Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Do wording and response option modifications to the EQ-HWB-9 affect measurement performance? Evidence from the Netherlands and the USA

Akanksha Akanksha

akanksha.akanksha@uts.edu.au

The University of technology Sydney

Andrea De Palma

Erasmus University Rotterdam

Fanni Rencz

EuroQol Research Foundation

Tessa Peasgood

University of Sheffield

Brendan Mulhern

The University of technology Sydney

Aureliano Paolo Finch

EuroQol Research Foundation

Research Article

Keywords: EuroQol Health and Wellbeing, EQ-HWB-9, psychometric properties, convergent validity, known-groups validity, cross-national

Posted Date: July 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10216340/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. TP, BM, FR, AF are members of the EuroQol Group. FR and AF are employed by the EuroQol Research Foundation. FR serves as Co-Editor-in-Chief of Health and Quality of Life Outcomes and had no involvement in the editorial or peer review process for this manuscript. The other authors declare no competing interests. The views expressed are those of the authors and do not necessarily reflect the views of the EuroQol Research Foundation.

Abstract

Background

The EuroQol Health and Wellbeing 9 (EQ-HWB-9) and EQ-HWB long form measures cover dimensions relevant for healthcare and social care users, and their caregivers. After initial testing of the experimental version (v1.1, 2022), areas for possible refinement were identified in item wording, response option labels, and item ordering. A modified experimental version (v1.2) was introduced in 2024. This study examines the psychometric performance of the EQ-HWB-9 v1.2 and v1.1 in the Netherlands (NL) and the United States (USA).

Methods

Data from 3,783 adult participants (NL n = 1,571; USA n = 2,212) were collected via an online panel using quota sampling. All participants completed EQ-HWB-9 v1.2 as part of a longer survey and a subset was randomized to one of four arms, including EQ-HWB long form, EQ-HWB-9 v1.1 and its variants. Response distribution was explored, with instrument-level ceiling and floor effects defined as > 15%. Known group validity was assessed using Mann-Whitney U test across five pre-defined subgroups, and convergent validity was assessed using Spearman's correlations between EQ-HWB-9 versions, EQ-HWB long form, and EQ-5D-5L level sum scores.

Results

Instrument level floor/ceiling effects were not observed across versions, except for a marginal ceiling effect above threshold in the NL v1.1 (15.3%). v1.1 presented a higher proportion of Level 1 responses, particularly for day-to-day activities (USA - v1.2: 51.9%; v1.1: 60.5%; NL - v1.2: 51.8%; v1.1: 60.1%), and control (USA - v1.2: 53.4%; v1.1: 63.0%, v1.2: 56.0%; v1.1: 59.5%). All versions significantly discriminated across all known groups in both countries with large effect sizes in four out of five groups. v1.2 showed very strong cross-version correlations with v1.1 (rho = 0.88 USA; rho = 0.85 NL). Convergent validity was also demonstrated with external measures (EQ-5D-5L, EQ VAS, global health satisfaction, SWLS), with slightly higher correlations in v1.2 in both countries. Item-level convergent validity between matched v1.2 and v1.1 items was very strong to strong in both countries (rho = 0.57 to 0.82) as well as with conceptually matched EQ-5D-5L items.

Conclusion

EQ-HWB-9 v1.2 demonstrated at least equivalent or slightly superior psychometric performance to v1.1 across all properties and consistently in both countries, supporting the modifications adopted in v1.2.

Introduction

The EuroQol Health and Wellbeing (EQ-HWB) is a new generic measure for use in settings and populations where the measurement of wellbeing in addition to health is relevant to decision making, for example, in social care or in measuring impacts on carers [1]. The instrument (currently under development) covers seven dimensions (activity, relationships, cognition, self-identity, autonomy, feelings, and physical sensations) and is available in two versions: a long form (EQ-HWB) and a short form (EQ-HWB-9). The development of the EQ-HWB involved rigorous qualitative and quantitative assessments across six countries and multiple population groups, including patient groups with physical and/or mental conditions, social care recipients and caregivers [2, 3]. Multiple studies have explored the psychometric performance of the EQ-HWB and EQ-HWB-9 across patient, caregiver, and general populations, with findings supporting the instrument's validity [4–9].

Despite this robust development process and promising psychometric results, recent evidence has highlighted areas for refinement. Content validity studies identified some challenges related to the frequency and difficulty scales, such as distinguishing between the "only occasionally" and "sometimes" labels [10, 11]. Additionally, the order of the first two items (getting around preceded by day-to-day activities) led some respondents to anchor their activities response to their prior getting around response [12, 13]. In response to these findings, modifications to the original 'experimental' version (v1.1) were suggested to improve clarity, and interpretability of the instrument, forming a new 'experimental modified' version (v1.2).

The EQ-HWB v1.2 introduced several modifications to v1.1, including changes related to response option labels, item stems, item ordering, and instrument instructions. Proposed modifications included changing response options 'Some difficulty' and 'Only occasionally' with 'Moderate difficulty' and 'A little of the time', respectively. The first two items were re-ordered so that day-to-day activities preceded getting around, and the stem was modified from 'In the last 7 days:' to 'In the last 7 days, how much difficulty did you have:'. The wording for the getting around item was revised from 'getting around inside and outside' to 'getting around inside or outside'. Finally, instrument instructions were condensed to 'These questions are about your health and wellbeing in the last 7 days. Please choose one answer for each question.' All proposed modifications are listed in Supplementary Table S1.

Comprehensive evidence across all phases of instrument development is recommended by instrument development guidelines [14–16]. Several qualitative and quantitative studies are currently ongoing [17] with a few published to date [18, 19]. Importantly, early evidence stems from assessing a single version in isolation rather than comparing versions directly [13, 19]. For instance, a Chinese general population study assessed the psychometric performance of the EQ-HWB-9 v1.2, but, in the absence of a v1.1 comparison, could not attribute their findings to the modifications themselves [18]. To date, no published study has directly compared v1.1 and v1.2 quantitatively within the same sample. Studies are particularly needed in countries and languages relevant to the instrument's future use. The Netherlands has been identified as a setting where broader wellbeing measurement is relevant to economic evaluation [20],

and the USA provided the first opportunity to test the English-language source version from which all EQ-HWB-9 translations derive. This study addresses this gap by directly comparing the psychometric performance of EQ-HWB-9 v1.2 and v1.1 in both countries.

Methods

Study design and data source

This study analyses data linked to the Wellbeing Multi-Instrument Comparison (W-MIC) study [21]. Following informed consent, participants completed the EQ-HWB-9 v1.2 alongside a series of health and wellbeing measures presented in a random order within the core W-MIC survey. Immediately upon completion, participants were invited to a second survey and randomized into one of four arms. To systematically disentangle the psychometric impact of specific modifications, these arms were defined as follows:

- Arm 1 (v1.2 long form): The 'experimental modified' version of the EQ-HWB long form (v1.2)
- Arm 2 (v1.1 MRO): The original 'experimental' EQ-HWB-9 v1.1 with modified response options only
- Arm 3 (v1.1 SOE): The original 'experimental' EQ-HWB-9 v1.1 with modified item stems, ordering and explanations only
- Arm 4 (v1.1): The original 'experimental' EQ-HWB-9 v1.1

This multi-arm design allowed for a direct comparison of the EQ-HWB-9 v1.1 and v1.2 instruments while isolating the independent effects of response modifications. Arms 2 and 3 were restricted to the USA sample only. Arms 2 and 3 were not required in the Dutch sample due to linguistic equivalence in the Dutch translations, where wording of the modified response options in v1.2 did not differ from v1.1.

Participants and recruitment

Participants aged ≥ 18 years residing in the Netherlands or the USA were recruited between March and September 2025 via an online panel (Dynata) using quota sampling. Recruitment from the W-MIC survey targeted subgroups including the general population, older adults, and individuals with self-reported chronic health conditions. Specific quotas were set for these populations to ensure sufficient variability in health and wellbeing levels for psychometric evaluation. Details on the recruitment and quality control protocols are described elsewhere [21].

The study obtained ethics approval (ETH2425-0073) from the Erasmus School of Health Policy & Management Research Ethics Review Committee. The research was conducted in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Measures

The EQ-HWB-9 contains 9 items: getting around, day-to-day activities, exhaustion, loneliness, concentrating/thinking, anxiety, sad/depressed, control, and pain, each scored on a 5-level scale [1, 22]. Item responses were summed to produce a Level Sum Score (LSS; range 9–45, higher = worse health and wellbeing). The EQ-HWB long form (study-specific v1.2; 25 items) administered in Arm 1 incorporated revisions to item wording, ordering, and response options, including those adopted in EQ-HWB-9 v1.2.

To establish convergent and known-groups validity, the following instruments were purposefully selected for analysis from the broader W-MIC survey:

- The EQ-5D-5L, a widely used generic health-related quality of life instrument covering five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each scored on a 5-level scale, along with the EQ visual analogue scale (EQ VAS; 0–100, higher = better health) [23].
- The Patient Health Questionnaire-15 (PHQ-15) is a 15-item somatic symptom severity scale (total score 0 to 30, higher = greater symptom burden) [24]
- The Satisfaction with Life Scale (SWLS) is a 5-item measure used to assess global life satisfaction [25]. Each item includes a 7-point Likert scale, with higher scores indicating higher life satisfaction (total score 5 to 35).
- A global item was also used to measure self-reported satisfaction with health, on a scale from 0 ("Not at all satisfied") to 10 ("Completely satisfied").

Statistical analysis

Response distributions, convergent validity and known-groups validity were assessed. Analyses were conducted separately for the Netherlands and the USA using the pooled sample for each country. Specifically, response distributions were evaluated at both the item and instrument levels to compare the EQ-HWB-9 versions, while convergent and known-groups validity were assessed to validate the measure against external criteria.

Response distribution

Ceiling effects (> 15% of respondents scoring LSS = 9, indicating highest health and well-being) and floor effects (> 15% scoring LSS = 45, indicating lowest health and well-being) were examined at instrument level [26]. At the item level, response option frequencies were examined. Any response option with a selection proportion of ≥ 50% or < 5% was noted to characterize the concentration of responses rather than an indicator of item failure based on criteria established in prior studies [7, 9]. To evaluate whether the modifications resulted in statistically significant marginal distributional shifts, a Stuart-Maxwell test of marginal homogeneity was conducted utilizing the within-subject paired data from Arm 4 (USA n = 243; NL n = 163).

Convergent validity

Convergent validity was assessed at both the item and instrument levels using Spearman's correlations. Correlations were interpreted according to Cohen's criteria: 0.10–0.29 (weak), 0.30–0.49 (moderate), 0.50–0.69 (strong), and ≥ 0.70 (very strong) [27]. v1.2 and v1.1 were compared for an initial assessment of cross-version measurement equivalence. The EQ-5D-5L, EQ VAS, global health satisfaction item and SWLS were used as external comparators for convergent validity. Thirteen *a priori* hypotheses regarding expected direction and magnitude of item-level correlations between EQ-HWB-9 v1.2, v1.1 and EQ-5D-5L items were formulated by the authors based on conceptual overlap and prior evidence [7, 28] and are presented in Supplementary Table S2. This approach allowed to explore separately whether different versions of the same instrument measured the same construct and whether these versions measured the intended constructs when compared to relevant, external measures.

Known-group validity

Known-group validity was assessed by comparing EQ-HWB-9 LSS across five pre-defined subgroups: EQ VAS score (< 80 vs ≥ 80) [29], presence vs absence of chronic health conditions, PHQ-15 somatic symptom severity (< 10 vs ≥ 10) [24], satisfaction with health global item (≤ 5 vs > 5), and SWLS total score (< 20 vs ≥ 20) [25]. Given the skewed distribution of LSS data, the Mann-Whitney U test was used to compare groups. Effect sizes were quantified using rank-biserial correlation ($r = z/\sqrt{n}$), interpreted as small (0.10–0.29), medium (0.30–0.49), and large (≥ 0.50) [27].

Results

Sample characteristics

The final sample comprised 3783 participants (USA: n = 2212, 58.5%; NL: n = 1571, 41.5%). In the USA, the mean age was 61.6 years (SD = 14.8), and 52.2% were female. In the Netherlands, the mean age was 59.3 years (SD = 14.7), and 50.7% were female. Most participants in both countries were married or living with a partner (USA: 58.9%; NL: 64.0%), and retired (USA: 47.6%; NL: 39.9%). Most reported being able to make ends meet at least fairly easily (USA: 67.6%; NL: 75.1%). The proportion with no health conditions was similar across countries (USA: 11.7%; NL: 11.6%), with the majority reporting two or more conditions (USA: 68.5%; NL: 64.5%). No statistically significant differences were observed across arms for age, EQ VAS, health conditions, PHQ-15, or SWLS, in either country (all $p > 0.05$). A marginally significant difference in sex distribution was observed across US arms ($p = 0.04$), though no health-related variables differed suggesting randomization was broadly successful and arm-specific comparisons are unlikely to be confounded by sample composition. Full sociodemographic characteristics are presented in Table 1. Detailed sample characteristics by arm are presented in Supplementary Table S3.

Table 1
Sociodemographic characteristics of the sample by country

Variable	USA (n = 2212)	NL (n = 1571)
Sex, n (%)		
Female	1154 (52.2)	796 (50.7)
Male	1057 (47.8)	773 (49.2)
Prefer not to respond	1 (0.05)	2 (0.1)
Age, mean (SD) [range]	61.6 (14.8) [18–93]	59.3 (14.7) [18–90]
Age group, n (%)		
18–69	1352 (61.1)	1098 (69.9)
70 or older	860 (38.9)	473 (30.1)
Relationship status, n (%)		
Married/living with partner	1304 (58.9)	1006 (64.0)
Single	340 (15.4)	248 (15.8)
Divorced/Separated	321 (14.5)	153 (9.7)
Widowed	192 (8.7)	87 (5.5)
In relationship, not cohabiting	55 (2.5)	77 (4.9)
Education, n (%)		
Less than high school	23 (1.0)	27 (1.7)
High school or equivalent*	593 (26.8)	304 (19.4)
Associate/college degree*	470 (21.2)	625 (39.8)
Bachelor's degree*	656 (29.7)	481 (30.6)
Graduate degree*	470 (21.2)	134 (8.5)
Employment status, n (%)		
Full-time employed	589 (26.6)	351 (22.3)
Part-time employed	161 (7.3)	240 (15.3)
Retired	1052 (47.6)	627 (39.9)

*Values are n (%). *Education categories differ between countries: US labels shown; NL equivalents used during data collection. **Other employment includes self-employed, looking after home or family, student, and not employed.*

Variable	USA (n = 2212)	NL (n = 1571)
Sex, n (%)		
Unable to work (disability/illness)	158 (7.1)	169 (10.8)
Other**	252 (11.4)	184 (11.7)
Ability to make ends meet, n (%)		
Easily	699 (31.6)	473 (30.1)
Fairly easily	796 (36.0)	706 (44.9)
With difficulty	521 (23.5)	330 (21.0)
With great difficulty	196 (8.9)	62 (3.9)
Receives social care, n (%)		
Yes (formal only)	139 (6.3)	176 (11.2)
Yes (informal only)	255 (11.5)	133 (8.5)
Yes (both)	125 (5.6)	105 (6.7)
No	1693 (76.5)	1157 (73.6)
Health conditions, n (%)		
None	259 (11.7)	183 (11.6)
1 condition	437 (19.8)	374 (23.8)
2 or more conditions	1516 (68.5)	1014 (64.5)

*Values are n (%). *Education categories differ between countries: US labels shown; NL equivalents used during data collection. **Other employment includes self-employed, looking after home or family, student, and not employed.*

Response distribution

LSS descriptives by version and country are presented in Table 2. In the USA, mean LSS were 18.60 (SD = 8.19) in v1.2 and 17.53 (SD = 7.72) in v1.1. When evaluating the single modifications, the mean LSS was 16.72 (SD = 7.47) for the v1.1 MRO version and 16.72 (SD = 7.31) for the v1.1 SOE version. In the Netherlands, the mean LSS was 16.81 (SD = 7.02) for v1.2 and 15.75 (SD = 6.58) for v1.1. Median LSS ranged from 13 to 16 across all versions and both countries.

Table 2
EQ-HWB-9 Level Sum Score (LSS) descriptives by version and country

	USA				NL	
	v1.2	**v1.1 MRO**	**v1.1 SOE**	**v1.1**	**v1.2**	**v1.1**
n	2212	243	242	243	1571	163
Mean (SD)	18.60 (8.19)	16.72 (7.47)	16.72 (7.31)	17.53 (7.72)	16.81 (7.02)	15.75 (6.58)
Median	16	14	14	15	15	13
Min-Max	9–45	9–37	9–38	9–40	9–43	9–35
Ceiling, n (%)	121 (5.5)	21 (8.6)	16 (6.6)	20 (8.2)	122 (7.8)	25 (15.3) ^
Floor, n (%)	2 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

LSS = sum of all 9 item responses; range 9 (best) to 45 (worst).

[a]*Ceiling effect defined as > 15% of respondents scoring LSS = 9 (highest well-being); floor effect defined as > 15% scoring LSS = 45 (lowest well-being).*

^*Exceeds the 15% ceiling effect threshold. v1.2 = EQ-HWB-9 experimental modified version; v1.1 MRO = v1.1 with modified response options only (USA only); v1.1 SOE = v1.1 with modified stems, ordering and explanations only (USA only); v1.1 = EQ-HWB-9 original experimental version.*

At the instrument level, no version exceeded the 15% ceiling effect threshold in either country. Only v1.1 in the Netherlands marginally exceeded the 15% threshold, where 15.3% of respondents scored LSS = 9 (highest well-being). Given that this result is only fractionally above the descriptive threshold, it does not indicate a substantial ceiling issue. All other versions remained well below the 15% threshold in both countries (5.5% to 8.6%).

At the item level, a response proportion of ≥ 50% was observed for multiple items across all versions and both countries, particularly for v1.1. In the USA, the highest response proportion at Level 1 (no difficulty/none of the time) was observed for the getting around item (v1.2: 66.9%; v1.1: 67.1%), day-to-day activities (v1.2: 51.9%; v1.1: 60.5%), and control (v1.2: 53.4%; v1.1: 63.0%). In the Netherlands, a broader range of items exceeded this threshold in both versions, including getting around (v1.2: 72.2%; v1.1: 75.5%), anxiety (v1.2: 68.3%; v1.1: 73.0%), loneliness (v1.2: 58.8%; v1.1: 62.0%), control (v1.2: 56.0%; v1.1: 59.5%), sad/depressed (v1.2: 52.3%; v1.1: 60.1%), and day-to-day activities (v1.2: 51.8%; v1.1: 60.1%). The exhaustion and pain items showed the most distributed responses across all levels in both versions and both countries. Most items maintained stable marginal distributions across versions, although v1.2 consistently yielded a lower proportion of respondents reporting at Level 1 across nearly all items in both countries. Statistically significant differences between v1.2 and v1.1 were observed in only two items in the USA (sad/depressed, $p = 0.04$; control, $p < 0.001$) and three items in the

Netherlands (day-to-day activities, p = 0.04; concentrating/thinking, p = 0.02; sad/depressed, p = 0.01). Sad/depressed was the only item to demonstrate a significant distributional shift in both countries. Full item-level response distribution v1.2 and v1.1 are presented in Fig. 1. Complete distribution data across all arms are provided in Supplementary Tables S4 to S9. Full results of the marginal homogeneity testing are provided in Supplementary Table S10.

Convergent validity

Instrument-level correlations

Instrument-level Spearman correlations are presented in Table 3. All correlations were statistically significant ($p < 0.001$). In terms of cross-version correlations, v1.2 showed a very strong correlation with v1.1 (USA: rho = 0.88; NL: rho = 0.85). v1.2 and v1.1 were also compared against established measures, demonstrating similar external convergent validity with slightly higher correlations in v1.2. A very strong correlation with the EQ-5D-5L LSS was observed (v1.2 - USA: 0.79; NL: 0.77; v1.1 - USA: 0.77; NL: 0.72) as well as a strong correlation with the EQ VAS (v1.2 - USA: -0.62; NL: -0.69; v1.1 - USA: -0.61; NL: -0.67), global health satisfaction (v1.2 - USA: -0.58; NL: -0.66; v1.1 - USA: -0.56; NL: -0.59), and SWLS (v1.2 - USA: -0.63; NL: -0.64; v1.1 - USA: =-0.59; NL: -0.54). Similar results were observed for the other EQ-HWB versions. As anticipated, these correlations were strong but not perfectly collinear, reflecting the conceptual overlap between the measures while maintaining the broader scope of the EQ-HWB-9. Figure 2 presents the individual-level scatter plot of v1.2 and v1.1 LSS by country, illustrating the strong agreement between versions.

Table 3
Convergent validity: instrument-level Spearman correlations by country

Correlation pair	USA rho	NL rho
v1.2 LSS vs v1.1 LSS	0.88	0.85
v1.2 LSS vs v1.1 MRO LSS	0.87	N/A
v1.2 LSS vs v1.1 SOE LSS	0.88	N/A
v1.2 LSS vs EQ-5D-5L LSS	0.79	0.77
v1.1 MRO LSS vs EQ-5D-5L LSS	0.78	N/A
v1.1 SOE LSS vs EQ-5D-5L LSS	0.78	N/A
v1.1 LSS vs EQ-5D-5L LSS	0.77	0.72
v1.2 LSS vs EQ VAS	-0.62	-0.69
v1.1 MRO LSS vs EQ VAS	-0.54	N/A
v1.1 SOE LSS vs EQ VAS	-0.65	N/A
v1.1 LSS vs EQ VAS	-0.61	-0.67
v1.2 LSS vs global health satisfaction	-0.58	-0.66
v1.1 MRO LSS vs global health satisfaction	-0.52	N/A
v1.1 SOE LSS vs global health satisfaction	-0.63	N/A
v1.1 LSS vs global health satisfaction	-0.56	-0.59
v1.2 LSS vs SWLS	-0.63	-0.64
v1.1 MRO LSS vs SWLS	-0.55	N/A
v1.1 SOE LSS vs SWLS	-0.60	N/A
v1.1 LSS vs SWLS	-0.59	-0.54
v1.2 LSS vs EQ-HWB long form v1.2 LSS	0.85	0.84

All correlations significant at p < 0.001. Spearman's rho reported. Direction: positive = higher LSS associated with higher score on comparator; negative = higher LSS associated with lower score on comparator (as expected for EQ VAS where higher = better health and global health satisfaction and SWLS, where higher = more satisfaction). Item-level correlations are reported in the supplementary material. EQ-HWB = EQ Health and Wellbeing long form (Arm 1 only). EQ VAS = EQ-5D visual analogue scale

v1.2 = EQ-HWB-9 experimental modified version; v1.1MRO = v1.1 with modified response options only (Arm 2, USA only); v1.1SOE = v1.1 with modified stems, ordering and explanations only (Arm 3, USA only); v1.1 = EQ-HWB-9 original experimental version.

N/A was reported for correlations that included EQ-HWB-9 versions that were not part of the Dutch survey.

Item-level correlations

Item-level correlations between matched v1.2 and v1.1 items are presented in Supplementary Table S11. In the USA, correlations ranged from strong to very strong across all items (0.57–0.82). The anxiety (0.82), sad/depressed (0.82), and concentrating/thinking (0.81) items showed the highest correlations. The getting around (0.57) and day-to-day activities (0.63) items showed the lowest correlation coefficients among matched pairs. In the Netherlands, correlations ranged from strong to very strong (0.62–0.75), with pain (0.75) and exhaustion (0.74) showing the highest correlations and getting around (0.62) the lowest. Item-level correlations between EQ-HWB-9 v1.2 and matched items in the EQ-HWB v1.2 long form (Arm 1) were strong to very strong across all items in both countries (USA: 0.65–0.78; NL: 0.61–0.73). The strongest correlations were observed for loneliness (USA: 0.78; NL: 0.73) and anxiety (USA: 0.78; NL: 0.69). Control item showed the lowest correlations (USA: 0.66; NL: 0.62), though remaining strong in both countries.

All 13 pre-specified hypotheses were confirmed in both countries (Supplementary Table S2), with observed correlations exceeding the expected magnitude for 8 of 13 pairs in the USA and 7 of 13 in the Netherlands. Figure 3 presents item-level correlations between v1.2, v1.1, and conceptually matched EQ-5D-5L items by country. v1.2 presented similar or higher correlations with conceptually matched item compared to v1.1. The strongest correlations were observed for pain/discomfort with physical pain (v1.2 - USA: 0.82; NL: 0.76; v1.1 - USA: 0.74; NL: 0.71), anxiety/depression with feeling anxious (v1.2 - USA: 0.79; NL: 0.64; v1.1 - USA: 0.79; NL: 0.59), and anxiety/depression with feeling sad/depressed (v1.2 - USA: 0.79; NL: 0.71; v1.1 - USA: 0.75; NL: 0.65). Usual activities with day-to-day activities showed very strong correlations in both countries (v1.2 - USA: 0.74; NL: 0.73; v1.1 - USA: 0.61; NL: 0.72). Partially overlapping pairs showed strong to moderate correlations, including anxiety/depression with loneliness (v1.2 - USA: 0.66; NL: 0.59; v1.1 - USA: 0.64; NL: 0.49) and self-care with getting around (v1.2 - USA: 0.54; NL: 0.43; v1.1 - USA: 0.56; NL: 0.44).

Known-group validity

Known-groups validity results are presented in Table 4. All versions of the EQ-HWB-9 significantly discriminated between all five known-groups in both countries (all $p < 0.001$), except for the health condition group for v1.1 in the USA ($z=-2.67$, $p = 0.01$). Overall, effect sizes were slightly larger in v1.2 compared to v1.1 in both countries. In the USA, rank-biserial correlations for v1.2 ranged in magnitude from r=-0.23 (health condition group) to r=-0.61 (PHQ-15 group). For v1.1-MRO, r ranged from − 0.22 (health condition) to -0.61 (PHQ-15). For v1.1-SOE, r ranged from − 0.28 (health condition) to -0.55 (PHQ-15). For v1.1, r ranged from − 0.17 (health condition) to -0.54 (PHQ-15). In the Netherlands, rank-biserial correlations for v1.2 ranged in magnitude from r=-0.35 (health condition) to r=-0.55 (EQ VAS). For v1.1, r ranged from − 0.35 (health condition) to -0.55 (EQ VAS).

Table 4

Known-groups validity: Mann-Whitney z-statistics and rank-biserial correlations by version and country

Known group	USA				NL	
	v1.2	v1.1 MRO	v1.1 SOE	v1.1	v1.2	v1.1
Mann-Whitney z						
EQ VAS (< 80 vs ≥ 80)	24.90	7.60	8.73	8.34	21.97	7.02
Health condition	-10.64	-3.50	-4.31	-2.67*	-13.95	-4.46
PHQ-15 (< 10 vs ≥ 10)	-28.78	-9.45	-8.55	-8.41	-21.63	-6.58
Sat. health (≤ 5 vs > 5)	23.18	6.73	8.44	7.47	19.78	6.28
SWLS (< 20 vs ≥ 20)	26.11	7.56	8.63	8.49	19.65	5.62
Rank-biserial correlation (r)						
EQ VAS (< 80 vs ≥ 80)	0.53	0.49	0.56	0.54	0.55	0.55
Health condition	-0.23	-0.22	-0.28	-0.17*	-0.35	-0.35
PHQ-15 (< 10 vs ≥ 10)	-0.61	-0.61	-0.55	-0.54	-0.55	-0.52
Sat. health (≤ 5 vs > 5)	0.49	0.43	0.54	0.48	0.50	0.49
SWLS (< 20 vs ≥ 20)	0.56	0.49	0.55	0.55	0.50	0.44

*All p < 0.001 unless otherwise indicated. *p = 0.01; small no-condition cell (n = 14) in this arm. Negative z and r indicate higher LSS (worse health/wellbeing) in the expected group (i.e. those with health conditions, lower EQ VAS, higher PHQ-15 scores, lower satisfaction with health, and lower SWLS scores). Rank-biserial r = z/√n; interpreted as: small (0.10–0.29), medium (0.30–0.49), large (≥ 0.50)* [27]. *PHQ-15 = Patient Health Questionnaire-15. SWLS = Satisfaction with Life Scale. EQ VAS = EQ visual analogue scale.*

v1.2 = EQ-HWB-9 experimental modified version; v1.1 MRO = v1.1 with modified response options only (Arm 2, USA only); v1.1 SOE = v1.1 with modified stems, ordering and explanations only (Arm 3, USA only); v1.1 = EQ-HWB-9 original experimental version.

Values are percentages. Color scale: dark blue=Level 1 (no difficulty/none of the time) to dark red=Level 5 (unable/most or all of the time). Level 5 shown in parentheses outside bar. v1.2 = EQ-HWB-9 experimental modified version; v1.1 = EQ-HWB-9 original experimental version.

All Spearman's rho values p < 0.001. Dashed line indicates the threshold for strong correlation (rho = 0.50; [27]. *v1.2 = EQ-HWB-9 experimental modified version; v1.1 = EQ-HWB-9 original experimental version. v1.2 vs EQ-5D-5L and v1.1 vs EQ-5D-5L correlations based on conceptually matched items only (mobility, usual activities, pain, anxiety, depression). v1.2 vs v1.1 correlations based on Arm 4 participants (USA n = 243; NL n = 163). v1.2 vs EQ-5D-5L correlations based on full country samples (USA n = 2212; NL n = 1571*

The PHQ-15 group produced the largest effect sizes across all versions and both countries. The health condition group produced the smallest effect sizes across all versions and both countries. The EQ VAS, satisfaction with health, and SWLS groups showed medium to large effect sizes across all versions and both countries.

Discussion

This study directly compared the psychometric performance of EQ-HWB-9 v1.2 and v1.1 across the Netherlands and the USA, providing the first multicounty within-sample cross-version comparison for EQ-HWB-9. Across all three psychometric properties assessed, v1.2 demonstrated equivalent or slightly superior performance to v1.1, with findings consistent across both countries.

In terms of response distribution, the absence of floor and ceiling effects at instrument level across both versions is consistent with prior EQ-HWB-9 psychometric studies in general population samples [5, 18]. The marginal ceiling effect observed for v1.1 in the Netherlands is likely attributable to the healthier profile of the NL respondents in Arm 4. This was not observed for v1.2 in the Netherlands, though this difference should be interpreted cautiously given substantial sample size difference between v1.2 and v1.1 in the Netherlands. Item-level response distributions clustering at Level 1 was observed, with most pronounced for getting around, day-to-day activities and control items in both countries and anxiety and loneliness in the Netherlands. Similar skewed response distributions for these EQ-HWB-9 items have been reported previously [5, 7, 29] and are likely attributed to using general population samples. However, lower clustering was observed in v1.2 compared to v1.1, suggesting modifications may have improved discrimination across the full response scale in the general population.

Convergent validity results confirmed both EQ-HWB versions reproduced expected relationships between cross-versions and external measures in both countries. The very strong correlation between v1.2 and v1.1 LSS support similarity between the two versions and suggest the modifications did not affect their overall assessments. In addition, the relationships with other measures of health and wellbeing (EQ-5D-5L LSS, EQ VAS, global health satisfaction, SWLS) are consistent with evidence observed in prior EQ-HWB studies [6, 7, 28, 29]. EQ-HWB v1.2 also showed strong correlations with EQ-HWB v1.2 long form, supporting the expected association between the short form and the long form and confirming both instruments cover the same underlying domains.

Convergent validity was also demonstrated at the item level between conceptually matched items, providing evidence to the conceptual alignment between matched items. This was observed both between v1.2 and v1.1 items, v1.2 and long form items and v1.2 and EQ-5D-5L items. For the latter, all 13 *a priori* hypotheses were confirmed in both countries, with observed correlations exceeding expected magnitudes for the majority of pairs. The stronger-than-expected correlations for partially overlapping pairs, including anxiety/depression with loneliness, control, and concentrating/thinking suggest that the EQ-5D-5L anxiety/depression item is to some extent sensitive to a broader range of psychological and cognitive constructs [28]. Item-level correlations between matched v1.2 and v1.1 items were slightly

lower for getting around and day-to-day activities than for other items, which is expected given these items underwent the most substantive modifications in v1.2 including reordering, wording changes, and response option revision.

Convergent validity results are consistent with prior evidence [7, 28] and support the construct validity of v1.2. However, compared to Xu and Rencz (2025) [18], who also tested v1.2 in a Chinese general population sample, stronger correlations between conceptually overlapping items were observed in our study. Since both studies used v1.2, version differences are unlikely to explain this discrepancy. Differences in translation are also an unlikely explanation, as a previous study using the original EQ-HWB-9 v1.1 in a Chinese sample of patients, carers and general public observed stronger correlations between overlapping EQ-HWB-9 and EQ-5D-5L items [30]. Xu & Rencz proposed several explanations for their finding, including limited response variability in their healthy sample and potential effects of specific v1.2 modifications, namely the change from 'and' to 'or' in the getting around (mobility) item and the relocation of 'how much difficulty' to the item stem. Our findings, drawn from two larger and heterogenous population samples, suggest that sample composition is the more likely explanation, with weaker correlations in the Xu and Rencz's being more plausibly attributable to lower response variability in a healthy general population than to the v1.2 modifications themselves.

In terms of known-group validity, both v1.2 and v1.1 significantly discriminated between all five known groups in both countries, with medium to large effect sizes consistent with previous studies [12, 18, 29]. Effect sizes were comparable across versions and between countries, indicating that the modifications introduced in v1.2 did not alter the instrument's ability to detect health differences across groups. The PHQ-15 group produced the largest effect sizes across all versions, reflecting the EQ-HWB-9's ability to capture somatic symptom burden. The health condition group produced the smallest effect sizes, consistent with the broad definition used (any vs no condition) and the limited sample size of the "no health conditions" group. The cross-national consistency of effect sizes across all versions supports the generalisability of the modifications to both the Dutch and US contexts.

This study has several strengths. The large cross-national sample enabled direct within-sample comparison across two countries, providing a comprehensive psychometric evidence base for the EQ-HWB-9 modifications. The a priori hypothesis testing framework, informed by prior evidence, strengthens the rigour of the convergent validity assessment in line with COSMIN recommendations [15]. While COSMIN guidelines do not explicitly address the evaluation of subtle linguistic modifications, defining expected correlation directions and magnitudes *a priori* provides a robust, transparent method for confirming that these minor structural and wording changes preserved the underlying measurement construct.

Several limitations should be noted. The comparison between the modified version v1.2 and the original v1.1 was based on Arm 4 only, limiting statistical power for subgroup analyses. Recruitment via an online panel with specific quotas for older people and people with health conditions means the study sample does not fully represent the general population. However, this was intentional as the EQ-HWB is

designed to measure broader health and wellbeing outcomes across healthcare, public health, social care, and informal caregiving settings, rather than a healthy general population. Slight demographic imbalances across arms, particularly in the USA, may have introduced selection bias in arm-specific comparisons, although no arm-specific patterns were observed that could indicate a selection bias. Our reliance on a global LSS for the EQ-HWB-9 introduces specific psychometric trade-offs. As demonstrated by Feng et al., non-parametric item response theory indicates that, while the 9-item set satisfies monotone homogeneity, it lacks an invariant item ordering and points to an alternative two-factor structure separating physical domains from psychosocial domains [31]. While this alternative scoring could have identified potential domain-specific variations, LSS remains an acceptable summary measure for cross-country and cross-version comparison until further psychometric work establishes the validity and reliability of a multi-score alternative. Additionally, all respondents completed the EQ-HWB-9 v1.2 prior to their randomly assigned variant. This fixed administration order may have introduced order effects, and the repeated completion of similar items within a single survey could have contributed to respondent fatigue or recall bias. Finally, as a cross-sectional study, test-retest reliability and responsiveness could not be assessed.

EQ-HWB-9 v1.2 demonstrated equivalent or somewhat superior psychometric performance to v1.1 across feasibility, convergent validity, and known-groups validity in both the Netherlands and the USA. These findings provide quantitative evidence supporting the modifications incorporated in v1.2, complementing existing qualitative evidence and contributing novel international evidence for the EQ-HWB-9.

Declarations

Ethics

The study obtained ethics approval (ETH2425-0073) from the Erasmus School of Health Policy & Management Research Ethics Review Committee. The research was conducted in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflicts of interest

TP, BM, FR, AF are members of the EuroQol Group. FR and AF are employed by the EuroQol Research Foundation. FR serves as Co-Editor-in-Chief of *Health and Quality of Life Outcomes* and had no involvement in the editorial or peer review process for this manuscript. The other authors declare no competing interests. The views expressed are those of the authors and do not necessarily reflect the views of the EuroQol Research Foundation.

Funding

The study was funded by a research grant from the EuroQol Group (2511-RA).

Author Contribution

All authors contributed to the study conception and design. Main manuscript text, tables and figures were prepared by AA and ADP. All authors worked on the methodology; AA and ADP prepared the first draft and all authors reviewed, edited and approved the final version of the manuscript.

Data Availability

W-MIC data publically availabel here: https://url.au.m.mimecastprotect.com/s/1bIOCNLJ8nfN1wmzrImfJcyWU6C?domain=doi.orgAdditional data utilised for analysis available upon reasonable request.

References

1. Brazier J, Peasgood T, Mukuria C, Marten O, Kreimeier S, Luo N, Mulhern B, Pickard AS, Augustovski F, Greiner W, Engel L, Belizan M, Yang Z, Monteiro A, Kuharic M, Gibbons L, Ludwig K, Carlton J, Connell J, Rand S, Devlin N, Jones K, Tsuchiya A, Lovett R, Naidoo B, Rowen D, Rejon-Parrilla JC. The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing and Key Results. Value Health. 2022;25:482–91. https://doi.org/10.1016/j.jval.2022.01.009.
2. Carlton J, Peasgood T, Mukuria C, Connell J, Brazier J, Ludwig K, Marten O, Kreimeier S, Engel L, Belizán M, Yang Z, Monteiro A, Kuharic M, Luo N, Mulhern B, Greiner W, Pickard S, Augustovski F. Generation, Selection, and Face Validation of Items for a New Generic Measure of Quality of Life: The EQ-HWB. Value Health. 2022;25:512–24. https://doi.org/10.1016/j.jval.2021.12.007.
3. Peasgood T, Mukuria C, Brazier J, Marten O, Kreimeier S, Luo N, Mulhern B, Greiner W, Pickard AS, Augustovski F, Engel L, Gibbons L, Yang Z, Monteiro AL, Kuharic M, Belizan M, Bjørner J. Developing a New Generic Health and Wellbeing Measure: Psychometric Survey Results for the EQ-HWB. Value Health. 2022;25:525–33. https://doi.org/10.1016/j.jval.2021.11.1361.
4. Bailey C, Dalziel K, Jones R, Hiscock H, Devlin NJ, Peasgood T. In collaboration with the Quality of Life in Kids: Key evidence to strengthen decisions in Australia (QUOKKA) Project team: The Validity of the EuroQol Health and Wellbeing Short Version (EQ-HWB-S) Instrument in Parents of Children With and Without Health Conditions. PharmacoEconomics. 42, 163–79 (2024). https://doi.org/10.1007/s40273-024-01351-5
5. Bailey C, Trapani K, Davies JN, Van Dam N, Galante J, Peasgood T. The psychometric performance of the EQ-HWB-9 for measuring health and wellbeing in a general population sample from Australia and New Zealand. Qual Life Res. 2025;34:3707–19. https://doi.org/10.1007/s11136-025-04061-3.
6. Kuharic M, Mulhern B, Sharp LK, Turpin RS, Pickard AS. Comparison of the EQ-HWB and EQ-HWB-S With Other Preference-Based Measures Among United States Informal Caregivers. Value Health. 2024;27:967–77. https://doi.org/10.1016/j.jval.2024.03.003.

7. Monteiro AL, Kuharic M, Pickard AS. A Comparison of a Preliminary Version of the EQ-HWB Short and the 5-Level Version EQ-5D. Value in Health. 25, 534–43 (2022). https://doi.org/10.1016/j.jval.2022.01.003
8. Pangestu S, Purba FD, Setyowibowo H, Azhar Y, Mukuria C, Rencz F. The Psychometric Properties of the EQ-HWB and EQ-HWB-S in Patients With Breast Cancer: A Comparative Analysis With EQ-5D-5L, FACT-8D, and SWEMWBS. Value Health. 2025;28:449–59. https://doi.org/10.1016/j.jval.2024.12.003.
9. Zhang G, Yang Z, Luo N, Busschbach J. Comparing the Measurement Properties of the Preliminary Version of the EuroQol Health and Well-Being and EQ-5D-5L in Patients, Healthy General Public, and Caregivers. Value Health Reg Issues. 2025;47. https://doi.org/10.1016/j.vhri.2025.101081.
10. Gibson AEJ, Longworth L, Bennett B, Pickard AS, Shaw JW. Assessing the Content Validity of Preference-Based Measures in Cancer. Value Health. 2024;27:70–8. https://doi.org/10.1016/j.jval.2023.10.006.
11. Masutti S, Falivena C, Purba FD, Jommi C, Mukuria C, Finch AP. Content validity of the EQ-HWB and EQ-HWB-S in a sample of Italian patients, informal caregivers and members of the general public. J Patient Rep Outcomes. 2024;8:36. https://doi.org/10.1186/s41687-024-00706-y.
12. Bailey C, Dalziel K, Constable L, Devlin NJ, Hiscock H, Skouteris H, Peasgood T. The performance of the EQ-HWB-S as a measure of quality-of-life of caregivers in families that have experienced adverse events. Eur J Health Econ. 2025;26:7–21. https://doi.org/10.1007/s10198-024-01688-w.
13. Bailey C, Peasgood T. Content Validity of the EQ-HWB in Caregivers of Children with Health Conditions. Patient. 2025;18:661–71. https://doi.org/10.1007/s40271-025-00749-3.
14. Food and Drug Administration (FDA): Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. (2009).
15. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, Terwee CB. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf, (2019).
16. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, Schwartz C, Revicki DA, Moinpour CM, McLeod LD, Lyons JC, Lenderking WR, Hinds PS, Hays RD, Greenhalgh J, Gershon R, Feeny D, Fayers PM, Cella D, Brundage M, Ahmed S, Aaronson NK, Butt Z. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22:1889–905. https://doi.org/10.1007/s11136-012-0344-y.
17. : EuroQol Research Foundation, Projects F. https://euroqol.org/research-at-euroqol/our-research-portfolio/funded-projects/
18. Xu RH, Rencz F. Psychometric Properties of the Modified EQ Health and Wellbeing and its Shorter Version in a General Chinese Population: A Comparative Analysis With EQ-5D-5L, SF-6Dv2, ICEpop CAPability measure for Adults, Short Version of Warwick-Edinburgh Mental Wellbeing Scale, and

World Health Organization Well-Being Index 5-Item. Value Health. 2025;28:1241–52. https://doi.org/10.1016/j.jval.2025.04.2151.

19. Metyovinyi Z, Plázár D, Becker K, Medvecz M, Rencz F. Qualitative testing of the EQ-HWB-9 with a blank EQ-HWB VAS and EQ-5D-5L with skin bolt-ons in a rare genetic skin disease population. Qual Life Res. 2026;35:69. https://doi.org/10.1007/s11136-026-04177-0.
20. Geuzinge HA, Alili ME, Enzing JJ, Veld LMH. in 't, Knies, S., Wit, G.A. de: The New Dutch Guideline for Economic Evaluations in Healthcare: Taking the Societal Perspective to the Next Level. Value in Health. 28, 930–935 (2025). https://doi.org/10.1016/j.jval.2025.03.002
21. De Palma A, Finch AP, van Exel J, Stolk E. Measuring Wellbeing For Economic Evaluations: An Overview Of The Wellbeing Multi-Instrument Comparison (W-MIC) and Its Open-Access, Large Dataset. Under review.
22. Carlton J, Peasgood T, Mukuria C, Connell J, Brazier J, Ludwig K, Marten O, Kreimeier S, Engel L, Belizán M, Yang Z, Monteiro A, Kuharic M, Luo N, Mulhern B, Greiner W, Pickard S, Augustovski F. Generation, Selection, and Face Validation of Items for a New Generic Measure of Quality of Life: The EQ-HWB. Value Health. 2022;25:512–24. https://doi.org/10.1016/j.jval.2021.12.007.
23. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20:1727–36. https://doi.org/10.1007/s11136-011-9903-x.
24. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med. 2002;64:258–66. https://doi.org/10.1097/00006842-200203000-00008.
25. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess. 1985;49:71–5. https://doi.org/10.1207/s15327752jpa4901_13.
26. Terwee CB, Bot SDM, De Boer MR, Van Der Windt DAWM, Knol DL, Dekker J, Bouter LM, De Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.
27. Cohen J. A power primer. Psychol Bull. 1992;112:155–9. https://doi.org/10.1037/0033-2909.112.1.155.
28. Kinchin I, Engel L, Rencz F. A Comparative Study of Health and Well-Being Measures in Ireland Using EQ Health and Wellbeing (EQ-HWB) and its Short Version, EQ-5D-5L, and ICEpop Capability Measure for Adults (ICECAP-A). Value Health. 2025;28:1268–79. https://doi.org/10.1016/j.jval.2025.04.2160.
29. McDool E, Mukuria C, Peasgood T. Psychometric Performance of the EQ Health and Wellbeing Short in a United Kingdom Population Sample. Value Health. 2024;27:1215–24. https://doi.org/10.1016/j.jval.2024.05.007.
30. Long C, Mao Z, Yang Z. A Head-to-Head Comparison of EQ-HWB and EQ-5D-5L in Patients, Carers, and General Public in China. Value Health. 2024;27:848–56. https://doi.org/10.1016/j.jval.2024.02.012.

31. Feng Y-S, Kohlmann T, Peasgood T, Engel L, Mulhern B, Pickard AS. Scoring the EQ-HWB-S: can we do it without value sets? A non-parametric item response theory analysis. Qual Life Res. 2024;33:1211–22. https://doi.org/10.1007/s11136-024-03601-7.

Figures

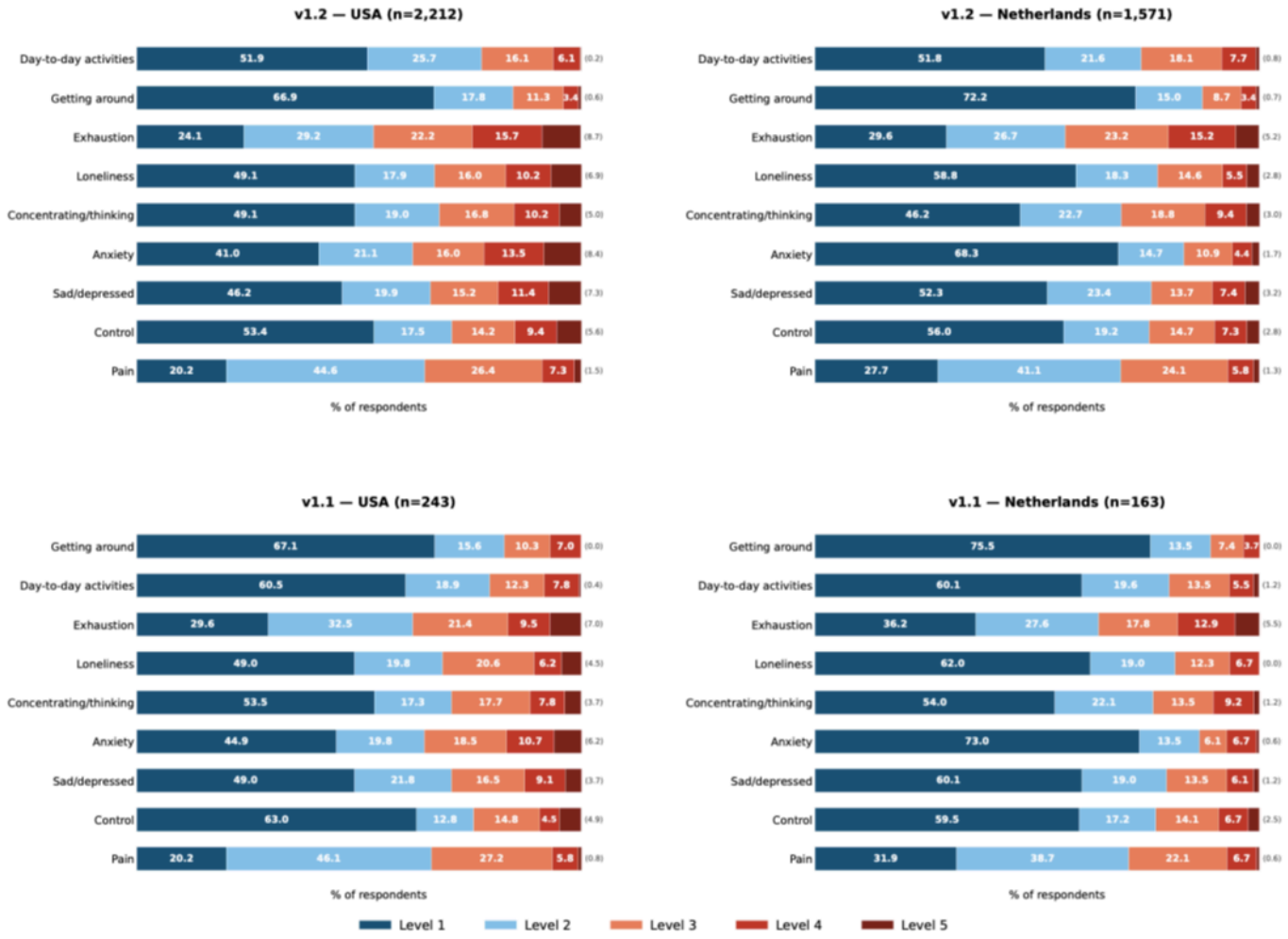


Figure 1

EQ-HWB-9 v1.1 and v1.2 response distribution by item, version and country

Values are percentages. Color scale: dark blue=Level 1 (no difficulty/none of the time) to dark red=Level 5 (unable/most or all of the time). Level 5 shown in parentheses outside bar. v1.2 = EQ-HWB-9 experimental modified version; v1.1 = EQ-HWB-9 original experimental version.

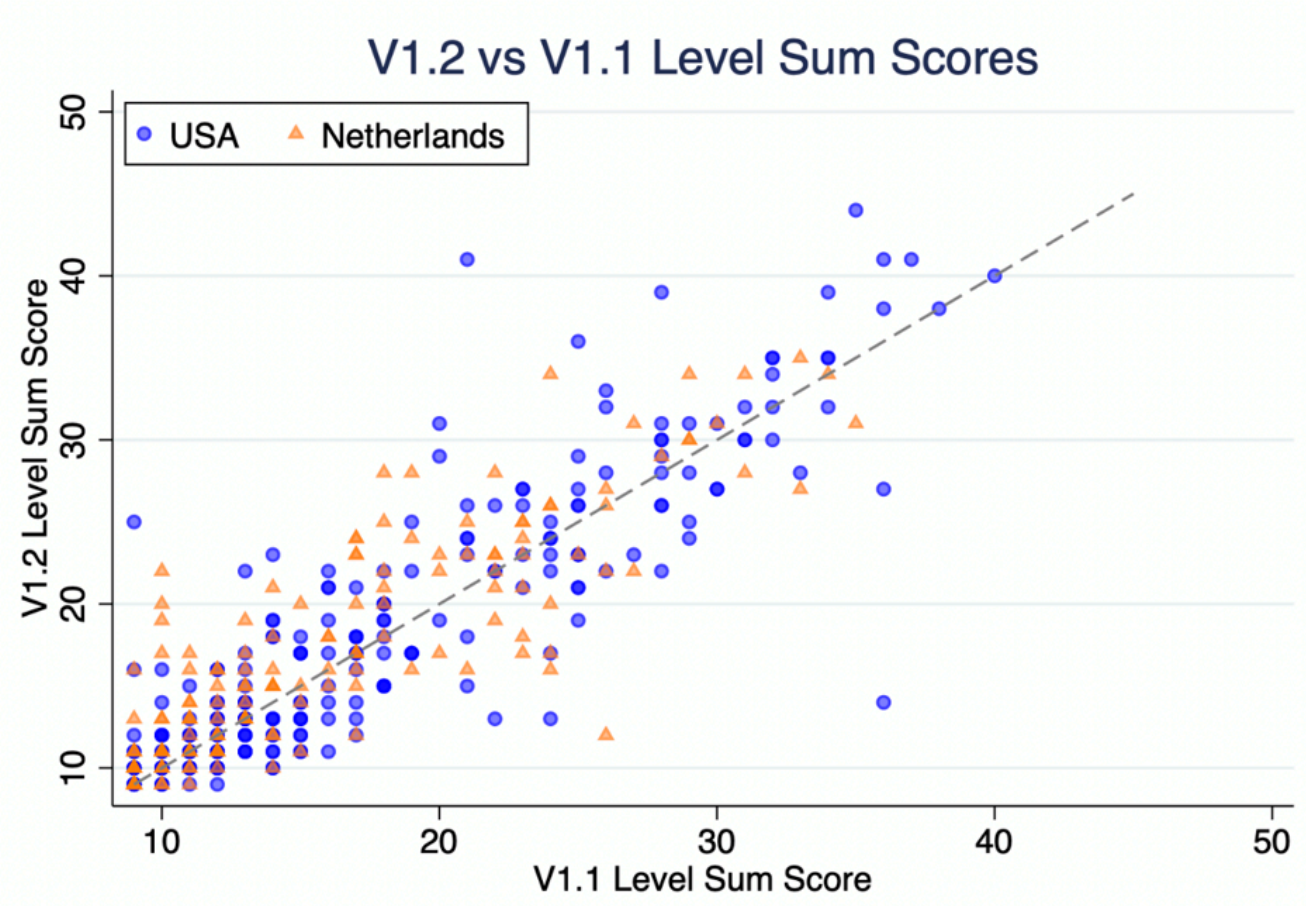


Figure 2

Correlation between EQ-HWB-9 v1.2 and v1.1 Level Sum Scores by country

Dashed line indicates perfect agreement. Spearman's rho: USA = 0.88; NL = 0.85. Both p<0.001. v1.2 = EQ-HWB-9 experimental modified version; v1.1 = EQ-HWB-9 original experimental version. Analysis restricted to Arm 4 participants who completed both versions (USA n=243; NL n=163).

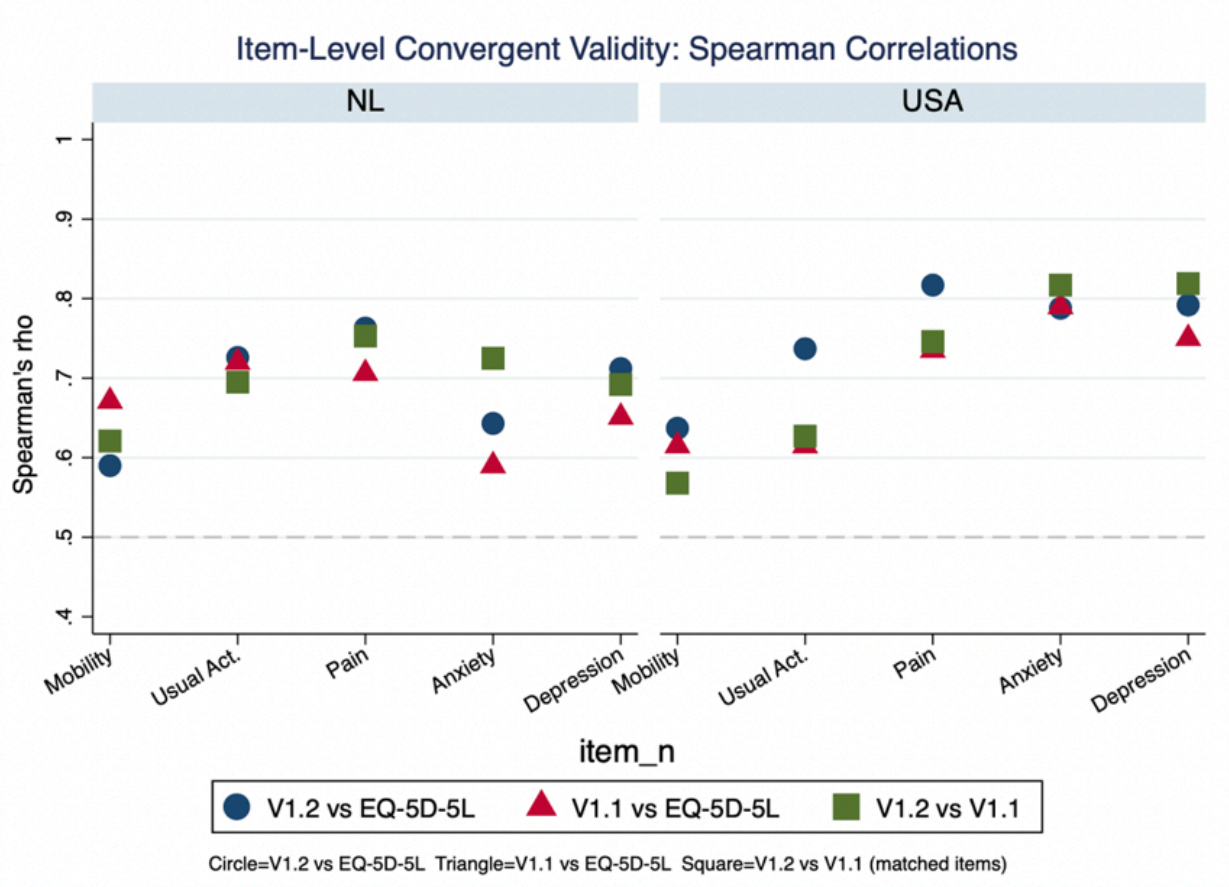


Figure 3

Item-level Spearman correlations between EQ-HWB-9 v1.2, v1.1, and EQ-5D-5L by country.

All Spearman's rho values p<0.001. Dashed line indicates the threshold for strong correlation (rho=0.50; [27]. v1.2 = EQ-HWB-9 experimental modified version; v1.1 = EQ-HWB-9 original experimental version. v1.2 vs EQ-5D-5L and v1.1 vs EQ-5D-5L correlations based on conceptually matched items only (mobility, usual activities, pain, anxiety, depression). v1.2 vs v1.1 correlations based on Arm 4 participants (USA n=243; NL n=163). v1.2 vs EQ-5D-5L correlations based on full country samples (USA n=2212; NL n=1571