The Validity and Reliability of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN Thai Version): Mood Disorders Section

Thawatchai Krisanaprakornkit MD*, Suchat Paholpak MD*,
Nawanant Piyavhatkul MD*, Jiraporn Khiewyoo PhD (Population and Development)**

** Department of Psychiatry, Faculty of Medicine, Khon Kaen University, Khon Kaen*
*** Department of Biostatistics and Demography, Faculty of Public Health, Khon Kaen University, Khon Kaen*

Background: *Mood disorders are one of the most prevalent psychiatric disorders that have early onset, are chronic and can cause functional impairment. It is therefore crucial to establish an accurate diagnosis for treatment and research purposes.*
Objectives*: The authors aimed to test the validity and reliability of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN Thai Version): Mood Disorders Section.*
Material and Method: *First, psychiatrists competent in the use of the schedules and their underlying objectives tests the linguistic clarity of the Thai psychiatric schedules in four regions of the country. Then between October 2003 and August 2004, Reliability of SCAN: mood disorder section was tested among 30 participants, including mood disordered patients and normal volunteers.*
Results: *Based on reactions from Thais and consultations from competent psychiatrists, content validity was established. The duration of interviews for the mood disorder sections averaged 63.6 min (SD = 17.1). The inter- and intra-rater reliability kappa were 0.80 (0.77-0.83) and 0.86 (0.84-0.88), respectively. The reliability of the majority of items reached substantial to almost perfect agreement; however, 1 item (1%) had only slight agreement and 15 items (12%) had moderate agreement, and needed clarification of the scoring method.*
Conclusion: *The Mood Disorders Sections of the WHO Neuropsychiatry Schedules for Clinical Assessment (SCAN Thai Version) were effective tools for diagnosing mood disorders among Thais.*

Keywords: *Semi-structured interview, Schedules for clinical assessment in neuropsychiatry, Validity, Reliability, Mood disorders, Mania, Depression, Dysthymia, Cyclothymia*

J Med Assoc Thai 2006; 89 (2): 205-11
Full text. e-Journal: *http://www.medassocthai.org/journal*

Diagnosing psychiatric conditions is the first step of the treatment process. The categorical classification systems have some limitations when used with psychiatric conditions which are subjectivity rather than objectivity, some overlapping between each disorders and where co-morbidities are frequently occurred[1]. Clinical judgment even from well trained physicians are not always in concordance. To develop a standardized interview schedule which is universally accepted and can be well adapted to specific culture is an important task and challenging task. Mood disorders are clinically heterogeneous and thought to result from multiple gene interactions triggered by environmental and developmental stressors.

Mood disorders are among the most burdensome of all classes of disease because of their early onset, prevalence, chronicity and co-morbidities[2,3]. The European Study of the Epidemiology of Mental Disorders (ESEMeD) Project 2004 found that 14% of population reported a lifetime history of any particular mood disorder[4]. Thavichachart et al found the prevalent lifelong mood disorders in Bangkok included

Correspondence to : Krisanaprakornkit T, Deparment of Psychiatry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand. Phone: 0-4334-8384, Fax: 0-4334-8384, E-mail: drthawatchai@yahoo.com

manic episodes (9.3%), major depressive episodes (19.9%) and dysthymia (1%)[5].

Despite the widespread use of psychiatric diagnostic interviews to diagnose mood disorders (*i.e.* Hamilton Rating Scale for Depression, Beck Depression Inventory, Zung Self Rating Scale for Depression, Young Manic Rating Scale, Diagnostic Interview Schedule for Children, Children's Depression inventory, Mood Disorder Questionnaire (MDQ) for bipolar disorder)[6],psychiatry lacks a comprehensive schedule covering all the psychopathological phenomena presenting in mood disorders. Moreover, no true gold standard exists against which to test the validity of any new psychiatric diagnostic technique(s), and the traditional use of the interview can produce variable diagnostic conclusions depending on the interviewer, the interviewee and the interaction between them.

The Schedules for Clinical Assessment in Neuropsychiatry (SCAN) constitute a semi-structured clinical interview for use by trained clinicians to assess and diagnose psychiatric disorders among adults.

SCAN's core is a Present State Examination (PSE) that has been tested globally for its validity and reliability. SCAN was developed within the framework of the WHO and the National Institute of Mental Health (NIMH) Joint Project on Diagnosis and Classification of Mental Disorders, Alcohol and Related Problems. The use of SCAN gives flexibility in the diagnosis of mental disorders based on the current International Classification of Disease (ICD), Diagnostic and Statistical Manual (DSM) systems and other diagnostic systems that may develop in the future. Additionally, SCAN allows worldwide comparisons of psychiatric diagnoses[7-9]; thus the uses of SCAN's semi-structured interview for epidemiological research of mood disorders will likely increase[10,11].

The authors aimed to test the validity and reliability of the Mood Disorders Section of SCAN's Thai version.

Material and Method

After translating the original English version of SCAN to Thai and back-translating to establish its validity (Paholpak *et al*., 2003), the SCAN mood disorders sections were extracted, including: Section 6: Depressed Mood And Ideation; Section 7: Thinking, Concentration, Energy, Interests; Section 8: Bodily Functions; And, Section 10: Expansive Mood And Ideation. These sections would be used to conduct interviews among psychiatric patients.

The Ethics Committee for KhonKaen University reviewed and approved the study protocols and informed consent was obtained from patients before conducting the interviews. Then between October 2003 and August 2004, the authors conducted semi-structured interviews using the Mood Disorders Section of the Thai version of SCAN both on mood disordered patients and normal volunteers at Srinagarind Hospital, Khon Kaen, Thailand.

The process of validity and reliability testing were accomplished as follows:

1. Content validity: Two psychiatrists well-versed in SCAN arrived at a consensus on the original meaning of each item and whether the Thai version conserved the original meaning. The comprehensibility of language was then tested among Thais from all four regions of the country. Reflections, comments and suggestions from the Thais interviewed were assessed then summarized during a consensus meeting of the two psychiatrists (PS and KT). With permission from WHO, the final Thai version was incorporated into the SCAN I-shell program (Fig. 1).

2. Reliability study: The presented sample size comprised of 30 subjects (15 mood disordered patients and 15 normal volunteers). The patients (from either the in- or out-patient departments) were identified using either the ICD-10 or DSM-IV criteria. All subjects had to be over 14 years of age, ethnic Thais (*i.e.* fluent in, and able to understand, spoken Thai). All subjects were interviewed by a psychiatrist familiar with SCAN. With permission from each subject, the interviews were recorded on digital video.

2.1 Inter-rater reliability: two psychiatrists (trained in the use of SCAN) independently rated the interviews; either live or on video; and,

2.2 Intra-rater reliability: one of the psychiatrists re-rated the video 2 weeks later.

Statistical analysis

Inter- and intra-rater reliability were determined from the agreement between raters; calculated using the kappa statistic ( ) for categorical data or the Intraclass Correlation for continuous data[12]. The simple percentage of agreement was used whenever the statistic could not be calculated. All statistics were done using STATA 7.0.

The predefined level for the degree of agreement was: 1 = poor agreement (< 0.00); 2 = slight

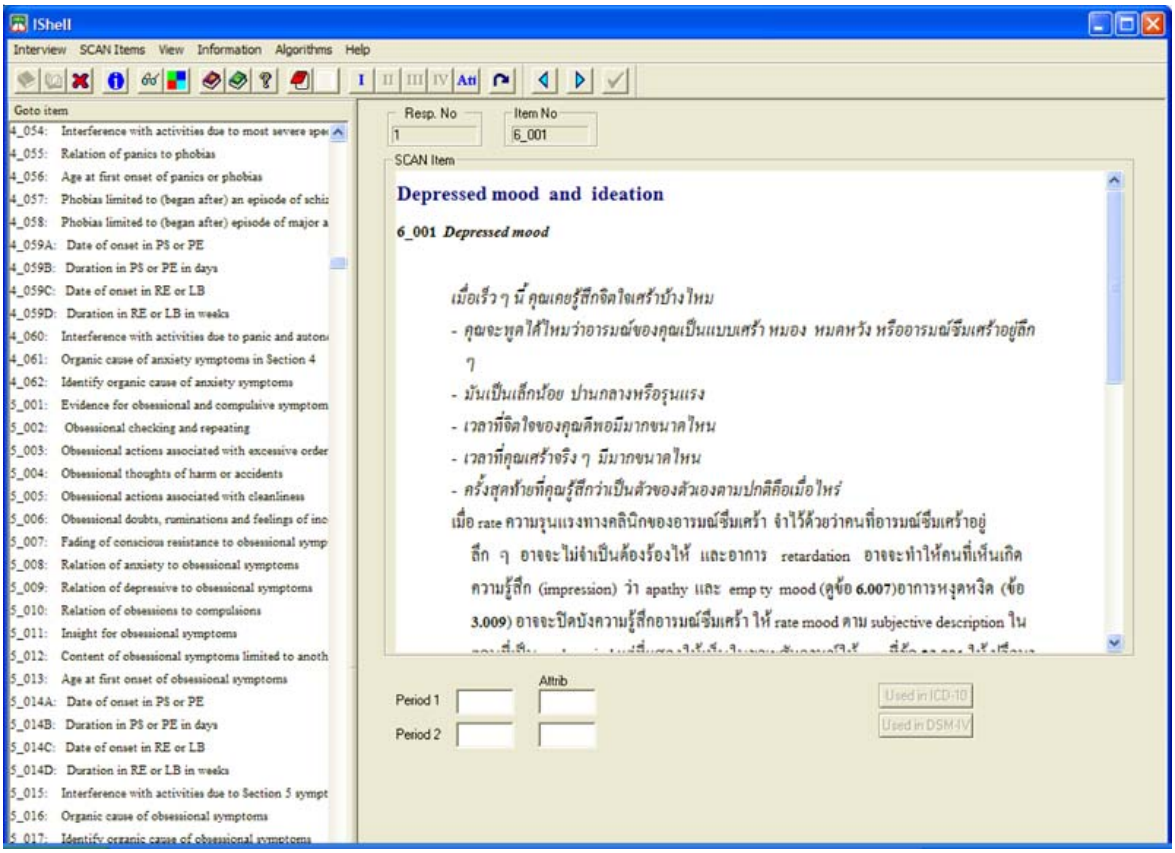

Fig. 1 WHO I-Shell program SCAN 2.1 (Thai version)

(: 0.00-0.20); 3 = fair (: 0.21-0.40); 4 = moderate (: 0.41-0.60); 5 = substantial (: 0.61-0.80); and, 6 = near perfect (K: 0.81-1.00)[13,14].

Results

Content validity was performed by two psychiatrists (PS and KT). Some adaptations were made to words or sequence of sentences describing symptoms to make them more understandable in the Thai (cultural and linguistic) context.

One of the researchers (KT) interviewed 80 volunteers, representing four regions of Thailand (20 volunteers per region), and tested their understanding of the terms used in the SCAN (Thai version). All of the comments and suggestions (*i.e.* for comparable meanings using local idioms) were gathered and the most suitable (*i.e.* understandable and conserving the original meaning) was chosen.

The reliability study commenced with 30 subjects (8 males [26.7%]; 22 females [73.3%]) between 15 and 60 years of age, including both mood disordered patients and normal volunteers to ensure a full range of scores and spectrum of symptoms. The interviews took between 21.7 and 111.1 minutes (average, 63.6 $\pm$ 17.1) and none of the subjects dropped out during the interviews.

SCAN Section 6: Depressed mood

The mean (SD) inter- and intra-rater reliability- for the 51 items in this section was 0.80 (0.18) and 0.85 (0.13), respectively. Almost all of the indicated near perfect to substantial agreement; except for the following six which had slight to moderate agreement: 6_002 (Masked Depression); 6_008 (Loss Of Reactivity); 6_009 (Morning Depression); 6_014 (Guilty Ideas Of Reference); 6_015 (Loss Of Self-Confidence With Other People); and, 6_018 (Delusion Of Guilt Or Worthlessness In Context Of Depression) (Table 1-3).

SCAN Section 7: Thinking, concentration, energy, interests

The mean (SD) inter- and intra-rater reliability- for the 9 items in this section was: 0.67 (0.12) and 0.84 (0.1), respectively. Again, the majority of the

Table 1. Agreement of each section determined by statistic

Section	No. of items	Inter-rater reliability (SD)	Intra-rater reliability (SD)
6	51	0.80 (0.18)	0.85 (0.13)
7	9	0.67 (0.12)	0.84 (0.1)
8	22	0.78 (0.14)	0.88 (0.12)
10	40	0.84 (0.19)	0.86 (0.13)
All sections (95%CI)	122	0.80 (0.77-0.83)	0.86 (0.84-0.88)

Table 2. Inter-rater reliability profile for each section

Section	Poor (<0.00)	Slight (=0.00-0.20)	Fair (=0.21-0.40)	Moderate (=0.41-0.60)	Substantial (=0.61-0.80)	Almost Perfect (=0.81-1.00)	Total (items)
6	-	1	-	5	16	29	51
7	-	-	-	2	6	1	9
8	-	-	-	3	7	12	22
10	-	-	-	5	9	26	40
All sections (%)	-	1 (0.8)	-	15 (12.2)	38 (31.1)	68 (55.8)	122 (99.9)

Table 3. Intra-rater reliability profile for each section

Section	Poor (< 0.00)	Slight (= 0.00-0.20)	Fair (= 0.21-0.40)	Moderate (= 0.41-0.60)	Substantial (= 0.61-0.80)	Almost Perfect (= 0.81-1.00)	Total (items)
6	-	-	1	-	12	38	51
7	-	-	-	-	3	6	9
8	-	-	-	1	3	18	22
10	-	-	-	1	12	27	40
All sections (%)	-	-	1 (0.8)	2 (1.6)	30 (24.6)	89 (73.0)	122 (100.0)

indicated near perfect to substantial agreement except for items 7_004 (Loss Of Interests) and 7_010 (Organic Cause Of Section 7 Symptoms), which were 0.59 and 0.46 (moderate agreement), respectively.

SCAN Section 8: Bodily functions

The mean (SD) inter- and intra-rater reliability- for the 22 items was 0.78 (.14) and 0.88 (0.12), respectively. The intra-class correlation was calculated for four items and for these the inter- and intra-rater reliability was 0.99-1.0 and 0.87-1.0, respectively. The reliability- for the majority of items indicate substantial to almost perfect agreement, but only moderate for items 8_006 (Loss Of Weight), 8_007 (Gain Of Weight) 8_029 (Organic Cause Of Sexual Dysfunction).

SCAN Section 10: Expansive mood and ideation

The mean inter- and intra-rater reliability- for the 40 items of this section was 0.84 (0.19) and 0.86 (0.13), respectively. The majority of items had almost

perfect to substantial agreement, except moderate agreement for 10_006 (Distractibility), 10_008 (Sharpened Thinking), 10_011 (Over-Optimism), 10_016 (Delusion Of Grandiose Abilities), 10_049 (2+ Year Periods Of Sub-Depression).

The overall inter-rater reliability- and 95%CI for the 122 items in all sections was 0.80 (0.77-0.83); indicating substantial agreement. The intra-class correlations for the 4 items in Section 8 were between 0.99-1.0.

The overall intra-rater reliability- for the 122 items in all sections was 0.86 (0.84-0.88); indicating almost perfect agreement. The intra-class correlations for the 4 items in Section 8 were between 0.87-1.0 (Table 1).

Discussion

Since the evaluation of the psychopathology of psychiatric illness is relatively subjective, it is imperative to have an instrument which is capable of detecting such symptoms. The primary advantage of SCAN is the use of a semi-structured interview and bottom-up approach, which group symptoms as much as possible before making a diagnosis. The second advantage is SCAN used cross-examination techniques which provide well-defined symptoms-criteria to help psychiatrists match their own clinically-relevant symptom-concepts with the symptoms experienced expressed by the patients.

In the content validity process of SCAN (Thai version), the authors adapted some of the sentences and words from the original SCAN (English version)[15] to make SCAN-Thai more understandable in the Thai linguistic and social context. These types of emendations are acceptable practice in translation science as long as the original meaning(s) is conserved. After testing SCAN in Taiwan, Cheng et al. concluded that cross-cultural implementation of SCAN was practicable[16].

The overall inter- and intra-rater reliability of SCAN (Thai version) Mood Disorder Section (Sections 6, 7, 8 and 10) were acceptable based on values ranging between 0.80 (0.77-0.83) (substantial agreement) and 0.86 (0.84-0.88) (near perfect agreement), respectively. The good agreement might be due to the authors use of psychiatrists well-versed in SCAN to rate (and re-rate) the interviews (Table 2).

The present results agree with a Spanish reliability study (of SCAN Spanish version), which also reported a high degree of reliability[17]. Even though SCAN should be used by experienced clinicians, Rijnders et al showed that less experienced (but well-trained) interviewers were able to reliably apply SCAN[18].

Only one (item 6_002: Masked Depression) of the 122 items had 'fair' agreement (*i.e.* < 0.4) (Table 3). Possibly, this item relies on interviewer judgment based on observation of patient behavior. This weakness was also noted by Rijnders et al, so, at the risk of neglect, special attention should be paid to items that have no explicit interview questions[18].

Similarly, some caution should be exercised when using the 15 (11.9%) (of 126) items with only moderate agreement (*i.e.* = 0.41-0.60) as the percent agreement varied between 63%-93%. Andrews et al, reported that when clinical judgment is involved in administering SCAN, agreement between the interviewer and observer is limited to moderate levels, which is less than that for CIDI - a highly structured interview[19]. The practical solution is to re-check the score and criteria used for rating (among raters) by consulting the SCAN glossary[20].

Limitations

1. A limitation to the present study is that during the reliability study, the authors recruited only participants from Srinagarind Hospital, Khon Kaen (Northeast Thailand);

2. Linguistic discrepancies might have persisted despite having conducted language pre-testing in the first phase; and,

3. The authors did not evaluate SCAN vs clinical interview in diagnostic agreement as such an analysis was beyond the objective of the present study; however, the authors believe further study of concurrent validity *should* be done.

Conclusion

The 'SCAN (Thai version): Mood Disorders Section' has good validity and reliability. Using this semi-structured interview would help in the correctness and consistent diagnosis among the interviewers. Training in the use of SCAN in Thailand should be set up to build familiarity with the terms and approach. Psychiatric research based on SCAN: mood disorder section, as an instrument in Thailand, should prove fruitful.

Acknowledgements

This study was funded by the National Research Council of Thailand. The authors wish to thank Mr. Bryan Roderick Hamman for his assistance with the English-language presentation.

References

1. Akiskal HS. The bipolar spectrum in psychiatric and general medical practice. Primary Psychiatry 2004; 11: 30-5.
2. Cross-national comparisons of the prevalences and correlates of mental disorders. WHO International Consortium in Psychiatric Epidemiology. Bull World Health Organ 2000; 78: 413-26.
3. Yerevanian BI, Koek RJ, Ramdev S. Anxiety disorders comorbidity in mood disorder subgroups: data from a mood disorders clinic. J Affect Disord 2001; 67: 167-73.
4. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl 2004; (420): 21-7.
5. Thavichachart N, Intoh P, Thavichachart T, Meksupa O, Tangwongchai S, Sughondhabirom A, et al. Epidemiological survey of mental disorders and knowledge attitude practice upon mental health among people in Bangkok Metropolis. J Med Assoc Thai 2001; 84: S118-26.
6. Miller CJ, Klugman J, Berv DA, Rosenquist KJ, Nassir Ghaemi SN. Sensitivity and specificity of the mood disorder questionnaire for detecting bipolar disorder. J Affect Disord 2004; 81: 167-71.
7. Aboraya A, Tien A, Stevenson J, Crosby K. Schedules for clinical assessment in neuropsychiatry (SCAN): introduction to WV's mental health community. W V Med J 1998; 94: 326-8.
8. Janca A, Ustun TB, Sartorius N. New versions of World Health Organization instruments for the assessment of mental disorders. Acta Psychiatr Scand 1994; 90: 73-83.
9. 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.
10. Michalak EE, Wilkinson C, Hood K, Srinivasan J, Dowrick C, Dunn G. Prevalence and risk factors for depression in a rural setting. Results from the North Wales arm of the ODIN project. Soc Psychiatry Psychiatr Epidemiol 2002; 37: 567-71.
11. Freudenstein U, Arthur A, Matthews R, Jagger C. Can routine information improve case finding of depression among 65 to 74 year olds in primary care? Fam Pract 2002; 19: 520-2.
12. Altman DG. Practical statistic for medical research. London: Chapman and Hall; 1991.
13. Feinstein AR. Principles of medical statistics. Washington, DC: Chapman&Hall/CRC; 2002.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-74.
15. WHO SCAN Advisory Committee. Schedules for clinical assessment in neuropsychiatry. Version 2.1: Interview. ed. Geneva: World Health Organization; 1999.
16. Cheng AT, Tien AY, Chang CJ, Brugha TS, Cooper JE, Lee CS, et al. Cross-cultural implementation of a Chinese version of the schedules for clinical assessment in neuropsychiatry (SCAN) in Taiwan. Br J Psychiatry 2001; 178: 567-72.
17. Vazquez-Barquero JL, Gaite L, Artal Simon J, Arenal A, Herrera Castanedo S, Diez Manrique JF, et al. Development and verification of the Spanish version of the "scanning system" psychiatric interview "Questionnaires for clinical evaluation in neuropsychiatry". Actas Luso Esp Neurol Psiquiatr Cienc Afines 1994; 22: 109-20.
18. Rijnders CA, van den Berg JF, Hodiamont PP, Nienhuis FJ, Furer JW, Mulder J, et al. Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). Soc Psychiatry Psychiatr Epidemiol 2000; 35: 348-52.
19. Andrews G, Peters L, Guzman AM, Bird K. A comparison of two structured diagnostic interviews: CIDI and SCAN. Aust N Z J Psychiatry 1995; 29:124-32.
20. Bertelsen A, Bruga T, Tien AY. Schedules for clinical assessment in neuropsychiatry. Version 2.1: Glossary. ed. Geneva: World Health Organization; 1999.

ความถูกต้องและความเชื่อถือได้ของ WHO Schedules For Clinical Assessment In Neuropsychiatry, ฉบับภาษาไทย: หมวดความผิดปกติทางอารมณ์

ธวัชชัย กฤษณะประกรกิจ, สุชาติ พหลภาคย์, นวนันท์ ปิยะวัฒน์กูล, จิราพร เขียวอยู่

ความผิดปกติทางอารมณ์เป็นปัญหาทางจิตเวชที่มีความชุกสูงในประชากรทั่วไปเริ่มเป็นตั้งแต่อายุน้อย เป็นเรื้อรัง และทำให้เกิดความบกพร่องในการทำหน้าที่ด้านต่างๆ การวินิจฉัยความผิดปกติทางอารมณ์ได้อย่างถูกต้อง จึงมีความสำคัญมาก ทั้งในด้านการดูแลรักษาและงานวิจัยทางจิตเวชศาสตร์

รายงานนี้เป็นการศึกษาความถูกต้องและความเชื่อถือได้ของ WHO Schedules for Clinical Assessment in Neuropsychiatry ฉบับภาษาไทย หมวดความผิดปกติทางอารมณ์ โดยทดสอบความถูกต้องเชิงภาษาในประชากรทั่วไปทั้ง 4 ภาคของประเทศไทยและความเห็นจากผู้เชี่ยวชาญ และทดสอบความเชื่อถือได้ในกลุ่มตัวอย่างที่ประกอบด้วยผู้ป่วยความผิดปกติทางอารมณ์และอาสาสมัครปกติ จำนวน 30 ราย

ผลการศึกษาทำให้ได้แบบสัมภาษณ์กึ่งโครงสร้างที่มีความถูกต้องเชิงภาษา ผู้ถูกสัมภาษณ์สามารถเข้าใจความหมายของข้อคำถามและรักษาความหมายได้ตรงกับต้นฉบับภาษาอังกฤษ ใช้เวลาในการสัมภาษณ์เฉลี่ย 63.6 นาที (SD = 17.1) และมีค่าความเชื่อถือได้จากการวัดความสอดคล้องตรงกันระหว่างผู้สัมภาษณ์ 2 คนเท่ากับ 0.80 (0.77-0.83) และความสอดคล้องตรงกันในผู้สัมภาษณ์คนเดียวกันที่ให้คะแนน 2 ครั้ง เท่ากับ 0.86 (0.84-0.88) ซึ่งเป็นระดับความสอดคล้องมากถึงความสอดคล้องเกือบสมบูรณ์ อย่างไรก็ตามพบว่ามีข้อคำถามที่มีค่าความสอดคล้องในระดับเล็กน้อย 1 ข้อ (ร้อยละ 0.79) และระดับปานกลางอยู่ 15 ข้อ (ร้อยละ 11.9) ซึ่งควรจะมีการตรวจสอบความเข้าใจของผู้สัมภาษณ์ก่อนการให้คะแนน ก็จะสามารถนำเครื่องมือนี้มาใช้ในการวินิจฉัยความผิดปกติทางอารมณ์ในคนไทยได้อย่างมีประสิทธิภาพ