3-1-2016

Measurement equivalence of the newly developed Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-55)

Mark A. Ferro
McMaster University

Shane W. Goodwin
Western University

Mark Sabaz
Sydney Children's Hospital, Randwick

Kathy N. Speechley
Western University, kspeechl@uwo.ca

FULL-LENGTH ORIGINAL RESEARCH

Measurement equivalence of the newly developed Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-55)

*†‡§Mark A. Ferro, ¶**Shane W. Goodwin, ††Mark Sabaz, and ¶**‡‡Kathy N. Speechley

Epilepsia, 57(3):427–435, 2016
doi: 10.1111/epi.13296

Dr. Mark Ferro is an Assistant Professor at McMaster University.

Summary

Objective: **The aim of this study was to examine measurement equivalence of the newly developed Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-55) across age, sex, and time in a representative sample of children with newly diagnosed epilepsy.**

Methods: **Data come from 373 children enrolled in the Health-related Quality of Life in Children with Epilepsy Study (HERQULES), a multisite prospective cohort study. Measurement equivalence was examined using a multiple-group confirmatory factor analysis framework, whereby increasingly stringent parameter constraints are imposed on the model. Comparison groups were stratified based on age (4–7 years vs. 8–12 years), sex (male vs. female), and time (measurement of health-related quality of life at diagnosis vs. 24 months later).**

Results: **The QOLCE-55 demonstrated measurement equivalence at the level of strict invariance for each model tested—age: χ^2 (3,123) = 4,097.3, $p < 0.001$; Comparative Fit Index (CFI) = 0.968; Root Mean Square Error of Approximation (RMSEA) = 0.042 (0.038, 0.045); sex: χ^2 (3,124) = 4,188.3, $p < 0.001$; CFI = 0.964; RMSEA = 0.044 (0.040, 0.047); and time: χ^2 (3,121) = 5,185.0, $p < 0.001$; CFI = 0.965; RMSEA = 0.046 (0.043, 0.048).**

Significance: **These findings suggest that items comprising the QOLCE-55 are perceived similarly among groups stratified by age, sex, and time and provide further evidence supporting the validity of the scale in children with epilepsy. Health professionals and researchers should be confident that group comparisons made using the QOLCE-55 are unbiased and that any group differences detected are meaningful; that is, not related to differences in the interpretation of items by informants. Future research replicating these findings is encouraged.**

KEY WORDS: **Factor analysis, Epilepsy, Health-related quality of life, Invariance, Measurement, Rating scales.**

Health-related quality of life (HRQoL) remains an important multidimensional clinical outcome and an active area of research in pediatric epilepsy. Numerous studies have shown that HRQoL, defined as the "subjective and objective impact of dysfunction associated with an illness or injury, medical treatment, and health care policy,"[1] is compromised in children with epilepsy.[2,3] Because HRQoL is recognized as one of the most important patient-reported outcomes,[4] considerable effort has been aimed at developing measures of HRQoL for children with epilepsy that are valid and

Accepted November 30, 2015; Early View publication January 10, 2016.

Departments of *Psychiatry and Behavioural Neurosciences; †Pediatrics; ‡Clinical Epidemiology and Biostatistics; §Offord Centre for Child Studies, McMaster University, Hamilton, Ontario, Canada; ¶Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada; **Children's Health Research Institute, Lawson Health Research Institute, London, Ontario, Canada; ††Department of Psychology, Sydney Children's Hospital (Randwick), Randwick, New South Wales, Australia; and ‡‡Department of Paediatrics, Western University, London, Ontario,Canada

Address correspondence to Mark A. Ferro, Department of Psychiatry and Behavioural Neurosciences, McMaster University, 1280 Main Street West, MIP 201A, Hamilton, Ontario, L8S 4K1, Canada. E-mail: ferroma@mcmaster.ca

Key Points

- Demonstrating measurement equivalence is a key component in scale development and evaluation
- The QOLCE-55 demonstrates equivalence across age, sex, and time in a representative sample of children with newly diagnosed epilepsy
- Group comparisons made using the QOLCE-55 are unbiased; any differences are meaningful and unrelated to differences in item interpretation

reliable.[5–9] Despite this progress, one aspect of scale development has largely been ignored: measurement equivalence.

As an approach to empirically test the external validity of a scale, measurement equivalence examines the extent to which the psychometric properties of observed scale items (i.e., questions) are generalizable across groups to determine if the same underlying construct, in this case, HRQoL, is being measured in the same way.[10] In other words, measurement equivalence tests whether questions asked of respondents are interpreted similarly across groups so that meaningful across-group comparisons can be made. Although often assumed, evidence suggests that response heterogeneity is common,[11] and violation of the assumption of measurement equivalence can lead to biased comparisons.[12] If measurement equivalence of a scale is demonstrated, differences, either among groups or over time, can be considered real and meaningful. However, without formally evaluating measurement equivalence, researchers and clinicians cannot be certain whether observed differences reflect true differences or whether they represent differences in the interpretation of items or structure of the underlying latent construct.[10] Evidence of measurement nonequivalence would provide the motivation to develop new or modify existing measures of HRQoL to result in more psychometrically sound measurement.

Although studies examining equivalence of HRQoL measures in children with epilepsy are nonexistent, some work has been done in other measures of health in children with epilepsy. One study found that the Attention and Thought Problems subscales of the Child Behavior Checklist and Youth Self Report demonstrated nonequivalence when comparing children with and without epilepsy.[13] Whereas modeling the association between epilepsy and these behavior problems with the nonequivalent scales suggested that children with epilepsy were at increased risk for attention problems ($\beta = 0.27$; $p < 0.001$), removal of nonequivalent items and computing the model with the equivalent scale negated this association ($\beta = 0.11$; $p = 0.417$). The consequences of such biases with the use of patient-reported outcomes can result in misinformed medical decision making among health professionals and families, and in turn, poorer health outcomes for children with epilepsy.

The objective of this study was to empirically test for measurement equivalence of the newly developed 55-item version of the Quality of Life in Childhood Epilepsy Questionnaire (QOLCE), an epilepsy-specific measure of HRQoL.[5] Measurement equivalence was tested between groups stratified by age (4–7 years vs. 8–12 years), sex (males vs. females), and time (measurement of HRQoL at diagnosis vs. 24 months later), as prior studies have shown associations between these variables and HRQoL in children with epilepsy.[3,14–16]

Methods

Data Source

Data come from 373 children enrolled in the Health-Related Quality of Life of Children with Epilepsy Study (HERQULES), a multisite prospective cohort of 4- to 12-year-old children newly diagnosed with epilepsy. Children were recruited from pediatric neurology outpatient clinics and followed for 24 months with data collected at diagnosis (baseline), and at 6, 12, and 24 months. At each of these data collection time-points, parents completed a postal survey that measured the HRQoL of their child. In addition, parents reported their own symptoms of depression, aspects of the family environment related to stressors, functioning, and resources, as well as sociodemographic information.

Pediatric neurologists provided clinical information including type of epilepsy syndrome and seizures, seizure frequency (low, moderate, high), and number of prescribed antiepileptic medications. Presence of cognitive or behavioral problems in children was based on neurologists' clinical experience measured using five-point and four-point Likert scales, respectively. Lower scores represented no/milder problems, and for this study, behavioral and cognitive problems were dichotomized as present or absent. In addition, neurologists provided an assessment of the overall severity of epilepsy using the Global Assessment of Severity of Epilepsy (GASE) scale.[17,18] The GASE is a validated, single-item, seven-point Likert scale that asks, "Taking into account all aspects of this patient's epilepsy, how would you rate its severity at his/her last visit?" Response options range from 1 = *extremely severe* to 7 = *not at all severe*. Additional details of the HERQULES sample and methodology are available.[3]

Instrument

The QOLCE-55 is a modified version of the original 76-item QOLCE[8,9] and provides an overall assessment of parent-reported HRQoL of children with epilepsy aged 4–18 years of age across the four domains of HRQoL (Fig. 1)—cognitive (22 items), emotional (17 items), social (seven items), and physical (nine items). A five-point Likert scale is used to rate items as follows: 0 = *very often*;

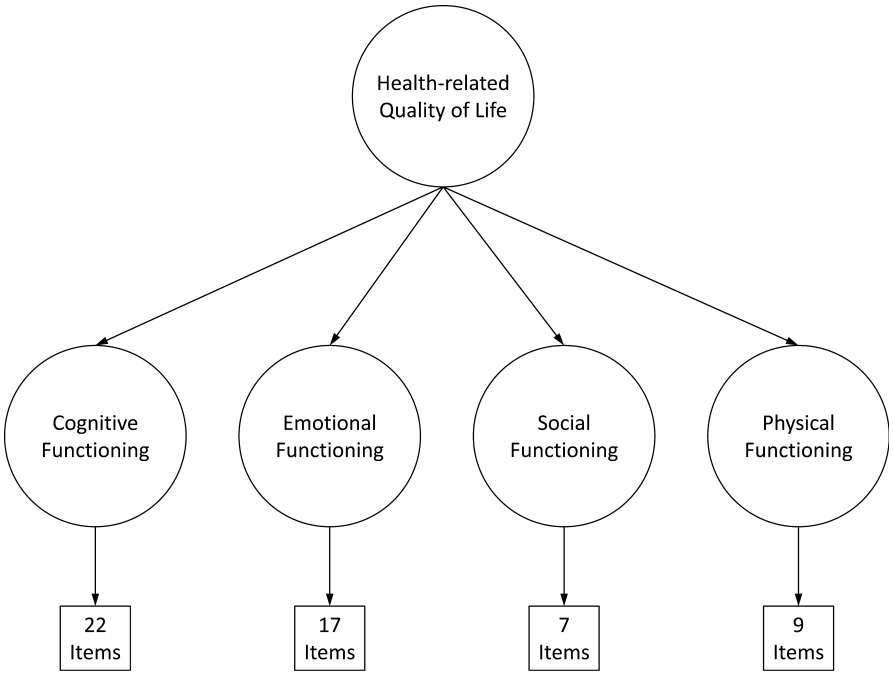

Figure 1.
Higher-order Factor Structure of the QOLCE-55. For simplicity, items loading onto the first-order factors are not shown. Individual items are presented in Table S1.
Epilepsia © ILAE

1 = *fairly often*; 2 = *sometimes*; 3 = *almost never*; and 4 = *never*. Ratings undergo linear transformation such that total scores can take values from 0 (low HRQoL) to 100 (high HRQoL). The QOLCE-55 has excellent internal consistency reliability ($\alpha = 0.96$). The QOLCE-55 and scoring instructions are freely available.[5]

Procedure

Measurement equivalence of the QOLCE-55 was investigated using multiple-group confirmatory factor analysis.[19] In this approach, increasingly stringent equality constraints are specified for model parameters between groups (younger vs. older age; male vs. female; and diagnosis vs. 24 months later). For the younger versus older age comparison, the sample was stratified to create groups that were as similar in size as possible. This is because the χ^2 difference test ($\Delta\chi^2$) used in measurement equivalence testing is affected by gross imbalances in group sizes that can result in biased inferences surrounding model fit and the equivalence of a measure.[10] For this study, younger children were those aged 4–7 years (n = 178) and older children (i.e., preadolescents) were those aged 8–12 years (n = 181). With regard to the longitudinal comparison, baseline data were compared to 24-month data for participants who completed both measurement occasions (n = 275). As this was the longest follow-up available in HERQULES, several participants (n = 86) transitioned from childhood to preadolescence, and thus our tests of measurement equivalence over time captured this important developmental milestone. For the sex comparison, there were 190 male and 169 female participants.

Guidelines for establishing measurement equivalence of higher-order factor models composed of categorical response items were followed.[19,20] (1) Configural invariance (model 1) imposed no equality constraints on parameters[19] and was the origin for subsequent tests.[21] (2) Weak invariance (constrained factor loadings) examined the extent to which the magnitude of the factor loadings (Λ) for particular items (model 2) and first-order factors (model 3) were equivalent between groups;[10] it is a prerequisite for making valid comparisons.[22] (3) Strong invariance (constrained item thresholds/intercepts) tested for evidence that item thresholds (ν; model 4) and first-order factor intercepts (τ) are equivalent between groups (model 5).[10] Strong invariance verifies whether mean differences at the item-level are fully explained by mean differences at the higher-order factor-level. (4) Strict invariance (constrained residual and factor variances) was performed to determine whether the variances (θ) of the regression equations for each item (model 6) and first-order factors (model 7) were equivalent across groups. Strict invariance is required for defensible item-score comparisons (i.e., average item scores) between groups.[23]

Evaluation and statistical analysis

Two criteria are required to establish measurement equivalence. First, model fit at each level of testing (i.e., configural, weak, etc.) must be adequate. It was determined a priori that two of the three fit indices (χ^2 goodness-of-fit, Comparative Fit Index [CFI], Root Mean Square Error of Approximation [RMSEA]) needed to meet established cut-points to declare adequate model fit.[13, 24] The cut-points were: χ^2 goodness-of-fit $p > 0.05$; CFI > 0.900 is acceptable, >0.950 is excellent; RMSEA < 0.08 is acceptable, <0.06 is excellent.[25–28] Second, *change* in fit indices from the less to the more constrained model (e.g., configural to weak), must not exceed established cut points for statistical or practical

significance. Statistical significance was based on the $\Delta\chi^2$ and practical significance on the ΔCFI and ΔRMSEA. Again, it was determined a priori that the $\Delta\chi^2$ and at least one of the ΔCFI or ΔRMSEA scores needed to meet this criterion to establish measurement equivalence at any given level of testing. The cut points for change in model fit indices were $\Delta\chi^2$ $p > 0.001$; $\Delta CFI \leq -0.010$; $\Delta RMSEA < 0.015$.[29] Statistical significance is supplemented by practical significance because the χ^2 difference is strongly influenced by sample size, and thus inferences related to measurement equivalence can be distorted.[10] To further reduce the effects of sample size on χ^2 difference test, we used the more conservative $\alpha = 0.001$ versus the traditional $\alpha = 0.05$ for statistical significance of equivalence testing.

Where measurement equivalence at a given level of testing was not established (i.e., significant worsening of model fit), modification indices were reviewed to identify constraints on relevant nonequivalent parameters that could be removed to improve model fit. The respecified model was then tested against the less constrained model and change scores for model fit computed. Freed parameters remained unconstrained as measurement equivalence testing proceeded. This approach, known as partial equivalence, argues that only a subset of model equivalent parameters are needed for substantive analyses comparing group mean differences.[30]

Because the QOLCE-55 contains ordered categorical response options, the confirmatory factor model used for measurement equivalence testing was estimated using a weighted least squares means and variance adjusted estimator.[31] This estimator uses a diagonal weight matrix and pairwise deletion of missing data to generate robust parameter estimates.[32] Fourteen participants (4%) were excluded from the analyses due to missing data on the QOLCE-55. Data analysis was conducted using M*plus* 6.11 (Muthén & Muthén).

Results

Sample characteristics

The mean age of children was 7.5 (standard deviation, 2.3) years and 53% were male. Most had partial seizures (62%), and cognitive and behavioral problems were reported in 20% and 15%, respectively. Epilepsy was relatively benign; mean scores on the GASE and QOLCE-55 were 5.4 (1.2) and 71.1 (14.3), respectively, at baseline. Based on the GASE, our sample would be described as consisting primarily of children with "somewhat severe" to "a little severe" epilepsy. Two thirds of children were currently prescribed at least one antiepileptic drug. Two thirds had seizure frequencies classified as low. Mothers were the most common caregiver informant (93%). Families' socioeconomic status was relatively high; most parents were in a partnered relationship (87%), had completed postsecondary education (66%), and were employed (67%). Nearly 40% had annual household incomes ≥$80,000. Sample characteristics by age and sex are described in Table 1. Parents of younger children were more likely to be younger ($p < 0.001$) and living in a partnered relationship ($p = 0.001$) compared to parents of older children. Compared to female children, male children were more likely to have partial seizures ($p = 0.013$) and behavioral problems ($p = 0.001$).

Measurement equivalence by age

Prior to equivalence testing, independent models for younger and older children were specified to verify adequate fitting baseline models with which to begin equivalence testing based on the established factor structure of the QOLCE-55 (Fig. 1). Data fit the factor model well for both the younger children: χ^2 (1,426) = 2,298.1, $p < 0.001$; CFI = 0.948; RMSEA = 0.059 (0.054, 0.063); and preadolescent groups: χ^2 (1,426) = 2,109.6, $p < 0.001$; CFI = 0.949; RMSEA = 0.051 (0.047, 0.056).

In the first step of testing, configural invariance, the factor structure of the QOLCE-55 was modeled simultaneously in both younger children and preadolescents with no parameter constraints. As shown in Table 2, model fit was adequate according to the prespecified limits: χ^2 (2,857) = 4,486.7, $p < 0.001$; CFI = 0.946; RMSEA = 0.056 (0.053, 0.060). At the level of weak invariance, constraints were placed first on item factor loadings and subsequently on first-order factor loadings. There was no appreciable worsening of model fit in both instances (models 2 and 3). In the next stage of testing, strong invariance, item thresholds, and then first-order intercepts were constrained between groups. Again, changes in the model-fit indices were within limits to continue equivalence testing (models 4 and 5). When item residuals were constrained at the level of strict invariance (model 6), there was evidence of nonequivalence [$\Delta\chi^2$ (55) = 129.6, $p < 0.001$]. Modification indices suggested that removing the constraints on following item residuals would improve model fit: *Had trouble understanding directions?* (cognitive); *Felt no one cared?* (emotional); and *Played with friends away from you or your home?* (physical). Once these constraints were removed, change in model-fit indices were acceptable [model 7: $\Delta\chi^2$ (52) = 86.6; ΔCFI = 0.002; ΔRMSEA = −0.001] and testing proceeded to constraining factor variances. Measurement equivalence was established at this final level of testing (model 8).

Measurement equivalence by sex

Baseline models for male: χ^2 (1,426) = 2,203.9, $p < 0.001$; CFI = 0.951; RMSEA = 0.054 (0.049, 0.058); and female participants: χ^2 (1,426) = 2,130.2, $p < 0.001$; CFI = 0.950; RMSEA = 0.054 (0.049, 0.059) indicated adequate fit to the factor model, and testing proceeded to establish measurement equivalence between sexes.

Table 1. Baseline sample characteristics stratified by age and sex

	Young children 4–7 years (n = 178)	Preadolescents 8–12 years (n = 181)	Males (n = 190)	Females (n = 169)
Child characteristics				
Age, years	5.5 (1.2)	9.5 (1.3)	7.4 (2.4)	7.6 (2.3)
Male, %	56.0	49.2	–	–
Partial seizures, %	62.3	58.6	67.4	54.8
Seizure frequency, %				
Low	64.1	67.6	66.0	66.5
Moderate	21.5	19.2	21.1	19.3
High	14.4	13.2	12.9	14.2
No. AEDs currently prescribed, %				
0	36.0	30.4	30.0	37.3
1	59.4	67.9	66.5	59.9
≥2	4.6	1.6	3.6	2.8
Cognitive problem, %	17.7	21.6	22.6	16.4
Behavior problem, %	14.1	15.7	21.2	8.4
Epilepsy severity, GASE	5.4 (1.2)	5.5 (1.2)	5.4 (1.3)	5.4 (1.1)
Health-related quality of life, QOLCE-55	70.5 (15.0)	71.8 (13.4)	70.9 (14.3)	71.7 (14.2)
Parent characteristics				
Age, years	36.0 (6.3)	39.3 (5.4)	37.3 (5.8)	38.1 (6.4)
Female, %	91.4	94.1	91.8	93.8
Living with a partner, %	91.4	81.8	89.2	83.7
Postsecondary education, %	66.7	66.3	69.2	63.5
Employed, %	64.9	69.2	67.4	66.7
Household income ≥$80,000, %	34.6	42.8	38.6	38.8

AEDs, antiepileptic drugs. Continuous variables are reported as mean (standard deviation).

Table 2. Tests of measurement equivalence by age

	χ^2 (df)	CFI	RMSEA (90% CI)	$\Delta\chi^2$ (df)	ΔCFI	ΔRMSEA
Configural invariance *no constraints*	4,486.7 (2,857)	0.946	0.056 (0.053, 0.060)	–	–	–
Weak invariance I *item factor loadings*	4,502.9 (2,908)	0.947	0.055 (0.052, 0.058)	64.5 (51)[a]	0.001	−0.001
Weak invariance II *first-order factors*	4,501.4 (2,911)	0.947	0.055 (0.052, 0.058)	3.7 (3)[a]	0.000	0.000
Strong invariance I *item thresholds*	4,556.3 (3,062)	0.950	0.052 (0.049, 0.055)	152.1 (151)[a]	0.003	−0.003
Strong invariance II *first-order intercepts*	4,490.4 (3,066)	0.953	0.051 (0.048, 0.054)	2.2 (4)[a]	0.003	−0.001
Strict invariance I *item residuals*	4,458.2 (3,121)	0.956	0.049 (0.046, 0.052)	129.6 (55)[b]	0.003	−0.002
Strict invariance I *item residuals* (θ_{CF1}, θ_{EF1}, θ_{PF1})	4,389.4 (3,118)	0.958	0.048 (0.044, 0.051)	86.6 (52)[a]	0.002	−0.001
Strict invariance II *factor variances*	4,097.3 (3,123)	0.968	0.042 (0.038, 0.045)	5.8 (5)[a]	0.010	−0.006

Children aged 4–7 years were in the "young child" group and those aged 8–12 in the "preadolescent" group. Italicized text describes the model parameters that were constrained at each step of the testing process. Constraints on the residuals of item 1 in each of the Cognitive (CF1), Emotional (EF1), and Physical Functioning (PF1) domains were removed in model 7 to establish partial equivalence and remained unconstrained in subsequent models. [a]Not statistically significant; [b]$p < 0.001$.

As shown in Table 3, model fit at the level of configural invariance was adequate according to the prespecified limits: χ^2 (2,857) = 4,463.3, $p < 0.001$; CFI = 0.946; RMSEA = 0.056 (0.053, 0.059). Similarly to tests between younger and older children, equivalence was established to the level of strong invariance (models 2 through 5). Again, item residuals were constrained at the level of strict invariance (model 6), there was evidence of nonequivalence [$\Delta\chi^2$ (55) = 110.9, $p < 0.001$]. Modification indices suggested that removing the constraints on following item residuals would improve model fit: *Felt down or depressed?* (emotional) and *Played with friends away from you or your*

Table 3. Tests of measurement equivalence by gender

	χ^2 (df)	CFI	RMSEA (90% CI)	$\Delta\chi^2$ (df)	ΔCFI	ΔRMSEA
Configural invariance *no constraints*	4,463.3 (2,857)	0.946	0.056 (0.053, 0.059)	–	–	–
Weak invariance I *item factor loadings*	4,482.3 (2,908)	0.947	0.055 (0.052, 0.058)	67.8 (51)[a]	0.001	−0.001
Weak invariance II *first-order factors*	4,479.1 (2,911)	0.947	0.055 (0.052, 0.058)	1.0 (3)[a]	0.000	0.000
Strong invariance I *item thresholds*	4,535.3 (3,062)	0.950	0.052 (0.049, 0.055)	157.8 (151)[a]	0.003	−0.003
Strong invariance II *first-order intercepts*	4,476.8 (3,066)	0.952	0.051 (0.047, 0.054)	3.4 (4)[a]	0.002	−0.001
Strict invariance I *item residuals*	4,404.0 (3,21)	0.957	0.048 (0.045, 0.051)	110.9 (55)[b]	0.005	−0.003
Strict invariance I *item residuals* (θ_{EF9}, θ_{PF1})	4,366.1 (3,119)	0.958	0.047 (0.044, 0.050)	85.4 (53)[a]	0.001	−0.001
Strict invariance II *factor variances*	4,188.3 (3,124)	0.964	0.044 (0.040, 0.047)	13.2 (5)[a]	0.006	−0.003

Italicized text describes the model parameters that were constrained at each step of the testing process. Constraints on the residuals of item 9 in the Emotional Functioning domain (EF9) and item 1 in the Physical Functioning (PF1) domain were removed in model 7 to establish partial equivalence and remained unconstrained in subsequent models. [a]Not statistically significant; [b]$p < 0.001$.

home? (physical). Once these constraints were removed, change in model fit indices were acceptable (model 7: $\Delta\chi^2$ (53) = 85.4; ΔCFI = 0.001; ΔRMSEA = −0.001), and testing proceeded to constraining factor variances that demonstrated measurement equivalence at this final level of testing (model 8).

Measurement equivalence over time

Baseline models at diagnosis—χ^2 (1,426) = 3,070.2, $p < 0.001$; CFI = 0.944; RMSEA = 0.058 (0.056, 0.061) and 24 months later: χ^2 (1,426) = 2,788.0, $p < 0.001$; CFI = 0.952; RMSEA = 0.059 (0.056, 0.062)—indicated adequate fit to the factor model and testing proceeded to establish measurement equivalence over time (Table 4).

At the level of configural invariance, model fit for equivalence over time was adequate: χ^2 (2,858) = 5,923.7, $p < 0.001$; CFI = 0.948; RMSEA = 0.058 (0.056, 0.061). Strong invariance was achieved (models 2 through 5); however, the constraints on a number of residuals needed to be removed to establish strict invariance (model 7): *Had trouble understanding directions?* (cognitive); *Felt no one cared? Felt excited or interested in something? Felt frustrated?* (emotional); *Limited his/her social activities (visiting friends, close relatives, or neighbors)? How limited are your child's social activities compared with others his/her age? Frightened other people?* (social); and *Played with friends away from you or your home?* (physical). Likewise, the constraint on the variance of Social Functioning domain needed to be removed to strict invariance at the higher order—model 9: $\Delta\chi^2$ (4) = 17.7; ΔCFI = 0.000; ΔRMSEA = 0.000. Parameter estimates for the final age, sex, and time-equivalent models are shown in the Table S1.

Discussion

Using data from a prospective cohort study of children newly diagnosed with epilepsy, measurement equivalence was demonstrated for the QOLCE-55—a newly developed, shorter version of the original scale measuring HRQoL in children with epilepsy. These findings suggest that items comprising the QOLCE-55 are perceived similarly between groups stratified by age and sex, as well as longitudinally among individuals assessed at diagnosis and again 24 months later, thus providing evidence to support the validity of the scale.

Full measurement equivalence was demonstrated through to the level of strong invariance (item thresholds/factor intercepts), and partial equivalence was demonstrated at the level of strict invariance (item residuals/factor variances). This finding is fundamental to establishing robust psychometric properties for the QOLCE-55. Previous research has shown that nonequivalent item thresholds or intercepts have large effects on the validity of mean scale comparisons between groups, whereas item residual effects are negligible.[33,34] This suggests that observed differences in QOLCE-55 scores are real, and not an artifact associated with performance of the scale or differential interpretation of items between comparison groups (e.g., males vs. females) or across repeated measurements over time. Although strict invariance is required for defensible item comparisons,[23] such investigations are typically exercised in psychometric analyses with minimal clinical relevance for health professionals and children with epilepsy and their families.

Of interest was the observation that the residuals associated with the Physical Functioning item—*Played with*

Table 4. Tests of measurement equivalence over time

	χ^2 (df)	CFI	RMSEA (90% CI)	$\Delta\chi^2$ (df)	ΔCFI	ΔRMSEA
Configural invariance *no constraints*	5,923.7 (2,858)	0.948	0.058 (0.056, 0.061)	–	–	–
Weak invariance I *item factor loadings*	5,943.1 (2,909)	0.949	0.057 (0.055, 0.059)	72.4 (51)[a]	0.001	−0.001
Weak invariance II *first-order factors*	5,940.9 (2,912)	0.949	0.057 (0.055, 0.059)	2.9 (3)[a]	0.000	0.000
Strong invariance I *item thresholds*	6,026.1 (3,066)	0.950	0.055 (0.053, 0.057)	208.3 (154)[a]	0.001	0.002
Strong invariance II *first-order intercepts*	6,043.6 (3,070)	0.949	0.055 (0.053, 0.057)	18.3 (4)[a]	−0.001	0.000
Strict invariance I *item residuals*	5,837.6 (3,125)	0.954	0.052 (0.050, 0.054)	198.1 (55)[b]	0.005	−0.003
Strict invariance I *item residuals* (θ_{CF1}, θ_{EF1}, θ_{EF12}, θ_{EF15}, θ_{SF1}, θ_{SF2}, θ_{SF7}, θ_{PF1})	5,655.8 (3,117)	0.957	0.051 (0.049, 0.053)	81.5 (47)[a]	0.003	−0.001
Strict invariance II *factor variances*	5,188.4 (3,122)	0.965	0.046 (0.043, 0.048)	25.8 (5)[b]	0.008	−0.005
Strict invariance II *factor variances* (θ_{SF})	5,185.0 (3,121)	0.965	0.046 (0.043, 0.048)	17.7 (4)[a]	0.000	0.000

Italicized text describes the model parameters that were constrained at each step of the testing process. Constraints on the residuals of item 1 in each of the Cognitive (CF1), Emotional (EF1), Social (SF1), and Physical Functioning (PF1) domains; as well as EF12, EF15, SF2, and SF7 were removed in model 7 and remained unconstrained in subsequent models. In model 9, constraints on the variance of SF domain was removed to establish partial equivalence. [a]Not statistically significant; [b]$p < 0.001$.

friends away from you or your home?—were nonequivalent in each model tested. This finding may be a function of the interpretation of the term "played," which can be quite different for preadolescents where play may or may not be equated with hanging out with friends or watching television compared to younger children where play may be related to toys or games. This may also explain why nonequivalence was observed in QOLCE-55 ratings over time (baseline vs. 24-months later) as nearly one fourth of our sample transitioned to preadolescence by the time of the 24-month follow-up. Likewise, the interpretation of play could also include gender-related undertones, whereby boys engage in more physically active play compared to girls.

Similarly, nonequivalent residuals associated with the Cognitive Functioning item, *Had trouble understanding directions?* and Emotional Functioning item, *Felt no one cared?* were found for age-related tests of equivalence (younger vs. older and longitudinal). This may be a function of parental expectations in the maturation of cognitive abilities of children during childhood. Matching this development are increased environmental expectations and demands on children in which cognitive difficulties may become more apparent. That is, the type of directions that younger children are expected to understand may be completely different from those in preadolescence, leading to nonequivalence. There is also some evidence to suggest that memory difficulties are associated with seizure type or epilepsy syndrome,[35] and thus they could potentially contribute to nonequivalence of the cognitive functioning item; however, there was no significant difference in the distribution of seizure types between younger children and preadolescents with epilepsy in our sample. In a related vein, feelings that no one cares, disconnectedness, or loneliness have been shown to be related to a lack of close friends, peer rejection, and victimization in older children and adolescents.[36] Such examples of stigmatization are commonly reported by children with epilepsy,[37] which may become more apparent as children mature through adolescence, thus contributing to nonequivalence.

Moving forward, use of the QOLCE-55 and replication of these findings in other samples of children with epilepsy is encouraged. As well, tests of measurement equivalence across cultural backgrounds and clinical aspects of epilepsy (e.g., seizure type, epilepsy severity, IQ, cognitive or behavioral comorbidities) are warranted, as these represent constructs whereby important differences in the interpretation of items comprising the QOLCE-55 could potentially emerge. Tests of equivalence between informants (e.g., mother vs. father) are also needed, given that maternal and paternal reports of child health and behavior often differ.[38,39] The current study could not address these tests of measurement equivalence due to: a lack of information on the cultural background of participants and formal diagnoses of cognitive or behavioral problems; small counts for specific seizure types (e.g., primary generalized, absence, and benign epilepsy of childhood with rolandic spikes); and few paternal informants (93% of informants were mothers). These limitations are not unique to the current study, but

reflect the current landscape of studies examining HRQoL in children with epilepsy. It is likely that researchers will need to pool data in an effort to achieve adequate statistical power to investigate these tests of measurement equivalence in the future. The QOLCE-55 is a parent-reported measure and as a result our study cannot address the phenomenon of child and parent discordance in reporting HRQoL in childhood epilepsy.[40] To do so would require a test of measurement equivalence between child and parent reports.

Conclusion

In summary, these findings provide evidence of measurement equivalence of the QOLCE-55 across age, sex, and time and contribute to its robust psychometric profile as a valid and reliable measure of HRQoL in children with epilepsy. Health professionals and researchers should be confident that such group comparisons made using the QOLCE-55 are unbiased and any differences detected are meaningful and not related to differences in the interpretation of items by informants.

Acknowledgments

We gratefully acknowledge the parents and physicians and their staff, without whose participation in HERQULES would not have been possible, as well as the Canadian Pediatric Epilepsy Network for effectively facilitating the participation of physician collaborators across the country. HERQULES was funded by a grant from the Canadian Institutes for Health Research (MOP-64311) to Dr. Speechley. Dr. Ferro is supported by the Research Early Career Award from Hamilton Health Sciences.

Disclosure

None of the authors has any conflict of interest to disclose. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

References

1. Spieth LE, Harris CV. Assessment of health-related quality of life in children and adolescents: an integrative review. *J Pediatr Psychol* 1996;21:175–193.
2. Wang J, Wang Y, Wang LB, et al. A comparison of quality of life in adolescents with epilepsy or asthma using the Short-Form Health Survey (SF-36). *Epilepsy Res* 2012;101:157–165.
3. Speechley KN, Ferro MA, Camfield CS, et al. Quality of life in children with new-onset epilepsy: a 2-year prospective cohort study. *Neurology* 2012;79:1548–1555.
4. Thurman DJ, Beghi E, Begley CE, et al. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia* 2011;52(Suppl 7):2–26.
5. Goodwin SW, Lambrinos AI, Ferro MA, et al. Development and assessment of a shortened Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-55). *Epilepsia* 2015;56:864–872.
6. Ronen GM, Streiner DL, Rosenbaum P, et al. Health-related quality of life in children with epilepsy: development and validation of self-report and parent proxy measures. *Epilepsia* 2003;44:598–612.
7. Cramer JA, Westbrook LE, Devinsky O, et al. Development of the quality of life in epilepsy inventory for adolescents: the QOLIE-AD-48. *Epilepsia* 1999;40:1114–1121.
8. Sabaz M, Cairns DR, Lawson JA, et al. Validation of a new quality of life measure for children with epilepsy. *Epilepsia* 2000;41:765–774.
9. Sabaz M, Lawson JA, Cairns DR, et al. Validation of the quality of life in childhood epilepsy questionnaire in American epilepsy patients. *Epilepsy Behav* 2003;4:680–691.
10. Brown TA. *Confirmatory factor analysis for applied research*. New York: The Guilford Press; 2006.
11. Muthén BO. Latent variable modeling in heterogeneous populations. *Psychometrika* 1989;54:557–585.
12. van de Schoot R, Lugtig P, Hox J. Developmetrics. A checklist for testing measurement invariance. *Eur J Dev Psychol* 2012;9:486–492.
13. Ferro MA, Boyle MH, Scott JG, et al. The Child Behavior Checklist and Youth Self-Report in adolescents with epilepsy: testing measurement invariance of the attention and thought problems subscales. *Epilepsy Behav* 2014;31:34–42.
14. Ronen GM, Streiner DL, Verhey LH, et al. Disease characteristics and psychosocial factors: explaining the expression of quality of life in childhood epilepsy. *Epilepsy Behav* 2010;18:88–93.
15. Devinsky O, Westbrook L, Cramer J, et al. Risk factors for poor health-related quality of life in adolescents with epilepsy. *Epilepsia* 1999;40:1715–1720.
16. Modi AC, Ingerski LM, Rausch JR, et al. Treatment factors affecting longitudinal quality of life in new onset pediatric epilepsy. *J Pediatr Psychol* 2011;36:466–475.
17. Speechley KN, Sang X, Levin S, et al. Assessing severity of epilepsy in children: preliminary evidence of validity and reliability of a single-item scale. *Epilepsy Behav* 2008;13:337–342.
18. Chan CJ, Zou GY, Wiebe S, et al. Global assessment of the severity of epilepsy (GASE) Scale in children: validity, reliability, responsiveness. *Epilepsia* 2015;13:337–342 doi:10.1111/epi.13216.
19. Millsap RE, Yun-Tein J. Assessing factorial invariance in ordered-categorical measures. *Multivariate Behav Res* 2004;39:479–515.
20. Chen FF, Sousa KH, West SG. Testing measurement invariance of second-order factor models. *Struct Equ Modeling* 2005;12:471–492.
21. Byrne BM. *Structural equation modeling with Mplus Basic concepts, applications, and programming*. New York: Taylor & Francis Group, LLC; 2012.
22. Bollen KA. *Structural equations with latent variables*. New York: John Wiley & Sons Inc; 1989.
23. Steinmetz H, Schmidt P, Tina-Booh A, et al. Testing measurement invariance using multiple group CFA: differences between educational groups in human values measurement. *Qual Quant* 2009;43:599–616.
24. Ferro MA, Boyle MH. Longitudinal invariance of measurement and structure of global self-concept: a population-based study examining trajectories among adolescents with and without chronic illness. *J Pediatr Psychol* 2013;38:425–437.
25. Hu L, Bentler PM. Cut-off criteria for fit indices in covariance structure analysis: conventional critieria versus new alternatives. *Struct Equ Modeling* 1999;6:1–55.
26. Schreiber JB, Nora A, Stage FK, et al. Reporting structural equation modeling and confirmatory factor analysis results: a review. *J Educ Res* 2006;99:323–337.
27. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;107:238–246.
28. Browne MW, Cudeck R. Alternative ways of assessing model fit. In Bollen KA, Long JS (Eds) *Testing structural equation models*. Newbury Park: Sage; 1993.
29. Chen FF. Sensitivity of goodness of fit indices to lack of measurement invariance. *Struct Equ Modeling* 2007;14:464–504.
30. Byrne BM, Shavelson RJ, Muthén B. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychol Bull* 1989;105:456–466.
31. Muthén LK, Muthén BO. *Mplus user's guide*. Los Angeles: Muthén & Muthén; 2010.
32. Asparouhov T, Muthén BO. *Weighted least squares estimation with missing data*. Los Angeles: Muthén & Muthén; 2010.
33. Steinmetz H. Analyzing observed composite differences across groups is partial measurement invariance enough? *Methodology* 2013;9:1–12.
34. Baumgartner H, Steenkamp JBEM. Multi-group latent variable models varying numbers of items and factors with cross-national and longitudinal applications. *Mark Lett* 1998;9:21–35.
35. Williams J. Learning and behavior in children with epilepsy. *Epilepsy Behav* 2003;4:107–111.
36. Qualter P, Vanhalst J, Harris R, et al. Loneliness across the life span. *Pers Psychol Sci* 2015;10:250–264.

37. Jacoby A, Austin JK. Social stigma for adults and children with epilepsy. *Epilepsia* 2007;48(Suppl 9):6–9.
38. Hechler T, Vervoort T, Hamann M, et al. Parental catastrophizing about their child's chronic pain: are mothers and fathers different? *Eur J Pain* 2011;15:515.e1–e9.
39. Weitkamp K, Daniels J, Rosenthal S, et al. Health-related quality of life: cross-informant agreement of father, mother, and self-report for children and adolescents in outpatient psychotherapy treatment. *Child Adolesc Mental Health* 2013;18:88–94.
40. Baca CB, Vickrey BG, Hays RD, et al. Differences in child versus parent reports of the child's health-related quality of life in children with epilepsy and healthy siblings. *Value Health* 2010;13:778–786.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Unstandardized parameter estimates for final age, sex, and longitudinal models.