

RESEARCH

Open Access



# Psychometric properties of the EQ-5D-5L for aboriginal Australians: a multi-method study

Pedro Henrique Ribeiro Santiago<sup>1,2\*</sup> , Dandara Haag<sup>1,2</sup>, Davi Manzini Macedo<sup>1</sup>, Gail Garvey<sup>3</sup>, Megan Smith<sup>4,5</sup>, Karen Canfell<sup>4,5</sup>, Joanne Hedges<sup>1</sup> and Lisa Jamieson<sup>1</sup>

## Abstract

**Introduction:** In Australia, health-related quality of life (HRQoL) instruments have been adopted in national population surveys to inform policy decisions that affect the health of Aboriginal and Torres Strait Islanders. However, Western-developed HRQoL instruments should not be assumed to capture Indigenous conceptualization of health and well-being. In our study, following recommendations for cultural adaptation, an Indigenous Reference Group indicated the EQ-5D-5L as a potentially valid instrument to measure aspects of HRQoL and endorsed further psychometric evaluation. Thus, this study aimed to investigate the construct validity and reliability of the EQ-5D-5L in an Aboriginal Australian population.

**Methods:** The EQ-5D-5L was applied in a sample of 1012 Aboriginal adults. Dimensionality was evaluated using Exploratory Graph Analysis. The Partial Credit Model was employed to evaluate item performance and adequacy of response categories. Area under the receiver operating characteristic curve (AUROC) was used to investigate discriminant validity regarding chronic pain, general health and experiences of discrimination.

**Results:** The EQ-5D-5L comprised two dimensions, Physiological and Psychological, and reliability was adequate. Performance at an item level was excellent and the EQ-5D-5L individual items displayed good discriminant validity.

**Conclusions:** The EQ-5D-5L is a suitable instrument to measure five specific aspects (Mobility, Self-Care, Usual activities, Pain/Discomfort, Anxiety/Depression) of Aboriginal and Torres Strait Islander HRQoL. A future research agenda comprises the investigation of other domains of Aboriginal and Torres Strait Islander HRQoL and potential expansions to the instrument.

## Introduction

It has been increasingly recognized by health policy makers, health practitioners and many population groups that biological parameters and clinical measures of disease are insufficient indicators of health status [1–4]. A growing body of evidence has documented the importance of subjective experiences and interpretation of health and illness on individuals' quality of life [5]. This idea is central to the concept of Health-Related Quality of Life (HRQoL), which encompasses individuals' evaluations

of physical, psychological, and social well-being associated with their health state. Health-related quality of life assessments are useful in the fields of research, clinical practice and policy making. From a research perspective, these instruments can be adopted to assess HRQoL in clinical and epidemiological studies in order to assess the impacts associated with health conditions on individual's well-being, as well as outcomes of healthcare interventions. In terms of policy implications, HRQoL assessments are useful for surveillance, as they support the development of evidence-based public health strategies, guiding the allocation of scarce resources. When adopted in healthcare settings, HRQoL measurements

\*Correspondence: [pedro.ribeirosantiago@adelaide.edu.au](mailto:pedro.ribeirosantiago@adelaide.edu.au)

<sup>1</sup> Adelaide Dental School, The University of Adelaide, Adelaide, Australia  
Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

are a useful communication tool for identifying and prioritizing patient problems and preferences.

Although the importance of such measures is well established, a longstanding question remains on the applicability of such instruments among Indigenous populations [6]. This challenge arises from the different frames of reference in conceptualising health among Indigenous populations and Western societies, and is intensified by the narrow focus of the multiple HRQoL measures in biological aspects of health. Researching the field indicates that, while HRQoL instruments claim to capture subjective perceptions related to health, multiple HRQoL tools are actually generic health status measures [7, 8].

In Australia, the National Aboriginal Community Controlled Health Organisation (NACCHO) defines health and wellbeing as:

*“... not just the physical well-being of an individual but the social, emotional and cultural well-being of the whole Community in which each individual is able to achieve their full potential as a human being thereby bringing about the total well-being of their Community. It is a whole of life view and includes the cyclical concept of life-death-life.”* [9]

This definition provides clear evidence that Aboriginal and/or Torres Strait Islander Australians have a concept of health and wellbeing that is distinct from the Western definition, encompassing elements such as community wellbeing and spiritual and cultural entities. Such conceptualization presents a holistic and multidimensional view of health and well-being. A comprehensive review of 95 articles by Butler, Anderson [6] identified nine domains of importance to the health and well-being of Indigenous Australians (autonomy, empowerment and recognition; family and community; culture, spirituality and identity; Country; basic needs; work, roles, and responsibilities; education; physical health; and mental health). Some of these domains are vaguely or not at all included in most Western HRQoL tools, such as spirituality, connection to Country and culture [6].

Theoretical criticisms for the use of HRQoL measures contribute further to the limitations in using such tools among Indigenous populations. One of the most important criticism relates to the strong emphasis on functional and role limitations placed by HRQoL measures, which may fail to assess the actual importance of these events on individuals' lives. This questions to what extent the meaning of the impacts of diseases are assessed according to individuals' personal beliefs, a central dimension for an Indigenous' holistic view of health. Equally important is the consideration that since population perceptions of health and well-being vary across time and space,

HRQoL tools are limited in the way they take these geographical and historical specificities into consideration. This is especially important for Aboriginal and Torres Strait Islander Australians, who are extremely diverse within themselves and who have been exposed to different geographical, political, economic and socio-historical determinants of health since colonisation [6].

Despite the abovementioned criticisms of the use of traditional HRQoL tools in the context of Indigenous health, these instruments have been largely uncritically adopted in national population surveys to inform policy decisions that affect the health of Aboriginal Australians. These include policy decisions such as the medication subsidy and health care performance evaluation (Department of Health: Canberra; 2016). To the best of our knowledge, there has been no HRQoL instrument developed that has been designed by and validated specifically for use among Australia's Aboriginal and Torres Strait Islander population. While a more comprehensive and culturally appropriate HRQoL tool is not yet developed, it is of paramount importance that those being currently employed in research and policy development that affects Australian Aboriginal and Torres Strait Islander peoples provide some meaningful insights into the HRQoL of this population.

### Present research

In the current study, in partnership with Aboriginal groups in South Australia and following recommendations for the cultural adaptation of instruments [10], an Indigenous Reference Group was established and consulted regarding the face and content validity of a prominent HRQoL instrument, the EQ-5D-5L. The EQ-5D is a 5-item measure for describing and valuing health [11]. Over the decades, it became the most well-known and commonly applied instruments to measure HRQoL, being used both in clinical and non-clinical populations and translated to more than 160 languages [12, 13]. While the EQ-5D original instrument had only three response categories, the EQ-5D-5L was later developed to include five response categories, showing better discriminant capacity, increased reliability and reduced ceiling effects [13].

The EQ-5D-5L evaluates HRQoL through the five domains of mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [13]. Upon examination of the instrument, the Indigenous Reference Group acknowledged the importance of capturing these HRQoL domains in Aboriginal and Torres Strait Islander populations, and in the absence of another suitable instrument being available endorsed its use. In addition to providing initial support for content and face validity, the Indigenous Reference Group advised that the EQ-5D-5L should

undergo further psychometric evaluation; it is necessary, for instance, to investigate other aspects of construct validity, such as dimensionality and criterion validity.

In non-Indigenous cultures, a key feature of the EQ-5D-5L has been the derivation of “value sets” to weight responses by patients. The combination of the EQ-5D-5L five items each with five response categories (5<sup>5</sup>) describes 3125 unique health states. Individuals can then be asked, using preference-based methods such *time-trade off* or *standard gamble*, to indicate which health states they would prefer and value them between 0 (dead) and 1 (full health). These valuations are then attributed to each one of the 3125 unique health states, creating a continuum regarding which states are the least desirable and constituting a population-specific “value set” that can be used to calculate several quantities of interest, such as quality-adjusted life years (QALYs) (for an in-depth discussion about EQ-5D-5L preference-based valuation, please refer to Devlin and Krabbe [12]). While for non-Indigenous groups the derivation of “value sets” can be obtained through specified research guidelines [12], we followed recommendations from Young, Yang [14] that the “first stage of deriving a preference-based single-index measure for use in calculating quality-adjusted life years (QALYs) is to derive a health-state classification system that is *amenable* [emphasis added] to valuation using a preference-elicitation technique”. That is, prior to the calculation of “value sets” and subsequent assignment of values to health states, the first step of a HRQoL instrument validation is to ensure that the instrument correctly measures the health states intended to be measured. Only after construct validity is established, the instrument has been shown to provide valid measurement of health states of mobility, self-care, usual activities, ‘pain/discomfort’, and ‘anxiety/depression’ and is, consequently, amenable to preference-based techniques. Instrument validation prior to the application preference-based methods seems particularly important in Indigenous populations, in which the EQ-5D “logical inconsistencies suggests that the health state valuation instrument lacks construct validity” [15] and other Western-developed HRQoL measures were previously found to be “unsuitable for use” [16].

Hence, we performed the psychometric validation according the steps recommended by Young, Yang [14], consisting of: (1) identification of the instrument dimensionality; and (2) evaluation of the functioning of individual items (including the adequacy of response categories). To establish the instrument dimensionality, we investigated whether the EQ-5D-5L captured at least one or more dimensions of Aboriginal and Torres Strait Islanders’ HRQoL. After dimensionality was established, we examined whether the EQ-5D-5L individual items

correctly measured mobility, self-care, usual activities, pain/discomfort, or anxiety/depression. Finally, we examined the instrument (3) criterion validity. To do so, we evaluated if EQ-5D-5L items scores could correctly identify Aboriginal and Torres Strait Islanders who had poor general health, were suffering from chronic pain or who experienced racial discrimination.

In summary, prior to the application of any preference-based techniques, this study examined whether the EQ-5D-5L could correctly measure in an Aboriginal population the health states intended to be measured (i.e. mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). Similarly to many other countries and cultures in which the EQ-5D-5L has been officially validated [17], the validation (and possible adaption) of the EQ-5D-5L for Aboriginal and Torres Strait Islanders will inform whether this instrument can be used in future research and policymaking. The availability of a validated instrument is crucial to the measurement of HRQoL among Aboriginal Australians over the next years and to produce evidence that can be compared with population levels of HRQoL among other groups (such as Non-Aboriginal Australians). This evidence is absent at the moment and it is not clear how Aboriginal Australians stand in terms of HRQoL in comparison to other population groups. Moreover, once a validated HRQoL instrument is available, future studies can use preference-based techniques to derive utilities and calculate the health and financial impact of public policies on Aboriginal health. This evidence is likely to be used to inform and guide government policy in Australia over the next years. The aim of this study was to evaluate the construct validity and reliability of the EQ-5D-5L in an Aboriginal Australian population.

## Methods

### Participants and procedures

Data were from an overarching study which, as its primary outcome, aimed to investigate population estimates of oncogenic genotypes of oral HPV infection in Aboriginal and Torres Strait Islander populations. Inclusion criteria included being aged 18+ years and identifying as being Aboriginal and/or Torres Strait Islander. Recruitment strategies included: establishing service agreements with key Aboriginal community-controlled health organisations in South Australia, liaising with community champions, and encouraging word-of-mouth [18]. The study had six project staff who were led by a senior Indigenous project manager. The three non-Indigenous staff undertook extensive cultural competency training. A sample of 1012 Aboriginal adults was recruited at baseline and included several distinct language groups: Adnyamathanha, Akenta, Amarak, Bungandidj, Diyari,

Erawirung, Kaurna, Kokatha Mula, Maralinga Tjarutja, Mirning, Mulbarapa, Narungga, Ngaanyatjarra, Ngadjuri, Ngarrindjeri, Nukunu, Parnkalla, Peramangk, Pitjantjatjara, Wirangu and Yankunjatjarra. All recruitment and data collection procedures were performed following the ethical standards laid down by the 1964 Declaration of Helsinki and its later amendments. Ethics approval was obtained from the University of Adelaide Human Research Ethics Committee (H-2016–246) and the Aboriginal Health Council of South Australia (04–17-729). All participants provided signed informed consent.

Missing responses of individual items ranged from 0.8% to 1.6%, meaning multiple imputation was not required [19]. Analyses were thus conducted with participants with complete questionnaire responses; that is, participants with responses to all EQ-5D-5L items. The final sample with complete questionnaire responses comprised 988 participants. These 988 participants were randomly and equally assigned into a test sample ( $n=494$ ) and validation sample ( $n=494$ ).

## Measures

### EQ-5D-5L

The EQ-5D is a generic preference-based measure of health that evaluates health-related quality of life (HRQoL) according to five dimensions: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression [11]. The EuroQoL Group recently introduced the EQ-5D-5L which expanded the original 3-category EQ-5D to include 5 categories [13]. In this study, the categories (response options) followed the format “no”, “slight”, “moderate”, “severe problems,” and “unable to”/ “extreme” for all dimensions.

### Self-rated general health and chronic pain

General health was measured with a single-item question that asked: “Would you rate your general health as: (1) Excellent; (2) Very good; (3) Good; (4) Fair; (5) Poor”. While there are methodological challenges inherent to the validation of single-item questionnaires [20], single-item self-report measures of general health are considered important since they provide a holistic and integrated perception of Aboriginal health, including biological, psychosocial and social factors into the judgement [21]. For instance, Lavrencic, Mack [21] showed that single-item self-report measures of general health are associated with several health outcomes among Aboriginal Australians such as chronic diseases (e.g. arthritis and kidney problems) and perceived resilience. Following Lavrencic, Mack [21], in addition to the 5-point measure of general health (1 = Excellent, 2 = Very good, 3 = Good, 4 = Fair, 5 = Poor), we also created a new dichotomous variable so scores from “Excellent” to “Fair” indicated

“good/fair health” and scores of “Poor” indicated “poor health”. The dichotomisation was done based on a ‘risk factor’ approach, aiming to identify the individuals with the highest risk (i.e. worst general health) [22]. This approach allows for the calculation of classification measures such as the area under the receiver operating characteristic curve (AUROC) [23], which can inform whether EQ-5D-5L scores correctly discriminate between individuals with good/fair and poor general health.

Chronic pain was measured with a single-item question that asked: “Do you now have significant pain that has lasted 6 months or more?” with response options “Yes” or “No”. Similar to previous research among Aboriginal Australians [24, 25], we evaluated self-report chronic pain instead of site-specific pain or clinical conditions causing pain. Previous research also showed that measures of self-report pain have been associated with criterion variables among Aboriginal Australians, such a high prevalence of multiple musculoskeletal conditions [25].

### Experiences of racial discrimination

Experiences of racial discrimination were measured by 9 items that evaluate the frequency of racial discrimination in different settings (i.e. work, home, education, recreation, legal, medical, governmental, services and public) [26]. These items were based on the Measure of Indigenous Racism Experiences (MIRE), originally developed for Aboriginal and Torres Strait Islanders [27]. Items were rated on a 5-point scale ranging from “Strongly disagree” to “Strongly Agree”. For the experiences of racial discrimination, we also followed a ‘risk factor’ approach to distinguish between individual at high risk of experiences of racial discrimination and individuals with lower risk [22]. Hence, in addition to the MIRE total score (ranging from 9 to 45), we also created a new variable by dichotomising the MIRE total score according to the median, thus indicating participants with lower or higher frequency of experienced racial discrimination.

## Statistical analysis

### Dimensionality

Exploratory Graph Analysis (EGA) [28] was used to investigate the EQ-5D-5L dimensionality in the test sample. EGA is a technique within the field of network psychometrics, a new scientific field dedicated to the study of psychological networks. Psychological networks are networks in which nodes represent items and edges represent the associations between items (e.g. partial correlations). In psychological networks, a *cluster* of items occurs when certain nodes are more strongly connected among each other compared to the rest of the network [29]. The aim of EGA is to identify these item clusters [30].

The first step of EGA is estimating a network model. The network model used in EGA was the Gaussian Graphical Model (GGM) [31], estimated with Least Absolute Shrinkage and Selection Operator (LASSO) [32] with turning parameter based on minimizing the Extended Bayesian Information Criteria (EBIC) [33]. After the network is estimated, EGA then employs a *walktrap algorithm* [34] to identify which items clustered in the psychological network. Since item clusters generate covariance patterns that are statistically equivalent to those produced by a latent variable [29], EGA can discover the instrument dimensionality by identifying the *number of item clusters*, in contrast to traditional factor analytical methods (e.g. Parallel Analysis) which identify the *number of latent variables* believed to connect the items. While an in-depth explanation of EGA is beyond the scope of this paper, accessible introductions to network psychometric and EGA can be found in Borsboom and Cramer [35] and Christensen, Golino [36], respectively.

Recent simulation studies showed that EGA performs as accurately as factor analytical procedures, such as the Automated Scree Test, Kaiser-Guttman eigenvalue greater than 1 rule and Parallel Analysis, and outperforms them in large sample conditions. For example, in sample sizes of 500 respondents (similar to the test sample in our study), EGA discovered the correct number of factors in 81% of all simulated cases, while traditional procedures such as Kaiser-Guttman eigenvalue greater than 1 rule discovered the correct number of factors only 70% of the time. Additionally, the EGA overall accuracy increases to 93% in samples of 5000 participants and its accuracy remained the highest among all methods to identify dimensionality independent of sample size [30].

The main output of EGA is a network plot in which nodes representing the five EQ-5D-5L items are coloured according to their identified dimensions. The network was plotted with the Fruchterman-Reingold algorithm [37], which arranges nodes more closely according to the strength of their associations (i.e. regularised partial correlations). One main reason we employed EGA over traditional methods is due to its graphical nature, providing an intuitive visual interpretation [30] of the associations established between the EQ-5D-5L items. Considering that EGA-identified dimensions are subject to sampling variation, we employed 2500 bootstrap samples to evaluate the stability of the identified dimensions and to ensure robustness of the results [38]. The analysis was conducted with R software [39] and the R package *EGAnet* [40].

#### Model fit

After the dimensional structure was identified by EGA in the *test* sample, we evaluated it with Confirmatory

Factor Analysis (CFA) in the *validation* sample. The dimensional structure was confirmed in a different sample (that is, in the validation sample) to avoid overfitting [41] due to capitalization on sampling variation [42]. We compared the dimensional structure identified by EGA with the 1-dimensional model, which assumes that all five EQ-5D-5L items constitute a single dimension. The 1-dimensional model is the most parsimonious and, if a single dimension cannot be rejected, there is no reason to evaluate more complex structures [43]. CFA models were estimated with weighted least squares with a mean- and variance-adjusted (WLSMV) test statistic [44]. To evaluate model fit, the scaled  $\chi^2$ , scaled CFI and scaled RMSEA were used. Values of CFI  $\geq 0.96$  and RMSEA  $\leq 0.05$  indicate good model fit [45], while RMSEA  $\leq 0.07$  indicates acceptable fit [46].

After the dimensions were identified, we calculated the corrected item-total correlation (CITC), which is the correlation between the item score and the total score without the item (i.e. restscore) [47]. The CITC needs to be calculated for each subscales, since items can only be summed into a score when they measure the same construct [48]. The CITC evaluates the degree to each item is coherent with the other items from the same subscale [49]. Given the ordinal nature of the data, the CITCs were calculated using non-parametric rank correlation Kendall's  $\tau$  [50] with bootstrapped CIs [51]. Items with CITCs higher than 0.30 were considered to be coherent with the subscale [52]. The Reliability was calculated with the Categorical Omega [53]. The advantage of the Omega coefficient is that it does not rely on restrictive assumptions of tau-equivalence as do other traditional coefficients, such as Cronbach's  $\alpha$  [54]. The analysis was conducted with R package *lavaan* [55]. All subsequent analysis, including the evaluation of the functioning of individual items (i.e. item analysis) and criterion validity, were also conducted on the validation sample.

#### Item analysis

After dimensionality and overall fit were established, we employed *item-response theory* to evaluate the EQ-5D-5L performance at an item level. We followed previous recommendations of the Rasch model (specifically, the Partial Credit Model for polytomous items) as a model of choice [14, 56, 57]. For example, the EuroQol Group employed the Partial Credit Model in the EQ-5D-5L initial validations [58], been later followed by other empirical research [59, 60]. In our study, the Rasch model (RM) for polytomous items, the Partial Credit model [61], was estimated with conditional maximum likelihood [62] and person parameters were estimated with weighted maximum likelihood (WML) [63]. As a means of sensitivity analysis, we also evaluated the Rating Scale model [64].

The Rating Scale model is a restricted version of the Partial Credit model which constrains the distance between item thresholds to be the same across all items. While the Rating Scale model is more parsimonious, the assumption of equal threshold distance across all items can be restrictive and incompatible with certain questionnaires in health sciences (see, for instance, Shea, Tennant [65]). Hence, a Likelihood Ratio Test (LRT) was conducted to determine whether the Partial Credit model or the Rating Scale model better explained the EQ-5D-5L item responses. The LRT null hypothesis is that differences between the restricted model, the Rating Scale model, and the Partial Credit model occur only due to sampling variation. Thus, a significant LRT indicates that Partial Credit model fitted the data better than the Rating Scale model [65].

Once the Partial Credit model or Rating Scale model were selected, fit to the RM was evaluated with the Conditional Likelihood Ratio (CLR) test [66]. Item fit was evaluated with *conditional* infit and outfit statistics [67] with bootstrapped standard errors [68]. We report item discrimination and threshold parameters. In the Partial Credit model and Rating Scale model, discrimination parameters are constrained to 1. The thresholds parameters indicate in the latent trait scale (i.e. HRQoL) the point of equal probability of choosing between two adjacent categories (e.g. “moderate” and “severe problems”) [69]. In addition, Item Characteristic Curves (ICCs) were plotted for a graphical inspection of item fit. Ideally, ICCs would display average observed item responses for each possible total score. However, since it is unlikely that the sample will contain a meaningful number of respondents for each possible total score, we created 5 class intervals [70].

Finally, we evaluated the adequacy of the EQ-5D-5L five categories through the visual inspection of Category Characteristic Curves. In polytomous items, it is expected that increasing amounts of the latent trait (i.e. poor health-related quality of life) will correspond to a monotonically increasing endorsement of response categories associated with poor HRQoL. For example, it is expected that respondents with poor HRQoL will have a higher probability of endorsing a category such as “I have extreme pain or discomfort” rather than “I have slight pain or discomfort”. Moreover, in case all EQ-5D-5L five categories are necessary to evaluate health-related quality of life, it is expected that each category will become the most probable for at least a certain range of HRQoL [71]. The analysis was conducted with DIGRAM v4.03 [72] and R package *iarm* [73].

### Criterion validity

To evaluate concurrent validity, we initially followed a ‘risk factor’ approach [22] and used dichotomized outcomes (e.g. good/fair health and poor health) to investigate whether the EQ-5D-5L scores could discriminate between individuals with high risk and low risk of poor general health, chronic pain and experiences of racial discrimination. To do so, we investigated the AUROC between EQ-5D-5L item scores (and subscale scores) and measures of general health and chronic pain. The AUROC indicates the probability that a randomly chosen participant with poor general health is correctly identified by having poor HRQoL (measured by EQ-5D-5L scores) compared to a randomly chosen participant with good/fair HRQoL. Similarly, the AUROC indicates the probability that a randomly chosen participant experiencing chronic pain is correctly identified by having poor HRQoL compared to a randomly chosen participant with no chronic pain. Thus, it was expected AUROC values higher than 50%, indicating that EQ-5D-5L scores identified participants with poor general health and chronic pain with a higher probability than random chance (50%).

We also inspected the AUROC between EQ-5D-5L item scores (and subscales scores) and the score derived for experiences of racial discrimination to evaluate discriminant validity. It is theoretically expected that, besides the domain of anxiety/depression, there is weak or no association between mobility, self-care, usual activities or pain/discomfort with experiences of racial discrimination. Thus, expected AUROC values should be closer to 50%, indicating that EQ-5D-5L scores were not able to identify participants who experienced more frequent episodes of racial discrimination better than chance alone.

Since dichotomous variable can lead to loss of information under certain circumstances [74], to evaluate the robustness of our findings, we also employed univariate linear regressions to evaluate the effect of the EQ-5D-5L individual items and the participants’ latent trait scores (i.e. person parameters) on the original variables of general health (i.e. 5-point measure of general health) and experiences of racism (i.e. MIRE total score) before dichotomisation. Since the 5-point measure of general health (1 = Excellent, 2 = Very good, 3 = Good, 4 = Fair, 5 = Poor) and MIRE total score (ranging from 9 to 45) are on different scales, we report standardized regression coefficients.

### Results

The participants’ characteristics are displayed in Table 1. The majority of participants had education up to finishing high school (67.4%), were unemployed or on benefits (74.9%) and did not have access to a health

**Table 1** Characteristics of study participants

	HPOVC Study (n = 988)	
	n	%
Age		
Mean (SD)	39.7 (15.1)	
Missing	7	0.7
Sex		
Male	334	33.8
Female	654	66.2
Missing	0	0.0
Location		
Regional	617	62.5
Metropolitan	369	37.3
Missing	2	0.2
Education		
Up to High School	666	67.4
Technical education or university	312	31.6
Missing	10	1.0
Employment		
Employed	239	24.2
Unemployed/Benefits	740	74.9
Missing	9	0.9
Access to health care card		
Yes	202	20.4
No	745	75.4
Don't know/Missing	41	4.2

Mean values, minimum, maximum and standard deviations; numbers and percentages

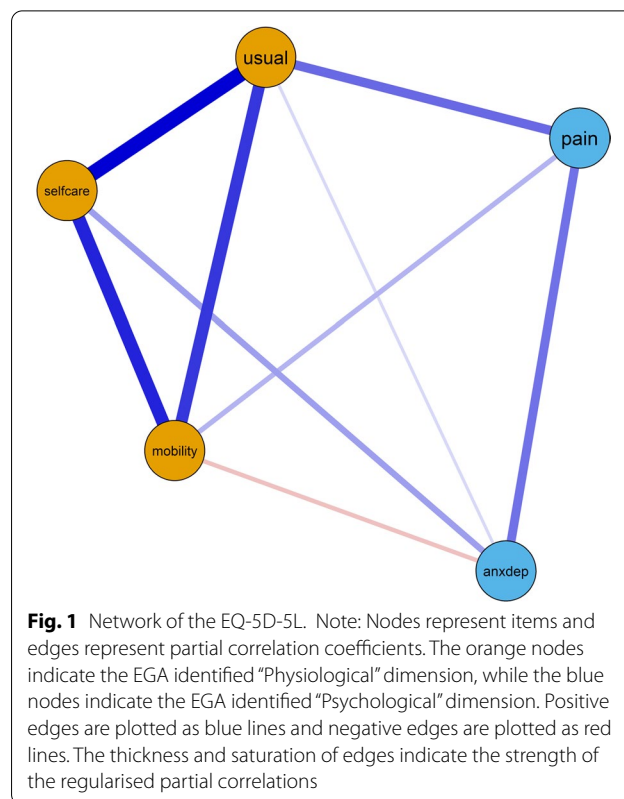
care card (75.4%). The average age was 39.7 years (Median = 37 years) and approximately 45 percent of the sample was aged 40 years or older. Two-thirds of the participants were female and more than 60 percent resided in non-metropolitan locations.

**Dimensionality**

The EGA indicated that the EQ-5D-5L has a two-dimensional structure. The first dimension, comprising items “Mobility”, “Self-care” and “Usual activities”, was named the “Physiological” dimension. The second dimension, comprising the items, “Anxiety/Depression” and “Pain”, was named the “Psychological” dimension. The network representation of the EQ-5D-5L is displayed in Fig. 1.

The network edges (i.e. regularised partial correlations between items) are displayed in Table 2.

The application of EGA to 2500 bootstrap samples showed that 2 dimensions were identified in 91.6% of the bootstrap samples, while 1 dimension was identified in 2.4% of the bootstrap samples and 5 dimensions were identified in 6.0% of the bootstrap samples. These results indicate that the 2-dimensional structure identified was



**Fig. 1** Network of the EQ-5D-5L. Note: Nodes represent items and edges represent partial correlation coefficients. The orange nodes indicate the EGA identified “Physiological” dimension, while the blue nodes indicate the EGA identified “Psychological” dimension. Positive edges are plotted as blue lines and negative edges are plotted as red lines. The thickness and saturation of edges indicate the strength of the regularised partial correlations

stable across the bootstrap samples; that is, it is unlikely that the identification of the 2-dimensional structure by EGA was merely a consequence of sampling variation.

**Model fit**

After the 2-dimensional structure was identified by EGA in the test sample, we compared it with the 1-dimensional structure in the validation sample to investigate which structure received more support from the data. Table 3 shows that the fit of the 1-dimensional structure was mixed since RMSEA (>0.07) had unacceptable values. On the other hand, the fit of the 2-dimensional structure was excellent since both CFI (>0.96) and RMSEA (<0.07) achieved desirable values.

The CITCs between the items “Mobility” (CITC=0.87—95% CI [0.84, 0.90]), “Self-care” (CITC=0.53—95% CI [0.47, 0.59]) and “Usual activities” (CITC=0.81—95% CI [0.77, 0.85]) with the “Physiological” subscale were moderate to strong. The CITCs between the items “Anxiety/Depression” (CITC=0.77—95% CI [0.74, 0.80]) and “Pain/Discomfort” (CITC=0.71—95% CI [0.68, 0.75]) with the “Psychological” subscale were strong. In both cases, the CITCs of all items were above suggested cut-off values (>0.30) indicating that the “Physiological” and “Psychological” subscales are constituted by a cohesive set

of items. Reliability of the Physiological scale was good ( $\Omega_c = 0.84$ —95% CI [0.79, 0.89]), while reliability of the Psychological subscale was adequate ( $\Omega_c = 0.70$ —95% CI [0.63, 0.74]).

**Item analysis**

The Likelihood Ratio test (LRT) indicated that Partial Credit model was a significantly better fit to the data compared to the Rating Scale model for the Physiological subscale ( $\chi^2(6) = 13.52, p = 0.03$ ) but there was no significant difference between the fit of both models to the Psychological subscale ( $\chi^2(3) = 1.42, p = 0.70$ ). While there was no significant difference in the fit of the Partial Credit model and the Rating Scale model to the Psychological subscale, the Rating Scale model was not appropriate for the Physiological subscale. Hence, to avoid employing different measurement models with distinct assumptions about the item response process for the two EQ-5D-5L subscales (i.e. Rating Scale

model for Psychological subscale and Partial Credit model for the Physiological subscale), the EQ-5D-5L item responses were evaluated with the measurement model that was adequate for *both* subscales, the Partial Credit model.

The Physiological subscale ( $\chi^2(11) = 2.55, p = 0.99$ ) and the Psychological subscale ( $\chi^2(7) = 8.74, p = 0.27$ ) achieved overall fit to the Partial Credit model. The fit to the Partial Credit model was also confirmed at the item level (Table 4). The observed infits and outfit were similar in magnitude (and not statistically different) from the expected value of 1 under the Partial Credit model. The CLR and observed infits and outfit further confirmed the adequacy of the Partial Credit model to model item responses to the Physiological and the Psychological subscales.

The graphical inspection of the Item Characteristic Curves confirmed that average observed item responses were in accordance with the item responses

**Table 2** Network edges of the EQ-5D-5L

	Mobility	Self-care	Usual activities	Pain/Discomfort	Anxiety/Depression
Mobility	1.00				
Self-care	0.27	1.00			
Usual activities	0.42	0.53	1.00		
Pain/Discomfort	0.20	0.00	0.16	1.00	
Anxiety/Depression	-0.12	0.23	0.04	0.34	1.00

The network edges indicate regularised partial correlations between items. The matrix of network edges is symmetrical so values above the diagonal were omitted

**Table 3** Model fit comparison of the 1-dimensional structure and the 2-dimensional structure identified by EGA

Model	$\chi^2$	df	p-value	RMSEA	90% CI	CFI
<i>EQ-5D-5L</i>						
1-dimensional structure	55.2	5	<0.001	0.143	[0.110, 0.178]	0.975
2-dimensional structure	10.3	4	0.04	0.057	[0.013, 0.100]	0.997

$\chi^2$  = scaled chi-square; df = degrees of freedom; RMSEA = scaled root mean square error of approximation; CFI = scaled comparative fit index

**Table 4** Item fit statistics for the EQ-5D-5L

Item	Conditional Outfit		Conditional Infit	
	Observed		p-value	Observed
<i>Physiological subscale</i>				
Mobility	0.994		0.940	1.005
Self-care	0.927		0.620	0.917
Usual movements	1.026		0.640	1.034
<i>Psychological subscale</i>				
Pain/Discomfort	1.000		0.960	1.008
Anxiety/Depression	1.000		0.960	1.008

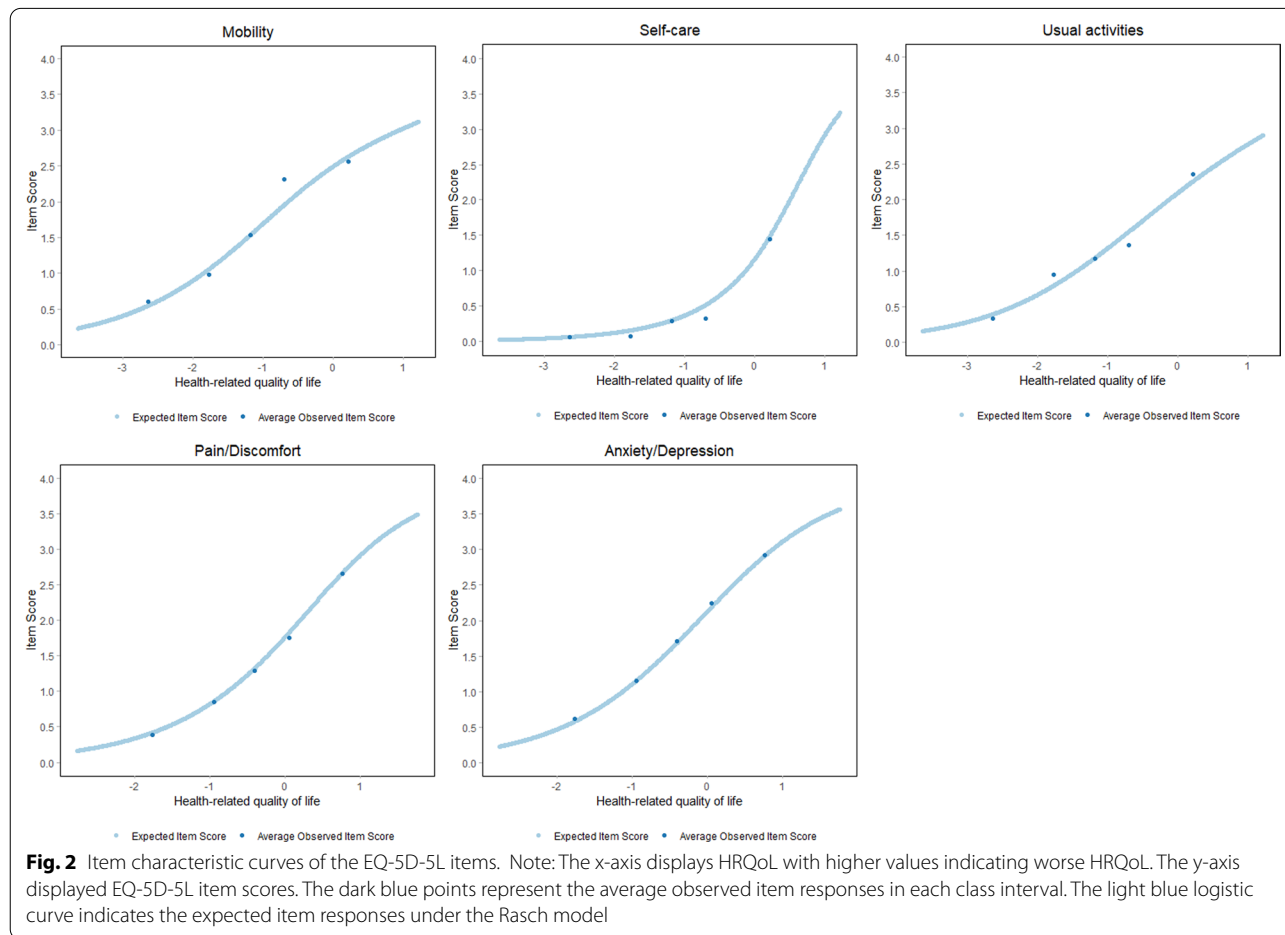
The Conditional Outfit and Conditional Infit statistics have expected values equal to one under the Rasch model



expectations (Fig. 2). Additionally, the average observed item responses *in general* monotonically increased given the values of HRQoL. That is, respondents with poorer HRQoL increasingly endorsed EQ-5D-5L categories indicating more problems with mobility, self-care, usual activities, pain/discomfort and anxiety/depression.

The item discrimination and threshold parameters are reported in Table 5.

The investigation of Category Characteristic Curves (Fig. 3) showed that the categories of all EQ-5D-5L items were correctly ordered and became the most probable for a specific range of HRQoL. The only exception was the Self-Care item which had disordered thresholds

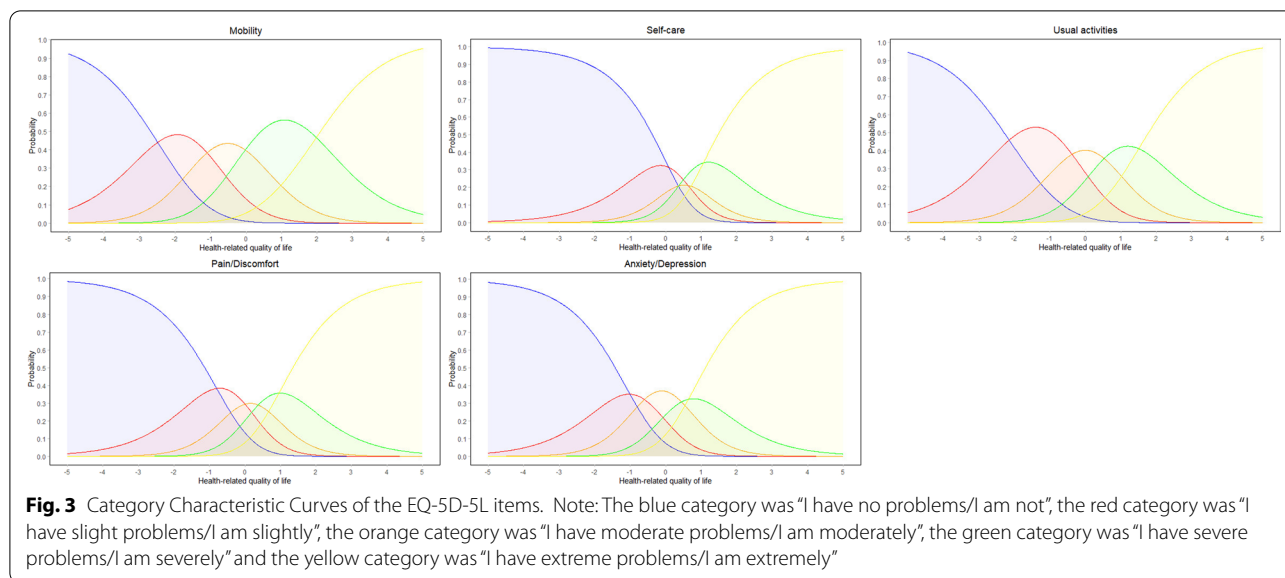


**Fig. 2** Item characteristic curves of the EQ-5D-5L items. Note: The x-axis displays HRQoL with higher values indicating worse HRQoL. The y-axis displayed EQ-5D-5L item scores. The dark blue points represent the average observed item responses in each class interval. The light blue logistic curve indicates the expected item responses under the Rasch model

**Table 5** Item parameters of the EQ-5D-5L

	Discrimination	Threshold 1	SE	Threshold 2	SE	Threshold 3	SE	Threshold 4	SE
<i>Physiological subscale</i>									
Mobility	1.00	-2.49	0.30	-1.10	0.28	-0.02	0.31	1.98	
Self-care	1.00	0.08	0.27	<b>0.63</b>	0.44	<b>0.36</b>	0.58	1.06	0.71
Usual movement	1.00	-2.15	0.28	-0.38	0.28	0.53	0.38	1.50	0.63
<i>Psychological subscale</i>									
Pain/discomfort	1.00	-0.80	0.15	0.00	0.16	0.37	0.23	0.96	0.36
Anxiety/depression	1.00	-1.01	0.17	-0.60	0.16	0.44	0.20	0.64	0.28

SE = standard error. The discrimination parameters do not have standard error since these parameters were constrained to 1 under the Partial Credit Model. Item disordered thresholds are highlighted in bold



(Table 5), so the middle category (“I have moderate problems washing or dressing myself”) never became the most probable category for the EQ-5D-5L respondents.

**Criterion validity**

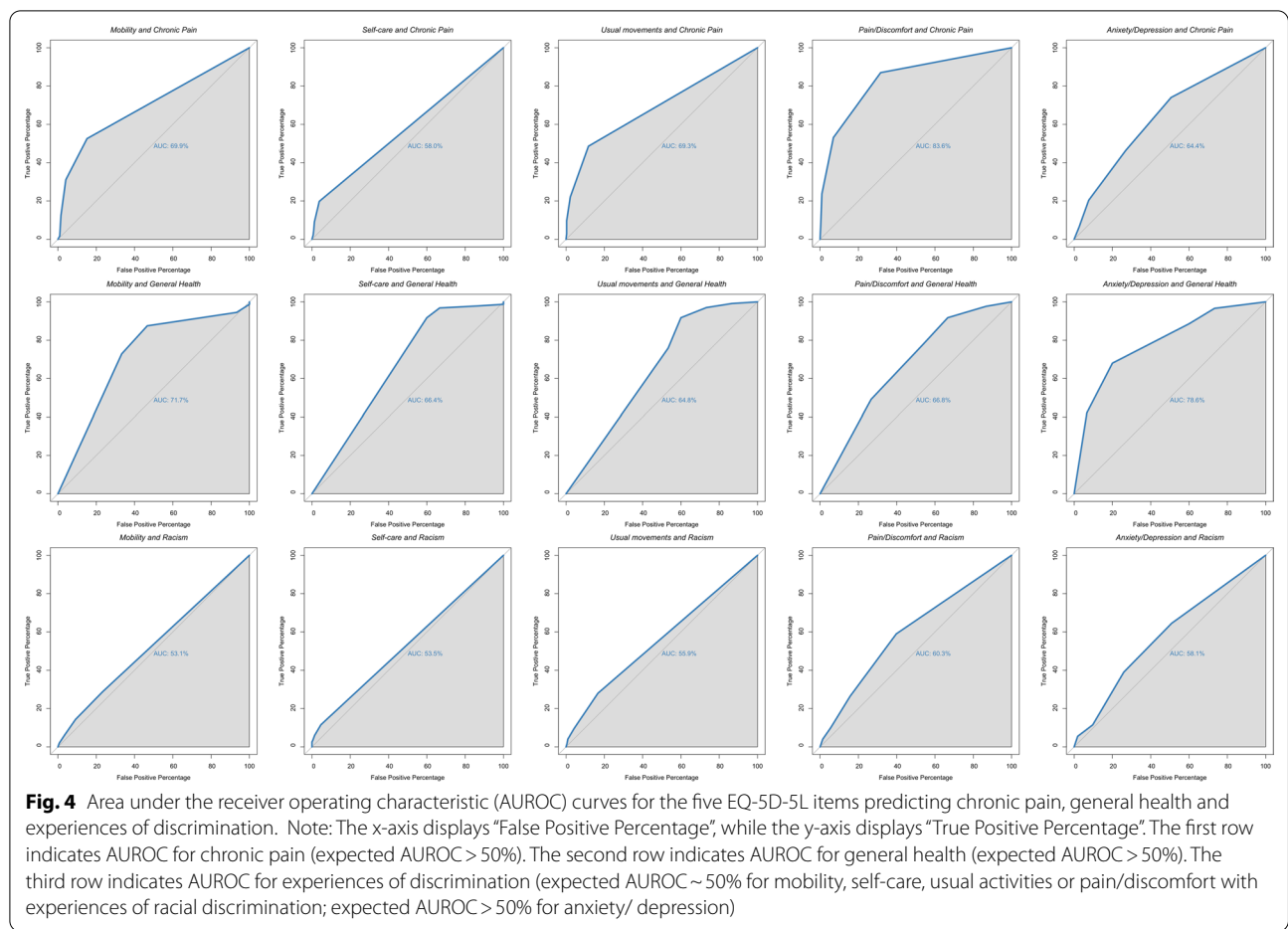
The AUROC values substantially above 50% indicated that EQ-5D-5L individual items were able to correctly identify participants with poor general health and chronic pain (Fig. 4). Figure 4 shows, for example, that the probability of identifying a participant experiencing chronic pain through high scores on the “Pain/Discomfort” item was 83.6% higher than if the participant was not experiencing chronic pain (first row, fourth column). These findings suggest good concurrent validity of the EQ-5D-5L five individual items.

Furthermore, when experiences of racial discrimination were considered, AUROC values of items from the “Physiological” subscale, such as mobility (AUROC = 53.1%), self-care (AUROC = 53.5%) and usual activities (AUROC = 55.9%), were close to 50%. That is, according to the expectations, scores of these three items had a weak association with experiences of racial discrimination and were able to identify participants that experienced racial discrimination just slightly better than random chance. The same was observed regarding the items from the “Psychological” subscale, such as anxiety/depression (AUROC = 58.1%) and pain/discomfort (AUROC = 60.3%), which were also poor predictors of racial discrimination. These results indicate that the EQ-5D-5L displayed reasonable discriminant validity.

The concurrent and discriminant validity and the AUROCs of the two identified dimensions, Physiological and Psychological subscales, are displayed in Fig. 5.

Figure 5 indicates that the dimensions of Physiological and Psychological subscales also displayed good concurrent and discriminant validity. For instance, high scores on the Physiological subscale score were able to identify more than 70% of the time participants experiencing chronic pain (AUROC = 72.2%) or poor general health (AUROC = 73.2%) compared to participants who were not. Moreover, when the Psychological subscale scores were used, these numbers increased to almost 80% for both chronic pain (AUROC = 79.4%) and poor general health (AUROC = 78.9%). Regarding discriminant validity, the Physiological subscale scores were able to identify participants who experienced racism 56% of the time (AUROC = 56.4%) and the Psychological subscale scores identified participants who experienced racism 61% of the time (AUROC = 60.9%), only marginally better than random chance. Both scales were poor predictors of individuals who experienced racism. These results indicate that the EQ-5D-5L good concurrent and discriminant validity not only on an item level but also on a dimension/subscale level.

Finally, the results were also consistent when the non-dichotomous variables were used, the 5-point measure of general health (1 = Excellent, 2 = Very good, 3 = Good, 4 = Fair, 5 = Poor) and MIRE total score (ranging from 9 to 45), were used. For instance, the effects of mobility ( $\beta=0.31-95\% \text{ CI } [0.22, 0.39]$ ), self-care ( $\beta=0.20-95\% \text{ CI } [0.12, 0.29]$ ), usual activities ( $\beta=0.34-95\% \text{ CI } [0.25, 0.42]$ ), pain/discomfort ( $\beta=0.33-95\% \text{ CI } [0.24, 0.41]$ ) and anxiety/depression ( $\beta=0.29-95\% \text{ CI } [0.21, 0.38]$ ) on



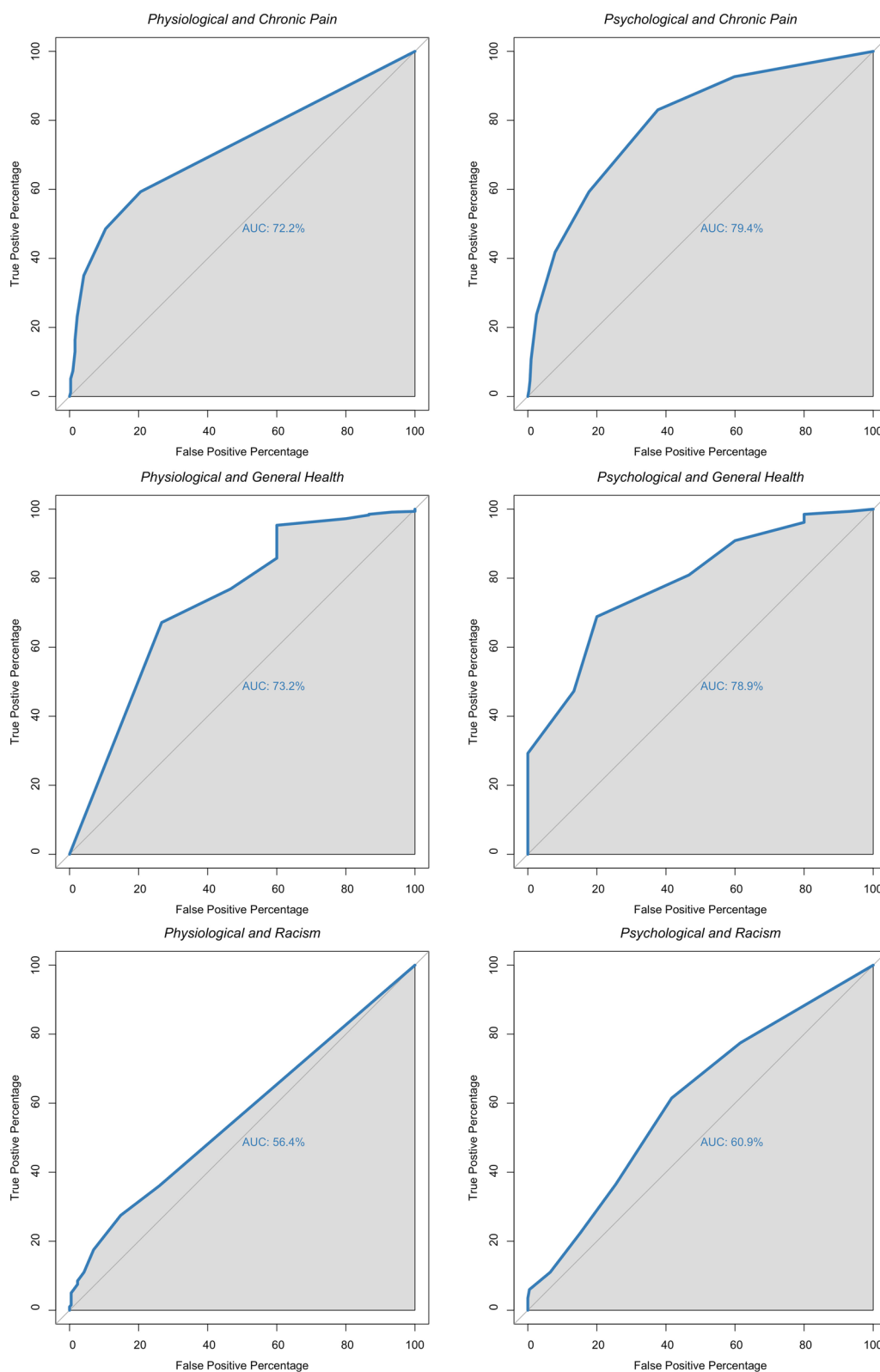
general health were stronger than the effects of mobility ( $\beta=0.13$ —95% CI [0.03, 0.24]), self-care ( $\beta=0.10$ —95% CI [−0.01, 0.20]), usual activities ( $\beta=0.12$ —95% CI [0.01, 0.23]), pain/discomfort ( $\beta=0.17$ —95% CI [0.07, 0.27]) and anxiety/depression ( $\beta=0.14$ —95% CI [0.04, 0.23]) on experiences of racism. Moreover, when evaluated at a dimension/subscale level, the effects of Physiological ( $\beta=0.36$ —95% CI [0.27, 0.44]) and Psychological ( $\beta=0.37$ —95% CI [0.29, 0.46]) latent trait scores on general health were also stronger than the effects of Physiological ( $\beta=0.14$ —95% CI [0.04, 0.25]) and Psychological ( $\beta=0.20$ —95% CI [0.10, 0.29]) latent trait scores on experiences of racism. The results further confirm the association between EQ-5D-5L items and subscales with general health indicating convergent validity, while the weak association between EQ-5D-5L items and subscale and experiences of racism support discriminant validity.

**Discussion**

This study aimed to evaluate the construct validity and reliability of the EQ-5D-5L for an Aboriginal Australian population. We employed a multi-method approach

to comprehensively evaluate the EQ-5D-5L psychometric properties in a large sample, both at an instrument level and at an item level. We also investigated whether EQ-5D-5L scores could correctly identify participants with poor general health and chronic pain. To the best of our knowledge, this is the first study to evaluate the EQ-5D-5L psychometric properties in any Indigenous population [16].

Our findings showed that EQ-5D-5L psychometric properties were excellent. The instrument is composed of two dimensions, Physiological and Psychological, and reliability was adequate. Moreover, the EQ-5D-5L provides a health-state classification system that is amenable to future valuation using preference-based techniques. Future research should also investigate whether the instrument can potentially be expanded to also incorporate other domains specific of Aboriginal Australians’ HRQoL, such as cultural health, knowledge and interaction with the health system, among others [75]. Implications for practice are provided.



**Fig. 5** Area under the receiver operating characteristic (AUROC) curves for the two EQ-5D-5L “Physiological” and “Psychological” dimensions predicting chronic pain, general health and experiences of discrimination. Note: The first row indicates AUROC for chronic pain. The second row indicates AUROC for general health. The third row indicates AUROC for experiences of discrimination

### Dimensionality

The findings indicated two overall dimensions, Physiological and Psychological, in an Aboriginal population. The distinction between a “physiological” and a “psychological” dimension is theoretically consistent with the current understanding about Aboriginal and Torres Strait Islanders’ SEWB. For instance, the nine domains that typically characterise Aboriginal SEWB include “physical health” and “mental health” [6]. These two domains have also been previously referred as “Connection to the body” and “Connection to mind and emotions”, respectively [76]. Finally, in a recent qualitative study, Aboriginal parents described the importance of the “physical” and “emotional” domains in the HRQoL of their children [75].

In other Indigenous groups, previous validations of HRQoL instruments also identified similar “physiological” and “psychological” dimensions. For example, in Native Americans, “Symptoms” and “Psychological Impact” (in addition to “Community and Social Restrictions”) were determined as dimensions of HRQoL [77]. In New Zealand, the distinction between “mental” and “physical” dimensions was also identified in Maori people, although older Maori (more than 45 years old) did not make the same differentiation. These results led Scott, Sarfati [78] to suggest that younger Maori perceived their HRQoL more similar to Western cultures, in which HRQoL questionnaires “measure largely independent dimensions of physical and mental health”.

In summary, our findings indicated that the EQ-5D-5L was comprised by two distinct dimensions, “physiological” and “psychological”. While these dimensions were consistent with theoretical understanding regarding Aboriginal and Torres Strait Islanders’ SEWB, there are two important things to be noticed. Firstly, it is unlike that the five items measuring mobility, self-care, usual activities, pain/discomfort and anxiety/depression are enough to *exhaust* the “physiological” and “psychological” dimensions of Aboriginal HRQoL. That is, there are also other factors that also potentially constitute the “physiological” and “psychological” dimensions of Aboriginal health. Secondly, “physiological” and “psychological” were the only two dimensions measured by the EQ-5D-5L. The EQ-5D-5L did not encompass other important dimensions of Aboriginal HRQoL. One of these dimensions, for example, is “cultural health”, which includes Aboriginal values, historical perspective in Australia and connection to the land [75]. We provide an in-depth discussion of this issue in the section “[Limitations and future directions](#)” and directions for future research.

### Item performance

Our findings showed robust evidence of the EQ-5D-5L validity at an item level. For instance, the fit of the items to the Rasch Model entails excellent measurement properties [79]. For our intended purposes, two important properties displayed by the EQ-5D-5L items were *monotonicity* and *adequacy of response categories*. Regarding monotonicity, the results showed that, on average, respondents with increasingly worse HRQoL monotonically endorsed higher scores on individual EQ-5D-5L items. Herein, we can be confident that higher scores on EQ-5D-5L items are correctly measuring higher values of the underlying construct (in this case, higher scores indicate worse HRQoL).

Moreover, the development of the EQ-5D-5L by the EuroQol Group (which has five response categories) occurred due to limited sensitivity of the EQ-5D-3L (which has three response categories) to detect changes in health, in part due to ceiling and floor effects [58]. Regarding the adequacy of response categories, our findings showed that all five categories (from “I have no problems/I am not” to “I have extreme problems/I am extremely”) were the most probable category of choice for a specific group of respondents according to their level of HRQoL. For example, patients with moderate HRQoL had a higher probability of endorsing the category “I have moderate problems/I am moderately”, while participants with very poor HRQoL were more likely to endorse the “I have extreme problems/I am extremely” category.

These findings are in accordance with previous research that the inclusion of five categories in the EQ-5D-5L was an improvement upon the EQ-5D-3L measurement properties in multiple countries [17, 80]. In our study, the only exception was the Self-Care item middle category (“I have moderate problems washing or dressing myself”), which never became the most probable category of choice for the Aboriginal respondents. Before changes in the instrument are made (such as collapsing categories), future independent studies with other Aboriginal populations should investigate whether problems with this category will re-appear. In summary, our investigation of the EQ-5D-5L five categories provided evidence that “category definitions are adequate (not too narrow in definition) and that responders have not been presented with overwhelming category options” [81].

### Implications for practice

The identification of two dimensions ensue implications for future use of the EQ-5D-5L in Aboriginal Australian populations. Our findings support that the EQ-5D-5L can be used as five independent items *or* as two broader “Physiological” and “Psychological” dimensions.

Firstly, the five EQ-5D-5L individual items showed good discriminatory power to identify respondents with poor self-rated general health and chronic pain. While fit to the Rasch Model indicated equal item discrimination regarding the latent trait (Physiological and Psychological), items can display different discrimination regarding outcomes, such as criterion variables [82]. The items' discriminatory power was mostly consistent with theoretical expectations. For example, the EQ-5D-5L item which better-discriminated participants who experienced significant pain over the last 6 months (or more) was the item "Pain/Discomfort", while the second item with highest discriminatory power was "Mobility". These findings provided support for the use of the EQ-5D-5L items as stand-alone items, which is the most common EQ-5D-5L usage [12]. Secondly, in case researchers are interested in the computation of total scores to evaluate the EQ-5D-5L at a *domain* level, the investigation of dimensionality showed that two subscales scores should be computed, one for the "Physiological" and one for the "Psychological" dimension (instead of one total score summing all five items).

#### Limitations and future directions

The SEWB of Indigenous Australians is a multidimensional and multifaceted construct. The holistic nature of SEWB, which includes several dimensions such as "family and community", "autonomy, empowerment and recognition", "work, roles and responsibilities", and "education" (Butler 2019), indicates that the health (and, consequently, HRQoL) of Aboriginal Australian populations is conceptualized differently to Western societies [83, 84]. Hence, many Western-developed HRQoL instruments (including the EQ-5L-5D) that are built upon more narrow conceptions of health are likely to overlook many aspects of health and wellbeing that are valued by Indigenous people [16].

In our study, in the initial consultation prior to the instrument application, the Indigenous Reference Group recommended the EQ-5D-5L as a potentially valid instrument to capture specific aspects of HRQoL until a broader instrument is available. Moreover, they also required that further evaluation should be conducted to investigate the EQ-5D-5L psychometric properties. Our findings showed that EQ-5D-5L items provided correct measurement of five aspects (mobility, self-Care, usual activities, pain/discomfort, anxiety/depression) of Aboriginal HRQoL and that these five aspects clustered into two overall dimensions ("Physiological" and "Psychological" dimensions). While the EQ-5D-5L was found to be an appropriate instrument

to measure these specific aspects of Aboriginal HRQoL, the EQ-5D-5L limitation in scope must always be considered. For instance, the EQ-5D-5L items do not exhaust Aboriginal Australians dimensions of "Physiological" and "Psychological" health or cover other relevant dimensions, such as "family and community" and "autonomy, empowerment and recognition".

Directions for future research include potentially expanding the EQ-5D-5L to encompass these other domains less common in Western conceptualizations of HRQoL [16] and, consequently, lead to an expanded instrument specific to Aboriginal Australian populations. These studies can implement, for instance, focus groups to investigate the relevant domains and develop new culturally-specific items. Then, these items can be piloted in a population and the psychometric properties of the expanded instrument studied [85].

#### Conclusions

In this study, we employed a multi-method approach to comprehensively evaluate the psychometric properties of the EQ-5D-5L in a large sample of Aboriginal Australians in South Australia. The evidence showed that the EQ-5D-5L displayed excellent psychometric properties and is a potentially valid instrument to measure five specific aspects (Mobility, Self-Care, Usual activities, Pain/Discomfort, Anxiety/Depression) of Aboriginal and Torres Strait Islander HRQoL. Moreover, the EQ-5D-5L provides a health-state classification system that is amenable to future valuation using preference-based techniques. A future research agenda comprises the investigation of other domains of Aboriginal and Torres Strait Islander HRQoL and potential expansions to the instrument.

#### Abbreviations

AUROC: area under the receiver operating characteristic curve; CITC: corrected item-total correlation; CFA: confirmatory factor analysis; CLR: conditional likelihood ratio; EBIC: extended Bayesian information criteria; EGA: exploratory graph analysis; GGM: Gaussian graphical model; HPV: human papillomavirus; HRQoL: health-related quality of life; LASSO: least absolute shrinkage and selection operator; LRT: likelihood ratio test; MIRE: measure of indigenous racism experiences; NACCHO: national aboriginal community controlled health organisation; SEWB: social and emotional wellbeing; QALYs: quality-adjusted life years; WML: weighted maximum likelihood.

#### Acknowledgements

Not applicable.

#### Authors' contributions

PHRS conceptualized the idea, conducted the psychometric analysis and wrote the first draft of the manuscript. DH and DMM provided intellectual contribution to the psychometric analysis and critically reviewed the manuscript. GG, MS, KC, JH supervised the development of work, provided intellectual contribution and critically reviewed the manuscript. LJ conceptualized the idea, supervised the development of work, provided intellectual contribution and critically reviewed the manuscript. All authors read and approved the final manuscript.

### Funding

This research was supported by a grant from the Australia's National Health and Medical Research Council project Grant (APP1120215). The funding body was not involved in the design of the study, data collection, analysis, interpretation of data, and writing of the manuscript.

### Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available since we do not have permission from the ethics committee to publicly release the datasets in either identifiable or de-identified form. The datasets are available from the corresponding author on reasonable request.

### Ethics approval and consent to participate

Ethics approval was obtained from the University of Adelaide Human Research Ethics Committee (H-2016–246) and the Aboriginal Health Council of South Australia (04–17-729). All participants provided signed informed consent. All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup> Adelaide Dental School, The University of Adelaide, Adelaide, Australia. <sup>2</sup> School of Public Health, The University of Adelaide, Adelaide, Australia. <sup>3</sup> Menzies School of Health Research, Darwin, Australia. <sup>4</sup> Cancer Council of NSW, Sydney, Australia. <sup>5</sup> School of Public Health, University of Sydney, Sydney, Australia.

Received: 23 October 2020 Accepted: 23 February 2021

Published online: 10 March 2021

### References

- Engel GL. The need for a new medical model: a challenge for biomedicine. *Science*. 1977;196(4286):129–36.
- Greenfield S, Nelson EC. Recent developments and future issues in the use of health status assessment measures in clinical settings. *Med Care*. 1992;MS23-MS41.
- Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care*. 1992;473–83.
- Starfield B. Basic concepts in population health and health care. *J Epidemiol Commun Health*. 2001;55(7):452–4.
- Feeny DH, Eckstrom E, Whitlock EP, Perdue LA. A primer for systematic reviewers on the measurement of functional status and health-related quality of life in older adults. 2013.
- Butler TL, Anderson K, Garvey G, Cunningham J, Ratcliffe J, Tong A, et al. Aboriginal and Torres Strait Islander people's domains of wellbeing: a comprehensive literature review. *Soc Sci Med*. 2019;233:138–57.
- Guyatt GH, Cook DJ. Health status, quality of life, and the individual. *JAMA*. 1994;272(8):630–1.
- Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA*. 1994;272(8):619–26.
- National Aboriginal Community Controlled Health Organisation. Constitution for the National Aboriginal Community Controlled Health Organisation. 2010.
- Geisinger KF. Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychol Assess*. 1994;6(4):304.
- Brooks R, Group E. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
- Devlin NJ, Krabbe PF. The development of new research methods for the valuation of EQ-5D-5L. Springer; 2013.
- Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36.
- Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Qual Life Res*. 2009;18(2):253.
- Perkins M, Devlin N, Hansen P. The validity and reliability of EQ-5D health state valuations in a survey of Māori. *Qual Life Res*. 2004;13(1):271–4.
- Angell B, Muhunthan J, Eades A-M, Cunningham J, Garvey G, Cass A, et al. The health-related quality of life of Indigenous populations: a global systematic review. *Qual Life Res*. 2016;25(9):2161–78.
- Janssen M, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*. 2013;22(7):1717–27.
- Hameson L, Garvey G, Hedges J, Mitchell A, Dunbar T, Leane C, et al. Human papillomavirus and oropharyngeal cancer among indigenous Australians: protocol for a prevalence study of oral-related human papillomavirus and cost-effectiveness of prevention. *JMIR Res Protocols*. 2018;7(6):e10503.
- Graham JW. Missing data analysis: Making it work in the real world. *Annu Rev Psychol*. 2009;60:549–76.
- Konrath S, Meier BP, Bushman BJ. Development and validation of the single item narcissism scale (SINS). *PLoS ONE*. 2014;9(8):e103469.
- Laurenco LM, Mack HA, Daylight G, Wall S, Anderson M, Hoskins S, et al. Staying in touch with the community: understanding self-reported health and research priorities in older Aboriginal Australians. *Int Psychogeriatr*. 2020;32(11):1303–15.
- Farrington DP, Loeber R. Some benefits of dichotomization in psychiatric and criminological research. *Crim Behav Ment Health*. 2000;10(2):100–22.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29–36.
- Wong A, Hyde Z, Smith K, Flicker L, Atkinson D, Skeaf L, et al. Prevalence and sites of pain in remote-living older Aboriginal Australians, and associations with depressive symptoms and disability. *Intern Med J*. 2020.
- Vindigni D, Griffen D, Perkins J, Da Costa C, Parkinson L. Prevalence of musculoskeletal conditions, associated pain and disability and the barriers to managing these conditions in a rural, Australian Aboriginal community. *Rural Remote Health*. 2004;4(3):1.
- Cunningham J, Paradies YC. Patterns and correlates of self-reported racial discrimination among Australian Aboriginal and Torres Strait Islander adults, 2008–09: analysis of national survey data. *Int J Equity Health*. 2013;12(1):47.
- Paradies YC, Cunningham J. Development and validation of the measure of indigenous racism experiences (MIRE). *Int J Equity Health*. 2008;7(1):9.
- Golino HF, Epskamp S. Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One*. 2017;12(6).
- Kruis J, Maris G. Three representations of the Ising model. *Sci Rep*. 2016;6:34175.
- Golino H, Shi D, Christensen AP, Garrido LE, Nieto MD, Sadana R, et al. Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychol Methods*. 2020.
- Lauritzen SL. Graphical models. Clarendon Press, New York; 1996.
- Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B (Methodol)*. 1996;58(1):267–88.
- Foygel R, Drton M, editors. Extended Bayesian information criteria for Gaussian graphical models. *Adv Neural Inf Process Syst*; 2010.
- Pons P, Latapy M, editors. Computing communities in large networks using random walks. International symposium on computer and information sciences; 2005. Springer. Berlin
- Borsboom D, Cramer AO. Network analysis: an integrative approach to the structure of psychopathology. *Annu Rev Clin Psychol*. 2013;9:91–121.
- Christensen AP, Golino H, Silvia PJ. A psychometric network perspective on the validity and validation of personality trait questionnaires. *Eur J Personal*. 2020.
- Fruchterman TM, Reingold EM. Graph drawing by force-directed placement. *Softw Pract Exp*. 1991;21(11):1129–64.
- Christensen AP, Golino H. Estimating the stability of the number of factors via bootstrap exploratory graph analysis: a tutorial. 2019.

39. R Core Team. R: A language and environment for statistical computing. 2013.
40. Golino H, Christensen A. EGAnet: Exploratory Graph Analysis: A framework for estimating the number of dimensions in multivariate data using network psychometrics. R package version 04. 2019.
41. Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect Psychol Sci*. 2017;12(6):1100–22.
42. Fokkema M, Greiff S. How performing PCA and CFA on the same data equals trouble. Hogrefe Publishing; 2017.
43. Kline RB. Principles and practice of structural equation modeling. New York, NY: Guilford publications; 2015.
44. Asparouhov T, Muthén B. Simple second order chi-square correction. Mplus technical appendix. 2010.
45. Yu C-Y. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Los Angeles: University of California; 2002.
46. Steiger JH. Understanding the limitations of global fit assessment in structural equation modeling. *Pers Individ Dif*. 2007;42(5):893–8.
47. Guilford JP. The correlation of an item with a composite of the remaining items in a test. *Educ Psychol Meas*. 1953;13(1):87–93.
48. Gardner PL. Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Res Sci Educ*. 1995;25(3):283–9.
49. Zijlman EAO, Tijmstra J, der Ark V, Andries L, Sijtsma K. Item-score reliability as a selection tool in test construction. *Front Psychol*. 2018;9:2298.
50. Kendall SM. Rank correlation. *Van Nostrand's Scientific Encyclopedia*. 1948.
51. Revelle WR. *psych: Procedures for personality and psychological research*. 2017.
52. Nunnally JC, Bernstein IH. *Psychometric theory*. New York: McGraw-Hill; 1967.
53. Green SB, Yang Y. Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika*. 2009;74(1):155–67.
54. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014;105(3):399–412.
55. Rosseel Y. Lavaan: An R package for structural equation modeling and more Version 05–12 (BETA). *J Stat Softw*. 2012;48(2):1–36.
56. Young TA, Rowen D, Norquist J, Brazier JE. Developing preference-based health measures: using Rasch analysis to generate health state values. *Qual Life Res*. 2010;19(6):907–17.
57. Brazier J, Rowen D, Mavranzeouli I, Tsuchiya A, Young T, Yang Y, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). NIHR Health Technology Assessment programme: Executive Summaries: NIHR Journals Library; 2012.
58. Van Hout B, Janssen M, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708–15.
59. Golicki D, Niewada M. EQ-5D-5L crosswalk value set for Poland. *Value Health*. 2013;16(7):A599.
60. Wahlberg M, Zingmark M, Stenberg G, Munkholm M. Rasch analysis of the EQ-5D-3L and the EQ-5D-5L in persons with back and neck pain receiving physiotherapy in a primary care context. *Eur J Physiother*. 2019;1–8.
61. Masters GN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47(2):149–74.
62. Andersen EB. Asymptotic properties of conditional maximum-likelihood estimators. *J R Stat Soc Ser B (Methodol)*. 1970;283–301.
63. Warm TA. Weighted likelihood estimation of ability in item response theory. *Psychometrika*. 1989;54(3):427–50.
64. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43(4):561–73.
65. Shea TL, Tennant A, Pallant JF. Rasch model analysis of the depression, anxiety and stress scales (DASS). *BMC Psychiatry*. 2009;9(1):21.
66. Andersen EB. A goodness of fit test for the Rasch model. *Psychometrika*. 1973;38(1):123–40.
67. Christensen KB, Kreiner S. Item fit statistics. *Rasch models in health*. 2012:83–104.
68. Müller M. Item fit statistics for Rasch analysis: can we trust them? *J Stat Distrib Appl*. 2020;7(1):1–12.
69. Masters GN, Wright BD. The partial credit model Handbook of modern item response theory. Berlin: Springer; 1997. p. 101–21.
70. Andrich D, Sheridan B, Luo G. Manual for the rasch unidimensional measurement model (RUMM2030). Perth: RUMM Laboratory; 2010.
71. Andrich D, DeJong J, Sheridan BE. Diagnostic opportunities with the Rasch model for ordered response categories. Applications of latent trait and latent class models in the social sciences. 1997;59:70.
72. Kreiner S, Nielsen T. Item analysis in DIGRAM 3.04: Part I: Guided tours. 2013.
73. Mueller M. *iarms: item analysis in rasch models*. 2020.
74. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods*. 2002;7(1):19.
75. Butten K, Newcombe PA, Chang AB, Sheffield JK, O'Grady K-AF, Johnson NW, et al. Concepts of health-related quality of life of Australian Aboriginal and Torres Strait Islander children: parent perceptions. *Appl Res Qual Life*. 2020:1–19.
76. Gee G, Dudgeon P, Schultz C, Hart A, Kelly K. Aboriginal and Torres Strait Islander social and emotional wellbeing. *Work Together Aborig Torres Strait Islander Mental Health Wellbeing Principles Pract*. 2014;2:55–68.
77. Gupchup GV, Hubbard JH, Teel MA, Singhal PK, Tonrey L, Riley K, et al. Developing a community-specific health-related quality of life (HRQOL) questionnaire for asthma: the asthma-specific quality of life questionnaire for native American Adults (AQLQ-NAA). *J Asthma*. 2001;38(2):169–78.
78. Scott KM, Sarfati D, Tobias MI, Haslett SJ. A challenge to the cross-cultural validity of the SF-36 health survey: factor structure in Māori, Pacific and New Zealand European ethnic groups. *Soc Sci Med*. 2000;51(11):1655–64.
79. Van Der Ark LA. Relationships and properties of polytomous item response theory models. *Appl Psychol Meas*. 2001;25(3):273–82.
80. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. *Value Health*. 2008;11(2):275–84.
81. Pickard AS, Kohlmann T, Janssen MF, Bonsel G, Rosenbloom S, Cella D. Evaluating equivalency between response systems: application of the Rasch model to a 3-level and 5-level EQ-5D. *Med Care*. 2007;812–9.
82. Vander Weele TJ. Causal inference and constructed measures: towards a new model of measurement for psychosocial constructs. [arXiv preprint https://arxiv.org/abs/2007.00520](https://arxiv.org/abs/2007.00520). 2020.
83. Garvey D. Review of the social and emotional wellbeing of Indigenous Australian peoples. 2008.
84. Santiago PHR, Roberts R, Smithers LG, Jamieson L. Stress beyond coping? A Rasch analysis of the Perceived Stress Scale (PSS-14) in an Aboriginal population. *PLoS ONE*. 2019;14(5):e0216333.
85. Fayers PM, Machin D. *Quality of life: the assessment, analysis and reporting of patient-reported outcomes*. John Wiley & Sons; 2015.

## Publisher's Note

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