Epilepsy & Behavior 90 (2019) 142–147

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Measurement equivalence of the short-form Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-16)

Shane W. Goodwin [a,*], Mark A. Ferro [a], Kathy N. Speechley [b,c,d]

[a] School of Public Health and Health Systems, University of Waterloo, Canada
[b] Department of Paediatrics, Western University, Canada
[c] Department of Epidemiology and Biostatistics, Western University, Canada
[d] Children's Health Research Institute, Lawson Health Research Institute, London, Ontario, Canada

ARTICLE INFO

Article history:
Received 19 July 2018
Revised 31 October 2018
Accepted 21 November 2018
Available online 7 December 2018

Keywords:
Pediatric epilepsy
Health-related quality of life
Measurement equivalence
Factor analysis
Validity

ABSTRACT

Objective: The aim of this study was to assess measurement equivalence in the 16-item short-form Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-16) across age, sex, and time in a sample of children with newly diagnosed epilepsy.
Methods: Data came from 373 children participating in the Health-related Quality of Life in Children with Epilepsy Study (HERQULES), a multicenter prospective cohort study. Measurement equivalence was assessed using multigroup confirmatory factor analysis methods. Comparison groups were stratified by age (4–7 years vs. 8–12 years), sex (male vs. female), and time (at diagnosis vs. 24 months postdiagnosis).
Results: The QOLCE-16 demonstrated measurement equivalence at the level of strict invariance for each comparison group tested — age: χ^2 (3, 274) = 429.6, $p < 0.001$; comparative fit index (CFI) = 0.985; root mean square error of approximation (RMSEA) = 0.056 (0.046, 0.066); sex: χ^2 (3, 271) = 430.5, $p < 0.001$; CFI = 0.984; RMSEA = 0.057 (0.047, 0.067); and time: χ^2 (3, 269) = 566.4, $p < 0.001$; CFI = 0.985; RMSEA = 0.059 (0.052, 0.066).
Significance: The findings provide support for the robust psychometric profile of the QOLCE-16 as a reliable and valid measure of health-related quality of life for children with epilepsy. Demonstrating good properties and a multidimensional structure, the QOLCE-16 is an appropriate short measure for both clinicians and researchers wanting to obtain health-related quality of life information on children with epilepsy.

1. Introduction

Health-related quality of life (HRQoL) is a critical clinical outcome for children with epilepsy [1] who, compared with their peers, report poorer HRQoL [2–5]. As a result, there is a need for robust, valid, and easily obtainable assessments. As health perspectives shift to long-term care and treatment, the ability to accurately obtain HRQoL information has become a priority for children living with epilepsy, parents caring for them, and clinicians hoping to establish ways to maintain or improve HRQoL over time [2].

One common measure of HRQoL in children with epilepsy is the Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-55) [6]. It is a 55-item, disease-specific, parent-reported measure of HRQoL for children with epilepsy [7]. The QOLCE-55 is multidimensional, with items providing insight into cognitive, emotional, social, and physical functioning domains. It has been shown previously to have good measurement properties [7–9]. Recently, a 16-item short-form version was created (QOLCE-16), with a goal to produce a measure that would minimize administrative burden in clinical settings while maintaining the strong properties of the longer version [10]. The QOLCE-16 provides a balanced measure of HRQoL by removing redundant or low information items and equalizing the number of items for each domain of HRQoL: cognitive, emotional, social, and physical functioning [10]. Nearly half of the items removed from the QOLCE-55 in creating the QOLCE-16 were those measuring aspects of cognitive functioning and item information curves suggested most of the removed items contributed negligibly to the explained variance of the overall HRQoL construct or were providing information already captured by the selected items [10]. The QOLCE-16 contains items that reflect the core components of HRQoL for computing an overall HRQoL score and tracking patient improvements in clinical settings. The QOLCE-16 is designed to focus on providing a valid and reliable measure of overall HRQoL, a construct relevant to patients, families, and health professionals.

In clinical practice, the feasibility of administering long-format questionnaires must be balanced by the knowledge obtained from using additional items. While there is a scarcity of research examining questionnaire length and administrative burden in pediatric clinical populations, a review of HRQoL measurement in pediatric clinical

* Corresponding author at: School of Public Health and Health Systems, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada.
E-mail address: sgoodwi6@uwo.ca (S.W. Goodwin).

https://doi.org/10.1016/j.yebeh.2018.11.024
1525-5050/

settings described questionnaire length as an additional burden on clinicians and administrators [11]. This sentiment was echoed in a recent systematic review in adult oncology [12,13]. Across all cases, the perception of time to administer, record, and interpret long and complex results were described as a challenge for clinical use. From a research perspective, the use of long-format questionnaires is particularly valuable when obtaining disease-specific information, while clinical practice benefits from short-form measures where the goal is to obtain indicators of improvement [11]. Having a measure that includes the core items of HRQoL that is short and easy to interpret has been described as a critical factor for clinical uptake [14].

Measurement equivalence examines the extent to which the psychometric properties of observed items are generalizable across groups [15]. The demonstration of measurement equivalence provides confidence that measurements obtained are assessing the same underlying construct and in the same way. Establishing measurement equivalence is a critical necessity for obtaining valid measurement [15,16]. The level at which nonequivalence is observed dictates the type of comparisons available to make. At the highest level, nonequivalence suggests that groups of individuals do not view the underlying constructs in a similar way, and no mean comparisons can be made, while at the lowest level, nonequivalence suggests that explained variance of each item is different, preventing item-level comparisons [15,16].

While the QOLCE-16 has demonstrated good measurement properties, measurement equivalence of the scale has yet to be established [10]. Measurement equivalence across age, sex, and time has been demonstrated previously in the QOLCE-55 [8,9]. The aim of the current research was to assess measurement equivalence using the short-form QOLCE-16. Comparison groups were stratified by age (4–7 years vs. 8–12 years), sex (males vs. females), and time (measurement of HRQoL at diagnosis vs. 24 months postdiagnosis).

2. Methods

2.1. Data source and participants

Data were obtained from the Health-Related Quality of Life in Children with Epilepsy Study (HERQULES). Health-Related Quality of Life in Children with Epilepsy Study was a multicenter prospective cohort study with the objective to investigate trajectories and determinants of HRQoL. Children aged 4–12 years with newly diagnosed epilepsy from across Canada were followed during their first two years postdiagnosis. Data were collected at four times: baseline (as close as possible to the time of diagnosis), 6, 12, and 24 months postdiagnosis. Participant inclusion occurred in two steps, with pediatric neurologists identifying eligible children and parents being asked to participate. A total of 456 children were identified by neurologists as eligible for participation, resulting in 373 parents (82%) completing a baseline self-administered questionnaire, 336 completing a 6-month questionnaire, 304 completing a 12-month questionnaire, and 282 completing a 24-month questionnaire. The current study used baseline data (N = 373) for comparisons of age and sex, and both baseline and 24-month data (N = 282) for comparisons across time. At each data collection point, physicians collected clinical information including type of epilepsy syndrome, frequency and severity of seizures, and antiepileptic drug (AED) use. Information on AEDs included current number of medications prescribed and an overall total number of AEDs that have been prescribed across the patient's history. The severity of cognitive and behavioral problems was reported by clinicians based on their clinical judgment and rated on five-point and four-point scales, respectively, where higher scores indicated more severe problems. In this study, these scores were dichotomized to indicate the presence or absence of cognitive problems and behavioral problems. Clinicians using the Global Assessment of the Severity of Epilepsy (GASE) scale, a single-item measure, reported severity of epilepsy. Scores range from 1 = extremely severe to 7 = not at all severe [17,18]. A more detailed description of the HERQULES methodology has been provided previously [3].

2.2. Instrument

The QOLCE-16 is a short-form version of the recently developed QOLCE-55 [7], created with the goal to provide an accessible measure to assess HRQoL with minimal respondent burden [10]. The QOLCE-16 measures child HRQoL based on parent report across four domains: cognitive, emotional, social, and physical functioning, with each domain assessed using four items rated on a 5-point Likert scale: 0 = *very often*, 1 = *fairly often*, 2 = *sometimes*, 3 = *almost never*, and 4 = *never* [10]. Ratings are linearly transformed, providing total scores from 0 (low HRQoL) to 100 (high HRQoL) [10]. The QOLCE-16 contains four subscales: cognitive, emotional, social, and physical functioning. Each domain consists of four items that are then used to calculate the subscale score. The QOLCE-16 has excellent internal consistency reliability ($\alpha = 0.90$). The QOLCE-16 and scoring instructions are freely available [10].

2.3. Statistical analyses

Analyses were conducted using Mplus software version 7 [19]. The established higher-order factor structure of the QOLCE-55 [7–9] was used as the basis for all equivalence-testing models. The QOLCE-16 was developed using a bifactor structure to take advantage of benefits available when using item-response theory methods for item selection, such as, the use of item characteristic curves to identify poor performing items [10]. A bifactor structure is an alternative conceptualization of a higher-order factor structure, where all items load onto a general factor (conceptualized as HRQoL in the QOLCE-16) in addition to loading on their respective domains [20]. There are a number of similarities between bifactor structures and higher-order factor structures, and it has been established that the two are often similar in estimates, or even equivalent under specific constraints [21,22]. Because of the ease of interpretation and common conceptualization in health research, the higher-order factor structure was used for all analyses.

Weighted least squares means and variance adjusted (WLSMV) estimation was used to obtain all estimates. In Mplus, the WLSMV produces and analyzes a polychoric correlation matrix and an associated weight matrix to provide robust unbiased estimation for categorical indicators. Further, WLSMV is robust against nonnormal data. Weighted least squares means and variance adjusted estimation using the theta parametization was implemented to obtain parameters for residual variances, allowing an additional investigation into a source for group differences [23]. Adequacy of model fit was evaluated based on expert recommendations [24–26]: Bentler's comparative fit index (CFI; where >0.900 is acceptable), root mean square error of approximation (RMSEA; where <0.08 is acceptable and <0.06 is excellent), and the weighted root mean square residuals (WRMR; where <1.00 is good) [24–26].

Measurement equivalence was evaluated using multiple-group confirmatory factor analysis. Groups for comparisons were age: 'young children' (4–7 years; n = 178) vs. 'preadolescents' (8–12 years; n = 181); sex: males (n = 190) vs. females (n = 169); and time: baseline (n = 373) and 24 months postdiagnosis (n = 275).

In this study, measurement equivalence was tested in a nested model approach as is common in equivalence research [27,28]. The steps used in this study are as follows: (1) establish the factor structure in each group separately; (2) test of configural invariance (*model 1* – no constraints on parameters); (3) test of metric or weak factor invariance (*model 2* – constraints added for invariance of factor loadings Λ for particular items; *model 3* – constraints added for invariance of factor loadings for first-order factors); (4) test of scalar or strong invariance (*model 4* – constraints added for invariance of item thresholds υ and first-order factor intercepts τ); and (5) test of strict invariance (*model 5* – constraints added for invariance of residual and factor variance θ). If at any level of analysis equivalence was not established, tests of partial invariance (placing specific constraints on individual items based on

statistical suggestions of model fit) are conducted. Any partial invariance (item-specific constraints) is then continued forward to the next sequential test.

All tests of equivalence were sequential. Model comparisons were done using Mplus DIFFTEST procedure for nested model comparisons under WLSMV. Partial invariance testing was informed by modification indices, which suggest equality constraints to remove to improve model fit. To establish measurement equivalence, two criteria were assessed: (1) model fit at each level of testing and (2) change in fit indices from the less constrained model to the more constrained model. Model fit was deemed adequate if two of three fit indices (χ^2 goodness-of-fit, CFI, RMSEA) met cutpoints that were established a priori [27,28]. The cutpoints were as follows: χ^2 goodness-of-fit, $p > 0.05$; CFI > 0.900; and RMSEA < 0.08. Furthermore, going from less constrained to more constrained models required that at least two of three change statistics met the following criteria: $\Delta\chi^2$, $p > 0.05$; ΔCFI ≥ −0.010; and ΔRMSEA < 0.015 [28–32]. Research has suggested that partial equivalence of two weak and a single strong invariance is sufficient for meaningful measurement scales [32].

3. Results

3.1. Sample characteristics

The mean age of children was 7.5 (2.3) years and 53% were male. Partial seizures were the most common diagnosis (62%). Severity of epilepsy was mild on average, with mean GASE scores of 5.4 (1.2) out of 7. Cognitive problems were reported in 20% of children while behavioral problems were reported in 15% of children. A majority of families reported annual incomes of at least $60,000 (56%) and 67% of parents having completed postsecondary education. Further details of the sample are found in Table 1.

3.2. Measurement equivalence by age

Baseline model fit adequacy was established for both children in the young group: $\chi^2(100) = 161.0$, $p < 0.001$; CFI = 0.991; RMSEA = 0.059 (0.041, 0.075) and preadolescent group: $\chi^2(100) = 168.1$, $p < 0.001$; CFI = 0.983; RMSEA = 0.061 (0.045, 0.077). In configural invariance (see Table 2, model 1), model fit was adequate: $\chi^2(203) = 448.6$, $p < 0.001$; CFI = 0.977; RMSEA = 0.082 (0.072, 0.092). At the level of weak invariance (models 2 and 3), model fit did not significantly worsen. At the level of strong invariance (model 4), model fit did not worsen; however, $\Delta\chi^2$ comparisons between models could not be computed due to the data distribution issues resulting in a singular matrix [31]. While statistical methods exist to attempt to alleviate the issue [33,34], the consequence of these methods on interpretation of results is unknown [33]. As such, given the successful fit statistics of our models, we did not apply techniques to allow $\Delta\chi^2$ comparisons between models. At the level of strict invariance (model 5), model fit did not worsen. Overall, no items were found to be nonequivalent between age groups.

3.3. Measurement equivalence by sex

Baseline factor models were adequate for both the males: $\chi^2(100) = 204.6$, $p < 0.001$; CFI = 0.980; RMSEA = 0.074 (0.060, 0.089) and females: $\chi^2(100) = 132.5$, $p < 0.016$; CFI = 0.993; RMSEA = 0.044 (0.020, 0.063). In configural invariance (see Table 3, model 1), model fit was adequate: $\chi^2(203) = 415.4$, $p < 0.001$; CFI = 0.979; RMSEA = 0.076 (0.066, 0.087). Model fit did not significantly worsen for weak invariance (model 2 and model 3) or strong invariance (model 4). There was evidence of nonequivalence at the level of strict invariance (model 5: $\Delta\chi^2$ (16) = 43.01, $p < 0.001$). Modification indices suggested removing the constraint on the residual of item 1 in the cognitive functioning domain: *Had trouble understanding directions?* Once the constraint was removed, invariance was achieved at the strict level (model 6).

3.4. Measurement equivalence by time (baseline and 24 months postdiagnosis)

Baseline and two-year data were adequate (baseline data: $\chi^2(100) = 221.6$, $p < 0.001$; CFI = 0.988; RMSEA = 0.058 (0.048, 0.069); two-year data: $\chi^2(100) = 206.8$, $p < 0.001$; CFI = 0.990; RMSEA = 0.062 (0.050, 0.074)). In configural invariance (see Table 4, model 1), model fit was adequate: $\chi^2(203) = 569.0$, $p < 0.001$; CFI = 0.982; RMSEA = 0.075 (0.068, 0.083). Model fit did not significantly worsen for weak invariance (model 2 and model 3). There was evidence of nonequivalence at the level of strong invariance (model 4: $\Delta\chi^2$ (40) = 105.76, $p < 0.001$). Modification indices suggested removing constraints on thresholds 3 and 4 of item 3 of the physical functioning domain: *Played freely outside the house like other children his/her age?* Removing the constraint established invariance at

Table 1
Child and parent characteristics.

	Baseline (n = 373)[a]	6 months (n = 336)	12 months (n = 304)	24 months (n = 282)
Child characteristics				
Age, years, mean (SD)	7.5 (2.3)	7.9 (2.4)	8.5 (2.3)	9.5 (2.3)
Sex, male	52.4	51.5	50.7	51.6
Severity of epilepsy				
Somewhat/a little/not at all	76.9	88.4	91.2	91.7
Seizure type				
Partial	59.6	59.0	58.4	57.8
Generalized	38.5	39.2	39.8	39.5
Unknown	1.9	1.8	1.8	2.7
Current AED use	67.1	81.0	81.8	76.5
Total AEDs prescribed, mean (SD)	0.8 (0.7)	1.2 (0.9)	1.3 (1.1)	1.4 (1.3)
Cognitive problems	20.0	23.0	25.5	28.4
Behavior problems	15.4	23.6	20.7	22.7
Parent characteristics				
Living with a spouse or partner	87	87	88	88
Household income ≥ $60,000	56.3	57.2	62.1	65.4
Age-primary caregiver, mean (SD)	37.7 (6.1)	38.2 (5.8)	39.1 (5.9)	40.3 (5.6)
Postsecondary school graduate	66.5	70.3	73.7	74.8
Employed	67.1	70.7	73.5	77.0

[a] Reported as percentages, unless otherwise stated.

Table 2
Tests of measurement equivalence by age.

	χ^2 (df)	CFI	RMSEA (90% CI)	$\Delta\chi^2$ (df)	ΔCFI	ΔRMSEA
1. Configural invariance	448.6 (203)	0.977	0.082 (0.072, 0.092)	–	–	–
No constraints						
2. Weak invariance I	451.3 (215)	0.978	0.078 (0.068, 0.088)	16.11 (12)[a]	0.001	−0.004
Item factor loadings						
3. Weak invariance II	442.9 (218)	0.979	0.076 (0.066, 0.086)	0.73 (3)[a]	0.001	−0.002
First-order factors						
4. Strong invariance	418.9 (258)	0.985	0.059 (0.048, 0.069)	–	0.006	−0.017
Item thresholds and First-order intercepts						
5. Strict invariance	429.6 (274)	0.985	0.056 (0.046, 0.066)	–	0.000	−0.003
Item residuals and First-order factor variance						

Children aged 4–7 years were in the 'young child' group and those aged 8–12 years in the 'preadolescent' group. Italicized text describes the model parameters that were constrained at each step of the testing process.
[a] Not statistically significant.

the strong level (model 5). There was evidence of nonequivalence at the level of strict invariance (model 6: $\Delta\chi^2$ (16) = 79.01, $p < 0.001$). Modification indices suggested removing the constraint on the residual of item 1 in the cognitive functioning domain: *Had trouble understanding directions?*, item 3 of the physical functioning domain: *Played freely outside the house like other children his/her age?*, and on item 1 in the social functioning domain: *How limited are your child's social activities compared with others his/her age?* Once the constraints were removed, invariance was achieved at the strict level (model 7).

A table of standardized factor loadings for each item at the level of strict invariance has been provided for each of the comparison groups (see Table 5).

4. Discussion

Our study suggests that the recently developed short-form QOLCE-16 is measurement equivalent. Our results indicate that items of the QOLCE-16 are similarly interpreted by children aged 4–7 years and those 8–12 years, by males and females, and longitudinally between baseline and at 24 months postdiagnosis. These findings are an important step towards confidence in the validity of the QOLCE-16. In all comparisons, measurement equivalence was reached to the level of strong invariance (item thresholds and factor intercepts), and partial equivalence was obtained at the level of strict equivalence (item residuals and factor variances). These findings align with tests of measurement equivalence using the QOLCE-55 [8]. While full equivalence was not reached across all comparison groups for all items, previous research suggests that reaching the level of strong invariance is adequate for making valid comparisons between groups [32,35]. Nonequivalence at the level of configural or weak/metric equivalence has been shown to impact mean total score comparisons between groups [32,35]. Strong invariance ensures that factor scores are equivalent across groups, an important property for establishing the validity of a mean score measure of HRQoL [36]. Having weak and strong invariance allows the understanding of how underlying external factors affect the mean score of the construct [36]. The need for strict invariance is primarily for situations where testing bias is of primary concern (an interest in ensuring equality of results) or comparisons at the item-level [36]. In the context of the QOLCE-16, strong measurement invariance is adequate.

Three items were identified as being nonequivalent between comparison groups (sex and across time) at the levels of strong and strict invariance: cognitive functioning item 1 (*Had trouble understanding directions?*), physical functioning item 3 (*Played freely outside the house like other children his/her age?*), and social functioning item 1 (*How limited are your child's social activities compared with others his/her age?*). In assessing measurement equivalence of the QOLCE-55, both cognitive functioning item 1 and social functioning item 1 were found to have nonequivalence issues on these same items (identified as having nonequivalent intercepts in QOLCE-55). In addition, a question on the physical functioning domain was found to be nonequivalent (*Played with friends away from you or your home?*) [8]. The similarity of findings is not surprising given that the QOLCE-16 contains a subset of items from the QOLCE-55. It was suggested at the time that the term "played" may be difficult to interpret in the HERQULES sample, in that over the two years postdiagnosis nearly one quarter of children transitioned from the young group to the preadolescent group [8]. Alternatively, the phrase "freely outside" may have different interpretations during transition to older ages. A second possible reason may be differences in parental restriction longitudinally, where parents at the time of diagnosis may be overly concerned about children playing away from home, but ease this restriction as the family adjusts to living with epilepsy. This is consistent with research suggesting strong

Table 3
Tests of measurement equivalence by sex.

	χ^2 (df)	CFI	RMSEA (90% CI)	$\Delta\chi^2$ (df)	ΔCFI	ΔRMSEA
1. Configural invariance	415.4 (203)	0.979	0.076 (0.066, 0.087)	–	–	–
No constraints						
2. Weak invariance I	422.3 (215)	0.979	0.073 (0.063, 0.084)	17.83 (12)[a]	0.000	−0.003
Item factor loadings						
3. Weak invariance II	416.5 (218)	0.980	0.071 (0.061, 0.082)	1.01 (3)[a]	0.001	−0.002
First-order factors						
4. Strong invariance I	421.0 (256)	0.984	0.060 (0.049, 0.070)	34.74 (38)[a]	0.004	−0.011
Item thresholds and First-order intercepts						
5. Strict invariance I	447.9 (272)	0.983	0.060 (0.050, 0.070)	43.01 (16)[b]	0.001	0.000
Item residuals and First-order factor variances						
6. Strict invariance I	430.5 (271)	0.984	0.057 (0.047, 0.067)	29.71 (15)[a]	0.000	−0.003
Item residuals (θ_{CF1}) and First-order factor variances						

Italicized text describes the model parameters that were constrained at each step of the testing process. Constraints on the residual of item 1 in the Cognitive Functioning domain (CF1) were removed to establish partial invariance and were unconstrained in subsequent models.
[a] Not statistically significant.
[b] $p < 0.001$.

Table 4
Tests of measurement equivalence by time.

	χ^2 (df)	CFI	RMSEA (90% CI)	$\Delta\chi^2$ (df)	ΔCFI	ΔRMSEA
1. Configural invariance	569.0 (203)	0.982	0.075 (0.068, 0.083)	–	–	–
No constraints						
2. Weak invariance I	562.8 (215)	0.983	0.071 (0.064, 0.079)	16.99 (12)[a]	0.001	−0.004
Item factor loadings						
3. Weak invariance II	552.2 (218)	0.983	0.070 (0.062, 0.077)	2.10 (3)[a]	0.000	−0.001
First-order factors						
4. Strong invariance I	622.2 (258)	0.982	0.067 (0.060, 0.073)	105.76 (40)[b]	0.001	−0.003
Item thresholds and First-order intercepts						
5. Strong invariance II	582.4 (256)	0.984	0.063 (0.057, 0.070)	62.65 (38)[a]	0.002	−0.004
Item thresholds ($\tau_{PF7\$3}$, $\tau_{PF7\$4}$) and First-order intercepts						
6. Strict invariance I	631.5 (272)	0.982	0.065 (0.058, 0.071)	79.01 (16)[b]	0.002	0.002
Item residuals and First-order factor variances						
7. Strict invariance II	566.4 (269)	0.985	0.059 (0.052, 0.066)	30.27 (16)[a]	0.003	−0.006
Item residuals (θ_{CF1}, θ_{PF7},θ_{SF3}) and First-order factor variances						

Italicized text describes the model parameters that were constrained at each step of the testing process. Constraints on the item thresholds and item residual of item 3 in the Physical Functioning domain (PF7), residual of item 1 in the Cognitive Functioning domain (CF1), and residual of item 1 in the Social Functioning domain (SF3) were removed to establish partial invariance and were unconstrained in subsequent models.

[a] Not statistically significant.

[b] $p < 0.001$.

parental protections and limitations regarding social activities reported by children with epilepsy [37,38]. This second reason may also explain residual nonequivalence with the social functioning item (limitations in social activities) as parents' perceptions change their view of what a limited social activity is at the time of diagnosis and after living with epilepsy for two years. Similar changes in parental perceptions/expectations across time may explain nonequivalence for cognitive functioning (understanding directions) [38]. Difficulties in memory are associated with clinical characteristics of epilepsy, and it is reasonable for parents' expectations of children's comprehension skills to change over time as the family adapts to living with epilepsy [39].

Overall, both QOLCE-16 and QOLCE-55 suggest measurement equivalence, and both show strong measurement properties. Despite the similarities of these two shortened versions of the QOLCE, each has unique strengths as measures of HRQoL. As additional items provide a better representation of each domain, the individual domains of HRQoL are better represented by the QOLCE-55. Thus, in cases warranting domain-specific information or more comprehensive assessments of HRQoL, the QOLCE-55 is the preferred measure. In contrast, the QOLCE-16 is more appropriate for a focus on an aggregate summary score. Its short length is well-suited for clinical settings, where tracking overall change is critical. The reduced number of items in the QOLCE-16 entails less administrative burden to facilitate greater clinical uptake and clinical utility. While our analyses have suggested nonequivalence for three items, we do not encourage revising these items because the level of nonequivalence for these items does not bias overall HRQoL comparisons. Furthermore, by not revising these items, we are ensuring consistency across all three versions of the QOLCE. Consistency between versions of QOLCE is paramount to allow long-term comparisons across studies and the ability to merge datasets using different versions of the measure.

There are several limitations in this study. While our sample included a large number of children with epilepsy, our group comparisons are considered small in terms of equivalence testing. Further, our sample may not adequately represent the full range of adolescence, as our upper range of child age is 12 years old. Our sample consists of families of relatively high socioeconomic status; however, variation exists across parental education, marital status, and household income, as well as children's clinical profiles. We do not believe a wider spectrum of

Table 5
Standardized factor loadings of the QOLCE-16 for each comparison group at the level of strict invariance.

	Standardized factor loadings[a]		
	Young/preadolescence	Male/Female	Baseline/24 months
Factor 1: cognitive functioning			
Had trouble understanding directions?	0.85 (0.01)/0.87 (0.02)	0.84 (0.01)/0.92 (0.01)	0.85 (0.01)/0.94 (0.01)
Had difficulty following complex instructions?	0.98 (0.01)/0.98 (0.01)	0.98 (0.01)/0.97 (0.01)	0.98 (0.01)/0.98 (0.01)
Had difficulty following simple instructions?	0.99 (0.01)/0.99 (0.01)	0.97 (0.01)/0.96 (0.01)	0.97 (0.01)/0.97 (0.01)
Had trouble remembering things people told him/her?	0.75 (0.03)/0.78 (0.03)	0.79 (0.03)/0.76 (0.03)	0.77 (0.03)/0.76 (0.03)
Factor 2: emotional functioning			
Felt nobody understood him/her?	0.74 (0.04)/0.76 (0.04)	0.75 (0.04)/0.76 (0.04)	0.74 (0.03)/0.75 (0.03)
Felt down or depressed?	0.65 (0.04)/0.68 (0.04)	0.64 (0.05)/0.66 (0.05)	0.69 (0.03)/0.69 (0.03)
Felt frustrated?	0.70 (0.04)/0.73 (0.03)	0.71 (0.04)/0.72 (0.04)	0.72 (0.03)/0.72 (0.03)
Felt confident?	0.68 (0.04)/0.71 (0.04)	0.70 (0.04)/0.71 (0.04)	0.74 (0.03)/0.75 (0.03)
Factor 3: social functioning			
How limited are your child's social activities compared with others his/her age?	0.73 (0.04)/0.78 (0.03)	0.75 (0.03)/0.77 (0.04)	0.75 (0.03)/0.89 (0.03)
Affected his/her social interactions at school or work?	0.87 (0.03)/0.90 (0.02)	0.88 (0.03)/0.88 (0.02)	0.90 (0.02)/0.89 (0.02)
Isolated him/her from others?	0.91 (0.02)/0.93 (0.02)	0.910 (0.02)/0.92 (0.02)	0.92 (0.02)/0.91 (0.02)
Made it difficult for him/her to keep friends	0.890 (0.03)/0.92 (0.03)	0.89 (0.03)/0.90 (0.03)	0.93 (0.02)/0.92 (0.02)
Factor 4: physical functioning			
Played freely in the house like other children his/her age?	0.67 (0.05)/0.69 (0.05)	0.72 (0.04)/0.65 (0.05)	0.64 (0.03)/0.61 (0.03)
Been able to do the physical activities other children his/her age do?	0.77 (0.05)/0.80 (0.05)	0.84 (0.04)/0.79 (0.05)	0.87 (0.03)/0.85 (0.03)
Played freely outside the house like other children his/her age?	0.14 (0.07)/0.15 (0.07)	0.17 (0.08)/0.14 (0.07)	0.14 (0.05)/0.34 (0.07)
Needs more supervision than other children his/her age?	0.77 (0.04)/0.80 (0.04)	0.78 (0.04)/0.72 (0.04)	0.86 (0.02)/0.83 (0.03)

[a] Values in parentheses are standard errors.

epilepsy would have changed the results of our analyses. While our results are an important step towards the validity of the QOLCE-16, further investigation in other samples and across other comparison groups is warranted. Comparisons across clinical characteristics and other sociodemographic indicators would advance our understanding of results obtained when using the QOLCE-16.

5. Conclusion

These findings provide initial evidence of measurement equivalence for the QOLCE-16 across age, sex, and time. Comparisons made using the QOLCE-16 are believed to be unbiased and differences between subjects meaningful. This lends additional support that the QOLCE-16 has acceptable measurement properties for use in assessing HRQoL in children with epilepsy. Pediatric HRQoL research in epilepsy may benefit from use of the combination of a more detailed disease specific instrument such as the QOLCE-55 in combination with a generic instrument, while clinical practice may benefit from use of a simple and short instrument such as the QOLCE-16 to allow translation of findings into improvements in clinical practice.

Acknowledgments

We gratefully acknowledge the parents and physicians and their staff, without whose participation, this study would not have been possible. The Canadian Paediatric Epilepsy Network effectively facilitated the participation of physician contributors across the country. Our thanks to HERQULES staff, Jane Terhaerdt and Wenyi Huang. This study was supported by a Canadian Institutes of Health Research operating grant (MOP-64311) to Kathy N. Speechley. Mark A. Ferro holds the Canada Research Chair in Youth Mental Health.

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] Thurman DJ, Beghi E, Begley CE, Berg AT, Buchhalter JR, Ding D, et al. Standards for epidemiologic studies and surveillance of epilepsy. Epilepsia 2011;52:2–26.
[2] Spieth LE, Harris CV. Assessment of health-related quality of life in children and adolescents: an integrative review. J Pediatr Psychol 1996;21:175–93.
[3] Speechley KN, Ferro MA, Camfield CS, Huang W, Levin SD, Smith ML, et al. Quality of life in children with new-onset epilepsy: a 2-year prospective cohort study. Neurology 2012;79:1548–55.
[4] Conway L, Smith ML, Ferro MA, Speechley KN, Connoly MB, Snead OC, et al. Correlates of health-related quality of life in children with drug resistant epilepsy. Epilepsia 2016;57:1256–64.
[5] Rodenburg R, Stams GJ, Meijer AM, Aldenkamp AP, Dekovic M. Psychopathology in children with epilepsy: a meta-analysis. J Pediatr Psychol 2005;30:453–68.
[6] Sabaz M, Cairns DR, Lawson JA, Nheu N, Bleasel AF, Bye AM. Validation of a new quality of life measure for children with epilepsy. Epilepsia 2000;41:765–74.
[7] Goodwin SW, Lambrinos AI, Ferro MA, Sabaz M, Speechley KN. Development and assessment of a shortened Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-55). Epilepsia 2015;56:864–72.
[8] Ferro MA, Goodwin SW, Sabaz M, Speechley KN. Measurement equivalence of the newly developed Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-55). Epilepsia 2016;57:427–35.
[9] Conway L, Widjaja E, Smith ML, Speechley KN, Ferro MA. Validating the shortened Quality of Life in Childhood Epilepsy Questionnaire in a sample of children with drug-resistant epilepsy. Epilepsia 2017;58:646–56.
[10] Goodwin SW, Ferro MA, Speechley KN. Development and assessment of the Quality of Life in Childhood Epilepsy Questionnaire (QOLCE-16). Epilepsia 2018;59:668–78.
[11] Varni JW, Burnwinkle TM, Lane MM. Health-related quality of life measurement in paediatric clinical practice: an appraisal and precept for future research and application. Health Qual Life Outcomes 2005;3:34.
[12] King S, Exley J, Parks S, Ball S, Bienkowska-Gibbs T, MacLure C, et al. The use and impact of quality of life assessment tools in clinical care settings for cancer patients, with a particular emphasis on brain cancer: insights from a systematic review and stakeholder consultations. Qual Life Res 2016;25:2245–56.
[13] Jacobsen PB, Davis K, Cella D. Assessing quality of life in research and clinical practice. Oncology 2002;16:133–9.
[14] Bezjak A, Ng P, Skeel R, Depetrillo AD, Comis R, Taylor KM. Oncologists' use of quality of life information: results of a survey of Eastern Cooperative Oncology Group Physicians. Qual Life Res 2001;10:1–13.
[15] van de Schoot R, Lugtig P, Hox J. Developmetrics. A checklist for testing measurement invariance. Eur J Dev Psychol 2012;9:486–92.
[16] Muthen BO. Latent variable modeling in heterogeneous populations. Psychometrika 1989;54:557–85.
[17] Speechley KN, Sang X, Levin S, Zou GY, Eliasziw M, Smith ML, et al. Assessing severity of epilepsy in children: preliminary evidence of validity and reliability of a single-item scale. Epilepsy Behav 2008;13:337–42.
[18] Chan CJ, Zou GY, Wiebe S, Speechley KN. Global Assessment of the Severity of Epilepsy (GASE) scale in children: validity, reliability, responsiveness. Epilepsia 2015;13:337–42.
[19] Muthén LK, Muthén BO. Mplus User's Guide. Eighth EditionLos Angeles, CA: Muthén & Muthén; 1998-2017.
[20] Gibbons RD, Bock RD, Hedeker D. Full-information item bifactor analysis of graded response data. Appl Psychol Measur 2007;31:4–19.
[21] Chen FF, West S, Sousa K. A comparison of bifactor and second-order models of quality of life. Multivar Behav Res 2006;41:189–225.
[22] Yung YF, Thissen D, McLeod LD. On the relationship between higher-order factor model and the hierarchical factor model. Psychometrika 1999;64:113–28.
[23] Muthen BO, Asparouhov T. Latent variable analysis with categorical outcomes: multiple-group and growth modelling in Mplus. Available from: http://www.statmodel.com/examples/webnote.shtml; 2002.
[24] Tabachnick BG, Fidell LS. Using multivariate statistics. Boston, MA: Pearson; 2014.
[25] Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 1992;21:230–58.
[26] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J 1999;6:1–55.
[27] Millsap RE. Statistical approaches to measurement invariance. New York, NY: Routledge; 2012.
[28] Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organization research. Organ Res Methods 2000;3:4–70.
[29] Ferro MA, Boyle MH, Scott JG, Dingle K. 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.
[30] Kimber M, Rehm J, Ferro MA. Measurement invariance of the WHODAS 2.0 in a population-based sample of youth. PLoS One 2015;10:e0142385.
[31] Chen FF, Sousa KH, West SG. Testing measurement invariance of second-order factor models. Struct Equ Model 2005;12:471–92.
[32] 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.
[33] Yuan KH, Wu R, Bentler PM. Ridge structural equation modeling with correlation matrices for ordinal and continuous data. Br J Math Stat Psychol 2011;64:107–33.
[34] Wothke W. Nonpositive definite matrices in structural modeling. In: Bollen KA, Long JS, editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 256–93.
[35] Steinmetz H. Analyzing observed composite differences across groups: is partial measurement invariance enough? Methodology 2013;9:1–12.
[36] Meredith W, Teresi JA. An essay on measurement and factorial invariance. Med Care 2006;44:69–77.
[37] Chong L, Jamieson NJ, Gill D, Singh-Grewal D, Craig JC, Ju A, et al. Children's experiences of epilepsy: a systematic review of qualitative studies. Pediatrics 2016;138:e20160658.
[38] Laxer KD, Trinka E, Hirsch LJ, Cendes F, Langfitt J, Delanty N, et al. The consequences of refractory epilepsy and its treatment. Epilepsy Behav 2014;37:59–70.
[39] Menlove L, Reilly C. Memory in children with epilepsy: a systematic review. Seizure 2015;25:126–35.