

Cross-Cultural Validation and Inter-Rater Reliability of the Personal and Social Performance Scale, Thai Version

Manit Srisurapanont MD*,
Suwanna Arunpongpaisal MD**, Sunanta Chuntaruchikapong MD***,
Chatchawan Silpakit MD****, Vira Khuangsirikul MD*****,
Nipatt Karnjanathanalers MD*****, Umpaikarnit Samanwongthai MD*****

* Department of Psychiatry, Faculty of Medicine, Chiang Mai University, Chiang Mai

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

*** Department of Psychiatry, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok

**** Department of Psychiatry, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok

***** Department of Psychiatry and Neurology, Pramongkutklao Hospital, Bangkok

***** Department of Psychiatry, Faculty of Medicine, Chulalongkorn University, Bangkok

***** Srithanya Hospital, Nonthaburi

Objective: To cross-culturally validate and examine the inter-rater reliability of the Personal and Social Performance scale (PSP), Thai version (Thai-PSP).

Material and Method: The authors translated, back translated, and conducted a panel review on the source, translated, and back translated version of the PSP. After nine psychiatrists, who were new to the PSP or the Thai-PSP, had received a three-hour session of training, they jointly watched four tape-recorded interviews and used the Thai-PSP for independent rating of the patients' functioning.

Results: The score ranges for items 1 (socially useful activities), 2 (personal/social relationships), and 3 (self-care) were between 0 (absent) and 4 (severe). The score range was between 0 (absent) and 3 (marked) for item 4 (disturbing/aggressive behavior). The total scores of four patients were rated between 2 (21-30 points) and 8 (81-90 points). The intraclass correlation coefficients (95% confidence intervals) of each item and total score were as follows: 0.63 (0.28-0.96) for item 1, 0.75 (0.42-0.98) for item 2, 0.69 (0.35-0.97) for item 3, 0.52 (0.17-0.94) for item 4, and 0.75 (0.41-0.98) for the total score.

Conclusion: The results of the present study confirm the reliability of PSP and Thai-PSP, as well as the ease of training.

Keywords: Schizophrenia, Functioning, reliability, Intraclass correlation coefficient

J Med Assoc Thai 2008; 91 (10): 1603-8

Full text. e-Journal: <http://www.medassocthai.org/journal>

As behavior cannot be measured in figures, functioning is an issue of concern for clinical practice and research in psychiatry. Impaired functioning has been used as an indicator of abnormal behavior. According to the Diagnostic and Statistical Manual for Mental Disorders, 4th edition (DSM-IV), functional impairment is a common criterion for almost all mental

disorders⁽¹⁾. In addition, levels of functioning area have to be described in Axis V, an important part, of the DSM-IV diagnostic system.

Advancements in the pharmacological treatment have contributed to the better clinical outcomes of psychiatric patients. The discovery of atypical antipsychotics plays an important role in raising the expectation of schizophrenic patients. It is now possible to bring a schizophrenic patient with severely impaired functioning (e.g., a life-threatening condition) back to his/her normal self. Due to the wide range of functioning

Correspondence to: Srisurapanont M, Department of Psychiatry, Faculty of Medicine, Chiang Mai University, Muang, Chiang Mai 50200, Thailand. Phone: 053-945-422, Fax: 053-945426, Email: msrisu@yahoo.com

in schizophrenic patients, a measure for the assessment of function in schizophrenic patients is desperately needed.

Global Assessment of Functioning (GAF) scale and Social and Occupational Functioning Assessment Scale (SOFAS) are two measures commonly used for functional assessment of mentally ill patients⁽¹⁾. However, several lines of evidence have shown that there are some limitations of these instruments. Firstly, the GAF scale takes into account the severity of psychiatric symptoms, which is not a part of functioning. Secondly, the intraclass correlation coefficients (ICCs) ranging between 0.56 and 0.59 on the GAF global score, symptoms component scale, and social functioning scale suggest that the inter-rater reliability of the GAF global and subscales are only at the moderate levels⁽²⁾. Lastly, the results of a study have shown that the SOFAS has poor concurrent validity with the Quality of Life Scale⁽³⁾ and lacks discrimination between homeless and not homeless patients⁽⁴⁾.

Recently, Morosini et al (2000) have developed a measure called 'the Personal and Social Performance scale or PSP'. The scale contents are drawn from 27 aspects widely accepted as important functioning of schizophrenic patients. Despite that, this scale is very concise by summarizing the functioning into 4 key areas, i.e., personal and social relationships, socially useful activities (including work and study), self-care, and disturbing and aggressive behavior. Choices (scores) of each area are: absent (0), mild (1), moderate (2), marked (3), severe (4), and very severe (5). The total PSP score of 0 to 100 is derived from the combined severity of those 4 key areas. Possibly due to the clear scoring system, the inter-rater reliability of this scale is very high (the ICC of the total score = 0.98). Compared to SOFAS, PSP has better face validity and psychometric properties⁽¹⁾. It was found to be a concise, valid, and reliable measure of patients' personal and social functioning.

To minimize the language and cultural barriers of this instrument, the authors proposed to cross-culturally validate and examine the inter-rater reliability of PSP, Thai version (Thai-PSP).

Material and Method

The present study was approved by the ethics committees of the Faculty of Medicine, Chiang Mai University and the Faculty of Medicine, Khon Kaen University. After the details of the study were fully explained, written informed consent was obtained from both patients and their relatives who were interviewed and tape-recorded.

The present study applied the first three steps of guidelines for cross-cultural adaptation of health-related quality of life measures, including, translation, back translation, and committee review⁽⁶⁾. A panel for the development of Thai-PSP comprised 7 psychiatric experts. Two panel members (MS and SA), who were familiar with the PSP, separately translated the English version of PSP into Thai (the forward translated versions of Thai-PSP). The other two members (SA and CS), who had never seen the PSP before, independently back translated the first drafts into English (the back translated versions of Thai-PSP). The panel reviewed the source, forward, and back translated versions by taking into account the instrumental concept and language/cultural appropriateness. Finally, the panel reached an agreement on the final version of Thai-PSP (see Appendix).

After being trained by United BioSource Corporation (formerly PharmaStar), an agency providing PSP training for research and practice, MS and SA interviewed five pairs of chronic schizophrenic patients and their close relatives. After finishing the tape-recorded interviews, the authors invited nine psychiatrists to participate in the joint inter-rater reliability study of the Thai-PSP. All of them had never been trained or used the PSP or the Thai-PSP before. The raters received a three-hour session of training, including the introduction to Thai-PSP and a case demonstration, instructed by MS and SA. After the providing of patients' background, these trained psychiatrists jointly watched the four tape-recorded interviews and independently rated the patients' functioning by the use of Thai-PSP.

The intraclass correlation coefficients with 95% confidence intervals (95% CI) (ICCs) of all four items (scores of 0-5) and the total score (10-point intervals of 0-9) of the Thai-PSP were calculated to determine the inter-rater reliability of the scale.

Results

All four patients, who were interviewed and tape-recorded, were diagnosed as chronic schizophrenia, paranoid type. Their ages were between 30 and 55 years old. None of them had an occupation. Their highest levels of education were primary school (2 patients), vocational school (1 patient), and university (1 patient). Their durations of illness were between 5-17 years. The raters included 4 male and 5 female psychiatrists across Thailand. Their mean (SD) duration of practice was 3.7 (2.9) years, respectively.

The score ranges for items 1 (socially useful activities), 2 (personal/social relationships), and 3

(self-care) were between 0 (absent) and 4 (severe). The score range was between 0 (absent) and 3 (marked) for item 4 (disturbing /aggressive behavior). The total scores of four patients were rated between 2 (21-30 points) and 8 (81-90 points). The ICCs (95% confidence intervals, p-value's) of each item and total score were as follows: 0.63 (0.28-0.96, p=000) for item 1, 0.75 (0.42-0.98, p=000) for item 2, 0.69 (0.35-0.97, p=000) for item 3, 0.52 (0.17-0.94, p=000) for item 4, and 0.75 (0.41-0.98, p=000) for the total score.

Conclusion

To the authors' knowledge, this is the first study of the inter-rater reliability of a translated PSP. Except the moderate reliability of disturbing/aggressive behavior item, the inter-rater reliability of all Thai-PSP items and the total score are good (or substantial)⁽⁷⁾.

Because the original study of PSP did not present the ICCs of each item, this respect cannot be compared between the original and the translated ones. However, the ICC of the total score of 10-point intervals in the former study (0.98) was relatively higher than that of the present study (0.75). The difference of ICCs between studies may be caused by the differences of methods used in the studies. Firstly, in comparison to the use of case-vignettes in the original study, the Thai-PSP study uses tape-recorded interviews, which was more similar to the real world practice. Secondly, the raters in a previous study, but not the present one, were trained by the PSP developers. Thirdly, while the raters of the former study were nurses, aides, and rehabilitation workers, those of the present one were psychiatrists. Lastly, it was also possible that the language/cultural adaptation for the Thai-PSP played a role in decreasing the inter-rater reliability.

A major limitation of the present study was the lack of subjects with impaired functioning at the very severe level. None of the studied patients had a score of 5 (very severe impairment of functioning) in the respects of socially useful activities, personal/social relationships, and self-care. In addition, all patients had scores of 4 (marked) or less for the disturbing/aggressive behavior. This limitation appears to be unavoidable because it is unlikely that anyone can do a tape-recorded interview on a patient with very severe impairment of functioning. Most schizophrenic patients who have very severe impaired functioning are those who also have severe psychopathology. The severe psychotic symptoms may cause uncooperativeness of the patient during the interview. In

addition, it would be unsafe to interview and tape-record a very aggressive patient.

Although the present study included only four interviews and nine raters for the assessment of the inter-rater reliability of Thai-PSP, the statistically significant agreements (p-value's = 0.000) of all items and the total score suggested that the sample size of the present study was adequate for the study of ICCs. The present study did not examine the validity, sensitivity to change, and other respects of reliability because it is still controversial to include these processes in a language/cultural adaptation of a measure⁽⁶⁾.

Despite some limitations, the results of the present study would be helpful for the future language/cultural adaptation of PSP in other languages. The substantial inter-rater reliability of Thai-PSP confirms the high agreement of the PSP in rating the functioning of mentally ill patients.

After a brief session of Thai-PSP training, Thai psychiatrists who are new to the scale can have a substantial agreement in rating schizophrenic patients' functioning. The results of the present study confirm the reliability of PSP and Thai-PSP, as well as the ease of training.

Acknowledgement

This study was supported by a grant from Janssen-Cilag (Thailand). We are grateful to all nine raters of this study.

References

1. American Psychiatric Association. Diagnostic and statistical manual. 4th ed. Washington DC: American Psychiatric Press; 1994.
2. Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: a review of measures of social functioning. *Am J Psychiatry* 1992; 149: 1148-56.
3. Lehman AF. A quality of life interview for the chronically mentally ill. *Eval Prog Plann* 1988; 11: 51-62.
4. Roy-Byrne P, Dagadakis C, Unutzer J, Ries R. Evidence of limited validity of the revised global assessment of functioning scale. *Psychiatr Serv* 1996; 47: 864-6.
5. Morosini PL, Magliano L, Brambilla L, Ugolini S, Pioli R. Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. *Acta Psychiatr Scand* 2000; 101: 323-9.
6. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures:

literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46: 1417-32.

7. Altman DG. *Practical statistics for medical research*. London: Chapman & Hall; 1991.

ความแม่นยำแบบข้ามวัฒนธรรมและความเชื่อมั่นระหว่างผู้ประเมินของแบบวัด *Personal and Social Performance* ฉบับภาษาไทย

มานิต ศรีสุรภานนท์, สุวรรณ อรุณพงศ์ไพศาล, สุนันทา ฉันทรุจิกพงศ์, ชัชวาลย์ ศิลปะกิจ, วีระ เขื่องศิริกุล, นิพัทธ์ กาญจนธนาเลิศ, อำไพชนิษฐ์ สมานวงศ์ไทย

วัตถุประสงค์: เพื่อจัดทำ *Personal and Social Performance scale (PSP)* PSP ฉบับภาษาไทยที่มีความแม่นยำแบบข้ามวัฒนธรรม และทำการตรวจสอบความเชื่อมั่นของ Thai-PSP

วัสดุและวิธีการ: คณะผู้นิพนธ์แปลไปข้างหน้า, แปลย้อนกลับ และร่วมกันพิจารณาแบบวัดต้นฉบับ, ฉบับร่างที่ได้จากการแปลไปข้างหน้า, และฉบับร่างที่ได้จากการแปลย้อนกลับ หลังจากที่ได้จิตแพทย์เก้าท่าน ซึ่งไม่เคยทราบเกี่ยวกับ PSP และ PSP ฉบับภาษาไทยมาก่อนได้รับการฝึกอบรมนาน 3 ชั่วโมง จิตแพทย์เหล่านี้ได้ร่วมกันดูการสัมภาษณ์ที่บันทึกเทปไว้ 4 ชุด และทำการประเมินการทำหน้าที่ของผู้ป่วยโรคจิตเภทเรื้อรังสี่รายโดยใช้ PSP ฉบับภาษาไทย

ผลการศึกษา: พิสัยของคะแนนสำหรับหัวข้อที่ 1 (กิจกรรมทางสังคมที่เป็นประโยชน์), หัวข้อที่ 2 (สัมพันธภาพบุคคลสังคม) และหัวข้อที่ 3 (การดูแลตนเอง) อยู่ระหว่าง 0 (ไม่มีปัญหา) และ 4 (มีรุนแรง). พิสัยของคะแนนในหัวข้อที่ 4 (พฤติกรรมรบกวน/ก้าวร้าว) อยู่ระหว่าง 0 (ไม่มีปัญหา) และ 3 (มีปรากฏอย่างชัดเจน) คะแนนรวมของผู้ป่วยทั้งสี่ราย ถูกประเมินว่าอยู่ระหว่าง 2 (21-30 คะแนน) และ 8 (81-90 คะแนน) ค่า *intraclass correlation coefficients (95% confidence intervals)* ของแต่ละหัวข้อและคะแนนรวมเป็นดังนี้: 0.63 (0.28-0.96) สำหรับหัวข้อที่ 1, 0.75 (0.42-0.98) สำหรับหัวข้อที่ 2, 0.69 (0.35-0.97) สำหรับหัวข้อที่ 3, 0.52 (0.17-0.94) สำหรับหัวข้อที่ 4 และ 0.75 (0.41-0.98) สำหรับคะแนนรวม

สรุป: ผลการศึกษานี้ยืนยันความเชื่อมั่นได้ของ PSP และ Thai-PSP รวมทั้งความง่ายในการฝึกอบรม

Appendix.

แบบประเมินความสามารถในการทำหน้าที่ส่วนบุคคลและทางสังคม
(Personal and Social Performance Scale, Thai Version, Thai-PSP)

ชื่อผู้ป่วย:

วันที่:

โปรดให้คะแนนเกี่ยวกับระดับความสามารถในการทำหน้าที่ของผู้ป่วยในระยะ 1 เดือนที่ผ่านมา หรือ ตั้งแต่พบกันครั้งสุดท้ายก่อนหน้า
ความสามารถในการทำหน้าที่มีอยู่ 4 ด้าน ดังนี้

	ไม่มี ปัญหา	มีบ้าง เล็กน้อย	มีพอ ปรากฏ	มีอย่าง ชัดเจน	มีรุนแรง	มีรุนแรง มาก
ก) กิจกรรมทางสังคมที่เป็นประโยชน์ รวมถึง การทำงาน, การเรียน, การทำงานบ้าน	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ข) ความสัมพันธ์กับคนอื่นและคนในสังคม	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ค) การดูแลอนามัยส่วนตัว	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ง) พฤติกรรมรบกวนผู้อื่นหรือก้าวร้าว	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

คะแนนรวม:

มีค่านิยามและเกณฑ์การตัดสินระดับความผิดปกติที่แตกต่างกันอยู่ 2 ชุด: ชุดแรกสำหรับข้อ ก-ค และชุดที่สองสำหรับข้อ ง

ชุดสำหรับข้อ ก-ค

- **ไม่มีปัญหา**
- **มีบ้างเล็กน้อย** หมายถึง ความบกพร่องสามารถรู้ได้จากคนใกล้ชิด
ผู้ป่วยมาก ๆ เท่านั้น
- **มีพอปรากฏ** หมายถึง ความบกพร่องเห็นได้ชัดจากทุกคน แต่ไม่รบกวน
ความสามารถในการทำตามบทบาทและหน้าที่รับผิดชอบในบริบททาง
สังคมและวัฒนธรรม, ที่สอดคล้องกับ อายุ เพศ และระดับการศึกษา
- **มีอย่างชัดเจน** หมายถึง ความบกพร่องที่มีผลกระทบอย่างมากต่อความ
สามารถในการทำตามบทบาทและหน้าที่ที่รับผิดชอบ อย่างไรก็ตามผู้ป่วย
ยังคงทำงานอย่างได้โดยปราศจากการช่วยเหลือจากสังคม หรือ นักวิชาชีพ
แต่ผลงานไม่สมบูรณ์ และ/หรือ ทำได้ชั่วคราวชั่วคราว แต่ถ้าได้รับการช่วยเหลือ
จากคนอื่น ผู้ป่วยอาจจะเพิ่มขีดความสามารถขึ้นมาถึงระดับก่อนป่วยได้
- **มีรุนแรง** หมายถึง ความบกพร่องมีผลทำให้ผู้ป่วยไม่สามารถทำตามบทบาท
หน้าที่ของตนเองได้ ถ้าไม่ได้รับการช่วยเหลือจากนักวิชาชีพ หรือ ความบกพร่อง
นั้นนำไปให้ผู้ป่วยเกิดความเสียหาย แต่ยังไม่ถึงกับเสี่ยงต่อการฆ่าตัวตาย
- **มีรุนแรงมาก** หมายถึง ความบกพร่องรุนแรงจนเป็นอันตรายต่อชีวิตของผู้ป่วย

ชุดสำหรับข้อ ง

- **ไม่มี**
- **มีบ้างเล็กน้อย** หมายถึง พฤติกรรมก้าวร้าว
สามารถรู้ได้จากคนใกล้ชิดผู้ป่วยมาก ๆ เท่านั้น
- **มีพอปรากฏ** หมายถึง ผู้ป่วยพูดเสียงดังเกินไป
หรือ พูดกับคนอื่น ๆ ด้วยกิริยาไม่สุภาพ ไม่เกรงใจ
หรือ รับประทานอาหารด้วยกิริยาไม่เป็นที่ยอมรับ
ทางสังคม
- **มีอย่างชัดเจน** หมายถึง พฤติกรรมโจมตีคนอื่น
ในที่สาธารณะ ทำลายสิ่งของพังพินาศ แสดงออก
บ่อยในสังคมอย่างไม่เหมาะสมแต่ไม่ถึงขั้นอันตราย
(เช่น เปลื้องผ้า ปัสสาวะกลางที่สาธารณะ)
- **มีรุนแรง** หมายถึง พูดจาข่มขู่ หรือทำร้ายร่างกาย
บ่อยครั้ง แต่ไม่ได้เจตนา หรือ ไม่มุ่งหวังให้บาดเจ็บ
รุนแรง
- **มีรุนแรงมาก** หมายถึง พฤติกรรมก้าวร้าว
เกิดขึ้นบ่อยโดยเจตนา และมุ่งหวังให้บาดเจ็บรุนแรง

หมายเหตุ:

- ถ้าพฤติกรรมก้าวร้าวเกิดขึ้นเป็นครั้งคราว การให้คะแนนอาจจะลดลง 1 ระดับ เช่นจากระดับมีรุนแรงเป็นมีอย่างชัดเจน
- กรณีเกิดขึ้นเป็นครั้งคราว หมายถึง เกิดขึ้น ตั้งแต่ 3 ครั้งขึ้นไปใน 1 เดือนที่ผ่านมา หรือ เกิดบ่อยกว่านี้แต่มีหลักฐาน และ/หรือ มีประวัติ
ในอดีตที่บ่งชี้ จนผู้ประเมินเชื่อว่ามีความเสี่ยงต่อการเกิดซ้ำในอนาคตอันใกล้

แนวทางการให้คะแนนในแบบประเมินความสามารถในการทำหน้าที่ส่วนบุคคลและทางสังคม	
100-91	ความสามารถทั้ง 4 ด้านทำได้ดีเยี่ยม ผู้ป่วยได้รับการพิจารณาว่ามีคุณภาพดี ปรับตัวได้อย่างเหมาะสม เมื่อเผชิญกับปัญหาชีวิต มีกิจกรรมและความสนใจหลากหลาย
90-81	ความสามารถทั้ง 4 ด้าน ทำได้ดี มีปัญหาทั่วไปบ้างเล็กน้อย
80-71	ความสามารถบกพร่องเล็กน้อยเพียง 1-2 ข้อ ในชุดข้อ ก-ค
70-61	ความสามารถบกพร่องพอปรากฏ แต่ไม่ถึงขั้นชัดเจน เพียง 1-2 ข้อ ในชุดข้อ ก-ค และมีความบกพร่องเพียงเล็กน้อยในชุดข้อ ง
60-51	ความสามารถบกพร่องอย่างชัดเจนในข้อใดข้อหนึ่งของชุดข้อ ก-ค หรือ ความสามารถบกพร่องพอปรากฏในชุดข้อ ง
50-41	ความสามารถบกพร่องอย่างชัดเจนตั้งแต่ 2 ข้อขึ้นไป ในชุดข้อ ก-ค หรือ ความสามารถบกพร่องรุนแรงในข้อใดข้อหนึ่งของชุดข้อ ก-ค ทั้งนี้อาจรวมกับความบกพร่องอย่างชัดเจนในชุดข้อ ง หรือไม่ได้
40-31	ความสามารถบกพร่องรุนแรงในข้อใดข้อหนึ่ง รวมกับความสามารถบกพร่องอย่างชัดเจน อย่างน้อย 1 ข้อในชุดข้อ ก-ค หรือ มีความบกพร่องอย่างชัดเจนในชุดข้อ ง
30-21	ความสามารถบกพร่องรุนแรงตั้งแต่ 2 ข้อขึ้นไปในชุดข้อ ก-ค หรือ ความบกพร่องรุนแรงในชุดข้อ ง รวมกับความสามารถบกพร่องในชุดข้อ ก-ค หรือไม่ได้
20-11	ความสามารถบกพร่องรุนแรงทั้ง 4 ด้าน (รวมข้อ ก-ง) หรือ มีความบกพร่องรุนแรงมากในชุดข้อ ง รวมกับมีความบกพร่องทั่วไปในชุดข้อ ก-ค ถ้าผู้ป่วยมีปฏิกิริยาต่อสิ่งเร้าภายนอกทันที ควรให้คะแนนระหว่าง 20-16 แต่ถ้าไม่ใช่ คะแนนควรอยู่ระหว่าง 15-11
10-1	ความสามารถขั้นพื้นฐานเสียไปอย่างหมดสิ้นความเป็นตัวของตัวเอง แต่พฤติกรรมไม่เสี่ยงต่อความตาย (ให้คะแนน 6-10) หรือ ถ้าเสี่ยงต่อความตาย เช่น ขาดอาหาร, ขาดน้ำ, ติดเชื้อ, ไม่สามารถรับรู้สถานการณ์ที่เป็นอันตรายได้ (ให้คะแนนระหว่าง 5-1)