



Measuring what matters to older persons for active living: part II cross-sectional validity evidence for OPAL measure across four countries

Nancy E. Mayo^{1,2,3} · Mohammad Auais⁴ · Ruth Barclay^{5,6} · Joan Branin⁷ · Helen Dawes⁸ · Ida J. Korfage⁹ · Kim Sawchuk¹⁰ · Eran Tal¹¹ · Carole L. White¹² · Zain Ayoubi¹³ · Ezinne Ekediegwu¹³ · Kedar Mate¹⁴ · Lyne Nadeau¹³ · Sebastian Rodriguez Duque¹¹ · Ayse Kuspinar¹⁵

Accepted: 11 June 2024

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024, corrected publication 2024

Abstract

Introduction Through interviews with 148 older persons from four countries and in four languages, the content for a 17-item measure of active living was developed. The purpose of this paper is to present further evidence of the extent to which this new measure, Older Persons Active Living (OPAL), is “fit-for-purpose” for measuring the extent of active living at one point in time.

Methods A cross-sectional study was carried out on a population aged 65 + and living independently, drawn from a participant panel, HostedinCanada, sampling people from Canada, United States, United Kingdom, and Netherlands. The survey instrument comprised the OPAL questionnaire rated on importance and frequency, sociodemographics, and information on physical and mental function. The argument-based approach to validity framed the analyses. Logistic regression, structural equation modeling, ordinary least-squares regression, and correlation were used to generate estimates for parameters underpinning validity evidence.

Results A total of 1612 people completed the survey, 100 to 400 people across the 6 country-language strata. The proportion of people rating the items as extremely or quite important ranged from 60 to 90%, with no important differences between men and women and few differences between strata. A single-factor structure was supported. The ordinality of the response options justified an additive total score yielding a near normal distribution (mean: 33.1; SD: 11.5; range 0–51). Correlations with other measures of converging constructs were of moderate strength (~0.50), and differences across groups known to affect functioning and health were observed, suggesting a Minimal Important Difference (MID) of 6 out of 51.

Conclusion The results of this study provide evidence that the 17-item OPAL measure is fit for the purpose of estimating the extent to which older persons are living actively at one point in time.

Keywords Aging · Active living · Measurement · Validity · Minimal important difference

Plain English summary

A 17-item measure of active living for older persons (OPAL) was developed by interviewing 148 older persons from four countries and in four languages, English, French, Spanish, and Dutch. Older persons identified active living as a way of being and the content of the measure is based on their input. To show that the measure is useable, it is necessary to test it in another sample of people and show that the items are important, hang together as a measure, and behave as

expected with other measures. A large sample of people, 1612 from four countries were surveyed. All the items were rated as important and there is room for improvement in how much people are living actively. The results showed that the score on the OPAL measure related to other measures of health and activity but was also different from them. The maximum score on the OPAL measure is 51 and the average value in the sample tested was 33. People with breathing problems, fatigue, or feeling they are in poor health scored around 10 points lower out of 51 than people without these symptoms. A difference of 6 points out was found to be an important difference. These results indicate that the OPAL measure can be used to measure people at one point in time.

Extended author information available on the last page of the article

Background

In 2020, the Canadian Institutes of Health Research (CIHR) funded an international team to develop a measure to inform and evaluate active living programs for older persons. The World Health Organization has spearheaded a movement towards active living or active aging [1] which is defined as ...*the process of optimizing opportunities for health, participation, and security in order to enhance quality of life as people age*. Active aging/living allows people to realize their full potential and to participate in society according to their needs, desires and capacities. How active living is operationalized into local programs or evaluated is often absent from the discussion.

An example of the challenges in evaluating community programs came from the experience of administering traditional measures of constructs related to health aspects of quality of life to a sample of older persons participating in a community outdoor walking program [2]. The participants were assisted to fill out the questionnaires and it was apparent from people doing the assisting that the items did not seem to relate to how this sample was currently living their lives. This resulted in many questions about how to respond to the items and even if there were right and wrong answers. There was confusion about terms that seemed similar (anxiety and stress), how to respond to questions about pain when pain is a background feature of their lives, and how to respond to questions about activities that they don't do. This was real world evidence that our existing measures for older persons were not fit for the purpose of assessing how older persons were living their lives outside of the clinical context nor for evaluating the impact of interventions or programs targeting active living. The need for a new measurement framework was considered a worthwhile effort given the growing older adult population and the number of programs being developed that require evaluation.

This experience was not unique and a group of researchers working in the field of senior's health and age studies became engaged in developing a new measure specifically designed to reflect how older persons were living their lives in their communities. The content development for this new measure previously has been described. Briefly, using best practices for content development a series of focus groups and individual interviews were conducted across four countries (Canada, United States, United Kingdom, Netherlands) and in four languages (English, French, Spanish, and Dutch).

The content development was framed from an anti-ableist perspective. The strong bias towards "fixing" aging and its inherent problems shares common ground with ableism [3], the bias towards considering people with

disabilities as inferior as they deviate from an ideal, able bodied, existence, and are in need of fixing [4]. Titchkosky postulates that disability can be an identity which "makes it possible to insert into the world alternative ways of being and of knowing" [5]. Our thesis was that this framing could be applied to the situation of aging—an identity making it possible to have alternate ways of being and knowing. As emphasized by thought leaders on aging in the 1990s, *Aging is not a problem to be solved rather a mystery to be lived*, [6].

The underlying conceptual model for the active living construct was formative [7–9], that is, the items are considered to form the construct. Based on this perspective, the content development phase identified 17 "ways of being" that older persons, across these 6 country-language strata, identified as contributing to active living as an older person. These 17 "ways of being" were each evaluated on a four-point ordinal scale based on frequency (always or almost always, often, sometimes, rarely or never) not extent, to avoid an ableist framework.

Persons were asked to indicate how often in the past six months they felt: active, confident, connected, useful, creative, encouraged, energetic, involved, happy, healthy mentally, healthy physically, independent, interested, mentally sharp, and motivated. The results of the content development phase provided evidence that existing measures did not meet users' needs. Many items were negatively worded and framed in an ableist view where persons are rated against an ideal. This first paper also demonstrated that the experiences of the target respondents were represented in the content of the measure.

Objectives

The purpose of this paper is to present further evidence of the extent to which this new measure, the Older Persons Active Living (OPAL) measure, is "fit-for-purpose" [10]. The objectives of this study are framed using American Educational Research Association's Standards for Educational and Psychological Testing. American Educational Research Association [11] using the argument based approach to validity [12, 13]. The specific objectives are to provide evidence that.

- (i) the items chosen are considered important to target responders and vary in frequency;
- (ii) the response categories are ordered in the expected linear fashion;
- (iii) the data derived from administering the items of the measure to the targeted sample fit the hypothesized structure of the measure;

- (iv) the scores from the measure under study correlate to the scores on a criterion measure or correlate with scores on measures that represent converging or diverging constructs; and
- (v) the scores on the index measure differ across categories of variables that are known to differ for this construct (e.g. age, sex, disease severity etc.).

Methods

This was a cross-sectional study of a population drawn from a participant panel, HostedinCanada (HIC). Briefly, the process is that HIC notifies individuals registered on the HIC platform of the survey and informs them of the eligibility criteria. Individuals who agree to receive more information about the study are provided electronically, by the HIC, with the study's consent form. Those who agree electronically sign the consent form and are directed to the survey which is hosted on the HIC platform. English, French, Spanish and Dutch versions of the consent form and survey instrument were provided. A description of the privacy protections for individuals and for the data, taken from the HIC website (<https://www.hostedincanadasurveys.ca/>).

Population

The target population was people over the age of 65 years living independently in the community, without an illness requiring ongoing hospital treatment. The population was balanced across country and language strata. The size of the sample was estimated based on using confirmatory factor analysis using the rule of thumb of > 10 persons per item [14]. The number of items to be tested was 17 resulting in a minimum sample size estimate of 170. As we wanted this minimum sample size per country, the total sample size was set in proportion to the size of the populations in the four countries and to have a minimum sample size of 100 for each language. The smallest sample size was for Canadian French ($n = 100$); Canada, UK, English USA were each to have a total sample size of 400 (300 Canadian English); US Spanish and the Netherlands were to have 200 each for a total sample size of 1600 people.

Measurement

The measurement strategy for this study was informed by aging models which are strongly influenced by the concept of frailty [15–17], more positively framed as intrinsic capacity [18], as well as models of health-related quality of life (HRQL) [19, 20].

The survey instrument comprised the items from the OPAL measure rated for importance on a 5-point ordinal

scale (extremely, quite a bit, moderately, a little bit, not at all) and rated on frequency on a 4-point ordinal scale (always or almost always, often, sometimes, rarely or never). These response options were chosen as they have been shown to have linear properties [21].

In addition to the items from the OPAL measure, the survey included content related to demographics, comorbidity, impairments and symptoms that are common in the older population, and indicators of well-being, physical and cognitive capacity, and health. Apart from single items, the survey included the Lower Extremity Functional Scale (LEFS [22]) and Communicating Cognitive Concerns (C3Q) [23], measures calibrated using Rasch analysis allowing for the use of shorter versions. Also used were the Well Being Index [24] and the EQ-5D-5L [25].

Analyses

Distributional parameters were used to characterize the sample. Distributions for the importance ratings and the frequencies were presented separately for men and women. As the sample size was large, only differences between men and women of 10% or more were considered of importance [26]. The effects of strata, age, and gender on importance ratings and frequencies were assessed using logistic regression. For strata, the referent category was US English; for age, it was the youngest category, 65–70 years, and for gender men were compared with women.

The ordinality of the item responses for frequency was tested by fitting the data to the Rasch Model yielding indications of disordered thresholds and by viewing the category probability curves. Differential item functioning (dif) was tested for strata (country/language segment), age, and gender using analysis of variance and visual inspection of the magnitude of the effects. The structure of the measure was assessed using item-to-item correlations for both importance ratings and frequencies. Structural equation modeling (SEM) using mPLUS was used to identify the structure of the measure. Exploratory factor analysis (EFA) was used and fit statistics were estimated for one to seven factor models using the interpretations from Schermelleh-Engel et al. [27]. To identify how to create a scoring method for the 17 items, a principal component analysis (PCA) was conducted.

The distribution of OPAL scores, calculated as an unweighted sum of the ordinal ratings, was assessed for floor and ceiling effects. The relationship between scores on the OPAL measure and the converging constructs of physical, emotional, and psychological health and HRQL were assessed using Spearman correlation. As the OPAL measure taps a completely new construct, correlations with converging constructs were expected to be moderate (~0.5). The effects of selected sociodemographic and impairment related variables on OPAL scores were described using calculated

differences and effect sizes and adjusted estimates of difference through linear regression adjusted for age, sex, and strata.

Results

Characteristics of the sample

Table 1 presents a description of the sample according to socio-demographic characteristics. Table 2 presents a

Table 1 Socio-demographic characteristics of the sample

Variable	<i>N</i>
Strata	
Canadian English	301 (18.7%)
Canadian French	100 (6.2%)
Netherlands	203 (12.6%)
United Kingdom	403 (25.0%)
United States English	404 (25.1%)
United States Spanish	201 (12.5%)
Men/women	765/846 (47.5%/52.5%)
Age group	
65–70	823 (51.1%)
71–75	471 (29.2%)
76–80	219 (13.6%)
81+	99 (6.1%)
Ethnicity (could report more than 1)	
White European Ancestry	1416 (87.8%)
Hispanic/Latino Origin	130 (8.06%)
Black or African Ancestry	44 (2.73%)
Asian	44 (2.73%)
First Nations	25 (1.55%)
Middle Eastern/North African	5 (0.31%)
Not specified	8 (0.5%)
Education	
High school or less	574 (35.6%)
Certificate	433 (26.9%)
Bachelor's degree	417 (25.9%)
Postgraduate degree	188 (11.7%)
Health conditions (can report more than 1)	
None of the above	877 (54.4%)
Cardiovascular disease	241 (15.0%)
Musculoskeletal	163 (10.1%)
Respiratory	159 (9.9%)
Gastrointestinal disease	127 (7.9%)
Mental health	91 (5.7%)
Neurological	91 (5.7%)
Diabetes	62 (3.8%)
Cancer	13 (0.8%)
Not always enough money to meet needs	170 (10.6%)

description of the sample on functional, health, and social support indicators. The sample was predominantly between 65 to 75 years of age and of white European ancestry (87%). The education of the sample was evenly distributed across high school or less, certificate programs, or bachelor's degree, with 11% with post graduate degrees. More than half (54%) declared they had none of the listed health conditions; the most common health conditions listed, ranging from 15 to 9%, were cardiovascular disease, musculoskeletal and respiratory diseases; 10% of people reported that they did not always have money to meet their needs.

Table 2 presents characteristics of the sample on functional indicators. A small proportion of people had low vision (0%) or low hearing (8%) and 27% reported pain interfering with daily activities. Sixty percent (60%) of the sample reported having indicators of positive well-being, a smaller proportion 45% reported they woke up feeling fresh and rested, and 37% reported feeling active and vigorous. For physical function, 30% reported they would have difficulty walking for 30 min.

Evidence for importance of content

Table 3 shows importance and frequency ratings for men and women separately. To simplify the presentation and the analysis, the 5-level response option for importance was dichotomized to extremely or quite important vs. moderately, a little bit, or not at all important. Similarly, the 4-level response option for frequency was dichotomized at always or almost always vs. sometimes or rarely/never.

Across the “ways of being” items, importance ratings ranged between 60 and 90%. There were no differences between men and women of magnitude 10% or more. The top 3 important “ways of being” for men and women were mentally sharp, independent, and healthy mentally. The least important items were creative and active. Average importance values across the 5-level ordinal rating scale ranged from a low of 2.66 (creative) to a high of 3.45 (mentally sharp) yielding an importance ratio of highest to lowest of 1.29.

As there were 17 items and 6 strata, there were a total of 17*5 item-strata comparisons of importance ($n = 85$). Of these, there were only 4 (4.7%) where the importance rating was higher for one strata compared to US English: encouraged, USA Spanish (OR 1.6; 95% CI 1.1–2.4); involved, Netherlands (OR 1.5; 95% CI 1.0 to 2.2); motivated, Canadian French (OR 1.8; 95% CI 1.0–3.3) and UK (OR 1.4; 95% CI 1.0 to 2.0). This suggests that the harmonization process yielded items that were equally relevant across these strata except for a few exceptions that might have been expected owing to the large number of comparisons made. There were six gender differences by statistical criteria but none of the crude gender differences were quantitatively different. There

Table 2 Characteristics of the sample on functional indicators

	Rating	N (%)
Impairments		
Low vision	Yes	176 (10.9%)
Low hearing	Yes	136 (8.4%)
Pain interference with daily activities	Moderately to extremely	437 (27.1%)
Out of breath with ordinary activities	Often to always	147 (9.1%)
Tired did not feel like doing anything	Mean days (SD)	1.3 (1.8%)
Well Being		
Calm and relaxed	Most or all of the time	1041 (64.6%)
Cheerful and in good spirits	Most or all of the time	977 (60.6%)
Life filled with things that interest me	Most or all of the time	891 (55.3%)
Wake up feeling fresh and rested	Most or all of the time	736 (45.7%)
Active and vigorous	Most or all of the time	603 (37.4%)
Health indicators (EQ-5D-5L)		
Problems with self-care	No	1432 (88.8%)
Problems with performing usual activities	No	1069 (66.3%)
Anxious or depressed	No	1018 (63.2%)
Problems with walking about	No	909 (56.4%)
Pain or discomfort	No	460 (28.5%)
Physical function		
Difficulty walking for 30 min	Yes	498 (30.9%)
Walking between rooms	No difficulty	1420 (88.1%)
Performing light household activities	No difficulty	1313 (81.5%)
Putting on your shoes and socks	No difficulty	1265 (78.5%)
Lifting an object like a bag of groceries	No difficulty	1166 (72.3%)
Getting in or out of a car	No difficulty	1165 (72.3%)
Getting into or out of the bath	No difficulty	1074 (66.63%)
Walking 2 blocks	No difficulty	1034 (64.14%)
Walking a mile	No difficulty	765 (47.46%)
Performing usual work or housework	No difficulty	920 (57.07%)
Going up and down 10 steps	No difficulty	911 (56.51%)
Performing recreational activities	No difficulty	884 (54.84%)
Performing heavy household activities	No difficulty	516 (32.01%)
Cognitive function		
Forget if I have already done something	Rarely	1246 (77.3%)
Forget tasks or activities I need to do	Rarely	1184 (73.45%)
Forget what I was about to do	Rarely	1027 (63.71%)
Lose focus on a conversation	Rarely	1263 (78.35%)
Lose focus when doing complex tasks	Rarely	1243 (77.11%)
Lose focus on verbal instructions	Rarely	1220 (75.68%)
Lose focus and too many thoughts	Rarely	1029 (63.83%)
Health perception		
Self-rated health	Excellent to very good	538 (33.3%)
Health compared to past year	About the same	1089 (67.56%)
	Much or somewhat better	228 (14.3%)
Social support		
People to visit in their homes	5 or more people	742 (46.03%)
Times in past week talk to friends etc	Daily	579 (35.92%)
Spend time people outside family	Daily +	147 (9.12%)
Someone you trust and can confide in	Yes	1479 (91.75%)
Lonely	Almost never	1036 (64.27%)

Table 3 Item specific importance and frequency ratings

	Importance: extremely or quite		Mean* (SD)	Frequency: always or almost always	
	Men	Women		Men	Women
Active	543 (71.0%)	589 (69.6%)	2.92	202 (26.4%)	216 (25.5%)
Confident	617 (80.7%)	686 (81.1%)	3.15	305 (39.9%)	282 (33.3%)
Connected	558 (72.9%)	657 (77.7%)	3.01	264 (34.5%)	278 (32.9%)
Useful	587 (76.7%)	680 (80.4%)	3.09	259 (33.9%)	275 (31.2%)
Creative	451 (58.9%)	511 (60.4%)	2.66	113 (15.7%)	134 (15.8%)
Encouraged	504 (65.9%)	608 (71.9%)	2.85	131 (17.1%)	131 (15.5%)
Energetic	519 (67.8%)	592 (70.0%)	2.83	129 (16.9%)	145 (17.1%)
Involved	525 (68.6%)	618 (73.1%)	2.91	180 (23.5%)	185 (21.9%)
Happy	614 (80.3%)	690 (81.6%)	3.16	255 (33.3%)	251 (29.7%)
Healthy mentally	640 (83.7%)	727 (85.9%)	3.31	368 (48.1%)	354 (41.8%)
Healthy physically	615 (80.4%)	705 (83.3%)	3.23	212 (27.7%)	231 (27.3%)
Independent	670 (87.6%)	758 (89.6%)	3.41	404 (52.9%)	470 (55.6%)
Interested	625 (81.7%)	725 (85.7%)	3.21	318 (41.6%)	347 (41.0%)
Mentally sharp	681 (89.0%)	758 (89.6%)	3.45	360 (47.1%)	357 (43.4%)
Motivated	579 (75.7%)	674 (79.7%)	3.07	233 (30.5%)	244 (28.8%)
Resilient	596 (77.9%)	698 (82.5%)	3.16	263 (34.4%)	304 (35.9%)
Self-sufficient	596 (77.9%)	698 (82.5%)	3.42	383 (50.1%)	457 (54.0%)

Bold items show statistically different ratings between men and women

* Mean rating across the 5-point ordinal scale from 0 to 4

CV is coefficient of variation (SD/Mean) and indicates how much variability there is in the values

were four age group differences with older age consistently endorsing higher importance. Full details are presented in Supplementary Table 1.

For the 85 item-strata frequency comparisons, there were 25 (29.4%) where there were differences across strata with 21 differences showing less endorsement of higher frequency of active living items in countries other than the USA. Full details presented in Supplementary Table 1.

Evidence for response categories

The response category for the frequency ratings was chosen from the list recommended by Mutebi et al. [21] as having linear properties. The ordinality and linearity was tested by fitting the response data to the Rasch Model. There were no disordered thresholds and visualization of the category probability curves and the score structure map supported linearity. Figure 1 shows how the different response options are distributed across a standardized OPAL score presented along the x-axis. The ordinal scoring structure has been converted to a 1 to 4 range rather than the original 0 to 3 structure (1: rarely or never; 2: sometimes; 3: often; 4: always or almost). With 4 response options there are three thresholds which are the boundaries between 1 and 2, 2 and 3, and 3 and 4. The width of the bars is quite similar showing that distance between the categories is quasi-equal.

Evidence for structure

Item-to-item correlations ($n = 136$) for importance ratings and frequencies are presented in Table 4. Thirteen pairs of importance items were correlated at 0.8 to < 0.87 and an additional 46 were correlated at 0.7 to < 0.8 ; the number of similarly strongly correlated frequency item-pairs were 12 and 47. Apart from “creative” all other items had some degree of strong correlation with other items. The item-to-item correlations for frequency inform the extent to which the items measure the same construct. The average correlation across items was 0.67 (SD: 0.08) with a range from 0.39 to 0.86.

Originally, the OPAL measure was conceptualized as measuring a construct similar to HRQL with domains related to physical, emotional, social, and psychological health. A confirmatory factor analysis (CFA) failed to provide a solution and hence an EFA was used. The scree plot, shown in Fig. 3, indicated that a two factor model was supported by the data with the first eigenvalue for the sample correlation matrix of 11.7, the second 1.1 and the third, 0.9. The 7-factor model showed good fit according to all the criteria and is presented in Table 5. This model showed only 1 item with a cross-factor loadings (> 0.3). There were four 3-item factors, two 2-item factors, and one 1-item factor. Interestingly, the item groupings made theoretical sense but the number of items per factor is below the recommended number for

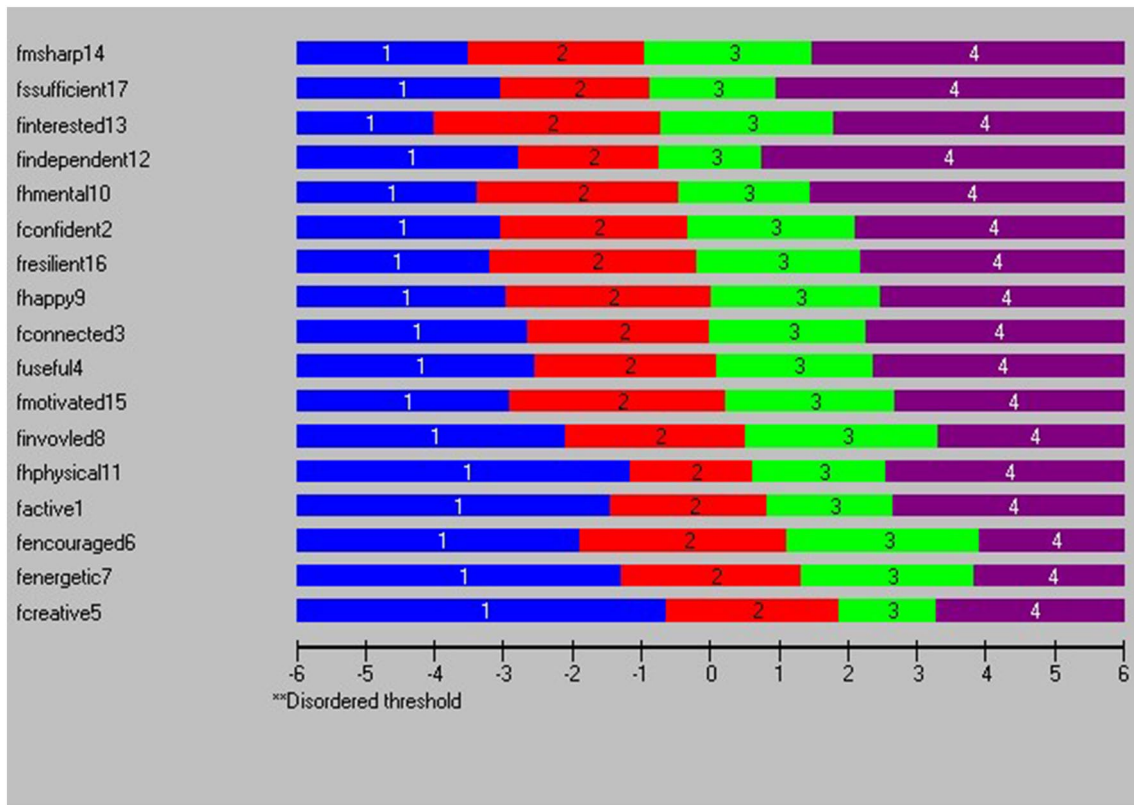


Fig. 1 Distribution of item categories over the OPAL items

Table 4 Item-to-item correlations among importance and prevalence ratings

Importance / Prevalence	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 Active		0.71	0.65	0.67	0.59	0.63	0.79	0.64	0.56	0.58	0.72	0.61	0.62	0.61	0.67	0.63	0.61
2 Confident	0.66		0.80	0.81	0.64	0.68	0.69	0.71	0.70	0.74	0.61	0.69	0.73	0.72	0.66	0.72	0.68
3 Connected	0.61	0.86		0.84	0.63	0.72	0.67	0.78	0.71	0.68	0.56	0.59	0.71	0.65	0.67	0.66	0.62
4 Useful	0.66	0.80	0.84		0.68	0.70	0.69	0.77	0.72	0.70	0.59	0.65	0.74	0.68	0.70	0.70	0.68
5 Creative	0.53	0.51	0.52	0.58		0.63	0.59	0.65	0.56	0.51	0.42	0.47	0.61	0.49	0.63	0.57	0.49
6 Encouraged	0.58	0.71	0.74	0.75	0.70		0.75	0.79	0.69	0.63	0.55	0.56	0.68	0.58	0.73	0.68	0.56
7 Energetic	0.82	0.70	0.66	0.71	0.60	0.73		0.80	0.72	0.70	0.72	0.62	0.68	0.65	0.74	0.69	0.63
8 Involved	0.64	0.73	0.80	0.81	0.61	0.80	0.78		0.77	0.72	0.58	0.61	0.77	0.67	0.75	0.71	0.65
9 Happy	0.55	0.77	0.76	0.73	0.51	0.73	0.68	0.75		0.85	0.62	0.64	0.74	0.73	0.65	0.66	0.66
10 Healthy mentally	0.52	0.77	0.73	0.70	0.48	0.65	0.62	0.71	0.85		0.65	0.70	0.73	0.84	0.65	0.71	0.72
11 Healthy physically	0.77	0.60	0.53	0.62	0.45	0.54	0.77	0.59	0.58	0.58		0.73	0.69	0.76	0.68	0.68	0.69
12 Independent	0.59	0.63	0.57	0.63	0.39	0.51	0.60	0.60	0.61	0.63	0.67		0.78	0.82	0.65	0.75	0.86
13 Interested	0.60	0.73	0.73	0.74	0.59	0.70	0.66	0.77	0.74	0.75	0.59	0.73		0.82	0.80	0.78	0.73
14 Mentally sharp	0.55	0.69	0.63	0.65	0.46	0.59	0.61	0.65	0.66	0.77	0.56	0.70	0.80		0.72	0.77	0.80
15 Motivated	0.70	0.73	0.71	0.73	0.60	0.75	0.77	0.77	0.75	0.73	0.67	0.65	0.85	0.77		0.84	0.73
16 Resilient	0.62	0.71	0.67	0.68	0.52	0.65	0.70	0.72	0.69	0.69	0.65	0.69	0.78	0.76	0.83		0.81
17 Self-sufficient	0.57	0.62	0.56	0.62	0.42	0.50	0.59	0.59	0.60	0.64	0.64	0.84	0.72	0.71	0.68	0.77	
Item-to-total	0.75	0.81	0.80	0.82	0.63	0.78	0.81	0.83	0.80	0.77	0.73	0.71	0.82	0.75	0.85	0.81	0.71

Darkest colours show correlation ≥ 0.8 considered strong in the context of measurement theory; lighter shading shows correlations between 0.70 and 0.79

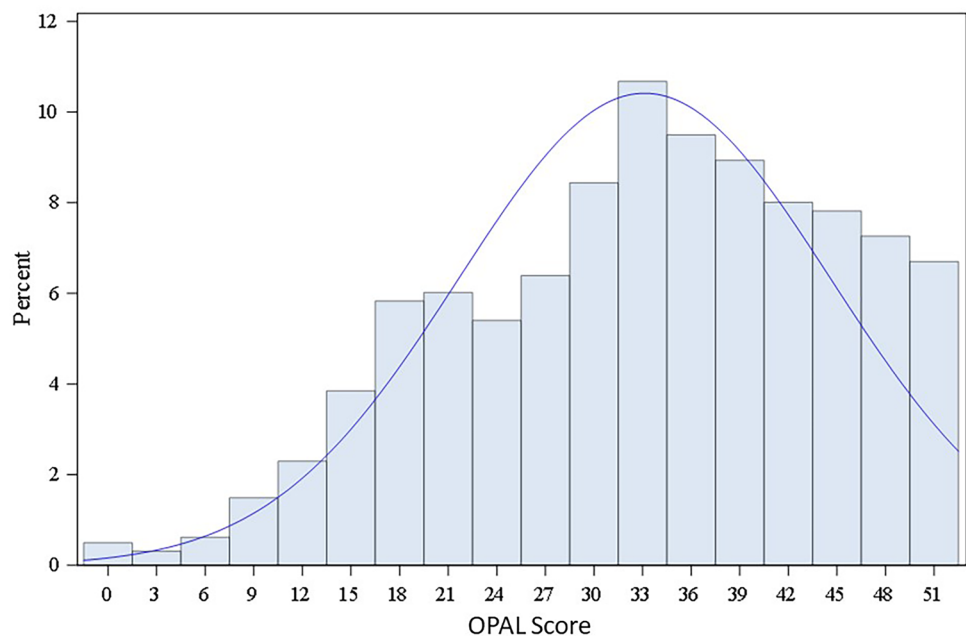
reliable domain representation [28, 29]. As the inter-item correlations presented in Table 4 were all high as were the item-to-total correlations mean 0.77), the evidence is not

overwhelming for a measure with multiple domains. The PCA showed that the first PC explained 68.9% of the variance, indicative of one strong component. The PCA weights

Table 5 Factor structure of the OPAL measure

Item	Ways of being	Number of factors						
		1	2	3	4	5	6	7
3	Connected	0.951	0.054	- 0.041	- 0.022	0.019	0.024	- 0.019
2	Confident	0.612	- 0.023	0.123	0.009	0.206	0.051	0.057
4	Useful	0.538	0.278	0.070	0.134	0.009	- 0.019	0.01
6	Encouraged	0.054	0.901	- 0.048	0	0.043	0.011	- 0.021
5	Creative	- 0.072	0.717	0.098	0.005	- 0.099	- 0.031	0.159
8	Involved	0.271	0.512	0.043	0.049	0.042	0.066	0.039
1	Active	0.244	0	0.774	0.022	- 0.076	- 0.051	0.123
11	Healthy physically	- 0.032	0.021	0.627	0.225	0.156	0.075	- 0.061
7	Energetic	0.048	0.355	0.557	- 0.017	0.056	0.095	- 0.016
12	Independent	0.042	0.028	0.03	0.911	0	- 0.012	0.012
17	Self-sufficient	- 0.008	- 0.015	0.013	0.631	0.019	0.309	0.036
10	Healthy mentally	0.024	0.031	0.017	0.013	0.862	- 0.032	0.15
9	Happy	0.176	0.262	- 0.001	0.027	0.505	0.044	0.002
16	Resilient	0.060	0.046	0.015	0.069	- 0.018	0.831	0.035
13	Interested	0.085	0.256	- 0.040	0.250	0.022	0.043	0.504
14	Mentally sharp	- 0.001	- 0.007	0.019	0.197	0.243	0.193	0.401
15	Motivated	0.011	0.252	0.199	- 0.027	0.073	0.280	0.345
Fit statistics								
	RMSEA (<0.05)	0.130	0.119	0.093	0.080	0.067	0.054	0.048
	RMSEA 90%CI	0.126–0.133	0.115–0.123	0.08–0.10	0.075–0.085	0.061–0.072	0.047–0.060	0.041–0.055
	CFI (0.97)	0.941	0.957	0.978	0.986	0.992	0.996	0.997
	SRMR (<0.05)	0.065	0.047	0.028	0.020	0.015	0.01	0.008

RMSEA root mean square error of approximation; CFI comparative fit index; SRMR standardized root mean square residual

Fig. 2 Distribution of OPAL scores

were similar across the 17 items: mean 0.82 (SD:0.05); range 0.67 to 0.90.

Evidence for OPAL scores and other constructs

Based on the linearity of the response options for frequency [21] (see Fig. 1) and the relatively similar importance ratings for each item (see Table 3), a simple linear sum was considered mathematically justified. Figure 2 shows that this approach yielded a distribution that was near normal.

Table 6 shows that correlations with the constructs related to physical function (LEFS), self-reported cognitive ability (C3Q), and HRQL (EQ-5D value) were all in the expected moderate range (~ 0.5).

Table 7 shows how the OPAL measure differs across constructs known to affect participation and HRQL. The OPAL score across all strata was 33.1 (SD: 11.5) with a maximum score of 51 (17 items scored on a 4-point scale, 0 to 3). There were only small differences in frequency across strata. There were no differences between men and women. The oldest groups (76+ years) scored slightly higher than the youngest group (65–70 years),

The adjusted differences between categories of variables hypothesized to influence active living ranged from -3.10 (low vision) to a high of -11.82 for being out of breath with ordinary activities. Given the overall SD was 11.5, these differences yield very large effect sizes. A medium effect size ($1/2$ SD) was observed for low hearing (-6.36) and having a health condition (-6.66) suggesting a Minimum Important Difference (MID) of 6. Large effect sizes, ≥ 0.8 were observed for pain, fatigue, and shortness of breath.

Discussion

The results of this study provide evidence that the 17-item OPAL measure is fit for the purpose of estimating the extent to which older persons are living actively at one point in time. Content representation was previously demonstrated and evidence presented in Table 3 shows homogeneity of importance ratings across population strata suggesting that the language versions were similar. The linearity of the response options was confirmed using item-characteristic curves, visualization of the threshold map (Fig. 1) and there was lack of differential item functioning across strata, age, and sex.

There was evidence that the items related to the same underlying construct as the item-to-item correlations (see Table 5) were majoritarily strong. The strong item-to-item correlations also suggest item redundancy and shorter versions may be possible and will be the topic of future research.

Although statistically a seven-factor model fit the data (see Table 4), each factor comprised very few items and item-to-item correlations were very high suggesting that one factor is plausible. In addition, although the OPAL measure is conceptualized as a formative model, all items fit the Rasch model suggesting unidimensionality.

These results provide evidence that a simple linear sum of the responses to the 17 items is a mathematically valid way of deriving a total score. While there were some differences in importance ratings, the ratio of the highest to lowest importance was 1.29.

The OPAL measure showed a low ceiling and floor effect with 6.2% of values at the maximum and only 5 values at 0. The distribution was near normal with a mean of 33.1 (out of 51) and a SD of 11.5 as shown in Fig. 2. Correlations with other measures of converging constructs were of

Table 6 Relationships between scores on OPAL and converging constructs

Measure (theoretical range)	Mean (SD) [Observed range]	CV	Correlation with OPAL (95% CI)
Older persons active living OPAL(0–51)	33.1 (11.5) [0 to 51]	0.35	–
Self-reported physical function LEFS (0–48)	39.6 (10.3) [0 to 48]	0.26	0.51 (0.47–0.54)
Self-reported cognitive ability C3Q (0–14)	11.8 (3.0) [0 to 14]	0.26	0.42 (0.38–0.46)
Well-being WHO (0–25)	16.0 (5.3)	0.33	0.78 (0.76–0.80)
HRQL (max 1.0) EQ-5D Value	0.82 (1.6) [– 0.15 to 0.95]	0.19	0.56 (0.53–0.59)
Degree of active living Rating (0–10)	7.0 (2.2) [0 to 10]	0.32	0.50 (0.46–0.53)

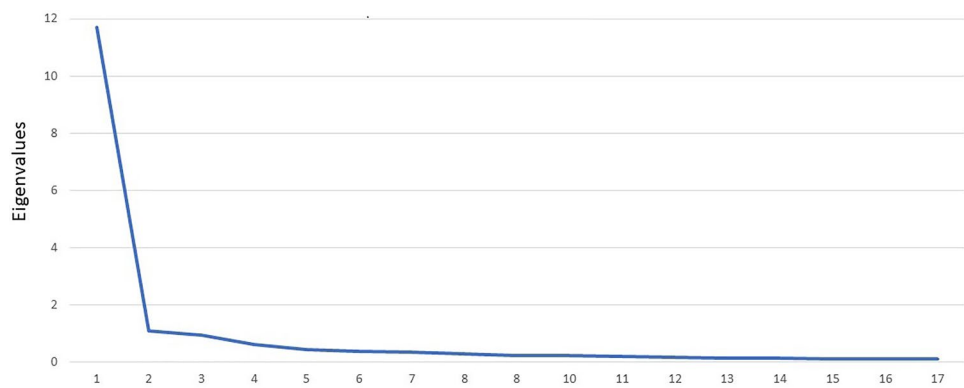
Table 7 Differences in OPAL scores across functional status indicators

Known group variables	N	Mean (SD)	Adjusted difference	95% CI
Overall OPAL Score (max. 51)	1612	33.1 (11.5)		
Canadian English	301	32.0 (11.4)		
Canadian French	100	33.6 (11.8)		
Netherlands	203	35.3 (11.0)		
United Kingdom	403	31.5 (11.2)		
United States English	404	33.9 (11.8)		
United States Spanish	201	34.0 (11.4)		
Men	765	33.7 (11.0)		
Women	846	32.6 (11.9)		
Age category				
65–70	823	32.5 (11.7)	Referent	
71–75	471	33.1 (11.7)	0.64	– 0.65, 1.94
76–80	219	34.4 (10.7)	1.83	0.12, 3.54
81 +	99	36.1 (10.3)	3.60	1.21, 5.99
Reported health condition				
No	1019	35.6 (11.0)	Referent	
Yes	593	28.9 (11.2)	– 6.66	– 7.77, – 5.54
Low vision				
No	1436	33.5 (11.2)	Referent	
Yes	176	30.4 (13.4)	– 3.10	– 4.88, – 1.31
Low hearing				
No	1476	33.6 (11.3)	Referent	
Yes	136	27.8 (12.2)	– 6.36	– 8.35, – 4.36
Pain interference with daily life				
No	1175	35.4 (10.8)	Referent	
Moderately, quite a bit, extremely	437	27.1 (11.1)	– 8.29	– 9.48, – 7.10
Out of breath with ordinary activities				
No	1465	34.2 (11.0)	Referent	
Yes	147	22.3 (10.3)	– 11.82	– 13.67, – 9.96
So tired did not feel like doing anything at least one day in a week				
No	851	37.7 (10.0)	Referent	
Yes	761	28.0 (10.9)	– 9.58	– 10.59, – 8.56
Do you have difficulty walking continuously for 30 min?				
No	1114	35.7 (10.7)	Referent	
Yes	498	27.4 (11.1)	– 8.59	– 9.73, – 7.45
Self-rated health				
Excellent	97	45.4 (5.8)		
Very good	441	39.4 (8.6)		
Good	644	33.3 (9.8)		
Fair	354	24.3 (9.9)		
Poor	76	20.4 (11.2)		
Excellent/very good	538	40.4 (8.4)	Referent	
Good/poor	1074	29.5 (11.1)	– 10.94	– 12.00, – 9.89

Differences are adjusted for age, sex, and segment

moderate strength (~0.50; see Table 6) and the differences across groups known to affect functioning and health were observed suggesting an MID of 6 out of 51 or near 12 when scores are converted to be out of 100 (see Table 7).

The findings from the validation analyses are typical of measures in the field. For example, respondents during the testing of the Older People's Quality of Life questionnaire (OPQOL-brief) [30], comprising 13 items rated on a 1–5 agreement scale (higher is better), showed an average

Fig. 3 Scree Plot from exploratory factor analysis

response value for each item ranging from 3.94 to 4.39. For the OPAL measure, with responses ranging from 0 to 3, the proportion responding 2 or 3 ranged from 15.5% (encouraged) to 55.6% (independent) as shown in Table 3 providing some evidence that the OPAL measure has a lower ceiling effect than the OPQOL-Brief. Item-to-total correlations for the OPQOL measure ranged from 0.36 (I get pleasure from my home) to a high of 0.67 (I enjoy my life overall). For the OPAL measure, the item-to-item correlations were higher ranging from 0.71 to 0.85, supporting the evidence from the factor analysis (Table 4) that a one-factor model is plausible. The total score on the OPQOL-brief was moderately correlated with variables hypothesized to influence QoL such as self-rated active aging (ρ 0.503) and self-rated health status (ρ 0.517). Our values for similar constructs were 0.50 for active living and 0.56 for EQ-5D. Similarly for physical function, correlation with OPQOL-brief was 0.43 and for OPAL and physical function, the correlation of 0.51 (Fig. 3).

The PROMIS-29 is another measure proposed for use among older adults with multiple chronic conditions [31]. The evidence they present is that there were strong differences in the physical and mental health scores (PHS; MHS) across groups expected to differ. Some of these differences make relevant comparisons to validity evidence presented for OPAL. Scores on PROMIS-29 decreased according to number of chronic conditions with similar magnitude to what was found for OPAL for people reporting a health condition (see Table 7). The other comparisons are less relevant as OPAL was developed for use by those living independently in their community. There was evidence presented that PROMIS-29 PHS correlated with physically associated scales (similar to item-to-total correlations) with ranges 0.63 to 0.81 similar to the item-to-item correlations for OPAL (see Table 4).

Not every type of validity evidence was possible to assess in this study. Further work is needed to derive evidence for test-retest reliability and how scores change in response to intervention. Generalizing the results of this study outside of these four settings and languages also needs the relevant

evidence. The development of a measure is an ongoing process and is based on the accumulation of evidence that the data arising from the measure is fit for intended purposes. “Validity” does not belong to the measure, it is a property of the data arising from the measure.

What construct does the OPAL measure? According to the ISOQOL dictionary [19] HRQL is *a term referring to the health aspects of quality of life, generally considered to reflect the impact of disease and treatment on disability and daily functioning*. This is in contrast to quality of life which, using the ISOQOL dictionary definition from the World Health Organization refers to an *individuals’ perception of their position in life in the context of the culture in which they live and in relation to their goals, expectations, standards and concerns*. The OPAL measure would seem to measure older persons active living-related quality of life (OPALQOL).

Conclusion

The results of this study provide evidence that the 17-item OPAL measure is fit for the purpose of estimating the extent to which older persons are living actively at one point in time. Further work is needed identify how the OPAL measure performs over time and as a way to assess an intervention and whether shorter forms would provide similar evidence. Research among older persons from other cultures and languages would provide needed evidence as to the applicability of this active living construct globally.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11136-024-03720-1>.

Author contributions NM, AK, MA, RB, JB, HD, IK, ET, KS, CW, KM contributed to the study conception and design. Material preparation, NM, AK, MA, RB, JB, HD, IK, ZA, EE, LN, CW. Data collection and analysis were performed by Nancy Mayo & Lyne Nadeau. The

first and final drafts of the manuscript was written by Nancy Mayo. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Canadian Institutes of Health Research (PJT-173538). Funding from Healthy Brains for Health Lives (McGill University) supported the Brain Health Outcomes Platform (BHOP) that was used for data entry.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose relative to this work.

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Research Ethics Board for the McGill University Health Center (21-06-1011; HRQL-OPAL/2021-7254) (OPAL-Survey Phase II/2023-9422).

Consent to participate Informed consent was obtained from all individual participants included in the study.

References


- World Health Organization. (2002). Active aging: A policy framework.
- Salbach, N. M., Barclay, R., Webber, S. C., Jones, C. A., Mayo, N. E., Lix, L. M., et al. (2019). A theory-based, task-oriented, outdoor walking programme for older adults with difficulty walking outdoors: Protocol for the Getting Older Adults Outdoors (GO-OUT) randomised controlled trial. *British Medical Journal Open*, 9(4), e029393.
- Leahy, A. (2023). Disability identity in older age? - Exploring social processes that influence disability identification with ageing. *Disability Studies Quarterly*. <https://doi.org/10.18061/dsq.v42i3-4.7780>
- Kaundinya, T., & Schroth, S. (2022). Dismantle ableism, accept disability: Making the case for anti-ableism in medical education. *Journal of Medical Education and Curricular Development*, 9, 23821205221076660.
- Titchkosky, T. (2003). *Disability, self, and society*. University of Toronto Press.
- Cole, T. R. (1992). *The journey of life: A cultural history of aging in America*. Cambridge University Press.
- Peterson, C. H., Gischlar, K. L., & Peterson, N. A. (2017). Item construction using reflective, formative, or Rasch measurement models: Implications for group work. *The Journal for Specialists in Group Work*, 42(1), 17–32.
- Stenner, A. J., Burdick, D. S., & Stone, M. H. (2008). Formative and reflective models: Can a Rasch analysis tell the difference? *Rasch Measurement Transactions*, 22(1), 1152–1153.
- Myszkowski, N., Storme, M., & Tavani, J.-L. (2019). Are reflective models appropriate for very short scales? Proofs of concept of formative models using the Ten-Item Personality Inventory. *Journal of Personality*, 87(2), 363–372.
- Food and Drug Administration (FDA). (2022). Patient-focused drug development: Selecting, developing, or modifying fit-for-purpose clinical outcome assessments.
- American Educational Research Association. (2014). *The standards for educational and psychological testing*. American Educational Research Association.
- Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457.
- Hawkins, M., Elsworth, G. R., Nolte, S., & Osborne, R. H. (2021). Validity arguments for patient-reported outcomes: Justifying the intended interpretation and use of data. *Journal of Patient-Reported Outcomes*, 5(1), 64.
- Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, 2018(9), 2207–2230.
- Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., et al. (2001). Frailty in older adults: Evidence for a phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 56(3), M146–M156.
- Rockwood, K., Song, X., MacKnight, C., Bergman, H., Hogan, D. B., McDowell, I., & Mitnitski, A. (2005). A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal*, 173(5), 489–495.
- Mayo, N. E., Aubertin-Leheudre, M., Mate, K., Figueiredo, S., Fiore, J. F., Jr., Auais, M., et al. (2023). Development of a frailty ladder using rasch analysis: If the shoe fits. *Canadian Geriatrics Journal*, 26(1), 133–143.
- Cesari, M., Araujo, D. C. I., Amuthavalli, T. J., Cooper, C., Martin, F. C., Reginster, J. Y., et al. (2018). Evidence for the domains supporting the construct of intrinsic capacity. *The Journals of Gerontology Series A*, 73, 1653–1660.
- Mayo, N. E. (2015). *Dictionary of quality of life and health outcomes measurement*. International Society for Quality of Life.
- Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *Journal of the American Medical Association*, 273(1), 59–65.
- Mutebi, A., Slack, M., Warholak, T. L., Hudgens, S., & Coons, S. J. (2016). Interpretation of verbal descriptors for response options commonly used in verbal rating scales in patient-reported outcome instruments. *Quality of Life Research*, 25(12), 3181–3189.
- Binkley, J. M., Stratford, P. W., Lott, S. A., & Riddle, D. L. (1999). The Lower Extremity Functional Scale (LEFS): Scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. *Physical Therapy*, 79(4), 371–383.
- Askari, S., Fellows, L. K., Brouillette, M. J., & Mayo, N. E. (2021). Development and validation of a voice-of-the-patient measure of cognitive concerns experienced by people living with HIV. *Quality of Life Research*, 30(3), 921–930.
- Topp, C. W., Ostergaard, S. D., Sondergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, 84(3), 167–176.
- EuroQol G. (2016). EQ-5D <http://www.euroqol.org/>.
- Austin, P. C. (2009). Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in Statistics - Simulation and Computation*, 38(6), 1228–1234.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219–246.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299.
- Bowling, A., Hankins, M., Windle, G., Bilotta, C., & Grant, R. (2013). A short measure of quality of life in older age: The performance of the brief Older People's Quality of Life questionnaire (OPQOL-brief). *Archives of Gerontology and Geriatrics*, 56(1), 181–187.
- Rose, A. J., Bayliss, E., Huang, W., Baseman, L., Butcher, E., Garcia, R. E., & Edelen, M. O. (2018). Evaluating the PROMIS-29 v20 for use among older adults with multiple chronic conditions. *Quality of Life Research*, 27(11), 2935–2944.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the

author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Nancy E. Mayo^{1,2,3}  · Mohammad Auais⁴ · Ruth Barclay^{5,6} · Joan Branin⁷ · Helen Dawes⁸ · Ida J. Korfage⁹ · Kim Sawchuk¹⁰ · Eran Tal¹¹ · Carole L. White¹² · Zain Ayoubi¹³ · Ezinne Ekediegwu¹³ · Kedar Mate¹⁴ · Lyne Nadeau¹³ · Sebastian Rodriguez Duque¹¹ · Ayse Kuspinar¹⁵

✉ Nancy E. Mayo
nancy.mayo@mcgill.ca

Mohammad Auais
mohammad.auais@queensu.ca

Ruth Barclay
ruth.barclay@umanitoba.ca

Joan Branin
Jbraninphd@gmail.com

Helen Dawes
h.dawes@exeter.ac.uk

Ida J. Korfage
i.korfage@erasmusmc.nl

Kim Sawchuk
kim.sawchuk@concordia.ca

Eran Tal
eran.tal@mcgill.ca

Carole L. White
whitec2@uthscsa.edu

Zain Ayoubi
zain.ayoubi@rimihc.ca

Ezinne Ekediegwu
ezinne.ekediegwu@mail.mcgill.ca

Kedar Mate
kedar.mate@mcgill.ca

Lyne Nadeau
lyne.nadeau@mcgill.ca

Sebastian Rodriguez Duque
sebastian.rodriguezduque@mail.mcgill.ca

Ayse Kuspinar
kuspinaa@mcmaster.ca

¹ Department of Medicine School of Physical and Occupational Therapy, McGill University, Montreal, Canada

² Divisions of Clinical Epidemiology, Geriatrics, Experimental Medicine, McGill University Health Center (MUHC), Montreal, Canada

³ Center for Outcomes Research and Evaluation (CORE), MUHC-Research Institute, 5252 de Maisonneuve, Montreal, QC H4A 3S5, Canada

⁴ School of Rehabilitation Therapy, Queen's University, Kingston, ON, Canada

⁵ Department of Physical Therapy, College of Rehabilitation Sciences, Rady Faculty of Health Sciences, University of Manitoba, R106-771 McDermot Ave, Winnipeg, MB R3E 0T6, Canada

⁶ Rehabilitation Hospital, RR323A, 800 Sherbrook St, Winnipeg, Canada

⁷ Center for Health and Aging, Pasadena, CA 91106, USA

⁸ Department of Public Health & Sports Sciences, Faculty of Health and Life Sciences, University of Exeter Medical School, University of Exeter, Medical School Building, College of Medicine and Health, St Lukes Campus, Heavitree Road, Exeter EX12LU, UK

⁹ Department of Public Health, Erasmus MC University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands

¹⁰ Concordia University, Montreal, QC, Canada

¹¹ Department of Philosophy, McGill University, Leacock Building, Room 933, Montreal, Canada

¹² School of Nursing, UT Health San Antonio, San Antonio, USA

¹³ Center for Outcomes Research and Evaluation (CORE), Research Institute of McGill University Health Center, Montreal, Canada

¹⁴ Faculty of Medicine, McGill University, Montreal, Canada

¹⁵ School of Rehabilitation Science, McMaster University, 1400 Main Street West, Hamilton, ON L8S 1C7, Canada