Validity and Reliability Study of the Thai Version of WHO Schedules for Clinical Assessment in Neuropsychiatry: Eating Disorders Section

Niramol Patjanasootorn MD*, Suchat Paholpak MD*, Thawatchai Krisanaprakornkit MD*

*Department of Psychiatry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

Objective: To determine the validity and reliability of the Thai version of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN) Version 2.1, Eating Disorders Section

Material and Method: The SCAN interview version 2.1 Eating Disorders Section (section 9) was translated into Thai. The content validity of the translation was verified by comparing the back-translation to the English original. Whenever inconsistencies were encountered, the Thai version was adapted so that it correctly conveyed the meaning of the original English version. The revised Thai version was then field-tested in four regions (Suanprung Psychiatric Hospital, Jitavej Khonkaen Hospital, Srithanya Hospital and Suansaranrom Psychiatric Hospital, each place comprised 20 volunteers) for comprehensibility of the relatively technical language. Re-edition of the Thai version was made in accordance with suggestions from the field trial. The reliability study was done between October 2005 and September 2008 in which 30 persons (4 males; 26 females) were recruited. Fifteen persons (1 males and 14 females) were eating disorders patients (4 anorexia nervosa and 11 bulimia nervosa patients) and the other 15 (3 males and 12 females) were normal. The number of years of formal education varied widely and occupations were diverse. Subjects were interviewed using SCAN eating disorders section by a psychiatrist competent in using the Thai version of SCAN. The interviews were recorded on video so that the material could be re-rated.

Results: Regarding validity of WHO SCAN eating disorders section, based on the response from Thai subjects and consultations with competent psychiatrists, the content validity was established. Regarding the reliability study, the time taken to interview an eating disorder patient averaged 23.80 ± 5.15 minutes (range, 18.35-34.26) versus 17.27 ± 4.05 minutes (range, 9.07-22.37) for a normal subject. The mean of inter-rater reliability (kappa) of the eating disorder section was 0.73, standard error (SE) = 0.02, 95% confident interval (CI) = 0.68-0.77. Seventy-eight point five percent of the items in the section had a substantial kappa. The mean of intra-rater reliabilities was 0.76, SE = 0.03, 95% CI = 0.70-0.82. Sixty-eight point four percent and 31.58% of the items had a substantial and almost perfect kappa respectively.

Conclusion: The Thai version of the Eating Disorders Sections of SCAN version 2.1 proved to be a valid and reliable tool for assessing eating disorders among Thai speakers.

Keywords: Eating disorder, Anorexia nervosa, Bulimia nervosa, SCAN, Schedules for clinical assessment in neuropsychiatry, Reliability, Semi-structured interview, Validity
inter-rater reliability to help psychiatrists interview, assess, measure, and classify psychopathology and behavior associated, according to the ICD-10 diagnostic system(11), with the major psychiatric disorders among adults. SCAN is the gold standard for verifying interview-diagnoses done through clinical trials and other forms of psychiatric research.

SCAN has a I-shell program, CATEGO, which is a set of programs for processing the SCAN data and generating output (diagnosis). SCAN is intended for use only by clinicians with an adequate knowledge of psychopathology who have taken the WHO-designated SCAN training. SCAN has broad international acceptability and has been translated into 26 major languages, including in the Peoples’ Republic of China, Japan, Turkey, and India(12-15).

Thailand has neither its own national nor a translated international standard, psychiatric, diagnostic instrument. There is currently a need for a reliable and valid measure of ED that is brief and economical to administer and that can be used to assess response to treatment. The authors determined to translate SCAN into Thai and planned its establishment as the gold standard for psychiatric diagnosing in Thailand.

Eating disorders of various kinds have been reported in up to 4% of adolescents and young adults. Women are affected more than men. Most common age of onset is between 14 and 18 years. Anorexia nervosa is found in 0.5 to 1% of adolescent girls, bulimia nervosa 1 to 3%(16). All the subjects suffering from an ED fulfilled diagnostic criteria for at least another DSM-IV axis I psychiatric disorder(17,18). Patients tend to have disturbance in impulse control, emotional tone, body image, social relationships, sexual attitudes(19-21).

Long term (7 to 19 years) follow-up revealed only 50% had fully recovered from ED, 26.9% were currently affected by an ED, 23.1% showed a persistent body image disturbance and/or the presence of compensatory behaviors(22).

As eating disorders are highly important, this particular sub-study focuses on the validity and reliability of SCAN ED section. With this instrument, the authors wished to further the knowledge of the disorders and treatment in the Thai context. Validity and reliability studies of some other sections of SCAN Thai version were reported separately(23-27).

Material and Method

The authors used a cross-sectional validity and reliability design. The Khon Kaen University Ethics Committee reviewed and approved the present study protocol. Regarding the validity study, the SCAN interview book version 2.1 was translated, with permission from the WHO, from English into Thai by SP. The content validity of the translation was verified by comparing the back-translation and adapting the Thai version in order to keep the meaning as the English original. The comprehensibility of the language was then tested by in-depth interview among a cross-section of Thais from four regions of the country (Chiang Mai, Suanprung Psychiatric Hospital, Khon Kaen, Jitavej Khon Kaen Hospital, Bangkok, Srithanya Hospital and Suratthani, Suansaranrom Psychiatric Hospital). Each region comprised 20 native volunteers including psychiatric patients and their normal relatives. Reflections, comments, and suggestions were assessed then summarized during a consensus meeting and SCAN Thai version was re-edited accordingly. The final Thai version was incorporated into the SCAN I-shell program and used for general testing.

Thirty (4 males and 26 females) volunteers were recruited between October 2005 and September 2008 for the reliability study. Inclusion criteria included Thai citizens, able to understand and speak Thai, at least 14 years of age, be a volunteer, and gave informed consent to participate in the present study. Each volunteer was given 200 Baht to cover overland travel expenses. Fifteen volunteers were ED patients (1 male and 14 females) and 15 (3 males and 12 females) were normal. The volunteer patients were from Srinagarind In-/Out-patient Departments and identified by ICD-10 or DSM-IV-TR criteria. The normal volunteers were faculty personnel with no previous psychiatric illness. The number of years of formal education of both patients and normal volunteers varied widely and occupations were diverse.

A psychiatrist trained in SCAN did the specific SCAN ED section interviews with all of the patients and normal persons. The interviews were recorded on video with a focus on the interviewee, not the interviewer. The interviews were recorded on video so that the material could be re-rated. To test the intra-rater reliability, a psychiatrist (trained in SCAN) used the Thai version of the ED Sections of SCAN to re-rate the videotaped interviews two more times, two weeks apart. Accordingly, the inter-rater reliability study was accomplished by two psychiatrists re-rating the video material simultaneously or at different times and comparing the results.

The authors probed for the presence and severity of ED symptoms in the Present State (PS). The
authors asked each volunteer whether he (she) used to have or was having the symptom in each item during the month before the date of examination. The PS may be part of a much longer present episode (PE), with onset years earlier. When rating the ED Sections the authors used Rating Scale I (a special rating scale for ED in SCAN) or an item-specific rating scale as indicated in each item. Twenty-eight questions probe the symptoms in the ED section. In order to have reliability study of as many questions as possible, even when the general probing question at the very beginning of each section got a negative answer, the authors continued to ask every detailed questions. However, whenever answers were unambiguous (i.e. a numerical length of time or a simple ‘yes’ or ‘no’), the rater’s judgment was not required. Consequently, the authors did not rate items that probed the duration of symptoms and age at onset. Finally, 19 items from ED section were rated.

Statistical evaluation

The inter- and intra-rater reliability was based on the agreement between raters by using descriptive statistics. Rating scales for ED section were treated as categorical data (i.e. 0, 1, 2, 3, 5, 8 and 9 indicating absence, presence in minor degree and not appropriate for classification, moderate severity, severe, presence of psychotic symptoms making the rating very difficult, not sure whether present or absent even with adequate examination, inappropriate to rate due to incomplete examination respectively. Calculation used the kappa (κ) statistic in STATA 7.0. The defined level for the degree of agreement was ‘poor’ (κ < 0), ‘slight’ (κ 0-0.20), ‘fair’ (κ = 0.21-0.40), ‘moderate’ (κ = 0.41-0.60), ‘substantial’ (κ = 0.60-0.80) and ‘almost perfect’ agreement (κ = 0.81-1.0)28-30.

Results

Regarding validity study, examining the back translation showed the ability to conserve the original meaning. During field trial, comments and suggestions for comparable meanings but re-phrasing by using Thai grammar and idioms were suggested from the participants. Therefore, some adaptation of the phraseology, wording, and sequencing of the sentences to make SCAN ED section understandable ‘Thai’ were done. Examples of these tasks can be seen in the appendix.

Regarding the reliability study 30 volunteers (4 males and 26 females), including 15 ED patients and 15 normal volunteers, were recruited and none of them withdrew. Eating disorders patients were 14 females and one male. They averaged 19.53 ± 2.03 years of age (range, 16-23) and averaged 12.9 ± 1.92 years (range 10-16) of formal education. Occupations included students (13), employees (1) and unemployed (1). The normal volunteers were 12 females and three males. They averaged 20.07 ± 2.31 years of age (range, 17-25) and averaged 12.8 ± 2.37 years (range 9-16) of formal education. Occupations included students (12), employees (2) and unemployed (1). The interview for an eating disorder patient required an average of 23.80 ± 5.15 minutes (range, 18.35-34.26) vs. 17.27 ± 4.05 minutes (range, 9.07-22.37) for a normal volunteer.

The inter-rater and intra-rater kappas, standard error (SE) and 95% confident interval (CI) for each symptom (item) of the ED Section are presented in Table 1. Classification of the κ value for each symptom as ‘almost perfect’, ‘substantial’, ‘moderate’, ‘fair’, ‘slight’ and ‘poor’ are presented in Table 2. Classification of the κ values as range, mean, median, mode, standard deviation, SE, 95% CI, minima and maxima for the whole ED section are presented in Table 3. Regarding inter-rater reliability, kappa values

<table>
<thead>
<tr>
<th>Item</th>
<th>Reliability (kappa value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inter-rater</td>
</tr>
<tr>
<td></td>
<td>κ</td>
</tr>
<tr>
<td>9.001</td>
<td>0.67</td>
</tr>
<tr>
<td>9.002</td>
<td>0.72</td>
</tr>
<tr>
<td>9.003</td>
<td>0.68</td>
</tr>
<tr>
<td>9.004</td>
<td>0.74</td>
</tr>
<tr>
<td>9.005</td>
<td>0.63</td>
</tr>
<tr>
<td>9.006</td>
<td>0.71</td>
</tr>
<tr>
<td>9.007</td>
<td>0.66</td>
</tr>
<tr>
<td>9.008</td>
<td>0.78</td>
</tr>
<tr>
<td>9.009</td>
<td>0.67</td>
</tr>
<tr>
<td>9.010</td>
<td>0.79</td>
</tr>
<tr>
<td>9.011</td>
<td>1.00</td>
</tr>
<tr>
<td>9.012</td>
<td>0.72</td>
</tr>
<tr>
<td>9.013</td>
<td>0.69</td>
</tr>
<tr>
<td>9.014</td>
<td>0.67</td>
</tr>
<tr>
<td>9.015</td>
<td>0.71</td>
</tr>
<tr>
<td>9.016</td>
<td>0.60</td>
</tr>
<tr>
<td>9.023</td>
<td>0.87</td>
</tr>
<tr>
<td>9.026</td>
<td>0.69</td>
</tr>
<tr>
<td>9.027</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Interviewing ED patients took an average of 24 minutes i.e. 6 minutes longer than normal control. SCAN ED Section is not lengthy and not time consuming. The high mean inter- and intra-rater reliability in this section was perhaps due to the 1) high validity, 2) high comprehensibility, 3) strict adherence to the rating criteria, or 4) good training in the use of the SCAN Glossary. SCAN is a clinician used interview schedule. Therefore, SCAN ED Section could help yielding a diagnosis of eating disorders. It is not like other questionnaires. Even Eating Attitudes Test (EAT) has been established as highly efficient as the sole means to identify eating disorders however, study has shown that the EAT can be an efficient screening instrument as part of a two-stage screening process in which those who score at or above a cut-off score of 20 are referred for a diagnostic interview. Bulimia Test-Revised had a particular attention to bulimic condition. Eating Disorder Diagnostic Scale even has a high clinical applicability yet it is a self-report. Bulimic Investigatory Test Edinburgh or BITE pays attention to binge condition. Eating Disorder Examination (EDE) is only a measure of the specific psychopathology of anorexia nervosa and bulimia nervosa. CIDI is a fully-structured interview but it was designed to be used by trained lay interviewers. The authors chose to validate and did a reliability test of the SCAN ED section as it is a scale back-translation, field trial process, and SCAN’s cross culture utility.

Respondents were able to understand and respond to the SCAN-ED section interview, thereby confirming both qualitative validation of the translation/
for clinician diagnosis and as a sub-protocol of the whole SCAN validity and reliability testing. With the proved validity and substantial level kappa, both intra-rater and inter-rater, will mean that any well-trained rater would obtain similar results and/or measurements or the resulting ratings would be representative of the subject’s score. Therefore, the user can apply SCAN with substantial confidence for both inter- and intra-rater rating. The authors expects, with the proved validity and reliability, different clinicians with adequate training in using Thai version of ED section of SCAN, to minimize the variability that occur in the diagnostic process and maximize the replicability of diagnoses and the discrimination of patients.

In conclusion, the Eating Disorders Section of the Thai version of SCAN was tested for its validity and reliability. Interviewing an ED patient using SCAN ED section did not take a long time. The inter- and intra-rater agreement assessments were consistently strong. Still, there was some limitation in the present study. During the reliability study, the authors recruited only the patients with anorexia nervosa or bulimia nervosa. Further reliability study on other types of eating disorders (atypical anorexia nervosa and atypical bulimia nervosa) might be needed.

Acknowledgments

The authors wish to thank the Department of Psychiatry, Faculty of Medicine, Khon Kaen University, and the WHO for their support and Emeritus Professor Dr. Somboon Srung Boonmee who made the final editing.

Potential conflicts of interest

None.

References

15. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1990; 47: 589-93.
<table>
<thead>
<tr>
<th>คำถามข้อที่</th>
<th>รายละเอียดของคำถาม</th>
<th>คำแปลที่เข้าใจยาก</th>
<th>คำแปลที่แก้ไขใหม่แล้ว</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.001</td>
<td>Have you worried, during the past year, about eating too much, or putting on weight, or getting too fat?</td>
<td>คุณเคยกังวลว่าจะกินมากเกินไปหรือกังวลว่าอ้วนเกินไปหรือกังวลว่ากินมากเกินไปในระหว่างปีที่แล้วบ้างไหม</td>
<td>คุณเคยกังวลว่าจะกินมากเกินไปหรือกังวลว่าอ้วนเกินไปหรือกังวลว่ากินมากเกินไปในระหว่างปีที่แล้วบ้างไหม</td>
</tr>
<tr>
<td>9.005</td>
<td>Symptoms present but little interference</td>
<td>มีอาการหลายอาการแต่มีผลกระทบเล็กน้อย</td>
<td>มีอาการหลายอาการแต่มีผลกระทบเล็กน้อย</td>
</tr>
<tr>
<td>9.012</td>
<td>Have you missed any period while you were keeping your weight down?</td>
<td>คุณเคยระดูไม่มาบ้างไหมในขณะที่คุณทำน้ำหนักตัวให้ลดไว้</td>
<td>คุณเคยระดูไม่มาบ้างไหมในขณะที่คุณทำน้ำหนักตัวให้ลดไว้</td>
</tr>
<tr>
<td>9.013</td>
<td>Have you had periods when you would eat abnormally large amounts of food within an hour or so, that is, binge eating?</td>
<td>คุณเคยมีหลายช่วงที่คุณกินอาหารจำนวนมากอย่างผิดปกติภายในเวลา1ชั่วโมง</td>
<td>คุณเคยมีหลายช่วงที่คุณกินอาหารจำนวนมากอย่างผิดปกติภายในเวลา1ชั่วโมง</td>
</tr>
</tbody>
</table>

**Appendix.** Examples of the comprehensibility difficulties and correction during field testing of the questioning items of Eating Disorders Symptoms Section of SCAN
การศึกษาความแม่นตรงและความเชื่อถือของ WHO Schedules for Clinical Assessment in Neuropsychiatry ฉบับภาษาไทย หมวดความผิดปกติของการกิน

นิรมา พจนะสุนทร, สุชาติ พหลภาคย์, ธวัชชัย กฤษณะประกรกิจ

วัตถุประสงค์: เพื่อศึกษาความแม่นตรงและความเชื่อถือของ WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN) Version 2.1 ภาคภาษาไทยหมวดความผิดปกติของการกิน

วิธีคัดและวิธีการ: คณะผู้วิจัยได้แปลบทสัมภาษณ์เกี่ยวกับความผิดปกติของการกินซึ่งเป็นบทที่ 9 ของ SCAN version 2.1 เป็นภาษาไทย คณะผู้วิจัยได้แปลกลับเป็นภาษาอังกฤษ และตรวจสอบความหมายกับภาคภาษาอังกฤษที่แปลดั้นมาแล้ว มีความหมายตรงกับความหมายเดิมหรือไม่ เอาไว้ดังกล่าวเป็นภาษาไทยแล้วคัดเป็นภาษาอังกฤษ มีความหมายตรงกับภาคภาษาอังกฤษทั้งหมด คณะผู้วิจัยได้นำ SCAN ภาคภาษาไทยที่แปลมาเป็นภาษาไทย อาศัยผู้มีภาษาส่วนที่ 4 ภาคได้แก่ โรงพยาบาลสวนปรุง โรงพยาบาลจิตวิทยาลัยขอนแก่น โรงพยาบาลชัยฤทธุนยา และโรงพยาบาลสวนราชบุรี สถานที่ละ 20 คน เพื่อตรวจสอบความถูกต้องของคำถามและการตรวจสอบคำแปลเป็นภาษาอังกฤษ มีความหมายตรงกับภาคภาษาอังกฤษที่แปลซึ่งผลการแปลได้แก่ SCAN ภาคภาษาไทยคนไทยสามารถเข้าใจคำถามได้ดี การศึกษาเกี่ยวกับความเชื่อถือของ SCAN ภาคภาษาไทยหมวดความผิดปกติของการกินได้กระทำตั้งแต่เดือนตุลาคม พ.ศ. 2548 ถึงเดือนกันยายน พ.ศ. 2551 ภาคภาษาไทยที่ตอบแบบสัมภาษณ์มี 30 คน เป็นชาย 4 คน หญิง 26 คน เป็นผู้ป่วยความผิดปกติของการกิน 15 คน (เป็นผู้ป่วย anorexia nervosa 4 คน bulimia nervosa 11 คน) (ชาย 1 คน หญิง 14 คน) และคนปกติ 15 คน (ชาย 3 คน หญิง 12 คน) กลุ่มตัวอย่างมีระดับการศึกษาและอาชีพเพิ่มเติมแต่ละกลุ่ม จิตแพทย์ที่ชำนาญในการใช้ SCAN ภาคภาษาไทยใช้ SCAN ภาคภาษาไทยหมวดความผิดปกติของการกินในมูลค่าของคำถามที่เข้าใจเพื่อการให้คะแนน ผู้ป่วยและคนปกติที่ไม่มีความรวมถึงการให้คะแนน

ผลการศึกษา: การศึกษาเกี่ยวกับความแม่นตรงและความเชื่อถือของ SCAN พบว่า SCAN ภาคภาษาไทยของอาการกินมีระดับที่แม่นตรงกับการศึกษาเกี่ยวกับความเชื่อถือ ได้พบว่าระยะเวลาที่ใช้ในการสัมภาษณ์บ่อยครั้ง คือ 23.80 ± 5.15 นาที (พิสัย 18.35-34.26 นาที) คนปกติ 17.27 ± 4.05 นาที (พิสัย 9.07-22.37 นาที) คาดว่ามีค่า standard error (SE) ค่าความเชื่องบทความละ 95% ของ inter-rater reliability (kappa) ของ SCAN หมวดความผิดปกติของการกิน คือ 0.73 0.02 และ 0.68-0.77 ค่าเดียวกัน ร้อยละ 78.5 ของคำถามได้ค่า kappa ในระดับ substantial ค่าเดียวกัน ร้อยละ 68.4 และร้อยละ 31.8 ของคำถามได้ค่า kappa ในระดับ substantial และ perfect ตามลำดับ

สรุป: SCAN ภาคภาษาไทยหมวดความผิดปกติของการกินเป็นเครื่องมือที่มีความแม่นตรงและความเชื่อถือได้ในการประเมินคนไทยที่มีความผิดปกติของการกิน