



Validation of the UCSD Performance-Based Skills Assessment (UPSA) for functional evaluation of people with severe mental illness in Hong Kong

Clement Chan^{a,*}, Kelvin P.Y. Leung^a, Cycbie C.M. Mok^a, Andrew M.H. Siu^b

^a Occupational Therapy Department, Castle Peak Hospital, Tuen Mun, New Territories, Hong Kong

^b Department of Health Sciences, College of Health, Medicine and Life Sciences, Brunel University of London, United Kingdom

ARTICLE INFO

Keywords:

Severe mental illness
Functional evaluation
Living skills
Assessment
Validation
Occupational therapy
UPSA

ABSTRACT

Background/hypothesis: This study aims to translate and adapt the University of California San Diego Performance-Based Skills Assessment Extended Range (UPSA-2-ER) to Chinese for assessing independent living skills of people with severe mental illness in Hong Kong.

Study design: An expert panel was conducted to evaluate the translation, content validity, and cultural relevance. The UPSA-2-ER (HK) was administered to 30 Hong Kong Chinese with severe mental illness (SMI) to examine the inter-rater reliability and 210 more to examine the internal consistency, structural validity and predictive validity. Sixty-six participants also completed the Chinese Functional Needs Assessment (CFNA) for concurrent validity.

Study results: Good content validity was established with modifications of items to address cultural relevance. Both inter-rater reliability (ICC ranged from 0.98 to 0.99) and internal consistency (Cronbach's $\alpha = 0.97$) of total score were excellent, while internal consistency of sub-domains ranged from fair to excellent (Cronbach's α ranged from 0.76 to 0.96). In Rasch analysis, raw variance explained by measures $\geq 40\%$ was interpreted as good in terms of unidimensionality (Sumintono & Widhiarso, 2015), study results provided modest evidence that the UPSA-2-ER(HK) has a unidimensional structure, as the variance explained by the items was 43.5% and the unexplained variance in the 1st contrast was 4.4%. Classification & regression tree analysis showed that the UPSA-2-ER(HK) correctly classified 78.5% of the 210 participants into their residential status of independent living, halfway house or long stay care home. For concurrent validity, the UPSA-2-ER(HK) demonstrated high correlation with the local golden standard, CFNA ($r = 0.82$).

Conclusions: The results showed that the instrument has satisfactory psychometric properties for assessing capacity to perform community living tasks.

1. Introduction

Severe mental illness (SMI) covers a range of mental disorders like schizophrenia spectrum disorders, bipolar disorder, and major depression, which often lead to significant dysfunction in areas of activities of daily living (ADL), work adjustment, and social skills (Zimmerman et al., 2018). Among these areas, dysfunction in activities of daily living is one of the most common barriers to independent living and community resettlement. Among people with schizophrenia, residual symptoms and neurocognitive deficits can significantly impact on one's performance on everyday functional tasks like use of transportation, shopping and managing money (Sin et al., 2022; Strassnig et al., 2015). Similarly for people with bipolar affective disorder, both clinical symptoms and

cognitive performance interact to influence functional impairment and quality of life (Chen et al., 2019). Living or life skills is commonly regarded as an outcome indicator of recovery, which encompass the practical abilities needed to manage daily life, maintain independence, and participate fully in the community. These skills include personal care, household management, financial management, and navigating social situations (Tungpunkom et al., 2012).

Comprehensive assessment of living skills is often conducted by mental health teams to monitor progress of recovery, as well as referring clients to a suitable type of supported housing (Fortuna et al., 2018). The supported housing approach suggested that people with SMI could gradually progress from hospital care, supervised accommodations (e.g. halfway houses, group homes), and then independent living (Mausbach

* Corresponding author at: Occupational Therapy Department, G/F, Block F, Castle Peak Hospital, 15 Tsing Chung Koon Road, Tuen Mun, NT, Hong Kong.
E-mail address: cc943@ha.org.hk (C. Chan).

<https://doi.org/10.1016/j.actpsy.2026.107430>

Received 2 September 2025; Received in revised form 20 May 2026; Accepted 7 July 2026

Available online 11 July 2026

0001-6918/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

et al., 2016; Richter & Hoffmann, 2017). People with SMI are expected to relearn skills and habits for independent living in the community and become more autonomous and empowered in the recovery process (Tungpunkom et al., 2012).

In Hong Kong (HK), Half-way Houses (HWHs) and Long Stay Care Homes (LSCHs) are the two standard supported housing services funded by the government for individuals with SMI (Social Welfare Department, The Government of the Hong Kong Administrative Region, 2023). They form a substantial portion of the local residential landscape for persons in mental recovery. Admission to these facilities is governed by a clinical triage system grounded on functional independence. For those who cannot reside at home with their carer or live independently, HWHs serve as transitional period of residential care to facilitate residents with higher functioning to return to employment and pursue independent living thereafter. Whereas LSCHs provide long-term maintenance care for those with chronic conditions and lower functional capacity. HWHs and LSCHs require lots of resource to run with professional support, and there is long waitlist for these supports (Cui & Wu, 2025; Research Office, Legislative Council Secretariat, 2023). There is a strong need to develop standardised tools to assess the functional capacity of clients and select the most suitable residential placement for them in a triage system.

While there were some attempts to develop and validate living skills assessment in Hong Kong (Au et al., 2005), the Chinese Functional Needs Assessment (CFNA) is regarded as the standard instrument for assessment of independent living skills in occupational therapy practice. Translated from the Functional Needs Assessment (FNA) (Dombrowski, 1990), the CFNA was designed to comprehensively review patient's functional capacity for independent living. However, there was no further research on the CFNA, except the validation conducted in a Master's thesis (Law, 1999). There is a growing need to develop an updated assessment of independent living in Hong Kong. The CFNA is mainly rated by assessor based on their daily observation of clients, as well as patient self-report on tasks of independent living. It does not cover some important areas of Instrumental Activities of Daily Living (IADL) like medication management, activity planning, or paying bills. To select a suitable and updated instrument for living skills assessment, the new instrument should: (1) Cover the assessment of ability to conduct everyday tasks, (2) Designed for assessment of people with severe mental illness, (3) Be a performance-based assessment, requiring clients to demonstrate their skills instead of relying mainly on self-report, (4) Has been validated and standardised for use, (5) Contribute to the recommendation of supported housing for the client with SMI.

According to these selection criteria, we reviewed several performance-based living skills assessments, including the Independent Living Scale (ILS; Loeb, 1996; Revheim & Medalia, 2004), the Canadian Objective Assessment of Life Skills (COALS; McDermid Vaz et al., 2013), and the University of California San Diego Performance-Based Skills Assessment in its Brief (UPSA-B; Mausbach et al., 2007) and Extended Range (UPSA-2ER; Haig et al., 2016; Patterson et al., 2001) versions (Moore et al., 2007; Revheim, 2024; Tungpunkom et al., 2012). Among these, the ILS was originally developed for geriatric and brain injury rather than for people with SMI. The COALS is an emerging living skills assessment, whereas the UPSA family are well validated and has been successfully adapted to different languages to use in different cultures (Garcia-Portilla et al., 2013; Kim et al., 2017; Kim et al., 2022; McIntosh et al., 2011; Olsson et al., 2012; Sumiyoshi et al., 2014). The cut-off scores established in some UPSA studies indicated that the instrument can effectively differentiate between patients who can live independently or require supported housing. The UPSA-B offers the advantage of administrative efficiency. However, our objective required more than a quick screening instrument. The UPSA-2ER is a full version assessment covering six domains of daily living skills. This allow clinicians to pinpoint specific skill deficits, which is essential for planning interventions and support for developing independence among people with SMI. Thus, the UPSA-2ER was selected as the chosen measurement

to validate and adapt in Hong Kong.

The objectives of this study are: 1) Translate and adapt UPSA-2ER for use with a Chinese population in Hong Kong, 2) Examine the content, structural and concurrent validity of the instrument, 3) Evaluate the internal consistency and inter-rater reliability of the instrument, 4) Examine the predictive validity of the instrument by using the UPSA scores to differentiate the functional ability of people with severe mental illness who are using different residential services.

2. Method

It is a cross-sectional validation study.

2.1. Participants

Two groups of participants subjects were recruited. Thirty participants were recruited using purposive sampling to examine the inter-rater reliability. A larger group of 210 participants were recruited using quota sampling for study of internal consistency, concurrent, structural, and predictive validity. A quota sample of 70 participants was recruited from three types of residential settings for people with mental illness, i.e., independent living, half-way house, and long stay care home. Quota sampling was chosen to ensure there are enough participants for comparison across groups.

A screening of potential participants was conducted based the inclusion criteria that they are: (1) Age between 18 and 64, (2) Diagnosed with severe mental illness, (3) Able to understand and follow verbal tests instruction for assessment. The exclusion criteria are: 1) History of organic brain disorders, substance abuse, 2) Diagnosis of intellectual disabilities, autistic spectrum disorders, 3) Inability to understand written or spoken Chinese.

2.2. Sample size

For study of inter-rater reliability, a sample size of 30 was required to reach the statistical power of 0.9 to detect an Intra-class Correlation Coefficient (ICC) of 0.8 versus the alternative hypothesis that ICC is 0.5 assuming p is below 0.05. The sample size was estimated using PASS12 (Hintze, 2013).

For study of structural validity, Rasch analysis would require the largest number of participants. As suggested by COSMIN study design checklist, a sample larger than 100 is considered sufficient in estimating internal consistency, and sample size of more than 200 is considered as "very good" in conducting analysis using the Rasch model (Mokkink et al., 2018).

For concurrent validity, we assume a moderate to high correlation ($r = 0.6$ to 0.8) between UPSA-2ER(HK) and the concurrent measure of Chinese Functional Needs Assessment (CFNA). Assuming α is 0.05, a sample size of 66 is adequate to achieve a statistical power of 0.8.

2.3. Instruments

2.3.1. The UCSD Performance-based Skills Assessment Extended Range (UPSA-2ER)

The UPSA-2ER is a performance-based measurement developed for assessing the capacity to perform community living tasks by persons with schizophrenia through role play. Assessment items are organized under the six domains of household chores, communication, finance, transportation, planning recreational activities and medication management. Raw scores of all domains are converted to a score between 0 and 20, and the domains scores are added up to give a total score ranging from 0 to 120 points (higher scores indicate better performance). The assessment takes approximately 45 min to administer. A systematic review of the UPSA summarized that the assessment tool has good psychometric properties across different language translations (Becattini-Oliveira et al., 2018). Test-retest reliability for one- and six-

Phase 1: inter-rater reliability

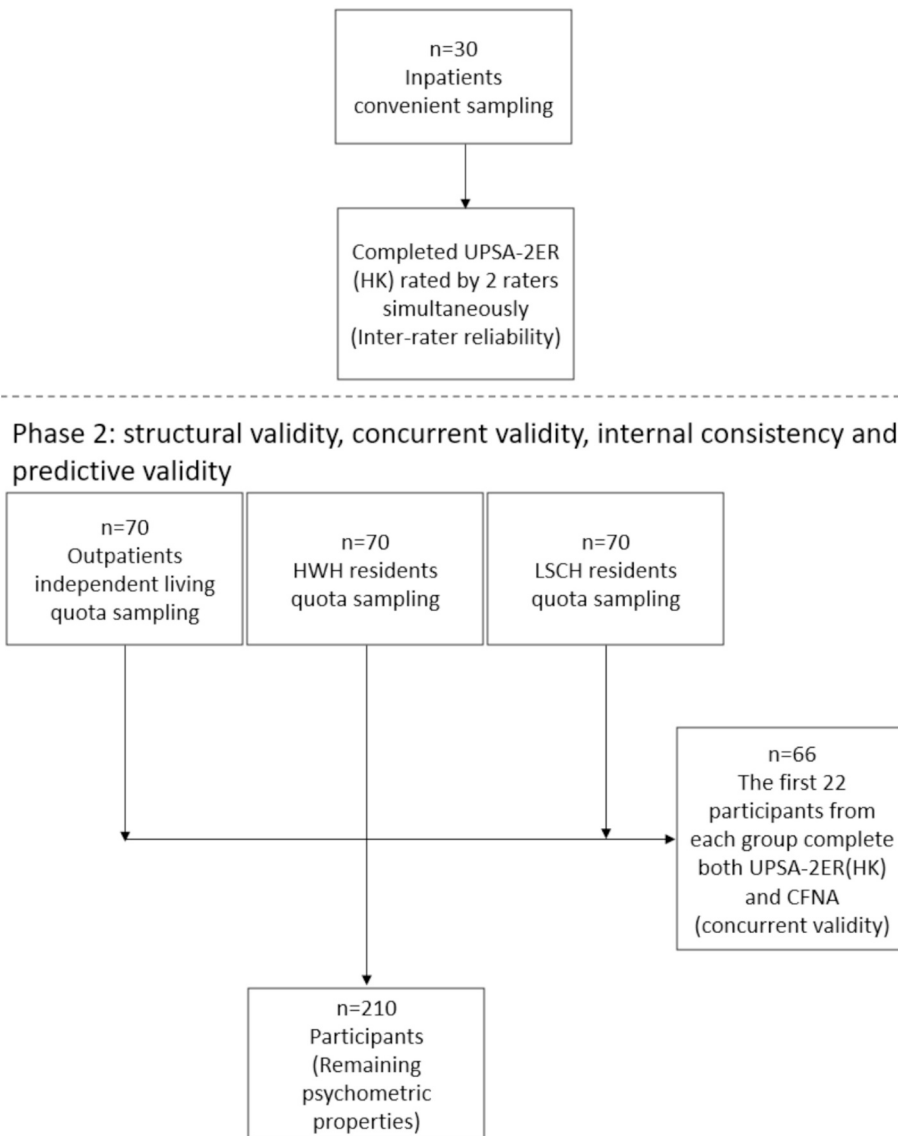


Fig. 1. Flow chart of subject recruitment.

Table 1

Inter-rater reliability, internal consistency and concurrent validity of the UPSA-2ER(HK).

UPSA-2ER(HK) sub-domains	ICC	Cronbach's α
Planning and organization	0.99	0.86
Finance	0.99	0.81
Communication	0.99	0.79
Transportation	0.99	0.88
Household management	0.99	0.76
Medication management	0.99	0.96
Total score	0.99	0.97

month intervals was acceptable ($r = 0.75$). In study of convergent validity, the UPSA showed a moderate to strong correlation with the MATRICS Consensus Cognitive Battery (MCCB), ($r = 0.65$ to 0.79). For concurrent validity, the instrument has high correlation with the Direct Assessment of Functional Status ($r = 0.86$) (Patterson et al., 2001). It achieved 69% sensitivity and 66% specificity when a cut-off score of 75 is used in differentiating individuals with severe mental illness who live

independently or residing in a care facility (Mausbach et al., 2008).

2.3.2. Chinese Functional Needs Assessment (CFNA)

The CFNA was translated and validated by Law (1999) and was widely used by occupational therapists to evaluate community living skills and functional capacity of people who with schizophrenia in Hong Kong. The CFNA requires the examinee to demonstrate the ability to perform certain tasks in response to instructions. The 130 assessment items were organized under 26 subscales, and the subscales were categorised under the two domains of self-care and community living. Each correct item rates a score of 5 and the total score of the 26 subscales could add up to a potential 650. The administration of the assessment takes approximately 45 min. The CFNA was regarded as the golden standard of functional assessment for people with schizophrenia in Hong Kong (Au et al., 2005). The CFNA subscales had moderate to excellent inter-rater reliability ($r = 0.65$ – 0.99). A two-factor structure was identified using factor analysis, and its internal consistency was high (Cronbach's $\alpha = 0.92$). Its scores could classify participants in independent living, HWH and LSCH with a 70% overall accuracy.

Table 2
Demographic characteristics of participants.

Variables ^{a,b}	All groups n = 210	IL n = 70	HWH n = 70	LSCH n = 70	p	Post-hoc analysis ^c		
Gender								
Male	135 (64.3%)	33 (47.1%)	52 (74.3%)	50 (71.4%)	0.001	Males are more likely to be in HWH and LSCH settings than IL.		
Female	75 (35.7%)	37 (52.9%)	18 (25.7%)	20 (28.6%)				
Age								
Mean (SD)	43.3 (12.3)	34.0 (7.5)	40.1 (9.4)	55.9 (7.4)	<0.001	Mean age is lowest in IL, followed by HWH, & LSCH.		
Diagnosis								
Schizophrenia spectrum disorder	189 (90.0%)	64 (91.4%)	61 (87.1%)	64 (91.4%)	0.56	–		
Bipolar affective disorder	15 (7.1%)	5 (7.1%)	5 (7.1%)	5 (7.1%)				
Major depression with psychotic symptoms	3 (1.4%)	1 (1.4%)	2 (2.9%)	0 (0.0%)	< 0.001	Participants with tertiary education are more likely in IL, while those with primary & junior secondary education are more likely in LSCH.		
Others	3 (1.4%)	0 (0.0%)	2 (2.9%)	1 (1.4%)				
Education level								
Primary 1–6	11 (5.2%)	0 (0.0%)	2 (2.9%)	9 (12.9%)	< 0.001	Participants with tertiary education are more likely in IL, while those with primary & junior secondary education are more likely in LSCH.		
Secondary 1–3	49 (23.3%)	4 (5.7%)	18 (25.7%)	27 (38.6%)				
Secondary 4–6	107 (51.0%)	36 (51.4%)	40 (57.1%)	31 (44.3%)				
Tertiary or above	43 (20.5%)	30 (42.9%)	10 (14.3%)	3 (4.3%)	<0.001	Participants in open or supported employed are more likely in IL or HWH. Unemployed participants are mostly in LSCH		
Employment status								
Open employment	61 (29.0%)	40 (57.1%)	21 (30.0%)	0 (0.0%)				
Supported employment	8 (3.8%)	4 (5.7%)	4 (5.7%)	0 (0.0%)				
Sheltered workshop	19 (9.0%)	0 (0.0%)	19 (27.1%)	0 (0.0%)				
Unemployed	113 (53.8%)	18 (25.7%)	25 (35.7%)	70 (100.0%)				
Study	6 (2.9%)	6 (8.6%)	0 (0.0%)	0 (0.0%)				
Others	3 (1.4%)	2 (2.9%)	1 (1.4%)	0 (0.0%)				

Notes. IL = Independent Living, HWH = Half-way House, LSCH = Long Stay Care Home.

^a Chi-square test.

^b ANOVA.

^c Post-hoc analysis using Bonferroni tests.

2.4. Procedures

The research protocol was approved by the Research Ethics Committee of the New Territories West Cluster, Hospital Authority (CREC ref. No: NTWC/REC/19087). Permission for translation and validation of a Hong Kong version was granted by NeuroCog Trials and Vera Science (former and current copyright holder). The original English version of the UPSA-2ER, including its administration and scoring manual, appendices, and props, was translated into traditional Chinese using an idiomatic approach by a bilingual translator holding a doctorate degree in translation. Following this, the translated version was comprehensively reviewed and synthesized jointly by the researcher and the translator to ensure the translated dialogues are natural and suitable for the Hong Kong population. An expert panel consisting of 5 bilingual occupational therapists then appraised the quality, equivalence, and clarity of the translation. The panel members all had a minimum of 8 years of clinical experience in psychiatric rehabilitation. Modifications were made according to the comments and suggestions from the panel members.

The revised version of the translated instrument was then reviewed for its content validity by an expert panel of 7 occupational therapists, each with at least 8 years of experience in psychiatric rehabilitation. This panel evaluated the content relevance and representativeness of

each assessment item. Panel members then met in a focus group to discuss and reach a consensus on the modifications needed to improve content validity or cultural relevance. The revised Hong Kong Chinese version of the UPSA-2ER was subsequently submitted to the copyright owner for review and approval of the translation and modifications. With the approval of the copyright owner, this version was officially named the UPSA-2ER adapted for Hong Kong (UPSA-2ER(HK)).

To examine the inter-rater reliability, 30 subjects were recruited from in-patients of a mental hospital. In every assessment session, one rater administered the UPSA-2ER(HK) while the other observes the performance. The participants' performance was rated by the two raters independently.

A total of 210 subjects were recruited for the study of structural validity, internal consistency, predictive validity, a quota sample of 70 participants were recruited from those who are living independently, staying in halfway house, or long-stay care home respectively. Sixty six out of 210 participants joined the study of concurrent validity, 22 participants from each group i.e., independent living, half-way house, and long stay care home, completed both the UPSA-2ER(HK) and CFNA (Fig. 1). Assessments of UPSA-2ER and CFNA were completed by two researchers and trained assistants. Training sessions were arranged to ensure the fidelity and consistency in administration and research procedures.

Table 3
Rasch item analysis of the UPSA-2ER(HK).

Item	Total score	Measure	Model SE	INFIT		OUTFIT	
				MNSQ	ZSTD	MNSQ	ZSTD
1	139	0.01	0.18	1.13	1.4	1.2	1.1
2	90	1.4	0.17	1.29	3.8	1.78	3.4
3	66	2.08	0.17	1.16	1.9	1.74	2.6
4	171	-1.18	0.21	1.32	2.3	1.52	1.5
5	132	0.22	0.17	1.12	1.4	1.12	0.7
6	86	1.51	0.17	1.08	1.1	1.24	1.2
7	42	2.89	0.2	1.1	0.9	1.16	0.6
8	199	-3.13	0.35	1	0.1	0.84	0
9	191	-2.35	0.28	1.16	0.9	0.69	-0.5
10	182	-1.74	0.24	1.17	1.1	0.8	-0.4
11	162	-0.79	0.2	0.98	-0.1	0.82	-0.6
12	142	-0.08	0.18	0.89	-1.3	0.87	-0.6
13	129	0.31	0.17	0.82	-2.3	0.69	-2.1
14	110	0.85	0.17	0.92	-1.1	0.83	-1.1
15	123	0.49	0.17	1.2	2.4	1.34	2
16	69	1.99	0.17	0.98	-0.2	0.85	-0.6
17	33	3.26	0.21	1.1	0.8	0.88	-0.2
18	67	2.05	0.17	1.3	3.6	2.69	5
19	178	-1.52	0.23	1.26	1.8	1.14	0.5
20	189	-2.2	0.27	0.94	-0.3	0.71	-0.5
21	189	-2.2	0.27	1.13	0.7	1.48	1
22	107	0.93	0.17	1.3	3.8	1.53	2.9
23	129	0.31	0.17	1.12	1.5	1.13	0.8
24	37	3.09	0.2	1.05	0.5	0.9	-0.2
25	182	-1.74	0.24	1.03	0.3	0.67	-0.8
26	190	-2.27	0.28	1.08	0.5	1.17	0.5
27	145	-0.18	0.18	1.11	1.2	1.5	2.3
28	135	0.13	0.17	1.02	0.3	1.17	1
29	136	0.1	0.18	1.29	3.1	1.45	2.3
30	187	-2.06	0.26	0.85	-0.9	0.42	-1.4
31	176	-1.42	0.23	1.03	0.3	0.85	-0.3
32	192	-2.43	0.29	1.12	0.7	0.93	0.1
33	65	2.11	0.17	1.7	7.2	2.1	3.5
34	62	2.21	0.18	1.15	1.8	1.31	1.2
35	183	-1.8	0.25	1.32	1.9	1.24	0.7
36	175	-1.37	0.22	0.83	-1.3	0.65	-1
37	119	0.6	0.17	0.9	-1.3	0.8	-1.3
38	102	1.07	0.17	0.93	-1	0.93	-0.4
39	93	1.32	0.17	0.83	-2.5	0.7	-1.8
40	47	2.7	0.19	0.99	-0.1	0.78	-0.7
41	171	-1.18	0.21	1.3	2.2	1.64	1.8
42	189	-2.2	0.27	1.25	1.4	1.79	1.5
43	112	0.8	0.17	1.02	0.3	1.57	3.1
44	105	0.99	0.17	0.87	-1.8	0.77	-1.5
45	136	0.1	0.18	0.8	-2.4	0.67	-2.1
46	163	-0.83	0.2	1.06	0.5	1.03	0.2
47	196	-2.8	0.32	0.87	-0.6	0.67	-0.4
48	95	1.26	0.17	0.86	-2	0.76	-1.4
49	92	1.34	0.17	0.89	-1.6	0.8	-1.1
50	138	0.04	0.18	0.77	-2.8	0.69	-1.9
51	80	1.68	0.17	0.82	-2.6	0.68	-1.6
52	83	1.59	0.17	0.86	-2.1	0.73	-1.4
53	42	2.89	0.2	0.75	-2.5	0.5	-1.8
54	53	2.5	0.18	0.84	-1.8	0.7	-1.1
55	166	-0.96	0.21	0.9	-0.8	0.72	-1
56	153	-0.46	0.19	1.35	3.1	1.72	2.8
57	144	-0.15	0.18	0.93	-0.7	0.89	-0.5
58	124	0.46	0.17	1.18	2.2	1.53	2.9
59	126	0.4	0.17	0.98	-0.2	1	0.1
60	65	2.11	0.17	1.13	1.6	1.96	3.2
61	143	-0.12	0.18	1.13	1.4	1.03	0.2
62	171	-1.18	0.21	1.09	0.7	1.04	0.2
63	70	1.96	0.17	1.14	1.7	1.17	0.8
64	161	-0.75	0.2	1.28	2.3	1	0.1
65	159	-0.68	0.2	1.31	2.6	1.04	0.2
66	58	2.33	0.18	1.04	0.5	1.25	1
67	195	-2.7	0.31	0.73	-1.4	0.25	-1.7
68	178	-1.52	0.23	0.93	-0.4	1.17	0.6
69	146	-0.22	0.18	1.3	2.9	1.57	2.5
70	159	-0.68	0.2	1	0	0.89	-0.4
71	151	-0.39	0.19	1.02	0.2	0.95	-0.2
72	124	0.46	0.17	1.18	2.1	1.2	1.2
73	134	0.16	0.17	1.06	0.7	1.09	0.6

Table 3 (continued)

Item	Total score	Measure	Model SE	INFIT		OUTFIT	
				MNSQ	ZSTD	MNSQ	ZSTD
74	146	-0.22	0.18	1.26	2.5	1.3	1.5
75	190	-2.27	0.28	0.68	-1.9	0.28	-1.9
76	175	-1.37	0.22	0.77	-1.8	0.82	-0.4
77	173	-1.27	0.22	0.74	-2.1	0.5	-1.7
78	158	-0.64	0.19	0.86	-1.3	0.84	-0.6
79	147	-0.25	0.18	0.75	-2.8	0.6	-2.3
80	136	0.1	0.18	0.85	-1.8	0.81	-1.1
81	135	0.13	0.17	0.85	-1.8	0.81	-1.1
82	191	-2.35	0.28	0.61	-2.4	0.24	-2
83	177	-1.47	0.23	0.86	-1	1.16	0.6
84	107	0.93	0.17	1.09	1.2	1.1	0.6
85	174	-1.32	0.22	0.61	-3.3	0.37	-2.3
86	166	-0.96	0.21	0.74	-2.3	0.66	-1.3
87	92	1.34	0.17	1.02	0.4	0.97	-0.1
88	145	-0.18	0.18	0.9	-1	0.86	-0.7
89	140	-0.02	0.18	0.94	-0.7	0.9	-0.5
90	87	1.48	0.17	0.91	-1.3	0.83	-0.9
91	127	0.37	0.17	1.09	1.2	0.99	0
92	192	-2.43	0.29	0.72	-1.5	0.25	-1.9
93	176	-1.42	0.23	0.8	-1.5	0.99	0.1
94	93	1.32	0.17	1.02	0.4	1.64	3
95	173	-1.27	0.22	0.61	-3.3	0.4	-2.2
96	165	-0.91	0.2	0.74	-2.4	0.68	-1.2
97	91	1.37	0.17	0.88	-1.8	0.73	-1.6
98	156	-0.56	0.19	0.68	-3.4	0.53	-2.4
99	146	-0.22	0.18	0.81	-2.1	0.67	-1.8
100	96	1.24	0.17	0.79	-3.2	0.65	-2.2
101	123	0.49	0.17	0.74	-3.5	0.61	-2.8
102	119	0.6	0.17	0.76	-3.5	0.62	-2.8
103	91	1.37	0.17	0.82	-2.7	0.69	-1.8
104	110	0.85	0.17	0.83	-2.4	0.71	-1.9
Mean	133	0	0.2	0.99	-0.1	0.98	-0.1
SD	44	1.51	0.04	0.2	2	0.42	1.6

Note. Items in bold are those which has an INFIT or OUTFIT values outside of acceptable range of 0.5 to 1.5.

2.5. Data analysis

Content validity was examined using mode index and percentage of agreement. Inter-rater reliability evaluated using the two-way random model of intra-class correlation coefficient (Koo & Li, 2016). Internal consistency was examined using Cronbach's α . Concurrent validity was examined by calculating correlation with the CFNA which measures the similar construct of independent living skills. The above analyses were conducted using IBM® SPSS® version 25.

Rasch analysis was conducted using WINSTEPS V3.73 (Linacre, 2009). We conducted the Principal Component Analysis of Residuals (PCAR) for exploration of dimensionality; estimated the INFIT mean-square (INFIT) and OUTFIT mean-square (OUTFIT) statistics, and finally the Wright map was produced. INFIT mean inlier-sensitive or information-weighted fit while OUTFIT refers to outlier-sensitive fit; in general INFIT reports underfitting or noises and OUTFIT reports overfitting for imputed responses; items with mean-square value ranged from 0.5 to 1.5 are regarded as productive for measurement, those outside of this range can be regarded as unproductive or distorting the measurement system (Linacre, 2002).

We used the Classification and Regression Tree (C&RT) analysis for study of predictive validity. The six sub-domain scores were used as predictors of functional level of participants, which were indicated by the residential services they are using. The C&RT programme is a module of IBM® SPSS® software version 25. It could determine the relative importance of sub-domains in predicting functional level, and to estimate cut-off scores for classifying subjects to the 3 groups. The decision tree would grow until all subjects are classified into a predicted subgroup of functional level. A series of classification criteria (cut-off scores) were generated after a resulting tree structure was formed, which could be adopted to predict the functional level of an individual.

Table 4
Misfit items of the UPSA-2ER(HK).

Subdomain	Item	INFIT		OUTFIT		Reason of retention
		MNSQ	ZSTD	MNSQ	ZSTD	
Planning and organization	2	1.29	3.8	1.78	3.4	Delete item 2 would affect overall difficulty and responses of item 8–14
	3	1.16	1.9	1.74	2.6	Delete item 3 would affect overall difficulty and responses of item 8–14
	4	1.32	2.3	1.52	1.5	Cannot administer item 5, 6 and 7 without item 4
	18	1.3	3.6	2.69	5	Delete item 18 would affect administration and prompts of item 8–14
Finance	22	1.3	3.8	1.53	2.9	Item 22 is the only item assessing change calculation, deleting this item would affect overall difficulty of this subdomain
	30	0.85	−0.9	0.42	−1.4	Delete item 30 would affect administration and prompts of item 29 and 31
Communication	33	1.7	7.2	2.1	3.5	Cannot administer item 34 without item 33
	41	1.3	2.2	1.64	1.8	Delete item 41 would affect administration and prompts of item 36–40
	42	1.25	1.4	1.79	1.5	Delete item 42 would affect administration and prompts of item 36–40
	43	1.02	0.3	1.57	3.1	Delete item 43 would affect administration and prompts of item 36–40
Household management	56	1.35	3.1	1.72	2.8	Cannot administer item 55 and 57 without item 56
	58	1.18	2.2	1.53	2.9	Cannot administer item 55 and 57 without item 58
	60	1.13	1.6	1.96	3.2	Cannot administer item 61–63 without item 60
Medication Management	67	0.73	−1.4	0.25	−1.7	Cannot administer item 67–74 without item 67
	69	1.3	2.9	1.57	2.5	Cannot administer item 67–74 without item 69
	75	0.68	−1.9	0.28	−1.9	Cannot administer item 75–81 without item 75
	82	0.61	−2.4	0.24	−2	Cannot administer item 82–91 without item 82
	85	0.61	−3.3	0.37	−2.3	Cannot administer item 82–91 without item 85
	92	0.72	−1.5	0.25	−1.9	Cannot administer item 92–104 without item 92
	94	1.02	0.4	1.64	3	Cannot administer item 92–104 without item 94
	95	0.61	−3.3	0.4	−2.2	Cannot administer item 92–104 without item 95

The default GINI method was used, and the maximum tree depth was set at four. The minimum numbers of cases in the parent node and child node were set at ten and one respectively. Pruning was applied based on the minimum error rule, aimed at developing a more complexed classification and regression tree with higher classification accuracy.

3. Results

3.1. Content validity and cultural adaptation

The expert panel commented that the sub-domains and items of the UPSA-2ER(HK) were relevant and representative to functional capacity of people with severe mental illness in Hong Kong. The emic approach was adopted to identify and adapt the cultural-specific facets in the UPSA-2-ER for use in the Hong Kong context (Squires et al., 2013). There was no need to remove any items, but the panel suggested many revisions due to cultural adaptation. There were changes in the sub-domains of planning and organization, finance, community skills, transportation, household management, and medication management. For example, it is necessary to change the currency from US dollars to Hong Kong dollars in the finance sub-domain, or the change of transportation means or underground routes in the Transportation sub-domain. In the Medication management, there is a need to replace the labels for medications, or notices for medical appointments.

3.2. Reliability

Among the 30 participants, sixteen (53%) were male and 14 (47%) were female. Their age ranged from 23 to 59 ($M = 43.17$, $SD = 11.85$). The mean year of education is 9.93 ($SD = 2.63$), with a range from 4 to 17 years. All the participants are in-patients and not working.

Inter-rater reliability was estimated using intra-class correlation coefficients (two-way mixed model), computed between the ratings made by the two raters for the 6 sub-domains. Across all the sub-domains, intra-class correlation coefficients were very high (0.99) indicating excellent inter-rater reliability (Table 1). The instrument also demonstrated excellent internal consistency, with Cronbach's $\alpha = 0.97$ for the whole scale. The sub-domains have moderate to excellent internal consistency (Cronbach's α range from 0.76 to 0.9) (Table 1).

3.3. Study of concurrent, structural, and predictive validity

Two hundred and ten participants were recruited for the study of validity (Table 2), 135 (64%) were male, 75 (36%) were female. They have a mean age of 43.3 ($SD = 12.3$), with a range of 18 to 64. Most (90.0%) participants had a diagnosis of schizophrenia spectrum disorder (covering schizophrenia and schizoaffective disorder). The majority (94.8%) received a Grade 6 education or above. When the demographic data were compared across the three groups, the LSCH group was significantly older than the half-way house (HWH) group; and the HWH group was significantly older than the independent living group ($F = 134.49$, $p < 0.001$). The Independent Living group has the highest rate (62.8%) of employment (including open and supported employment, followed by an employment rate of 32.8% in the HWH group. All participants in the LSCH group subjects were not working.

3.3.1. Concurrent validity

The UPSA-2ER(HK) total score had high correlation ($r = 0.82$) with the concurrent measure of CFNA total score (Table 1).

3.3.2. Structural validity

We conducted a Rasch item analysis of the UPSA-2ER(HK) using the WINSTEPS program (Linacre, 2009). INFIT and OUTFIT mean-square statistics were performed to examine the fitness of the 104 items with the Rasch rating scale model (Table 3). We found that 21 items (Table 4) did not fall into the acceptable range of INFIT and OUTFIT mean-square values (0.5 to 1.5) (Linacre, 2002). Item 2, 3, 4 and 18 in the planning and organization sub-domain, item 22 in the finance sub-domain, item 41, 42 and 43 in the communication sub-domain, item 56, 58 and 60 in the household management sub-domain, as well as item 69 and 94 in the medication management sub-domain attained OUTFIT mean-square values above 1.5, ranged from 1.52 to 2.69. Item 30 in the finance sub-domain, item 67, 75, 82, 85, 92 and 95 in the medication management sub-domain attained OUTFIT mean-square values below 1.5, ranged from 0.24 to 0.42. Item 33 in the communication sub-domain was the only item with both INFIT and OUTFIT mean-square values fell outside the range of 0.5–1.5, with INFIT = 1.7 and OUTFIT = 2.1. Differential item functioning (DIF) by diagnosis was completed (Table 5) (Fig. 5).

We also conducted principal component analysis of residuals (PCAR) to examine the dimensionality of the UPSA-2ER(HK). The percentage of

Table 5
Differential item functioning (DIF) of the UPSA-2ER(HK) by diagnosis.

DIF Measure (diff.)	PERSON DIF CLASS			
Item	1	2	3	*
1	0.48	0.08	-0.36	0.01
2	1.61	1.97	0.38	1.4
3	2.38	2.13	1.44	2.08
4	-0.4	-0.93	-1.61	-1.18
5	0.22	0.65	-0.28	0.22
6	1.47	1.97	0.91	1.51
7	2.87	3.3	2.28	2.89
8	-1.64	-2.44	-3.69	-3.13
9	-3.4	-2.87	-2.1	-2.35
10	-3.31	-1.33	-1.8	-1.74
11	-1.2	-0.71	-0.76	-0.79
12	-1.2	0	0.21	-0.08
13	-1.61	0.31	1.11	0.31
14	0.1	1.05	1.21	0.85
15	1.06	0.37	0.13	0.49
16	2.04	2.13	1.68	1.99
17	3.16	3.15	4.46	3.26
18	2.65	1.81	1.32	2.05
19	-2.38	-1.67	-1.34	-1.52
20	-1.64	-3.59	-2	-2.2
21	-1.64	-2.87	-2.1	-2.2
22	1.54	0.85	0.38	0.93
23	0.24	0.65	-0.04	0.31
24	3.48	2.68	2.7	3.09
25	-2.38	-1.88	-1.61	-1.74
26	-2.38	-2.44	-2.21	-2.27
27	0.24	0	-0.6	-0.18
28	0.36	0	0.13	0.13
29	0.69	0.44	-0.68	0.1
30	-3.35	-2.87	-1.7	-2.06
31	-2.38	-1.49	-1.26	-1.42
32	-2.38	-2.43	-2.43	-2.43
33	3.66	1.19	1.11	2.11
34	2.65	1.97	1.68	2.21
35	-2.38	-1.05	-2.21	-1.8
36	-1.64	-2.13	-1	-1.37
37	0.48	0.6	0.73	0.6
38	0.48	1.25	1.44	1.07
39	0.24	1.74	1.96	1.32
40	2.79	2.3	3.73	2.7
41	-0.87	-1.05	-1.34	-1.18
42	-1.2	-1.88	-2.56	-2.2
43	0.24	0.72	1.44	0.8
44	0.48	1.19	1.21	0.99
45	-1.2	-0.24	1.01	0.1
46	-0.87	-0.83	-0.83	-0.83
47	-3.43	-2.87	-2.69	-2.8
48	0.79	1.32	1.81	1.26
49	0.98	1.39	1.81	1.34
50	-0.87	-0.16	0.64	0.04
51	1.15	1.68	2.96	1.68
52	1.15	1.81	1.96	1.59
53	2.52	3.3	3.73	2.89
54	2.38	2.21	4.43	2.5
55	-2.37	-1.49	-0.44	-0.96
56	0.48	-0.42	-1	-0.46
57	-0.05	-0.61	0.21	-0.15
58	0.69	0.46	0.29	0.46
59	0.59	0.3	0.38	0.4
60	2.17	2.3	1.68	2.11
61	-0.4	0	-0.12	-0.12
62	-1.2	-1.33	-1.09	-1.18
63	1.96	2.05	1.81	1.96
64	-0.87	-0.08	-1.34	-0.75
65	-0.87	0	-1.26	-0.68
66	2.31	2.58	1.96	2.33
67	-3.42	-3.59	-2.44	-2.7
68	-0.87	-2.13	-1.43	-1.52
69	0.59	-0.24	-0.68	-0.22
70	-0.4	-0.42	-1	-0.68
71	0.1	-0.24	-0.76	-0.39
72	0.89	0.65	-0.12	0.46
73	0.24	0.08	0.21	0.16

Table 5 (continued)

DIF Measure (diff.)	PERSON DIF CLASS			
Item	1	2	3	*
74	-0.22	-0.24	-0.22	-0.22
75	-3.39	-3.59	-1.9	-2.27
76	-1.2	-2.87	-0.92	-1.37
77	-1.64	-2.13	-0.84	-1.27
78	-0.05	-1.33	-0.44	-0.64
79	-0.62	-0.51	0.13	-0.25
80	0.1	-0.16	0.38	0.1
81	0.1	-0.24	0.55	0.13
82	-3.4	-3.59	-2	-2.35
83	-0.87	-1.49	-1.61	-1.47
84	0.89	1.12	0.73	0.93
85	-2.38	-1.49	-1.09	-1.32
86	-0.62	-1.05	-1	-0.96
87	1.31	1.39	1.32	1.34
88	-0.05	0.16	-0.6	-0.18
89	0.24	0.37	-0.6	-0.02
90	1.31	1.81	1.21	1.48
91	0.69	0.58	-0.12	0.37
92	-3.4	-3.59	-2.1	-2.43
93	-0.62	-2.44	-1.26	-1.42
94	0.89	1.6	1.44	1.32
95	-1.64	-1.88	-0.92	-1.27
96	-0.62	-1.33	-0.76	-0.91
97	1.15	1.19	2.11	1.37
98	-0.62	-0.71	-0.44	-0.56
99	0.24	-0.42	-0.28	-0.22
100	0.59	1.53	1.55	1.24
101	0.49	0.37	0.64	0.49
102	0.69	0.44	0.73	0.6
103	1.23	1.37	1.55	1.37
104	1.06	0.65	0.91	0.85

variance explained by the measures was 43.5%. This was almost 10 times of the unexplained variance in the first contrast, which was only 4.4%. The results of PCAR provided modest evidence that the UPSA-2ER (HK) is unidimensional (Table 3).

Raw scores of the UPSA-2ER(HK) of all participants were transformed from Rasch model to linear score, using the Wright map (Fig. 2). This was to enable comparison of an individual's functional capacity with all the participants in this study, for the purposes of interpretation. The Wright map showed that the difficulty of items largely consistent with expectations and matched well with the distribution of functional ability of participants.

3.3.3. Predictive validity

The results of the CRT analysis are showed in Figs. 3 and 4. The classification tree has five levels, and the scores of sub-domains 1, 2, 3, 4 and 6, were identified as important in the classification of the three groups of participants. Cut-off scores were identified to predict the membership of the participants in the residential placements. Backward pruning was conducted, a total of 8 nodes (node 15, 16, 23, 24, 25, 26, 27 and 28) were pruned under the minimum error pruning approach (Patel & Upadhyay, 2012) retaining the overall accuracy of the tree and optimizing the tree size.

The correct classification rate for independent living is highest (66 out of 70, 94.2%). Three of them were misclassified as HWH level while the remaining 1 was misclassified as LSCH level. In the HWH group, 52 out of 70 (74.3%) were correctly classified, 16 of subjects of this group were misclassified as independent living level, while the remaining 2 were misclassified as LSCH level. In the LSCH group, (67.1%), 47 of the 70 (67.1%) of participants were correctly classified, 5 of them were misclassified as independent living level, while the other 18 subjects were misclassified as HWH level (Table 6). The overall correct classification rate is 78.5%. With the results of the confusion matrix, sensitivity and specificity per class of the 3 groups were calculated. For independent living group, the sensitivity per class was 94.3%, while the specificity per class was 85.0%. For HWH group, the sensitivity per class was

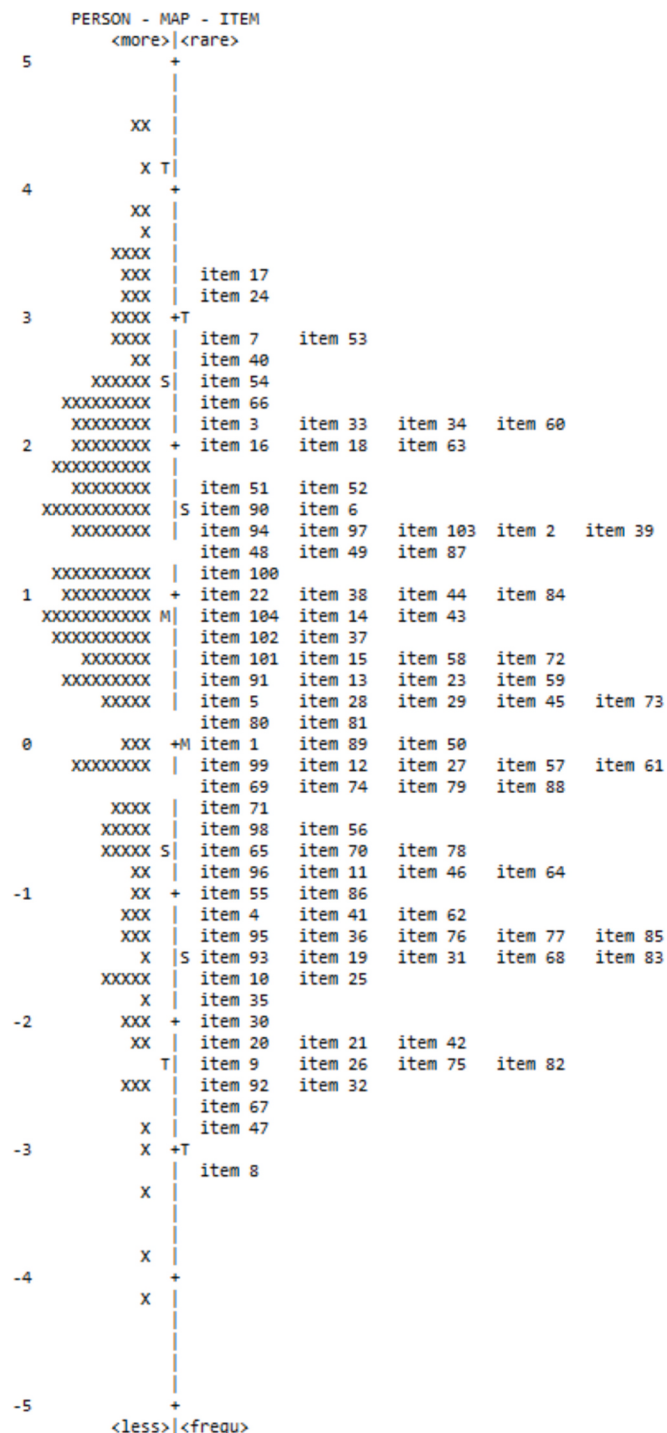


Fig. 2. Person-item difficulty map of the UPSA-2ER(HK).

74.3%, while the specificity per class was 85.0%. For LSCH group, the sensitivity per class was 67.1%, while the specificity per class was 97.9%. The average sensitivity by class was 78.6%, while the average specificity per class was 89.3% (Table 7). To evaluate model stability, a 10-fold cross-validation was conducted. The model achieved a pooled cross-validation risk estimate of 0.400 and a standard error of 0.034.

4. Discussion

This validation study showed that the UPSA-2ER(HK) is a reliable and valid tool for assessing the independent skills of people with mental

illness in Hong Kong. The translation method adopted procedures and standards commonly used in other cross-cultural validation studies (Chan et al., 2006; Chan et al., 2007; Law, 1999; Mok et al., 2005) in Hong Kong. The combination of a qualified translator, pre-panel review alignment synthesis, expert panel reviews and the endorsement by the copyright owner for the translation should address potential individual bias from the use of a single translator and ensure cross-cultural equivalence. The expert panel reviews suggested assessment items were relevant and representative of independent living in Hong Kong, and the quality of translation is very good. Only minor revisions to the translated version were suggested to address cultural or contextual issues like the change of currencies, name of transportation means, or medication labels. We agreed with some experts that ongoing update of assessment items in the UPSA is needed, such as the use of technology (like mobile phones, emails, social media, or contactless payment, smartphone apps). Items in the finance sub-domain, like hand calculation, count banknotes/coins or write bank cheque, would soon become obsolete.

Internal consistency of the sub-domains is good to very good, while inter-rater reliability is excellent. When compared with previous validation of the English full or brief version, the Chinese version had similar range in internal consistency (Becattini-Oliveira et al., 2018). The inter-rater reliability estimation was similarly excellent compare to the first UPSA study by Patterson et al. (2001). The high correlation between UPSA-2ER(HK) and the local gold standard measure of independent living skills (CFNA) indicated satisfactory concurrent validity. This result is also consistent with previous studies of concurrent validity with neurocognitive performance (attention, memory, executive function) or functional measures (Personal and Social Performance and Global Assessment of Functioning) (Mantovani et al., 2015). This study is also one of the first to examine the dimensionality of the UPSA using Rasch analysis. The PCAR in Rasch analysis showed that the UPSA-2ER (HK) could be regarded as having a unidimensional structure (Sumintono & Widhiarso, 2015). This implies that the total score could be used to indicate the functional independence of people with severe mental illness. The cutting score in five sub-domains can predict the residential status of three groups of participants with high accuracy. This supports the predictive validity of the instrument.

There are several important implications from the results. First, the UPSA-2ER(HK) could provide a more updated functional assessment, and covers skills not assessed by the CFNA, such as medication management. The UPSA-2ER(HK) also covers a wider range of functional disabilities among people with severe mental illness than the CFNA, which is designed to assess people with chronic schizophrenia only. Second, the C&RT analysis suggested that the UPSA-2ER(HK) assessment results could predict supported housing status with high accuracy. The instrument has potential to delineate client groups with different functional capacities. The cut-off scores for five sub-domains, i.e., 1 (planning and organization), 2 (finance), 3 (communication), 4 (transportation), and 6 (medication management) could be used in providing recommendations for entry into different levels of residential or rehabilitation services.

Third, Rasch analysis could be a powerful tool in validation of performance assessment in mental health practice. It could reveal if assessment items are able to delineate abilities among participants and examine the dimensionality of the instrument. Future validation studies of functional assessment could consider using Rasch item analysis method as an alternative approach in validation.

In further studies of the UPSA-2ER(HK), researchers could review if there is a need to add new daily living tasks assessment items like use of technology, or some items may become obsolete (e.g. writing checks). Further data collection is also needed to evaluate test-retest reliability. The UPSA-2ER(HK) could be used as an outcome measure in intervention studies of mental health recovery, to evaluate if the tool could be used to monitor changes in functional capacity or independent living skills.

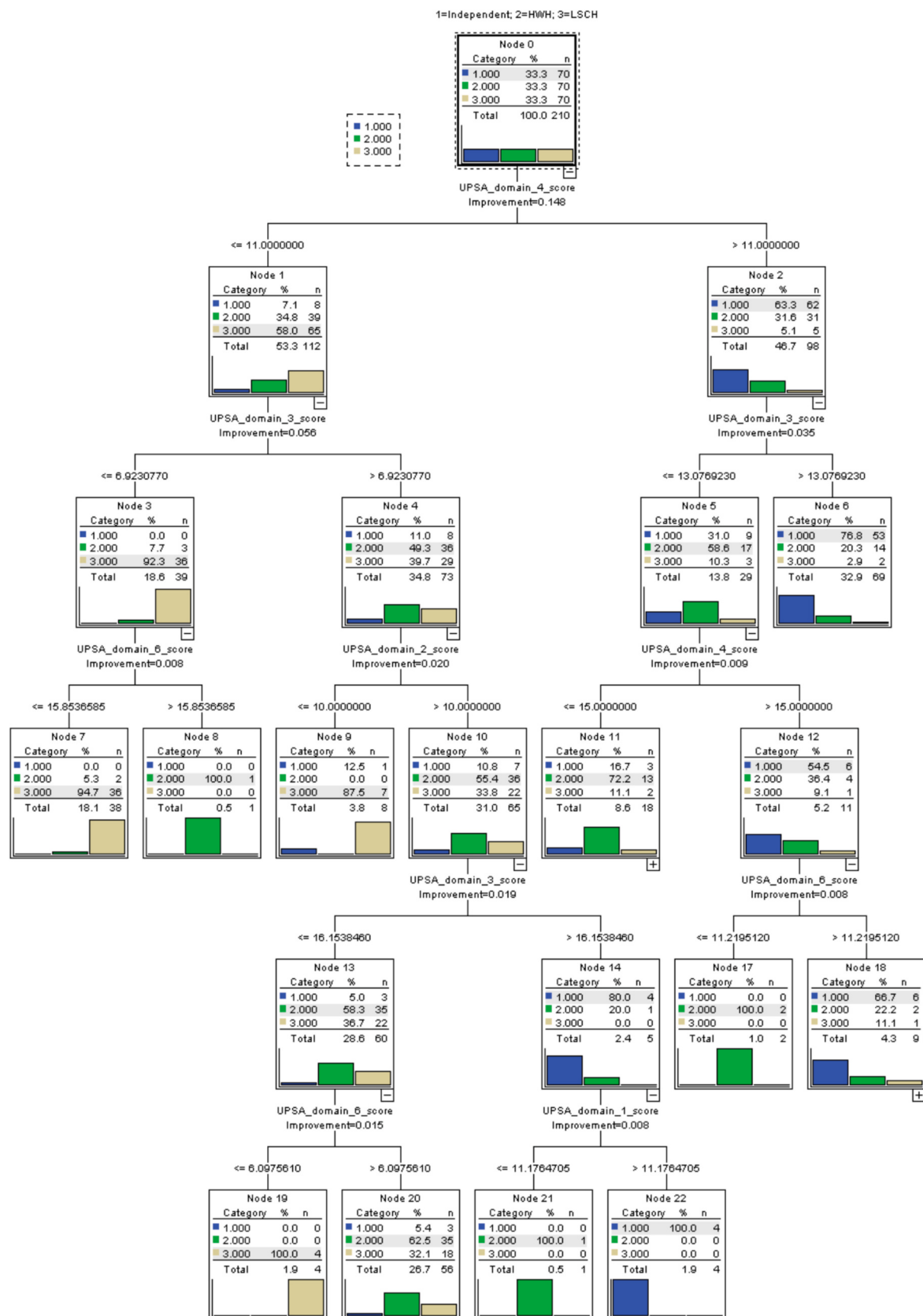


Fig. 3. Results of the classification and regression tree analysis of the UPSA-2ER(HK).

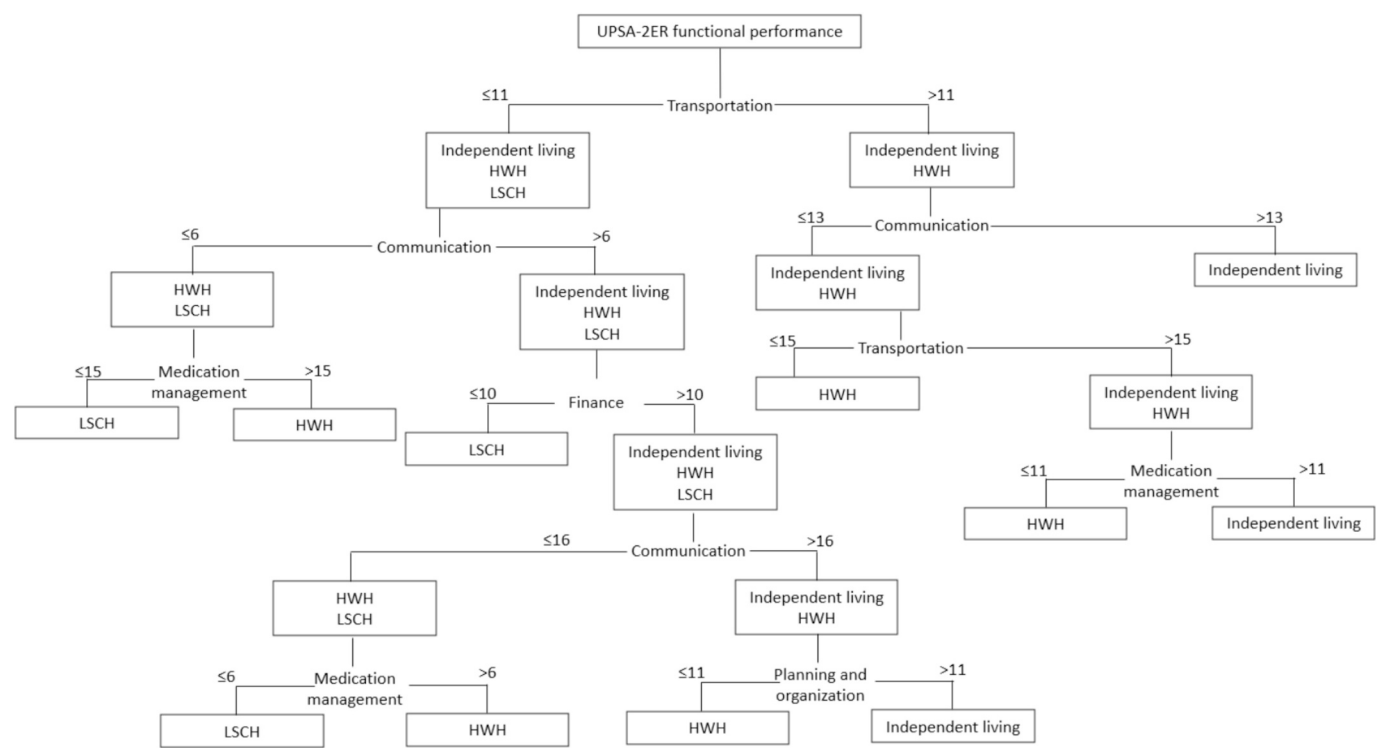


Fig. 4. Decision tree results of the UPSA-2ER(HK).

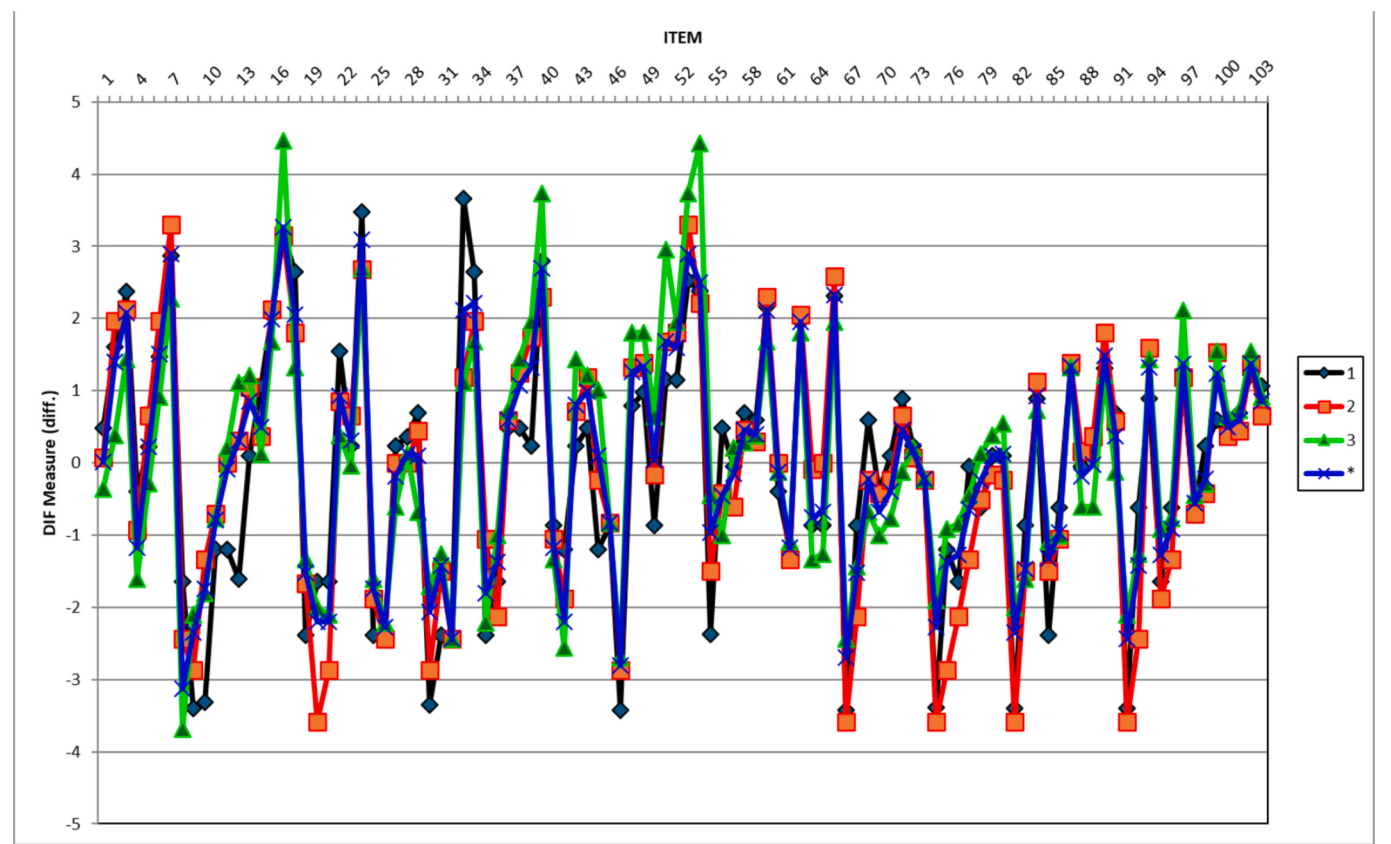


Fig. 5. Differential item functioning (DIF) measure of the UPSA-2ER(HK).

There are several limitations in this study. First, we found that 21 items did not fit well with Rasch rating scale model. However, many of these misfit items were part of a set of assessment items. Deleting some items of the set could create great difficulty in administration and

Table 6

Results of the Classification and Regression Tree analysis showing the % accuracy of the UPSA-2ER(HK) in classifying the residential (functional) status of the participants.

Observed	Predicted			Correct classification rate
	1. Independent living	2. Half-way house	3. Long stay care home	
1. Independent living	66	3	1	94.2%
2. Half-way house	16	52	2	74.2%
3. Long stay care home	5	18	47	67.1%
			Overall Accuracy	78.5%

Table 7

Sensitivity and specificity per class of the UPSA-2ER(HK).

	TP	TN	FP	FN	Sensitivity	Specificity
Independent living	66	119	21	4	0.942857143	0.85
HWH	52	119	21	18	0.742857143	0.85
LSCH	47	137	3	23	0.671428571	0.978571

scoring of the assessment. For example, the misfit items 67, 69, 75, 82, 85, 92, 94 and 95 belong to the sub-domain of medication management. Deleting misfit items deemed to have a negative impact on the reliability of the assessment. For instance, Cronbach's α will decrease from 0.81 to 0.75 if we remove the misfit items 18 and 22 in the finance sub-domain.

Second, there are significant differences in the demographic characteristics of the three groups, including age, gender proportions, education. The independent living group are younger and have higher education level, while the LSCH group are older, stayed in institutional care, and do not participate in employment. The three types of residential settings also have eligibility criteria related to functional capacity. These are confounding factors that could affect validity of predictive and classification tree analysis. Moreover, with the age differences across groups, the C&RT classification accuracy should be interpreted with the understanding that functional capacity and age-related decline are intertwined in this population. Future studies controlling for age (e.g., through covariate-adjusted analyses) would clarify the instrument's independent predictive contribution. Third, this study used functional capacity measured by the UPSA-2ER(HK) to predict residential placement with a high correct classification rate. In clinical practice, the determination and assignment of residential placement could be affected by other factors beyond functional capacity, including treatment compliance, supervision requirements, familial support, and the risk of challenging behaviours. These factors likely interact with functional capacity, future predictive studies should investigate their relative and combined contributions to placement outcomes.

Fourth, the classification and regression tree was pruned using the minimum error rule, despite it attained modest classification accuracy, sensitivity and specificity by class, the model might result in risk of over-fitting.

Last, the participants were recruited from a limited number of residential settings, and this limits the external validity of the study. Only 5 of 7 LSCHs joined the study, and we also recruit from only a small number of half-way houses in Hong Kong without systematic sampling. There are also smaller number of supported group housing settings which are not included in the study. Further studies could apply a more systematic sampling of participants from all types of residential settings.

5. Conclusion

The UPSA-2ER was translated and culturally adapted in Hong Kong.

The psychometric properties of the adapted version were evaluated using a sample of 210 participants with SMI. The UPSA-2ER(HK) demonstrated good to very good internal consistency, and excellent inter-rater reliability. Concurrent validity was satisfactory, supported by a high correlation with the CFNA. Rasch analysis results supported a unidimensional structure with good model fit. There were unfit items, but they were not removed due to their close connection to set of test questions under different topics. Using C&RT analysis and discriminant analysis, the UPSA-2ER sub-domain scores were effective in differentiating functional capacity levels linked to residential placements, including independent living, HWH and LSCH, at a high rate of correct classification. The analyses generated cutting scores which could provide guidelines for predicting functional independent and residential placements. The validated tool provides an updated assessment tool for assessing functional independence of people with SMI.

CRedit authorship contribution statement

Clement Chan: Writing – review & editing, Writing – original draft, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kelvin P.Y. Leung:** Writing – review & editing, Writing – original draft, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Cybie C.M. Mok:** Writing – review & editing, Resources. **Andrew M.H. Siu:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank NeuroCog Trials and Vera Science (Part of WCGclinical) for their permissions to translate, culturally adapt and validate the UPSA-2ER. We would like to thank all the participants of the study, Dr. Simon LUI and his research team, the involved NGOs including Caritas Hong Kong, Hong Kong Sheng Kui Hui Welfare Council Limited, Mental Health Association of Hong Kong, New Life Psychiatric Rehabilitation Association, and Tung Wah Group of Hospitals for the assistance in participant recruitment and facilitation of data collection. We would like to thank our expert panel members for their opinion provided in content validity and translation quality. We would like to thank the Occupational Therapy Department of Castle Peak Hospital for all the support given to this research project. Special thanks go to Ms. Catherine Li, Ms. Riva Cheng, Ms. Sandy Pang, Ms. Natasha Shiu, Ms. Doris Tse, Ms. Miki Lee, Mr. Alex So and Ms. Tammy Tong for their contribution in data collection.

Data availability

Data will be made available on request.

References

- Au, R. W. C., Tam, P. W. C., Tam, G. W. C., & Ungvari, G. S. (2005). Cross cultural validation of the St. Louis inventory of community living skills for Chinese patients with schizophrenia in Hong Kong. *Psychiatric Rehabilitation Journal*, 29(1), 34–40. <https://doi.org/10.2975/29.2005.34.40>
- Becattini-Oliveira, A. C., Dutra, D. d. F., Spenciende de Oliveira Campos, B., de Araujo, V. C., & Charchat-Fichman, H. (2018). A systematic review of a functional assessment tool: UCSD performance-based skill assessment (UPSA). *Psychiatry Research*, 267, 12–18. <https://doi.org/10.1016/j.psychres.2018.05.005>
- Chan, H., Li-Tsang, C. W. P., Chan, C., Lam, C. S., Hui, K. L., & Bard, C. (2006). Validation of Lam assessment of employment readiness (C-LASER) for Chinese injured workers.

- Journal of Occupational Rehabilitation*, 16(4), 697–705. <https://doi.org/10.1007/s10926-006-9050-3>
- Chan, S. C. C., Chan, C. C. H., Siu, A. M. H., & Poon, P. K. K. (2007). Stage of change in self-management of chronic diseases: Psychometric properties of the Chinese version of the University of Rhode Island Change Assessment (C-URICA) scale. *Rehabilitation Psychology*, 52(1), 103–112. <https://doi.org/10.1037/0090-5550.52.1.103>
- Chen, M., Fitzgerald, H. M., Madera, J. J., & Tohen, M. (2019). Functional outcome assessment in bipolar disorder: A systematic literature review. *Bipolar Disorders*, 21(3), 194–214. <https://doi.org/10.1111/bdi.12775>
- Cui, J., & Wu, J. (2025). Connecting to home in transitional accommodation: Exploring the perspectives of people with mental health issues in Hong Kong. *Health & Place*, 95, Article 103500. <https://doi.org/10.1016/j.healthplace.2025.103500>
- Dombrowski, L. B. (1990). *Functional needs assessment for chronic psychiatric patients. Therapy Skills Builder*. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Dombrowski%2C+L.+B.+%281990%29.+Functional+needs+assessment+program+for+chronic+psychiatric+patients.+&btnG=
- Fortuna, K. L., Lohman, M. C., Bruce, M. L., & Bartels, S. J. (2018). Utility of functioning measures in the prediction of independent living status in older adults with serious mental illness. *International Journal of Geriatric Psychiatry*, 33(2), 423–431. <https://doi.org/10.1002/GPS.4764>
- Garcia-Portilla, M. P., Gomar, J. J., Bobes-Bascaran, M. T., Menendez-Miranda, I., Saiz, P. A., Muñoz, J., ... Goldberg, T. E. (2013). Validation of a European Spanish version of the University of California performance Skills Assessment (Sp-UPSA) in patients with schizophrenia and bipolar disorder (p. 1). *Schizophrenia Research*, 150(2–3), 421–426. <https://doi.org/10.1016/j.schres.2013.07.049>
- Haig, G., Wang, D., Othman, A. A., & Zhao, J. (2016). The α 7 nicotinic agonist ABT-126 in the treatment of cognitive impairment associated with schizophrenia in nonsmokers: Results from a randomized controlled phase 2b study. *Neuropsychopharmacology*, 41(12), 2893–2902. <https://doi.org/10.1038/npp.2016.101>
- Hintze, J. (2013). *PASS 12*. Kaysville, UT: NCSS, LLC.
- Kim, K., Yim, S.-J., Lee, W.-H., Lee, K. H., Kim, Y. J., Chae, J.-H., & Lee, J.-I. (2022). Validation of the Korean version of the University of California San Diego performance-based skills assessment for high-functional capacity in the elderly. *International Journal of Gerontology*, 16(2), 116–121. [https://doi.org/10.6890/IJGE.202204_16\(2\).0008](https://doi.org/10.6890/IJGE.202204_16(2).0008)
- Kim, S.-J., Kim, J.-M., Shim, J.-C., Seo, B.-J., Jung, S.-S., Ryu, J.-W., Seo, Y.-S., Lee, Y.-C., Moon, J.-J., Jeon, D.-W., Park, K.-D., & Jung, D.-U. (2017). The Korean version of the University of California San Diego Performance-based Skills Assessment: Reliability and validity. *Clinical Psychopharmacology and Neuroscience*, 15(3), 261–268. <https://doi.org/10.9758/cpn.2017.15.3.261>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Law, K. (1999). Validation of functional needs assessment (Chinese version) for people with chronic schizophrenia. <https://theses.lib.polyu.edu.hk/handle/200/5021>
- Linares, J. M. (2009). *Winsteps® Rasch measurement computer program (version 3.73) [computer software]*. Portland, Oregon: Winsteps.com. <https://www.winsteps.com/>
- Linares, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, 16(2), 878.
- Loeb, P. A. (1996). *Independent Living Scales (ILS)*. Psychological Corporation.
- Mantovani, L. M., Machado-de-Sousa, J. P., & Salgado, J. V. (2015). UCSD Performance-Based Skills Assessment (UPSA): Validation of a Brazilian version in patients with schizophrenia. *Schizophrenia Research*, 2(1), 20–25. <https://doi.org/10.1016/j.schres.2014.12.002>
- Mausbach, B. T., Bowie, C. R., Harvey, P. D., Twamley, E. W., Goldman, S. R., Jeste, D. V., & Patterson, T. L. (2008). Usefulness of the UCSD performance-based skills assessment (UPSA) for predicting residential independence in patients with chronic schizophrenia. *Journal of Psychiatric Research*, 42(4), 320–327. <https://doi.org/10.1016/j.jpsychres.2006.12.008>
- Mausbach, B. T., Harvey, P. D., Goldman, S. R., Jeste, D. V., & Patterson, T. L. (2007). Development of a brief scale of everyday functioning in persons with serious mental illness. *Schizophrenia Bulletin*, 33(6), 1364–1372. <https://doi.org/10.1093/schbul/sbm014>
- Mausbach, B. T., Tiznado, D., Cardenas, V., Jeste, D. V., & Patterson, T. L. (2016). Validation of the UCSD Performance-based Skills Assessment (UPSA) in Hispanics with and without schizophrenia. *Psychiatry Research*, 244, 388–393. <https://doi.org/10.1016/j.psychres.2016.08.027>
- McDermid Vaz, S. A., Heinrichs, R. W., Miles, A. A., Ammari, N., Archie, S., Muharib, E., & Goldberg, J. O. (2013). The Canadian Objective Assessment of Life Skills (COALS): A new measure of functional competence in schizophrenia. *Psychiatry Research*, 206(2–3), 302–306. <https://doi.org/10.1016/j.psychres.2012.10.020>
- McIntosh, B. J., Zhang, X. Y., Kosten, T., Tan, S. P., Xiu, M. H., Rakofsky, J., & Harvey, P. D. (2011). Performance-based assessment of functional skills in severe mental illness: Results of a large-scale study in China. *Journal of Psychiatric Research*, 45(8), 1089–1094. <https://doi.org/10.1016/j.jpsychres.2011.01.012>
- Mok, C. C. M., Siu, A. M. H., Chan, W. C., Yeung, K. M., Pan, P. C., & Li, S. (2005). Functional disabilities profile of Chinese elderly people with Alzheimer's disease – A validation study on the Chinese version of the Disability Assessment for Dementia. *Dementia and Geriatric Cognitive Disorders*, 20(2–3), 112–119. <https://doi.org/10.1159/000086612>
- Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Moore, D., Palmer, B., Patterson, T., & Jeste, D. (2007). A review of performance-based measures of functional living skills. *Journal of Psychiatric Research*, 41(1–2), 97–118. <https://doi.org/10.1016/j.jpsychres.2005.10.008>
- Olsson, A.-K., Helldin, L., Hjärthag, F., & Norlander, T. (2012). Psychometric properties of a performance-based measurement of functional capacity, the UCSD Performance-based Skills Assessment - Brief version. *Psychiatry Research*, 197(3), 290–294. <https://doi.org/10.1016/j.psychres.2011.11.002>
- Patel, N., & Upadhyay, S. (2012). Study of various decision tree pruning methods with their empirical comparison in WEKA. *International Journal of Computers and Applications*, 60(12), 20–25. <https://doi.org/10.5120/9744-4304>
- Patterson, T. L., Goldman, S., McKibbin, C. L., Hughes, T., & Jeste, D. V. (2001). UCSD performance-based skills assessment: Development of a new measure of everyday functioning for severely mentally ill adults. *Schizophrenia Bulletin*, 27(2), 235–245. <https://doi.org/10.1093/oxfordjournals.schbul.a006870>
- Research Office, Legislative Council Secretariat. (2023). Support for carers of persons with disabilities and chronic diseases (Statistical Highlights ISSH35/2023). http://app7.legco.gov.hk/rpdb/en/uploads/2023/ISSH/ISSH35_2023_20231208_en.pdf
- Revheim, N. (2024). The independent living scales. In *Assessments in occupational therapy mental health* (pp. 383–396). Routledge. <https://doi.org/10.4324/9781003522645-28>
- Revheim, N., & Medalia, A. (2004). The Independent Living Scales as a measure of functional outcome for schizophrenia. *Psychiatric Services*, 55(9), 1052–1054. <https://doi.org/10.1176/appi.ps.55.9.1052>
- Richter, D., & Hoffmann, H. (2017). Independent housing and support for people with severe mental illness: Systematic review. *Acta Psychiatrica Scandinavica*, 136(3), 269–279. <https://doi.org/10.1111/ACPS.12765>
- Sin, W. Y., Wong, W. N., Michelle Lo, Y. T., Elizabeth Tse, U. T., Helen Wong, H. W., Elizabeth Tung, K. Y., ... Chien Kuo, M. C. (2022). Cognition, symptomatology, and community living skills in schizophrenia. *Current Psychiatry Research and Reviews*, 18(3), 220–228. <https://doi.org/10.2174/2666082218666220801095440>
- Social Welfare Department, The Government of the Hong Kong Administrative Region. (2023). Residential care services for persons in mental recovery. https://www.swd.gov.hk/en/pubsvr/rehab/cat_residcare/resserexmental/
- Squires, A., Aiken, L. H., van den Heede, K., Sermeus, W., Bruyneel, L., Lindqvist, R., ... Matthews, A. (2013). A systematic survey instrument translation process for multi-country, comparative health workforce studies. *International Journal of Nursing Studies*, 50(2), 264–273. <https://doi.org/10.1016/j.ijnurstu.2012.02.015>
- Strassnig, M. T., Raykov, T., O'Gorman, C., Bowie, C. R., Sabbag, S., Durand, D., ... Harvey, P. D. (2015). Determinants of different aspects of everyday outcome in schizophrenia: The roles of negative symptoms, cognition, and functional capacity. *Schizophrenia Research*, 165(1), 76–82. <https://doi.org/10.1016/j.schres.2015.03.033>
- Sumintono, B., & Widhiarso, W. (2015). *Application of Rasch modeling in educational assessment*. Trim komunikata.
- Sumiyoshi, C., Takaki, M., Okahisa, Y., Patterson, T. L., Harvey, P. D., & Sumiyoshi, T. (2014). Utility of the UCSD Performance-based Skills Assessment-Brief Japanese version: Discriminative ability and relation to neurocognition. *Schizophrenia Research*, 150(2–3), 137–143. <https://doi.org/10.1016/j.schres.2014.08.002>
- Tungpunkom, P., Maayan, N., & Soares-Weiser, K. (2012). Life skills programmes for chronic mental illnesses. *The Cochrane Database of Systematic Reviews*, 2012(1), Article CD000381. <https://doi.org/10.1002/14651858.CD000381.pub3>
- Zimmerman, M., Morgan, T. A., & Stanton, K. (2018). The severity of psychiatric disorders. *World Psychiatry*, 17(3), 258–275. <https://doi.org/10.1002/wps.20569>