

Article

A Cross-Cultural Adaptation of the ICECAP-O: Test–Retest Reliability and Item Relevance in Swedish 70-Year-Olds

Helena Hörder ^{1,*}, Susanne Gustafsson ², Therese Rydberg ¹, Ingmar Skoog ¹ and Margda Waern ¹

¹ Neuropsychiatric Epidemiology Unit, Institute of Neuroscience and Physiology, Sahlgrenska Academy, Centre for Ageing and Health—Agecap, University of Gothenburg, Mölndal 43141, Sweden; Therese.Rydberg@neuro.gu.se (T.R.); Ingmar.Skoog@neuro.gu.se (I.S.); Margda.Waern@neuro.gu.se (M.W.)

² Division of Occupational and Physical Therapy, Institute of Neuroscience and Physiology, Sahlgrenska Academy, Centre for Ageing and Health—Agecap, University of Gothenburg, Gothenburg 40530, Sweden; Susanne.Gustafsson@gu.se

* Correspondence: helena.horder@neuro.gu.se; Tel.: +46-31-7760403

Academic Editor: Kristine Crondahl

Received: 8 August 2016; Accepted: 26 September 2016; Published: 30 September 2016

Abstract: Background: While there is a plethora of Quality of Life (QoL) measures, the Investigating Choice Experiments for the Preferences of Older People—CAPability index (ICECAP-O) is one of the few that taps into the concept of capability, i.e., opportunities to ‘do’ and ‘be’ the things that one deems important in life. We aimed to examine test–retest reliability of the ICECAP-O in a Swedish context and to study item relevance. **Methods:** Thirty-nine 70-year-olds who took part in a population-based health study completed the Swedish version of the ICECAP-O on two occasions. We analyzed the test–retest reliability for the index and for the individual items. Participants also rated the relevance of each item on a visual analogue scale (0–100). **Results:** Test–retest reliability for the index score was in good agreement with an ICC of 0.80 (95% CI 0.62–0.90). However, Kappa was low for each item and ranged from 0.18 (control) to 0.41 (role). For attachment, we found a systematic disagreement with lower ratings at the second test occasion. Participants gave their highest relevance rating to attachment and lowest to enjoyment. **Conclusion:** The Swedish version of the ICECAP-O had good test–retest agreement, similar to that observed for the English version. Item level agreement was problematic, however, highlighting a need for future research.

Keywords: capability; older persons; ICECAP-O; ICECAP; quality of life; well-being; questionnaire; reliability

1. Introduction

The concept of capability provides an alternative approach to the study of human well-being. Capability can be described as people’s ability to perform actions in order to reach goals that they have reason to value [1]. At the core of this theoretical framework are a person’s functionings (beings and doings) and capabilities (the genuine opportunities or freedoms to realize these functionings). The distinction between functionings and capabilities is between achievements, on the one hand, and freedoms or valuable opportunities on the other [1]. The concept of capability has been highly influential in welfare economics, philosophy, and social and political science [2].

Although the capability approach is difficult to apply empirically because of the subjectivity of the concept, a capability-based measure of general Quality of Life (QoL) has recently been developed. The Investigating Choice Experiments for the Preferences of Older People—CAPability

index (ICECAP-O) [3–5] is conceptually linked to Amartya Sen’s capability approach [3,6] and designed to reflect QoL in its broadest sense [7].

To date, there is no Swedish language QoL questionnaire that taps into the concept of capability. As the ICECAP-O encompasses areas identified as highly valued in the wellbeing of community-dwelling [8], as well as frailer older persons [9,10] living in Sweden, we wanted to adapt the questionnaire to the Swedish setting.

The main aim of this paper, which is the first in a series, is to evaluate test–retest reliability of the ICECAP-O in a Swedish context. A further aim is to study participants’ perceptions of the relevance of the individual ICECAP-O items.

2. Methods

2.1. Setting and Participants

Participants were recruited from the ongoing comprehensive population-based study of health among older persons, the Gerontological and Geriatric Population study in Gothenburg, Sweden (H70) [11]. A convenience subsample ($n = 40$) from a birth-cohort of 70-year-olds born in 1944 and examined in 2015 took part in the test–retest procedure and rated the relevance of the individual ICECAP-O items.

2.2. Ethics, Consent, and Permissions

The study was approved by the Regional Ethical Review Board (Dnr 869-13 and T139-15). Written informed consent was obtained after the participants had received a complete description of the study.

2.3. The ICECAP-O

The ICECAP-O is a five-item instrument (attachment, security, role, enjoyment, and control) with four-level response options that are described as statements representing: none, a little, a lot, and full capability [5]. The instrument was developed based on findings from rigorous qualitative and quantitative research with older persons in the UK [3,4,7,12]. Values are anchored with a best–worst scaling, ranging from 1.00 (full capability) to 0.00 (no capability). A total index score, based on a tariff computed from population-based values in the UK, is thus obtained [4,13]. The ICECAP-O questionnaire, is freely available at the University of Birmingham’s webpage [5]. The Swedish version could also be accessed through a direct link: <http://www.birmingham.ac.uk/Documents/college-mds/haps/projects/icecap/Iccap-o/SWEDISH-ORIGINAL-ICECAP-O-Helena-Hörder.docx>.

2.4. Procedure

We followed guidelines for cross-cultural adaptation of health-related quality of life measures, including forward–backward translation, committee review, and pre-test [14]. For assessment of test–retest reliability, we administered the final Swedish version of the ICECAP-O on two occasions: first as one of several self-report questionnaires included in the above described H70-study, and second as a single postal questionnaire, sent 1- to 2-weeks after the research appointment. The time interval was chosen to minimize recall bias as well as bias due to actual change.

At the second occasion, participants were asked to “[r]ate the relevance of each item for [their] well-being” on a 0–100 visual analogue scale (VAS), ranging from “not important at all” to “could not be more important.”

2.5. Analysis

We analyzed the test–retest reliability for both the index and the individual items. For the index (continuous data), we calculated the intraclass correlation coefficient (ICC), using a two way mixed model of absolute agreement. The ICC can range from 0.00 (no stability/agreement) to 1.00 (perfect agreement). An ICC of 0.70 is considered to be acceptable [15]. For each item, we calculated percentage

agreement (PA), $PA \pm 1$ (proportion of persons with answers within a scale score range of ± 1) and Cohen's weighted kappa [16]. For a deeper understanding, we also calculated systematic disagreements in relative position (RP) using a rank-based statistical method for paired ordinal data [17]. RP indicates to which extent the distribution between the two test occasions is systematically shifted towards higher or lower scale categories. Values can range from -1 to 1 , with a value close to 0 representing a small systematic disagreement. Statistically significant values are indicated by a 95% confidence interval (CI) that does not include the zero value.

VAS ratings of the relevance of each of the individual ICECAP-O items are presented as mean (± 1 SD) and range.

3. Results

All participants, 21 women and 19 men, had complete data on all five items on both occasions. One person was not included in the analyses due to an actual change in conditions that seriously affected capability, yielding a total of 39 paired ratings.

The mean ICECAP-O index score was 0.86 (SD 0.10) at test occasion one and 0.84 (SD 0.11) at test occasion two. The ICC for the index was 0.80 (95% CI 0.62–0.90). For details, see Figure 1.

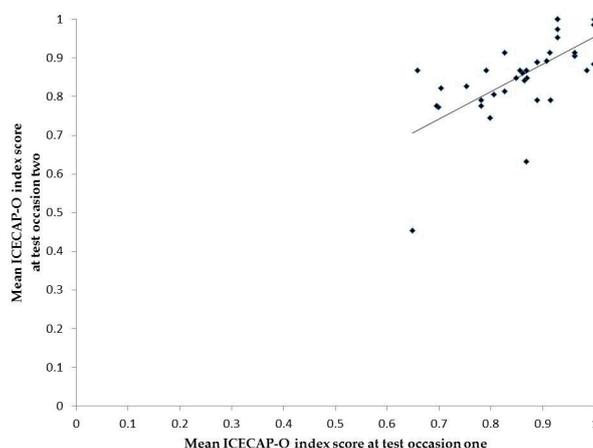


Figure 1. Correlation between ICECAP-O index scores at test occasion one (x-axis) and test occasion two (y-axis). Possible range from 1.00 (full capability), to 0.00 (no capability).

Test-retest ratings were consistent for all five items for eight participants (21%). Seven persons (18%) reported a lower rating on one or more items at the second test occasion, and eight (21%) a higher rating on one or more items. A mix of higher and lower ratings was observed in 16 persons (41%).

The agreement on item level is presented in Table 1. The agreement was highest for Role (Kappa 0.41, PA 67%), followed by Attachment, Enjoyment, Security, and Control. The $PA \pm 1$ was $\geq 95\%$ for all items. A systematic change in RP was seen for Attachment only. For this item, the pattern of change in position indicated a small but significant probability of lower scores at the second test occasion.

Table 1. Test-retest reliability of the ICECAP-O on item level in 70-year-olds ($n = 39$).

	Kappa	Agreement (%)	Agreement ± 1 (%)	Systematic Disagreement in Relative Position (95% CI)
Attachment	0.34	60%	100%	-0.17 (-0.35–-0.03)
Security	0.22	50%	100%	0.05 (-0.11 – 0.20)
Role	0.41	65%	95%	0.00 (-0.16 – 0.16)
Enjoyment	0.24	52%	100%	-0.02 (-0.19 – 0.14)
Control	0.17	50%	95%	-0.13 (-0.32 – 0.05)

Percentages of agreement of less than 60% indicate poor agreement. Relative position values close to 0 represent a high level of reliability. Values in bold indicate significant changes. CI = confidence interval.

Table 2 shows details on changes in positions for each item. Disagreements were mainly due to response shifts from “full” (rating 4) to “a lot” (rating 3). Sub-analyses indicated that men were more likely than women to report a lower capability in Attachment at the second test occasion (results not shown).

Table 2. Distributions of change in positions between test occasion one and test occasion two for ICECAP-O items (1 = no capability to 4 = full capability), $n = 39$.

ICECAP-O item	1 (no capability)	2 (little capability)	3 (a lot capability)	4 (full capability)
Attachment				
4 (full capability)			3	13
3 (a lot capability)		1	9	10
2 (little capability)		2	1	
1 (no capability)				
Security				
4 (full capability)			5	4
3 (a lot capability)		5	11	4
2 (little capability)	1	5	4	
1 (no capability)				
Role				
4 (full capability)		1	2	7
3 (a lot capability)	1	2	17	4
2 (little capability)		2	3	
1 (no capability)				
Enjoyment				
4 (full capability)			2	3
3 (a lot capability)		6	14	7
2 (little capability)		4	3	
1 (no capability)		1		
Control				
4 (full capability)			4	8
3 (a lot capability)		3	12	5
2 (little capability)			5	2
1 (no capability)				

x-axis = test occasion one, y-axis = test occasion two. In bold = absolute agreement at test occasion one and two.

Results of the VAS ratings for item relevance are shown in Figure 2. The highest rating was observed for Attachment and the lowest for Enjoyment.

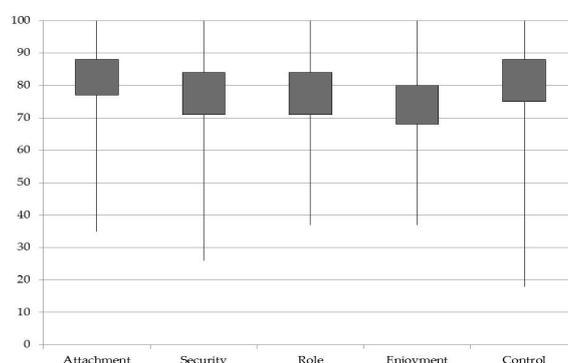


Figure 2. VAS ratings (0–100) of relevance of individual items included in ICECAP-O (± 1 sd and range). 0 = not important at all, 100 = could not be more important.

4. Discussion

This cross-cultural adaptation of the ICECAP-O indicates that the index has good test–retest reliability, similar to that observed in a Dutch study that focused on frail older adults [18]. On the other hand, the absolute agreement for each item was low to moderate in our study.

Good reliability for the index despite the low absolute agreement for individual items could be explained by the fact that most item changes involved shifts from “full” (level 4) to “a lot” (level 3), and these levels have very similar weighting in the tariff, in contrast to the much larger difference in weighting between the two lowest levels [4]. Further, about 40% had *both* lower and higher ratings on individual items at the second test occasion, resulting in a relatively consistent index score.

This is the first study to examine test–retest agreement for individual ICECAP-O items. A partial explanation for the observed item inconsistency might be related to the age of our participants. In a general population-based British study that utilized the ICECAP-A (adult), higher age was associated with inconsistent item ratings [19].

Another possible explanation for test–retest item inconsistency might be related to differences in available time for completion of the questionnaire. On the first occasion, the ICECAP-O was included in an extensive questionnaire packet that was administered in connection with a comprehensive health examination. In contrast, the retest was completed at home at the participant’s leisure. This meant that participants were free to take their time and reflect on ICECAP-O item response options, which could impact on interpretation and choice of response. In a think aloud study, persons were shown to vary in interpretations when rating capabilities [20]. Previous general population-based research has shown slightly lower reliability for the ICECAP index compared to EuroQoL [19,21]. As highlighted by others [20], a more thorough guiding might be one way to achieve more consistent interpretations of capabilities.

Participants’ rating of item relevance showed that attachment was valued highest, followed by control, role, security, and enjoyment. This order is similar to that observed in the original UK study on older persons [4].

Some important study limitations need to be mentioned. The sample was relatively homogenous and small in size. Reliability testing is necessary for other age groups. A wider variation in medical conditions and functional abilities would be anticipated in older age groups, and this would be expected to result in a larger variation in item responses.

5. Conclusions

The ICECAP-O provides a promising self-report capability approach to the assessment of QoL. The Swedish version presented here showed high test–retest reliability for the index score, but agreement for individual items was problematic. Further testing is needed to better understand effects of testing environments, contexts, and time frames.

Acknowledgments: This work was supported by the Forte-Centre Aging and health: Centre for capability in ageing (AGECAP). It was also supported by grants from the Swedish Research Council for Health, Working Life and Welfare (2004-0145, 2006-0596, 2008-1111, 2010-0870), Epilife 2006-1506, AGECAP 2013-2300, 2013-2496), the Alzheimer’s Association Stephanie B. Overstreet Scholars (IIRG-00-2159, 2008-1229), the Alzheimer’s Association Zenith Award (ZEN-01-3151), the Swedish Research Council (no. 11267, 2005-8460, 825-2007-7462, 825-2012-5041, 2013-8717, 2015-02830), Sahlgrenska University Hospital (ALF), the Bank of Sweden Tercentary Foundation, Swedish Brain Power, Stiftelsen Gamla Tjänarinnor, Handlanden Hjalmar Svenssons Forskningsfond, Systarna Greta Johansson and Brita Anderssons minnesfond, Axel Linders stiftelse, Stiftelsen Ragnhild o Einar Lundström, Stohnes stiftelse, Lundgrenska stiftelserna and Svensk-Franska stiftelsen.

Author Contributions: H.H. was the Principal Investigator and designed the study. I.S. was the Principal investigator of the larger H70 study in which this sub-study was carried out. H.H. and T.R. contributed to the acquisition of data. H.H., S.G. and M.W. cooperated on the final translation of the questionnaire. H.H. conducted the analyses and drafted the manuscript. All authors contributed to the interpretation of data, commented on drafts of the manuscript and read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Sen, A. *The Idea of Justice*; The Belknap Press of Harvard University Press: Massachusetts, MA, USA, 2009.
2. Fukuda-Parr, S. The human development paradigm: Operationalizing Sen's ideas on capabilities. *Fem. Econ.* **2003**, *9*, 301–317. [[CrossRef](#)]
3. Grewal, I.; Lewis, J.; Flynn, T.; Brown, J.; Bond, J.; Coast, J. Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? *Soc. Sci. Med.* **2006**, *62*, 1891–1901. [[CrossRef](#)] [[PubMed](#)]
4. Coast, J.; Flynn, T.N.; Natarajan, L.; Sproston, K.; Lewis, J.; Louviere, J.J. Valuing the ICECAP capability index for older people. *Soc. Sci. Med.* **2008**, *67*, 874–882. [[CrossRef](#)] [[PubMed](#)]
5. University of Birmingham. ICECAP-O. Birmingham; 2016. Available online: <http://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/HE/ICECAP/ICECAP-O/index.aspx> (accessed on 5 February 2016).
6. Sen, A. *Choice, Welfare and Measurement*; Harvard University Press: Massachusetts, MA, USA, 1982.
7. Coast, J.; Peters, T.J.; Natarajan, L.; Sproston, K.; Flynn, T. An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. *Qual. Life Res.* **2008**, *17*, 967–976. [[CrossRef](#)] [[PubMed](#)]
8. Horder, H.M.; Frandin, K.; Larsson, M.E. Self-respect through ability to keep fear of frailty at a distance: Successful ageing from the perspective of community-dwelling older people. *Int. J. Qual. Stud. Health Well-Being* **2013**, *8*, 20194. [[CrossRef](#)] [[PubMed](#)]
9. Ekelund, C.; Mårtensson, L.; Eklund, K. Self-determination among frail older persons—A desirable goal older persons' conceptions of self-determination. *Qual. Ageing Older Adults* **2014**, *15*, 90–101. [[CrossRef](#)]
10. Ebrahimi, Z.; Wilhelmson, K.; Moore, C.D.; Jakobsson, A. Frail elders' experiences with and perceptions of health. *Qual. Health Res.* **2012**, *22*, 1513–1523. [[CrossRef](#)] [[PubMed](#)]
11. Rinder, L.; Roupe, S.; Steen, B.; Svanborg, A. Seventy-year-old people in Gothenburg. A population study in an industrialized Swedish city. *Acta Med. Scand.* **1975**, *198*, 397–407. [[CrossRef](#)] [[PubMed](#)]
12. Flynn, T.N.; Louviere, J.J.; Marley, A.A.; Coast, J.; Peters, T.J. Rescaling quality of life values from discrete choice experiments for use as QALYs: A cautionary tale. *Popul. Health Metr.* **2008**, *6*, 6. [[CrossRef](#)] [[PubMed](#)]
13. Flynn, T.N.; Louviere, J.J.; Peters, T.J.; Coast, J. Best-worst scaling: What it can do for health care research and how to do it. *J. Health Econ.* **2007**, *26*, 171–189. [[CrossRef](#)] [[PubMed](#)]
14. 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*, 1417–1432. [[CrossRef](#)]
15. Munro, B.H. *Statistical Methods for Health Care Research*, 4th ed.; JB Lippincott: Philadelphia, PA, USA, 2000.
16. Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* **1968**, *70*, 213–220. [[CrossRef](#)] [[PubMed](#)]
17. Svensson, E. Application of a rank-invariant method to evaluate reliability of ordered categorical assessments. *J. Epidemiol. Biostat.* **1998**, *3*, 403–409.
18. van Leeuwen, K.M.; Bosmans, J.E.; Jansen, A.P.; Hoogendijk, E.O.; van Tulder, M.W.; van der Horst, H.E. Comparing measurement properties of the EQ-5D-3L, ICECAP-O, and ASCOT in frail older adults. *Value Health.* **2015**, *18*, 35–43. [[CrossRef](#)] [[PubMed](#)]
19. Al-Janabi, H.; Flynn, T.N.; Peters, T.J.; Bryan, S.; Coast, J. Test–Retest Reliability of Capability Measurement in the UK General Population. *Health Econ.* **2015**, *24*, 625–630. [[CrossRef](#)] [[PubMed](#)]
20. Al-Janabi, H.; Keeley, T.; Mitchell, P.; Coast, J. Can capabilities be self-reported? A think aloud study. *Soc. Sci. Med.* **2013**, *87*, 116–122. [[CrossRef](#)] [[PubMed](#)]
21. Couzner, L.; Crotty, M.; Norman, R.; Ratcliffe, J. A comparison of the EQ-5D-3L and ICECAP-O in an older post-acute patient population relative to the general population. *Appl. Health Econ. Health Policy* **2013**, *11*, 415–425. [[CrossRef](#)] [[PubMed](#)]

