

CHAPER III

METHODOLOGY

The purposes of this study were to develop and establish psychometric properties of the Thai Family Health Routine scale (TFHR scale). The TFHR scale was developed based on the structural domain of the Family Health Model (Denham, 2002; 2003a). This chapter presented the methodology used for constructing the scale and testing its validity and reliability. The following was a discussion of the research design, research setting, population and sample, protection of human subject, data collection, instrument development, and data analysis procedures.

Research design

This is a scale development and psychometric study; both qualitative and quantitative approaches were used to answer the research questions. In particular, the guidelines for scale development in Burns & Grove (2001) and DeVillis (2003) were integrated and applied into two phases; scale construction composed of pilot study for generating the item pool and pretest study for constructing the TFHR scale, and psychometric testing as a main study for testing validity and reliability of the TFHR scale.

Research setting

Research setting of this study was the central region of Thailand which composed of 27 provinces including Bangkok metropolis (Ministry of Interior, 2000). The central region and Bangkok metropolis were selected as research setting because they are living area of various families who migrated from the other regions of Thailand for earning jobs, especially Bangkok metropolis (Ministry of Public Health, 2003). Therefore, the families in this area have potential for representing the target population of Thai family.

Population and Sample

In this study, the target population was Thai families; whereas the samples were Thai families living in the central region of Thailand. Population size was estimated by a number of family households living in Thailand approximately 15,877,200 families (National Statistical Office, 2000).

The samples were selected based on criteria that identifying themselves as Thai family living in Thailand, being able to use Thai verbal communication, and willing to participate in this study. Exclusion criterion for selecting samples was families that had member(s) experiencing physical or mental illness within their own families. In case of respondents, mothers/wives were theoretically accepted as representatives of the families (Denham, 2003a), who were assigned to answer the TFHR scale.

As noted in research design, this study consisted of three sections; pilot study, pretest study, and field study. Sample size estimation and sampling methods were separately operated as follows:

1. Samples of pilot study

A qualitative approach with family in-depth interview was employed in pilot study to get knowledge about routine health behaviors of Thai family. This information was used for wording item statements of item pool of the TFHR scale. Recruiting the participants, convenient sampling was employed to select families who were willing to participate in the interview from both nuclear and extended family types. A number of participants continued up to data saturation or when no new information emerged from the in-depth family interview. The number of participants in this qualitative approach was thirteen (N=13).

2. Samples of the pretest study

In this study, item analysis and item review were conducted to pretest study. At least 50 were required for a number of samples for item analysis, (Nunnally and Bernstein, 1994), whereas one tenth the size of samples proposed for main study was required for pretest study (Pett, Lackey, and Sullivan, 2003).

According to sample size of the main study were initially approximated to be 1,450 families. This number was calculated from ten times a number of items in the first draft scale which consisted of 145 items. To obtain 145 samples, Saraburi province and Bangkok metropolis were purposively selected to recruit 145 families from convenient sampling (N=145). The samples were selected from both rural and urban areas in order to represent same characteristics of samples in the main study. In case of item review, the participants were fifteen mothers/wives of the families selected by convenient sampling method (N=15).

3. Samples of the main study

The main study was to test construct validity and internal consistency reliability on the TFHR scale which consisted of 85 items. Samples of the study were separately selected according to testing procedures; confirmatory factor analysis, criterion related validity and reliability, and contrasted-groups approach.

3.1 Samples for testing confirmatory factor analysis

A relative large sample size was required for confirmatory factor analysis. In order to meet the criteria of reducing sampling error (Nunnally and Bernstein, 1994), the minimum ratio of samples per item (Dixon, 2001), and the excellent sample size for factor analysis (Comrey and Lee cited in Pett, Lackey, and Sullivan, 2003), the sample size at least 1,000 families was required.

The samples of this study were families living in the central region which composed of 27 provinces officially recorded by the Ministry of Interior (2000). A multi-stage sampling method, as shown in figure 3, was employed to recruit the samples. To meet the criterion at least 1,000 families by using a four-stage sampling method, four provinces; Bangkok, Chai Nat, Saraburi, and Phra Nakhon Si Ayutthaya, were randomly selected. From each of the four provinces, four amphurs/districts were also purposively selected. In each of sixteen ampurs/destricts, sixty five families were conveniently sampled. The total samples in field study were 1,040 families.

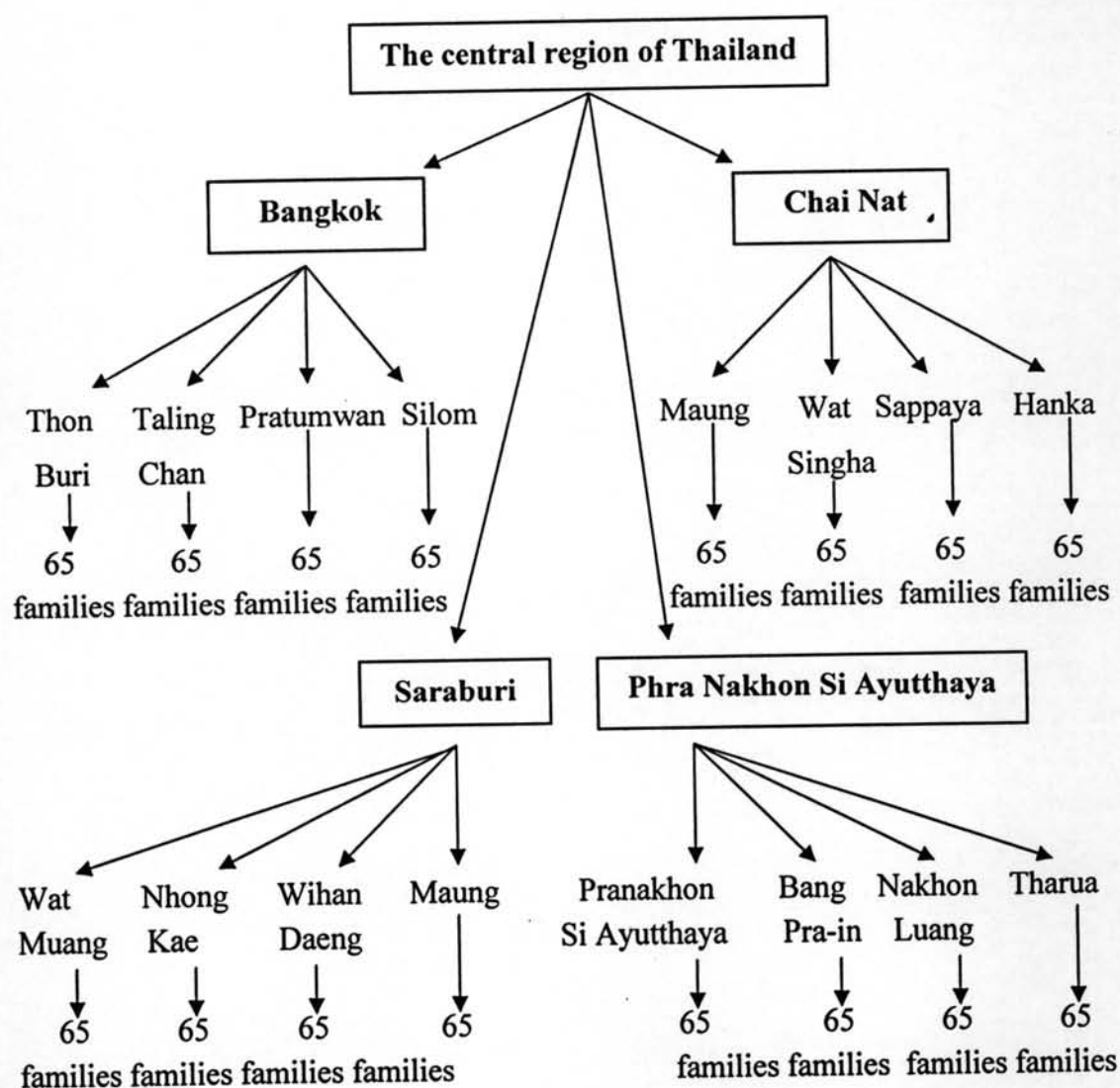


Figure3. A process of multi-stage sampling

3.2 Samples for testing criterion related validity and reliability

The number of sample size for testing criterion related validity and reliability was estimated based on the significant criteria at .05, power analysis = 0.8, and medium effect size for a significance of product-moment coefficient (r). The necessary sample size for those criteria would be 85 subjects (Cohen, 1992). With criteria for medium effect size, questionnaires which spared for incomplete answering were sent to 100 families. According to returning rate was 100%, all of 100 families were used to be the samples of both studies. In this case, the significance of product-moment coefficient was also approximately fallen in range of the medium effect size.

3.3 Samples of contrasted-groups approach

The samples used for operating the contrasted-groups approach were sixty families divided into two contrasting groups equally. Based on the Family Health Model (Denham, 2003a: 279), characteristics of healthy family were identified as nurturing acts, emotional support, caring attributes, and member interactions that produce and outcome that results in individuation, unity, and identity that satisfy members' needs. In addition, health of the family could be identified by the family itself (Denham, 2003a). Therefore, having ill members, low income, domestic abused, and levels of family health, which had potential to alter characteristics of healthy family, were used as criteria for identifying which families were healthy or unhealthy.

There was special procedure for selecting extreme contrasting groups of healthy and unhealthy families incorporate to the characteristics of healthy family. For healthy family groups, the characteristics of the family were identified as having no ill member in the families, and scoring very healthy on the question at the end of the TFHR scale stated that "what is the level of your family health". For unhealthy

family group, the characteristics of the family were identified as having at least one of these problems; low income, domestic abused, or having member(s) who experienced physical or mental illness, and scoring quite unhealthy and unhealthy on the same question about level of family health.

Scale development

The following sections addressed the development of the Thai Family Health Routines (TFHR) scale. In scale development procedures, there were five major tasks; identifying scale's format, generating item pool, investigating content validity, constructing the second draft of the TFHR scale, and testing psychometric properties of the TFHR scale. Each of these tasks was examined in details.

1. Identifying a format of the scale

The intended use of the TFHR scale was to measure the health of Thai families for clinical and research purposes. The scale was design as a self-report instrument with 6 subscales consisting of both positive and negative item wordings. Each item of these subscale was equally weighted by degree of frequency in a four-point rating scale ranging from 1 = never, 2 = rarely, 3 = sometimes, and 4 = always.

2. Generating item pool

Generating item pool was started with identifying operational definition of the Thai family health routines concept and its six health routines; self-care, safety and prevention, mental health behavior, family care, family caregiving, and illness care routine.

An item pool of the TFHR scale was generated from reviewing literature and family interview based on operational definitions of the six routines (Denham, 2002; 2003a). In case of reviewing literature, existing knowledge relevant to

family health routines was clustered into the six routines. For family interview, in-depth interview with 13 Thai families employed to enrich knowledge about their routine health behaviors. Information from the reviewing literature and the in-depth interview was integrated for constructing item statements of the item pool composed of 206 items (Appendix E).

Each item was constructed by writing a short declarative statement reflecting the routine health behaviors of Thai families. In order to cover all aspects of the operational definitions of the six routines, items were constructed as the large item pool that was expected to be a representative the universe items of the TFHR scale. For six constructs; there were 63 items for self-care routines, 45 items for safety and prevention routines, 40 items for mental health behavior routines, 16 items for family care routines, 26 items for family caregiving routines, and 16 items for illness care routines. Therefore, the item pool consisted of 206 items reflected all aspects of the six constructs of the family health routines concept.

3. Content validity

Content validity concerns whether the scale and the items it contains are representative of the content domain that the researcher intends to measure (LoBiondo-Wood and Haber, 2002). Validating content, two key issues; whether individual item are relevant and appropriate in term of the construct and whether the items adequately measure all dimensions of the construct (Polit and Hungler, 1999) were examined by a panel expert. DeVellis (2003) and Mishel (1998) suggested that asking for feedback in relation to accuracy, appropriateness, relevant to the test specification, and readability of each item, all of these were recommended. The results

from content validity were to identify the items that should be refined, changed, or deleted following comments or suggestions of a panel expert.

Regarding a number of experts, at least three experts were recommended, but a larger numbers also advised if the construct is complex (Polit and Hungler, 1999). Therefore, six family nursing experts (Appendix A) were invited to validating content of the 206-item pool of the TFHR scale because of complexity of the construct. Five of the six experts were nursing instructors with expertise in family nursing and family research. The other expert was a community nurse who has been closely worked with Thai families in community settings. One of the five instructors was the author who has been developed the Family Health Model (Denham, 2003a) used as the conceptual framework of this study. Accordingly, the author was an American citizen, so a back translation process was necessary. Two linguistic experts were conducted the translation of the draft scale from Thai into English and English into Thai back and forth, to make sure that each item from both English and Thai version have the same meaning.

Before conducting content validity, operational definition of the Thai family health routines identified by researcher were sent to the author to make sure that the definitions represented all aspects of the concept and the six family health routines. Then, the item pool was generated, and back translation was conducted. The item pool in English version was sent to the American expert, and the Thai version was sent to Thai experts for evaluating content validity.

The content validity in this study was a qualitative procedure guided by the test specification, called the Content Validity Form, which contained 2 important issues: clarity of expression and relevance in relation to the family health routines of Thai families. In case of clarity of expression, adequacy in each item in terms of the language

used, offensiveness or appearance of bias, and redundancy were examined. The six experts were asked toward the Content Validity Form with the respect of placing each item in one of four-point scale that would reflect to: 1) Relevance to the operational definition and content domain using the four-point rating scale: 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = very relevant, and 2) Clarity of items using the four-point rating scale: 1 = not clear, 2 = somewhat clear, 3 = quite clear, 4 = very clear.

In process of the content validity, the six experts proposed various comments and suggestions. First expert was an American expert who has been developed the Family Health Model proposed that a 206 item-pool covering the six constructs of family health routines concept, but rating score on each item in test specification was not performed due to cultural differences. For the other experts, they commented that the item pool was too long, some items were too much in details, some certain statements were ambiguous and not necessary, and many of them were required correction. Therefore, rated score on the Content Validity Form, comments, and suggestions on each item by five experts were used to consider whether the item would be corrected or deleted.

Item deletion was performed when three of the five experts recommended that it would be deleted, or rated item's score on relevancy to operational definition were less than 3. Comments and suggestions on each remained item by the five experts were clustered. Items, which their statements were commented on double-barrel or ambiguity by at least three of five experts, were considered revising. And whenever there were redundant items, only the best one was selected.

Even though, there was no CVI which indicated level of content validity, following comments and suggestions of the experts can maximize the content

validity of the scale. The 206 items of the pool were revised, reshaped, deleted or added together following comments and suggestions of the experts. According to redundant meaning, there were 54 items which were deleted and 10 items were grouped into 4 items. There was no item deleted due to irrelevant to the operational definitions.

After completing content validity process, the number of 206 items in the item pool was reduced to 145 items in this step and used to construct the first draft of the TFHR scale (Appendix G). The first draft scale composed of 145 items still covered the six constructs of family health routines and reflected all aspects of routine health behaviors of Thai family provided in the operational definitions.

4. Constructing the second draft of the TFHR scale

The pretest study was conducted to construct the second draft of the TFHR scale. Before researcher printed out items in a form for a main study in a large sample size, it was a good idea to try out the items on a small group of samples (Crocker and Algina, 1986). In this study, the pretest was conducted to; 1) determine the amount of item that took to complete the scale, 2) establish the scale if its instructions were unclear, and 3) identify clarity and appropriateness of scale use if participants found anything objectionable or inappropriate about the scale (Pett, Lackey, and Sullivan, 2003).

To meet purposes of the pretest study, the first draft of the TFHR scale which composed of 145 items was examined by using item analysis and item review. Item analysis was employed to obtain statistical data; while item review by mothers/wives of Thai families was used to obtain qualitative data concerning comments and suggestions about scale use.

4.1 Item analysis

Item analysis was employed to select the appropriate items which were a representative of a sample domain of the item universe for constructing the second draft scale. Therefore, descriptive statistics of each item, item-total correlation, item-item correlation, and Chronbach's alpha coefficient were examined.

Descriptive statistics of each item were examined by using mean, standard deviation, skewness, and kurtosis. The items which represented normal distribution were selected. Therefore, criteria for selecting the appropriate items were skewness values falling inside the range of -1 to +1 (Hair, Aderson, Tatham, and Black, 1998), magnitude of kurtosis was less than 2 (Wagner, Schnoll, and Gipson, 1998).

Corrected item-total correlation was proposed in term of precision of the item indicated how strongly an individual item reflected the total scale. Regarding a common rule of thumb, the corrected item-total correlation should be between 0.30 and 0.70. Those less than 0.30 were not contributing much to measurement of the concept while those greater than 0.70 were probably redundant (Polit and Hungler, 1995). Therefore, items with corrected item-total correlation less than 0.30 would be deleted, and the paired items with item-item correlation greater than 0.70 were considered keeping the best one of each paired item.

Chronbach's alpha coefficient, which represented an internal consistency of the scale, was used as the criterion for keeping appropriate items. When the alpha coefficient if any items were deleted was less than 0.7, those items would be retained. In addition, Chronbach's alpha coefficient of the first draft scale should be at least 0.70 for new developed instrument (Nunnally and Bernstein, 1994).

4.2 Item review

Item review was employed to determine appropriateness and clarity of each item wording of the first draft of the TFHR scale. The first draft scale which composed of 145 items was reviewed by 15 mothers/wives of Thai families. Results of the review were used for improving items that were difficult to understand or answer.

In pretest study, both statistical and qualitative data were used as criteria for selecting, revising and improving items appropriately to construct the second draft scale. After completing pretest study, the number of 145 items in the first draft scale was reduced to 85 items in this step and used for constructing the second draft of the TFHR scale. The second draft scale composed of 85 items still covered the six constructs of family health routines and reflected all aspects of routine health behaviors of Thai family provided in the operational definitions.

5. Psychometric testing

The psychometric testing phase was operated to test validity and reliability of the TFHR scale. This section consisted of four steps. First, confirmatory factor analysis was used to test construct validity of the TFHR scale on a large group of samples in the field study. Second, criterion related validity, using the Chulalongkorn Family Inventory (CFI) developed by Umaporn Trangkasombutb (1997) as a concurrent criterion, was conducted to test second type of construct validity of the scale. Third, contrasted-groups approach, comparing between two groups of healthy and unhealthy families, was operated to test the third type of construct validity of the scale. The last step of psychometric testing phase was an investigation of internal consistency reliability. The expected outcome of this phase was a valid and reliable research instrument for measuring family health through routine health behaviors of Thai family.

The procedures of developing The TFHR scale could be summarized as shown in Figure 4.

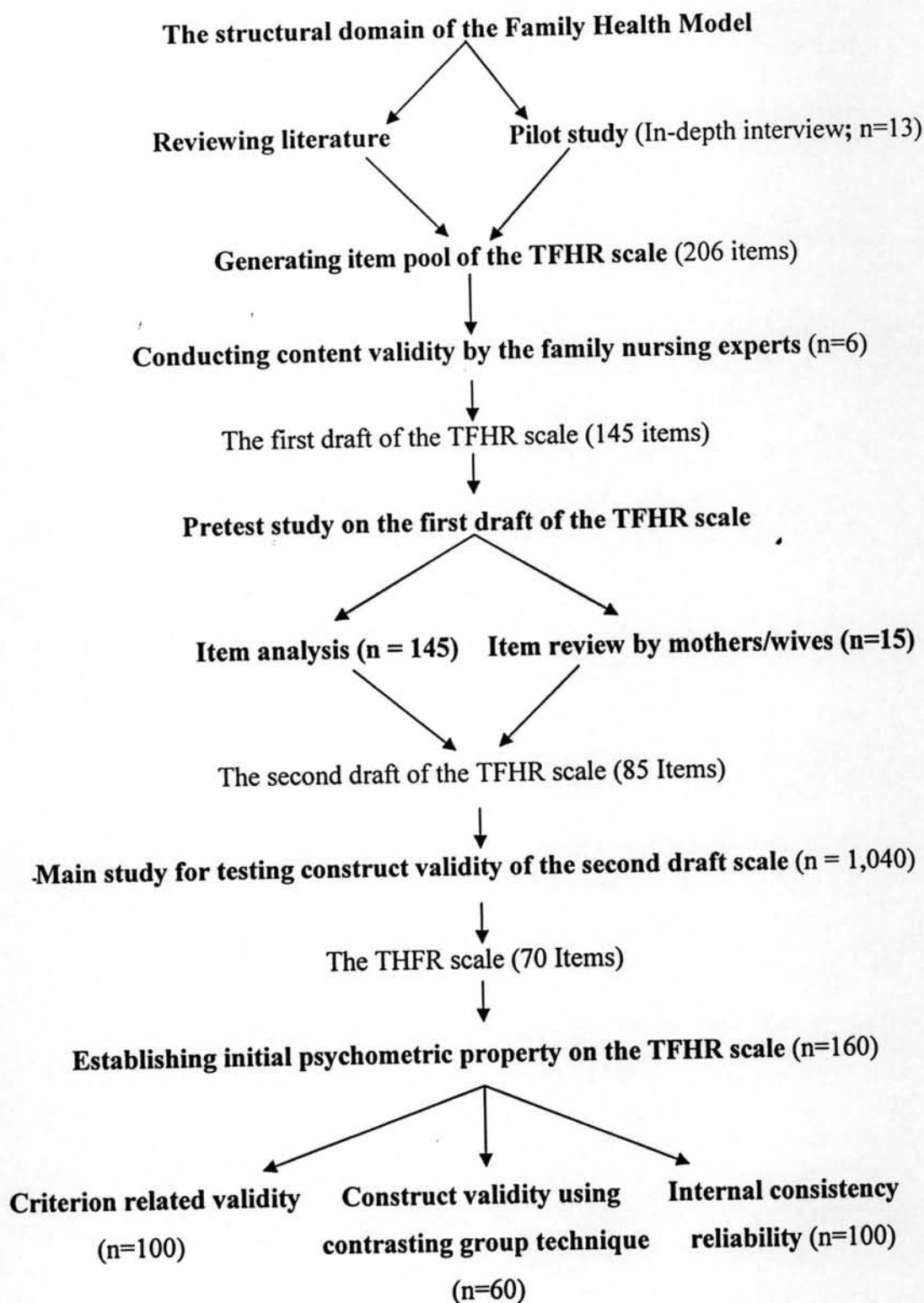


Figure 4 The flow chart of the TFHR scale development procedures

Research instrument

From the Thai Family Health Routine Scale (TFHR scale) that reported earlier in part of scale development, two others self-report instruments, demographic data sheet, and the Chulalongkorn Family Inventory (CFI; Tranksombut, 1997) were selected to use in this study.

1. Demographic data sheet (Appendix E) was used to collect basic information of samples, such as age of respondents, family type, marital status, major occupation of the family, family income, religion, status of house occupying, living area, and health problems among family members.

2. The Chulalongkorn Family Inventory: CFI (Appendix E) was used to examine the TFHR scale's criterion-related validity. This instrument is a Thai family measure which was developed by Umaporn Trankasombat (1997) based on the McMaster Model of Family Functioning. The CFI was applied from the Family Assessment Device (FAD) into Thai culture. The CFI is a self-report with a 4-point rating scale designed as a screening tool to distinguish healthy and unhealthy families. The CFI composes of 36 items in seven subscales: problem solving, communication, role, affective responsiveness, affective involvement, and overall general functioning. Each item consists of both positive and negative statements which express how well family functioning reflected by family members.

The CFI has been widely used to measure family functioning of Thai families that had member(s) experiencing psychological problems. Reliability of the CFI established in previous studies was ranged from 0.85 to 0.87 (Pummapichat Seangkhyo, 1999; Ulis Sombutkaew, 2002; Thada Charoenkusol, 2002). The evidences

demonstrated that CFI was a reliable research instrument and widely used to measure health of Thai families.

The reasons for using the CFI as an appropriate criterion for testing validity of the TFHR scale in this study were: 1) the CFI and the TFHR were used to measure the same concept, health of Thai family and 2) the Family Health Model defined that the family health routines in structural domain and family functioning in functional domain were correlated with each other and the both domains also described phenomena of family health. Before testing criterion related validity in this study, reliability of the CFI was tested on internal consistency. The Chronbach alpha of the CFI in this study was 0.86 (N=30).

Protection of human subjects

This procedure was performed before collecting data in both qualitative and quantitative approach in order to explain that there was not risk to be the participants or samples in this study. Approval of study plan for protection of human subjects was obtained from the Ethical Review Committee for Research Involving Human Subjects and/or Use of Animal in Research, Health Science Group of Faculties, Colleges and Institutes, Chulalongkorn University before collecting data started.

Before collecting data, the participants/samples were given an information sheet which described about purpose of the study, a chance to ask questions and expressed concerns before they signed their names on informed consent sheet. During the data collection, the process would be stopped whenever the participants/samples needed. After completing data collection, all data were kept anonymous through the use of name codes. Questionnaire, tape-transcribed notes, name codes, and addresses or telephone numbers were stored in a locked cabinet.

Data collection

In this study, data collection was divided into three sections. Section one was data collection for the pilot study in order to generate the item pool of TFHR scale. In the pilot study, qualitative approach with in-depth interview was performed between June and September 2005. Section two was data collection for pretest study performed between September and November 2006 in order to construct the TFHR scale. Lastly, data collection for the main study was undertaken between December 2006 and February 2007 in order to test psychometric properties of the TFHR scale.

1. Pilot study

In the pilot study, data was collected by researcher. The in-depth interview, with semi-structural interview guide (Appendix D), was employed to enrich knowledge about family health routines in order to generate the item pool. The six categories of family health routines (Denham, 2003a); self-care, safety and prevention, mental health behavior, family care, family caregiving, and illness care routines, were used as the conceptual framework to create the major questions of the interview guide.

The interview process was started when researcher visited each family at their house. Participants of this study were families that had both husband and wife, and single-parent families that had only mother. The participants were invited to go through interview process together at the same time in their houses. The researcher informed the participants about personal information of researcher, purposes and usefulness of the study, interview method, and protection of human right. In addition, the researcher let them know whenever they did not want to participate in this study; they could withdraw from the study at any time.

After mutual agreement between participants and researcher, the participants signed their names in consent sheets, and the researcher asked the participants to use a tape recorder, papers, and pencils for collecting interview data, then the interview process started.

During the interview process, researcher used the semi-structure interview guide, probing technique, and specific questions to gain rich information about routine health behavior from each family. The conversation was begun in close sympathy between researcher and participant. Then, a general statement, such as "Please tell me what do you think about meaning of family health?" and "To meet family health that you told me, what are activities in daily life of your family after getting up until going to bed?", would then be introduced. And the matter of listening, probing, and encouraging, all would be used in all participants.

The interview time was taken between 45 and 60 minutes. When there was no any new information to be identified, the interview ended. The participants were interviewed at least once. Second interview was carried out when some information from the first one was unclear. Each interview was recorded by using audiotape, pencil and paper. The data on recorded audiotapes were transcribed verbatim. Data about the participants' background were recorded before the in-depth interview took place. The process of data collection on the qualitative approach in this study started in June 2005 and continued until September 2005.

2. Pretest study

There were two purposes for pretest study on the second draft of the TFHR scale. The first purpose was to performed item tryout analyzed by using item analysis. The second one was to determine appropriateness and clarity of each

item wording by using item review. Data in pretest study were collected by researcher or research assistants. The respondents were mothers/wives who were representatives of Thai samples.

The pretest study was stated after receiving the permit letter from the Ethical Review Committee, Chulalongkorn University, Thailand, the collecting process were took place. According to research settings of pretest study, two provinces were selected, Bangkok and Saraburi. The researcher had to send the official letters, authorized by dean of the Faculty of Nursing, Chulalongkorn University, to the provincial governor of Bangkok and Saraburi for allowing the researcher to collect data at their supervision areas. Then, the researcher and research assistants started to collect data after having the permission from legal authorities.

In case of collecting data for item analysis, two research assistants were instructed by the researcher about purposes of the study, theoretical framework, application of the questionnaires, informed consent of the study, and methods of data collection. After obtaining permission to collect data from the provincial governor of each province, sample codes and house locations of the samples were recorded by researcher or research assistants. Then, the questionnaires consisting of informed consent sheet, demographic data sheet, and the second draft of the TFHR scale were sent by hands to respondents (mothers/wives) of each family by researcher or research assistants.

After mutual agreement between respondents and researcher, the respondents signed their names in consent sheets, and the researcher or research assistants explained the objectives and usefulness of the study before collecting data. During answering the questionnaires, the respondents could refuse to answer the questionnaires whenever they wanted. After each respondent completed the questionnaire, the researcher or research assistances checked its completeness and kept in a locked cabinet, and only researchers could have access to these data.

For second purpose, determining appropriateness and clarity of items' wording, item review was conducted with 15 mothers/wives. Data, reactions of each respondent while answering questionnaire, were collected by researcher. Many behaviors, such as long pauses, answer changing, confusing conditions about specific items, and length of time used were noted. After completing the questionnaire, the respondents were invited to ask and discuss about their problems while responding to the statements of the questionnaire that included clarity of language and format, and ease of understanding as noted earlier. Comments on items, and suggestions from the respondents were used to improve the second draft scale. Additionally, timing to complete the questionnaire was then evaluated.

3. Main study

Data collection, in the main study was generally similar to those in pretest study, but research settings and questionnaires were quite different.

Research settings of the main study were four provinces. Therefore, the official letters, authorized by dean of the Faculty of Nursing, Chulalongkorn University, were sent to the provincial governors of Bangkok, Chai-Nat, Phra Nakhon Si Ayutthaya, and Saraburi for allowing the researcher to collect data in their supervision areas.

Questionnaires, used for the main study in order to test construct validity using conducting confirmatory factor analysis, consisted of informed consent sheet, demographic data sheet, the 85-item TFHR scale, and the Chulalongkorn Family Inventory (Appendix I), were sent by hands to respondents (mothers/wives) of each family by researcher or research assistants. In case of testing criterion related validity, construct validity using contrasted-groups, and internal consistency reliability, the 70-item TFHR scale was used to test those psychometric properties.

Data analysis

The scale development in this study composed of both qualitative and quantitative approach; therefore, data from two approaches were separately analyzed as follows:

1. Qualitative data

Data from in-depth interview with 13 Thai families were analyzed using content analysis proposed by Graneheim and Lundman (2003). After extracting data from the recorded tapes by researcher, content analysis procedures were employed in order to generate the item pool of TFHR scale from the interview data. Content analysis procedures in this study compose of three steps, identifying meaning unit, condensed meaning unit, and coding and categorizing the condensed meaning units based on the six categories of family health routines (Denham, 2002; 2003a).

The content of interview data which described health behaviors of Thai families that they regularly used to regain, sustain and promote health individually and their family as a whole in any aspects of family health routines, all were included in specifications (N =13). The data of each family was extracted and brought together into one text, which constituted the unit of analysis. The text was highlighted to identify meaning units. Then, the meaning units were shortened to be condensed meaning units. The condensed meaning units were abstracted and labeled with a code, later.

The whole context was considered after condensing and labeling meaning units with codes. The various codes were grouped based on differences and similarities. They also included the operational definition by six categories of family health routines. The codes were sorted into the six categories; self-care, safety and

prevention, mental health behavior, family care, family caregiving and illness care routines. The item pool was generated from many condensed meaning units of the interview data.

2. Quantitative data

Data from quantitative approach was analyzed using the Statistical Package for the Social Science (SPSS for Windows version 14), and LISREL 8.52 was used for testing validity using confirmatory factor analysis. Before conducting data analysis, data from the pretest and field study was screen through descriptive analysis for detecting missing data, and scores on negative items had to recode.

2.1 Demographic data of samples

Characteristics of the sample were assess by descriptive statistics consisted of frequency and percentage, mean, standard deviation, and range.

2.2 Item description

Descriptive statistics including mean, standard deviation, skewness, and kurtosis provided information of outliers, and normal distribution. Two statistic indicators, represented normal distribution, were skewness and kurtosis. The skewness represented symmetry of the distribution, whereas the kurtosis demonstrated distribution peak. If skewness and kurtosis of any items are zero, their distributions are normal (Wegner, Schnoll, and Gipson, 1998). The items which their skewness values falling inside the range from -1 to +1 would represent fair normal distribution (Hair, Aderson, Tatham, and Black, 1998).

2.3 Item analysis

Item analysis, in this study, was to select the best item for constructing the TFHR scale before testing its construct validity. The analyses involved

descriptive statistic, Cronbach's alpha, corrected item-total correlation, and item-item correlation. The results of the various analyses were used as the criteria for eliminating poorly performing items.

Corrected item-total correlation was proposed in term of precision of the item by Brink and Wood (1998). The Pearson product-moment correlation between scores of each item and a total score were used to indicate how strongly an individual item reflected the total scale. Theoretically, corrected item-total correlation, the items that have the highest correlation coefficients, should be retained. Practically, a common rule of thumb was used as the criteria for retaining or eliminating items. The rule suggests that inter-item correlation of each item should be between .30 and .70 (Jacobson, 1997). The items which have correlation coefficients below .30 indicate that they do not contribute much to measurement of the concept, and those above .70 are probably redundant (Jacobson, 1997; Polit and Hungler, 1995). Therefore, the items that have a correlation coefficient below .30 or above .70 would be eliminated, and alpha of total scale were accepted at least .70 for an early developed instrument (Nunnally and Bernstein, 1994).

2.4 Construct validity

In confirmatory factor analysis, validity is defined as "the extent to which a measure performs in accordance with theoretical expectations" (Carmines and Zeller, 1979: 27). The construct validity of the measurement model of the TFHR scale was also examined through confirmatory factor analysis (CFA). For meaningful results to be obtained in the CFA, correlation between 25 indicators was assessed to investigate whether it has sufficient correlations to determine the appropriateness to perform factor analysis (Dixon, 2001; Hair, Anderson, Tatham, and Black, 1998; Pett, Lackey, and Sullivan, 2003). In addition, Bartlett test of sphericity,

Measure of Sampling Adequacy (MSA), and Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO), all had to be examined in order to determine the appropriateness of the matrix, correlation coefficients (Pett, Lackey, and Sullivan, 2003).

2.4.1 Confirmatory factor analysis (CFA)

CFA was used to test the proposed measurement model of the TFHR scale. Due to the complexity of the factor structure of the TFHR scale which composed of 2 unidimensional and 4 multidimensional factors, the proposed model of the TFHR scale had to test at two steps of construct validity. First step was testing on measurement model of each sub-dimension of the 4 factors using first order factor analysis. Then, item parceling was employed to combine individual items under each sub-dimension after the model of sub-dimension was confirmed by the first order factor analysis. Second step was testing construct validity on the overall factor structure model of the TFHR scale by using a second order factor analysis.

In case of model assessment procedures, the proposed model was assessed using four procedures: (1) model specification, (2) model identification and estimation, (3) assessing model fit, (4) model modification. Model modification provides the state of availability for estimating model. Assessing model fit determines whether the model is acceptable to the sample data. When results of model were not acceptable to the sample data, model modification was considered to improve the overall model fit.

Model specification: Specification of the confirmatory factor model require statement about (1) the number of common factors (2) the number of observed variables (3) the variances and covariances among the common factors (4) the relationships among observed variables and latent factors (5) the relationship among unique factors and observed variables and (6) the variances and covariances among the unique factors (Long, 1983).

Model identification and estimation: In order to examine the hypothesized measurement model, model identification was important to test whether there are sufficient numbers of observed variances and covariance to allow estimation of all the model parameters (Byrne, 2001). There are three forms of identification status. Firstly, just-identified, the model has equal number of parameters and observation. Secondly, over-identified, the model has fewer parameters than observations. Lastly, under identified, the model has more parameters than observation (Nonglak Wiratchai, 1999). In this study, the hypothesized model was over-identified model. In case of model estimation, there are two methods that have the advantage of estimating model fit (Long, 1983); Generalized Least Squares (GLS) and Maximum Likelihood (ML). In this study, ML was used for estimating model fit.

Assessing model fit: The issue of model fit refers to whether the model is consistent with the sample data. There are several indices for assessing overall model fit to the sample data. There are three types of good-of-fit measures (Byrne, 2001; Hair and others, 1998; Kline, 1998). Firstly, absolute fit measures include Chi-square (χ^2) statistics, Goodness of fit index (GFI), Root mean square residual (RMSR), Root mean square error of approximation (RMSEA), and Expected cross-validation index (ECVI). Secondly, increment fit measures include Adjust goodness-of-fit index (AGFI), Non-normed fit index (NNFI), Normed fit index (NFI), and Comparative fit index (CFI). Lastly, parsimonious fit measures include Parsimonious normed fit index (PNFI), Parsimonious good-of-fit index (PGFI), and Akaike information criterion (AIC).

The specific model fit indices used for measuring the overall model fit in this study were Chi-square (χ^2) statistics, Goodness of fit index

(GFI), and Root mean square error of approximation (RMSEA), Comparative fit index (CFI), and Parsimony normed fit index (PNFI).

1) The Chi-square (χ^2) statistics was used to test hypothesized that whether the unrestricted population variance/covariance matrix of the observed variables is equal to the model-implied variance/covariance matrix (Mueller, 1996). Therefore, a smaller value of the chi-square (χ^2) statistics indicates that the model reflects better model fit to the sample data which come from the large enough sample size. The chi-square statistic (χ^2); however, is unrealistic for identifying a well-fitting hypothesized model in empirical research because of the sensitivity of the large sample size and excessive kurtosis (Byrne; 2001; Diamantopoulos and Siguaw, 2000; Kline, 1998). To reduce the sensitivity of the chi-square (χ^2) statistic to sample size, it is suggested that the value of chi-square (χ^2) is divided by degree of freedom (χ^2/df). A χ^2/df ratio less than 3 is an acceptable recommended threshold (Kline, 1998).

2) The goodness-of-fit index (GFI) is “analogous to a squared multiple correlation in that it indicates the proportion of the observed covariances explained by the model-implied covariances” (Kline, 1998: 128). The value of GFI which is roughly 0.90 or greater is an acceptable recommended threshold (Byrne, 2001; Hair, et al., 1998; Munro, 2001).

3) The root mean squared error of approximation (RMSEA) focuses on the discrepancy between covariance matrix of reality and covariance matrix of samples per degree of freedom (Diamantopoulos and Siguaw, 2000). The small value of this index reflects a better fit of the model to the data. The value of equal to or less than 0.08 represents an acceptable model fit to the data (Byrne, 2001).

4) The comparative fit index (CFI) is adapted for measuring an incremental fit; how much better the model fits relative to an independent model (Tate, 1998). A recommended threshold of CFI is 0.90 or greater, which represents an acceptable model fit (Kline, 1998).

5) The parsimony normed fit index (PNFI) was used as parsimonious fit measure for this study. The parsimonious fit measures are related to the issue of model parsimony as achieving higher degree of fit per degree of freedom used. The PNFI “takes into account the number of degrees of freedom used to achieve a level of fit” (Hair and others, 1998: 658). Higher value of PNFI represent that the model is more parsimony. An acceptable threshold of a value of PNFI as indication of close model fit is larger than 0.60 (Hu and Bentler, 1999).

The recommended threshold for goodness-of-fit indices used in this study was demonstrated in Table 3. The model fit indices were evaluated to determine whether the initial hypothesized model was acceptable. If the initial hypothesized model was acceptable as good fit, its result could be reported. If not, the hypothesized model would be modified and reevaluated.

Table 3 Goodness-of-fit indices and recommended threshold values

Goodness-of-fit indices	Recommended thresholds
χ^2/df ratio	Less than 3.00
RMSEA	0.08 or lesser
GFI	Roughly 0.09 or larger
CFI	Roughly 0.09 or larger
PNFI	Larger than 0.60

Model modification: The purpose of model modification is to improve model fit by revisions of model misspecification. In this study, the modification index for each paired indicator is a useful aid for identifying possible relationship between error terms of the pair indicator to be added to a model with poor fit (Byrne, 2001; Joreskog and Sorbom, 1988). No modification index or a small value indicates the hypothesized model fit is acceptable; a large modification index value suggests the overall model fit could be improved.

2.4.1.1 First order factor analysis

First order factor analysis was conducted to confirm raw items which belonged to each sub-dimension of the four multidimensional factors were their actual indicators before parceling those items.

In the measurement model of the TFHR scale, there were 6 common factors. Four of the 6 factors; self-care, safety and prevention, mental health behavior, and family caregiving routine were multidimensional factors. Combining several individual raw scores into a single sub-dimension (an item parcel) of each multidimensional factor was necessary for further testing on construct validity using the second order factor analysis. The composite scores would be used as representative scores of the groups of original items.

In this study, item parceling (combining individual items into small groups of items within scales or subscales) was used (Bandalos, 2002) to overcome the well-known limitations of applying second order factor analysis technique to individual item scores (Allinson and Hayes, 2000); the complexity of factor structure of the measurement model, and the large number of raw items. Yuan et al. (1997 cited in Nasser and Takahashi, 2003) suggested performing item

parcels based on empirical knowledge that has already indicated which variables were good indicators of a latent factor.

For example, self-care routine factor was multi-dimensions consisting of five sub-dimensions (i.e., dietary practice, sleep and rest pattern, hygiene care, exercise and physical activity, and sexuality). The raw items of each of five sub-dimensions had to parcel to be five indicators of the self-care routine factor. Before parceling raw items, first order CFA was conduct to confirm the model of each sub-dimension. If the model were overall fit, parceling items were permitted. In this process, indicators which had low factor loading ($b < 0.2$) or non-statistically significant loading ($p < .05$) would be deleted before parceling the items.

Since the high differences in factor loadings of each sub-dimension model, computing factor scores using multiple regression method through data reduction of SPSS version 14.0 was appropriately used for generating the scores of composed item (Kim and Mueller, 1978; Nonglak Wiratchai, 1999). This score were standardized of which their mean were "0" and standard deviations were "1"

2.4.1.2 Second order factor analysis

The second order factor analysis was conducted to test construct validity of the overall model of factor structure of the TFHR scale. In this case the hypothesized factor structure was proposed and specified based on the conceptual framework of the structural domain of the Family Health Model (Denham, 2003a). If the result of testing the hypothesized model was misfit with model data, model modification was performed. Then, the modified model was tested on overall model fit and measurement model fit.

2.4.2 Contrasted-groups approach

There were two contrasting groups, healthy and unhealthy family. Each group composed of 30 families. All families in both groups were treated as comparing samples. The TFHR scales were administered to each group. Scores on the TFHR scale of these two groups were tested on differences using independent-sample t-test. The result of testing should be significantly difference which indicated that the scale could distinguish two groups of healthy and unhealthy families.

2.5 Criterion-related validity

Criterion related validity of the TFHR scale was tested against the CFI which has been previously validated as a standard goal. The scores of the TFHR scale were expected to significantly correlate with the CFI's scores. In this case, Pearson's correlation coefficients between the total score of the TFHR scale and the CFI would be calculated.

2.6 Reliability

Reliability refers to "the extent to which a variable or set of variables is consistent in what it is intended to measure" (Hair et al., 1998: 90). Internal consistency reliability was used to examine the extent to which all of the instrument's items or subscale measuring the same attribute. Internal consistency using Cronbach's alpha method was evaluated for the second draft of the TFHR in the pretest study, and the TFHR scale in a process of psychometric property testing. A value above .70 for the reliability coefficients was considered satisfactory (Nunnally & Bernstein, 1994).

3. Management of missing data

The data were examined prior to data analysis to resolve any problems associated with missing data, to identify outliers caused by data entry mistakes. After checking data for each record, questionnaire which had missing values more than 10% of all items, were excluded from data analysis. The questionnaires that had randomly missing data on items less than 10% of overall records were retained, and those missing data could be replaced by the item mean, median, or mode depending on the score distribution (Polit and Hungler, 1999). In this study the missing data were replaced by mean value of the item score. There were 28 questionnaires which had to discard because incomplete answered items were more than 10% of total items of the questionnaires.

In pretest study, it was found that 18 of 145 records (12.41%) were not completed by which one item score of each record was missed. For testing confirmatory factor analysis, 101 of 1040 records (9.71%) were not completed. Each record composed of 85 items. Missing data of each 101 records ranged between one item score and three items scores (0.11- 0.35%). In case of testing criterion related validity, contrasting group technique, and reliability, there was no missing data.

In summary, this chapter provided details of the research methodologies for constructing the FHR scale and testing its construct validity and internal consistency reliability. In order to test the proposed conceptual model of this study, instrument was developed and estimated with the collected data. The results of qualitative approach and statistical analysis from the pretest study, and the main study would be reported in chapter 4.