M., Reis, S., 2007. Evaluating the impact of an evidence-based medicine educational intervention on primary care doctors' attitudes, knowledge and clinical behaviour: a controlled trial and before and after study. J Eval Clin Pract 13(4), 581-98.

Smith Peter G., Morrow Richard H., 1996. Field Trials of Health Interventions in developing countries: Macmillan education LTD, London.

Stanton, F., Grant, J., 1999. Approaches to experiential learning, course delivery and validation in medicine. A background document. Med Educ 33(4), 282-97.

Sutton, S., 2005. Another nail in the coffin of the transtheoretical model? A comment on West (2005). Addiction 100(8), 1043-6; author reply 1048-50.

Tamblyn R, Battista R, 1993. Changing clinical practice: which interventions work? Journal of Continuing Education in the Health Professionals 13, 273-288.

Timmermans, S., Mauck, A., 2005. The promises and pitfalls of evidence-based medicine. Health Aff (Millwood) 24(1), 18-28.

Tomson, T., (2005). Telephone support for smoking cessation; the Swedish example. Social Medicine. Stockholm, Karolinska Institutet. PhD Desertation: 41.

Vicente, B., Kohn, R., Levav, I., Espejo, F., Saldivia, S.,Sartorius, N., 2007. Training primary care physicians in Chile in the diagnosis and treatment of depression. J Affect Disord 98(1-2), 121-7.

Wahlstrom, R. (1997). Bridging the gap between guidlines and clinical practice Public Health Sciences. Stockholm, Karolinska Institutet. PhD Desertation: 128.

WHO (2001). Psychiatric disorders in general medical practice. World health day 2001. EMRO, WHO. Geneva,.

WHO (2006). Country cooperation strategy at a glance; Islamic republic of Iran. Geneva: WHO: 2006.

7 APPENDICES

7.1 Analysis for assessing the external validity

Difficulty indices for items with one correct alternative worth a single point, simply denote the percentage of students who answer an item correctly. In this case, it is also equal to the item mean. The item difficulty index ranges from 0 to 100; the higher the value, the easier the question. When an alternative is worth other than a single point, or when there is more than one correct alternative per question, the item difficulty is the average score on that item divided by the highest number of points for any one alternative. Item difficulty is relevant for determining whether students have learned the concept being tested. It also plays an important role in the ability of an item to discriminate between students who know the tested material and those who don't. The widely-used *item discrimination index* reports the difference between the proportions of high and low scorers answering a dichotomous item correctly. High values are reputed to flag good items, low values bad.

Following the conventional approach, we extracted from the sample G of persons two equal-sized sub-groups, the high scorers UG (Upper group) and the low scorers LG (Lower group). Typically, following Kelley (1939), the distribution of persons is treated as normal and UG and LG are the upper and lower 27%. Then the item discrimination index is $d = p$ (UG) - p(LG), where p(UG) and p(LG) are the proportions of correct answers by UG and LG respectively. The maximum value of d, Max(d), is 1.0 and occurs when all the UG group succeed and all the LG group fail on an item (Lord 1952).

7.2 Knowledge questionnaire

1-What are the main signs and symptoms of depressive disorders. Please list below:

2-Choosing the most appropriate answer:

2-1 You counsel a depressed patient who is taking 20mg of Fluxitine

20mg /per day. She has also suffered from allergic rhinitis. Which of the following drugs should not be prescribed?

A- Terphenadine   B-Loratadine

C-Cholorpheniramine   D-Diphenhydramine

2-2 A 40-year- old lady came to your office. She is complaining of apathy, insomnia , loss of energy and, appetite and being disappointed. Once, five years ago she had the same symptoms. How long should her treatment be continued?

A- Three months   B-Six months

C- One year   D- An unlimited time

2-3 A patient comes to your office, she is taking an antidepressant drug but she forgets its name. When she has started her treatment,s' she suffers some side effects including: dry mouth, constipation and ……Which of the following drugs did she probably use?

A- Lithium   B-Imiperamine

C-Fluxitine   C-Citalopram

2-4 You see a 54-year- years old man who is suffering from major depression and is displaying suicidal tendencies. She also does not have any psychotic symptoms. He has hypertension and cardiovascular disorders. Which of the following drugs do you prescribe for him?

A- Fluxitine   B- Amitriptilen

C- Imiperamine   D-Doxipine

2-5 What is the <u>effective minimum dose</u> of Nortriptilin for treatment of major depression in adolescence?

A- 25mg   B-50mg

C- 75mg   D-100mg

2-6 Which of the following drugs is most effective in the treatment of depressive disorders more rapidly?

A- Fluxitine   B- Imiperamine

C- Tetrazodone   D-Electroshock

2-7 If an antidepressant is not having an effect on a depressed patient, how long should the treatment should be continued?

A- Two weeks   B- Four weeks

C- Eight weeks   D- Twelve weeks

3- To what extent do you believe each of the following statements is <u>true</u> or <u>false?</u> On the line next to each statement, circle one number (from 1 to 5) for the answer that is closest to your own.

Definitely True=1   Mostly True=2   Don't know=3   Mostly false=4   Definitely false=5

a. The maintenance phase of treatment for major

   depression focuses on preventing recurrence.   1 2 3   4   5

b. If psychotherapy for major depression has no effect within

   6 weeks of regular sessions, medication is recommended. 1   2   3   4   5

c. An appropriate trial of antidepressant medication for major

   depressive disorder requires use of therapeutic dosage

   daily for at least 4-6 months.   1   2   3   4   5

d. Anti depressant drugs have wide ranges of side effects

   on patients taking these drugs.   1   2   3   4   5

e. Medication and psychotherapy are efficacious for

   depression in elderly adults as well as for the non elderly. 1   2   3   4   5

f. Evidence suggests that primary care clinicians prescribe

   appropriate dosage of antidepressant to fewer than a third

   patients with a current major depressive disorder.   1   2   3   4   5

g. Dysthymic disorders is mild, brief depression.　　1　2　3　4　5

h. In general, antidepressant medication should be discontinued

　　after 4-9 months for patients with a single major depressive

　　episode who no longer have symptoms of depression.　　1　2　3　4　5

i. Anoxiolytic and sedatives (minor tranquilizers) have equivalent

　　efficacy in major depression as antidepressant medications.　　1　2　3　4　5

j. Psychotherapy with a trained therapist is appropriate as the sole

　　treatment for moderate major depression that is not

　　　chronic, psychotic or melancholic.　　1　2　3　4　5

k. Tricyclic antidepressants and SSRIs have equivalent

　　side effect profiles.　　1　2　3　4　5

Vignettes

Patient A (lost husband)

Mrs A is 75 years old. Her husband died six weeks ago. She says she is depressed because of his death. She is dressed in black, is tearful, and lethargic. She complains of her loneliness, difficulties in sleeping at night and not being able to get interested in anything.

Diagnosis:

Management:

Patient B (gastrointestinal complaints)

Mr B is a 38-year-old accountant. He presents with abdominal discomfort, diarrhoea, dry mouth, and sleeplessness (with walking in the night and early waking in the morning). This has gone on for over 2 months and he has lost 5 kilograms in weight. He has not had mucus nor blood in his stools. There is no past history of gastrointestinal disturbance. He is not coping as he used to at work and is uncharacteristically anxious and irritable. He has lost interest in sport and hobbies.

Diagnosis:

Management:

Patient C (recurrent depression)

Mrs C is 45 years old. She presents with a recurrence of depression over 2 months. This is like other episodes she has had over the last 15 years. She had hoped it would just go away. She is depressed, tearful, feels hopeless and beyond help and does not show much animation during the interview. She lacks energy and is not keeping up with household tasks. She is slow in her speech and says she is slow in her activities and thoughts as well. Her appetite is down and she says she has lost quite a lot of weight. She feels worse in the morning. Her sex drive has gone and she has no joy in her life. She is tense with episodic panic, palpitations and shortness of breath. She has felt too tired to go out or visit friends, but felt forced to come to see you, as she is so down. She does not believe life is worth living any more and fears she may harm herself. She has many old tablets from previous courses of treatment, and there is a firearm in the house. She feels this illness is punishment for not having done all she could have for others. She asks for help.

Diagnosis:

Management:

7.3 Attitude Questionnaire

This questionnaire is designed to measure your attitude to the management of depression in general practice. For each statement please circle the answer that reflects your degree of agreement with the statement.

1) Biochemical abnormality is the basis of severe depression.

Strongly agree   Agree   Neutral   Disagree   Strongly disagree

2) It is difficult to know if patients are unhappy or have a clinical depressive disorder needing treatment.

Strongly agree   Agree   Neutral   Disagree   Strongly disagree

3) I feel comfortable dealing with a depressed patient's needs.

Strongly agree   Agree   Neutral   Disagree   Strongly disagree

4) Depression is a patient response which cannot be changed.

Strongly agree   Agree   Neutral   Disagree   Strongly disagree

5) Becoming depressed is part of being old.

Strongly agree   Agree   Neutral   Disagree   Strongly disagree

6) Working with depressed patients is heavy going.

Strongly agree   Agree Neutral   Disagree   Strongly disagree

7) There is little to offer depressed patients who do not respond to what GP's do.

Strongly agree   Agree   Neutral   Disagree   Strongly disagree

8) Psychotherapy tends to be unsuccessful with depressed patients.

Strongly agree   Agree Neutral   Disagree   Strongly disagree

9) Depressed patients needing antidepressants are better off with a psychiatrist than a GP.

Strongly agree   Agree Neutral   Disagree   Strongly disagree

10) If psychotherapy was freely available, it would be more beneficial than antidepressants for most patients.

Strongly agree   Agree   Neutral   Disagree   Strongly disagree

7.4 Scenarios

1-Assessing the risk of suicide and appropriate management

Mr A is a single 25 years man, divorced 2 years ago. He was a drug abuser 2 years ago but now he just smokes cigarettes. He smokes two boxes each day. He has no children. He graduated from high school. He is unemployed. He is complaining of insomnia. He often wakes up early in the morning and has difficulty falling asleep again. His appetite is down and he has lost 10 kg in weight. It is having an adverse effect on his relationship with his family and friends. He does not like his life and he does not predict a good future for himself.

2-The differential diagnosis of normal and pathologic grief

Mrs B is a 25-year- old widow. Her husband died six months ago in a car accident, when he was 30 years old. She cried a lot. She has two children. She doesn't take such good care of her children and household. She goes to sleep late at night. Her appetite is normal. She is irritable and gets angry very quickly. She is preoccupied with memories of her husband.

3-Somatization, somatic symptoms and depression

Mrs C is a 40 year old married lady. She has two children. She is complaining of headache and tiredness. She also has some muscular pains. Her muscular pains are located in different parts of her body. When she gets angry her pain severity increases. She is not capable of doing her housework properly. Her appetite is down and she has lost 5kg in weight. She cries easily. She is irritable and argues with her husband and children.

4-The difference between hypothyroidism and depression

Mrs S is a 25- year- old lady, who complains of fatigue, loss of energy and hypersomnia. She always feels tired. She has gained around 5kg in weight. She is not capable of doing her daily work. She went to the endocrinologist three months ago. Her disease was diagnosed as hypothyroidism and she is on Levothyroxine. Although, she feels a bit better, she is still suffering from some of the symptoms.

5-Bipolar disorder diagnosis and management

Mr D is a 27-year- old single man. He has a good appetite and has had hypersomnia for 2 months. During the day he usually stays at home and does not want to go out. He is so disappointed and he thinks he'd be much better off dead. He was previously hospitalized for hyperactivity and early anger attacks. He was put on medication in the hospital and for a short period after that but he has forgotten the names of the drugs. After treatment he feels better and he does not use any drugs at the present time.

7.5 Checklists

Risk of suicide – first part

Multiply	Items	Yes	No
3	Did doctor ask about sadness and/or mood swings?		
2	Did doctor ask about changes in appetite or weight?		
2	Did doctor ask about changes in sleep patterns?		
2	Did doctor ask about fatigue and decreased energy levels?		
2.5	Did doctor ask something about anhedonia or loss of interest?		
3	Did doctor ask about death thoughts?		
3	Did doctor ask about any previous attempts to commit suicide?		
2.5	Did doctor ask about any previous plans to commit suicide?		
2	Did doctor ask about feelings of excessive or inappropriate guilt regarding the last attempt to commit to suicide?		
2	Did doctor ask about previous history of psychiatric disorders?		
3	Did doctor talk about planning future visits?		
2	Did doctor ask about the possible severe diseases that could be the reason for the dysfunction?		
2	Did doctor ask about previous treatments or response to the treatment?		
2	Did doctor ask about drug abuse?		
2	Did doctor ask about any previous major somatic disorders?		
2	Did doctor ask about symptoms of bipolar disorder, like insomnia, excessive activity, etc?		

Risk of suicide - Second part

Multiply	Items	Correct	Mistake
2.5	Did doctor consider an appropriate treatment strategy?		
1.5	Did doctor consider an appropriate drug, dose and duration of treatment?		
4	Did doctor refer the patient to a psychiatrist?		
2	Quality of the referral paper		

Grief- first part

Multiply	Items	Yes	No
3	Did doctor ask about sadness and/or mood swings?		
2	Did doctor ask about changes in appetite or weight?		
2	Did doctor ask about changes in sleep patterns?		
3	Did doctor ask about fatigue and decreased energy levels?		
3	Did doctor ask about anhedonia or loss of interest?		
3	Did doctor ask what and when things happened that caused your low mood (grief)?		
3.5	Did doctor ask about death thoughts?		
3.5	Did doctor ask about any previous plans to commit suicide?		
3	Did doctor ask about previous history of psychiatric disorders?		
3	Did doctor ask about the possible severe diseases that could be reason for the dysfunction?		
2.5	Did doctor ask about previous treatments and/or response?		
3.5	Did doctor diagnose your problem as normal grief or depressive disorder ?		
3	Did doctor ask about previous major somatic disorders?		
3	Did doctor ask about symptoms of bipolar disorder like insomnia, excessive activity, etc?		

Grief- second part

Multiply	Items	Correct	Mistake
2.5	Did doctor consider an appropriate treatment strategy?		
7.5	Did doctor consider an appropriate drug, dose and duration of treatment?		

Headache and hypothyroidism- first part

Multiply	Items	Yes	No
3	Did doctor ask about sadness and/or mood swings?		
2	Did doctor ask about changes in appetite or weight?		
2	Did doctor ask about changes in sleep patterns?		
2.5	Did doctor ask about fatigue and decreased energy levels?		
3	Did doctor ask about anhedonia or reduction or loss of interest?		
3	Did doctor ask about death thoughts?		
3	Did doctor ask about any previous attempts to commit suicide?		
3	Did doctor ask about previous history of psychiatric disorders?		
3	Did doctor ask about possible severe diseases that could be reason for the dysfunction?		
2.5	Did doctor ask about previous treatments or response to treatment?		
2	Did doctor ask about drug abuse?		
2	Did doctor ask about previous major somatic disorders?		
2	Did doctor ask about symptoms of bipolar disorder like insomnia, excessive activity, etc?		

Headache and hypothyroidism - second part

Multiply	Items	Correct	Mistake
2.5	Did doctor consider an appropriate treatment strategy?		
5.5	Did doctor consider an appropriate drug, dose and duration of treatment?		
2	Did doctor order the examinations?		

Bi-polar disorder- first part

Multiply	Items	Yes	No
3	Did doctor ask about sadness and/or mood swings?		
2	Did doctor ask about changes in appetite or weight?		
2	Did doctor ask about changes in sleep patterns?		
2.5	Did doctor ask about fatigue and decreased energy levels?		
3	Did doctor ask about anhedonia or reduction or loss of interest?		
3	Did doctor ask about death thoughts?		
3	Did doctor ask about any previous attempts to commit suicide?		
3	Did doctor ask about previous history of psychiatric disorders?		
3	Did doctor ask about possible severe diseases that could be reason for the dysfunction?		
2.5	Did doctor ask about previous treatments or response to treatment?		
2	Did doctor ask about drug abuse?		
2	Did doctor ask about previous major somatic disorders?		
2	Did doctor ask about symptoms of bipolar disorder like insomnia, excessive activity, etc?		

Bi-polar disorder - second part

Multiply	Items	Correct	Mistake
2.5	Did doctor consider an appropriate treatment strategy?		
1.5	Did doctor consider an appropriate drug, dose and duration of treatment?		
4	Did doctor refer patient to a psychiatrist?		
2	The referral paper		