University of Belgrade

Faculty of Medicine

Dejan S. Stevanovic

Applying the principles of good research practice for cross-cultural adaptation to pediatric health-related quality of life questionnaires

Doctorial dissertation

Belgrade, 2014

Univerzitet u Beogradu

Medicinski fakultet

Dejan S. Stevanović

Primena principa dobre istraživačke prakse za kulturološku adaptaciju na pedijatrijske upitnike o kvalitetu života

Doktorska disertacija

Beograd, 2014

Mentor:

Aneta Lakić, MD, PhD

University of Belgrade,

Faculty of Medicine

_____

Examining committee:

Tatjana Pekmezović, MD, PhD

University of Belgrade,

Faculty of Medicine

President

_____

Jasna Jančić, MD, PhD

University of Belgrade,

Faculty of Medicine

_____

Nataša Bogavac Stanojević, PhD

University of Belgrade,

Faculty of Farmacy

_____

Doctorial dissertation defended on

Acknowledgements

Many people, from my family, close collaborators and friends, journal editors and unknown reviewers, helped me with this research. I thank to all of them. Especially, I would like to thank to my mentor prof. Aneta Lakic, she who could understand all my ideas, to prof. Jasna Jancic, she who is open-minded and supportive to all my progress, and to my wife Jelena, she who has spent many days waiting for me to finish my studies with support and love.

Conflict of interest

I declare no conflict of interest.

Applying the principles of good research practice for cross-cultural adaptation to pediatric health-related quality of life questionnaires

Summary

Although still debated about its definition and conceptualization, the health-related quality of life (HRQOL) concept is simply defined as the patient's evaluation of the impact of a health condition and its treatment on daily life. HRQOL is a multidimensional construct that covers physical, emotional, mental, social, and behavioral components of well-being and functioning. HRQOL is possible to conceptualize through objective indicators. A great deal of attention has been paid to ensure reliable and valid HRQOL measurements through the development of questionnaires with sound psychometric properties.

Continually from the past decade, HRQOL is more frequently used in prevention, treatment, and rehabilitation both, nationally and internationally. This requires that HRQOL measures are available across different nations/languages. With this in mind, questionnaires with good psychometric properties need to be simultaneously developed across different nations and cultures or, a slightly different but equally successful approach, the translation, subsequent analysis, and adaptation of existing and accepted measures into other languages is performed while considering aspects of the cultural settings. So far, cross-cultural adaptation of questionnaires has been recognized as one of the priority in HRQOL research.

It is early recognized that HRQOL assessments have some specific characteristics when children and adolescents are considered. The development of pediatric HRQOL measurement followed specific pathways with several important issues: specific HRQOL domains, age and developmental characteristics, self- and proxy-rating, generic and disease specific approaches to assessments, and psychometric considerations. Up to date, no clear guidelines for the cross-cultural adaptation of pediatric HRQOL questionnaires were developed. The aim of this thesis was to operationalize a model of the cross-cultural adaptation of pediatric

HRQOL questionnaires. Taking a systematic approach, the cross-cultural adaptation process was described as should be followed for pediatric HRQOL questionnaires – including the procedures of translation and cultural adaptation, pretesting, psychometric evaluation, and assessing equivalence with bias elimination.

The operational model of the cross-cultural adaptation process for pediatric HRQOL questionnaires was developed following the previously established guidelines for translation and cultural adaptation, psychometric evaluation, and equivalence testing with bias elimination. The operational model has five consecutive phases: pre-translation, translation, pre-testing, psychometric, and finalization phase. Within the cross-cultural adaptation model, five aspects of investigating equivalence are recognized to claim levels of equivalence achievement: conceptual, semantic, items, operational, and psychometric. The operational model was presented considering four generic and two epilepsy specific questionnaires: KINDL questionnaire, KIDSCREEN questionnaire, Pediatric Quality of Life Inventory Version 4.0 Generic Core Scales – PedsQL, Quality of Life Enjoyment and Satisfaction Questionnaire Short form – Q-LES-Q – SF, Health-Related Quality of Life Measure for Children with Epilepsy - CEQOL-25, and Quality of Life in Epilepsy Inventory for Adolescents - QOLIE-AD-48.

The cross-cultural adaptation process of a HRQOL questionnaire to a new language/culture is a complex process involving several consecutive steps in order to ensure that the HRQOL concept represented by the questionnaire is appropriately transferred to that new language/culture. The essence of the process actually involves weighting between the altering of the source questionnaire's items literally (i.e. translation) and removing, changing, adding, supplementing and/or modifying those items that deal with behavior that does not generalize equivalently in the target culture (i.e. cultural adaptation). Qualitative evaluations (i.e. pre-testing) and quantitative evaluations of the target questionnaires (i.e. psychometric evaluations) are added to confirm that the measuring concept represented by the questionnaire is appropriately transferred to the target culture/language.

The operational model adopted in this thesis for the cross-cultural adaptation of pediatric HRQOL questionnaires offers new opportunities and challenges for pediatric HRQOL research. However, it would be important to evaluate in follow-up studies the importance and relevance of each step proposed in the phases using different approaches. Additionally, it would be important to develop specific, consensus-based checklists for the cross-cultural adaptation of HRQOL questionnaires and other patient-reported outcomes measures.

Key words –quality of life; translation; cross-cultural; psychometrics; children; adolescents.

Scientific field – Medicine; subfield - Child Psychiatry

Primena principa dobre istraživačke prakse za kulturološku adaptaciju na pedijatrijske upitnike o kvalitetu života

Sažetak

Iako se još uvek debatuje o definiciji i konceptu, kvalitet života povezan sa zdravljem (eng. health-related quality of life - HRQOL) bi jednostavno označavao kako pacijent procenjuje uticaj svog zdravstvenog stanja i primenjenih tretmana na svakodnevni život. Ovo je multidimenzionalni konstrukt, koji uključuje fizičke, emocionalne, mentalne, i socijalne komponente blagostanja i funkcionisanja pacijenta. Kao subjektivini pojam, HRQOL je moguće shvatati kroz objektivne indikatore i kroz kvalitativne i kvantitativne procene. Tokom poslednje dve decenije, posebna pažnja je posvećena obezbeđivanju pouzdanih i validinih procena HRQOL, kroz razvoj upitnika, koji imaju dobre psihometrijske osobine.

HRQOL se učestalo istražuje u prevenciji, lečenju i rehabilitaciji različitih poremećaja kako nacionalno tako i internacionalno, što zahteva da su upitnici dostupni za više jezika/nacija. Imajući ovo u vidu, HRQOL upitnici sa dobrim psihometrijskim karakteristikama moraju biti razvijeni simultano kroz više kultura/nacija ili da se prevode i adaptiraju već postojeći upitnici poštujući kulturološke odlike jezika na koji se prevode. Na ovaj način, unakrsna kulturološka adaptacija upitnika je prepoznata kao jedan od prioriteta u HRQOL istraživanjima.

Procena HRQOL kod dece i adolescenata ima određene specifičnosti, na primer HRQOL domeni specifični samo za ovu populaciju, godine i razvojne karakteristike, samo-procna i proksi-procena, generički i specifični upitnici, kao i psihometrijske odlike. Međutim, ne postoje jasne smernice za prevod i kulturološku adaptaciju pedijatrijskih upitnika za HRQOL. Cilj ove teze je da se koncipira model za kulturološku adaptaciju pedijatrisjkih upitnika. Pristupajući na sistematičan način procesu kulturološke adaptacije HRQOL upitnika predloženo je nekoliko procedura: prevod, kulturološka adaptacija, pre-testiranje, psihometrijska evaluacija, uz procenu nivoa slaganja originalnog i prevedenog upitnika (ekvivalentnost upitnika). Operativni model za kulturološku adaptaciju pedijatrijskih upitnika za HRQOL je razvijen usvajanjem postojećih vodiča i

preporuka za prevod i kulturološku adaptaciju, psihometrijsku evaluaciju i procenu ekvivalentnosti raznih upitnika. Model ima pet konsekutivnih faza: procesi pre prevoda, sam prevod, pre-testiranje, psihometrijska procena i finalizacija prevoda. Pet aspekata procene ekvivalentnosti upitnika su izdvajena: konceptualna, sematička, pitanja, operacionalna i psihometrijska. Operativni model je predstavljen na četiri generička i dva upitnika specifična za epilepsiju: KINDL, KIDSCREEN, Pediatric Quality of Life Inventory Version 4.0 Scales – PedsQL, Quality of Life Enjoyment and Satisfaction Questionnaire Short form – Q-LES-Q – SF, Health-Related Quality of Life Measure for Children with Epilepsy - CEQOL-25, i Quality of Life in Epilepsy Inventory for Adolescents - QOLIE-AD-48.

Proces kulturološke adaptacije HRQOL upitnika za decu i mlade na drugi lezik je složen process, koji uključuje niz konsekutivnih koraka kojim se obezbeđuje da HRQOL koncept koji je u osnovi originalnog upitnika bude adekvatno "prenet" na novu verziju, to jest drugi jezik. Suštinu procesa čini odmeravanje između "literarnog preuređivanja" i menjanja, adaptiranja, dodavanja i/ili modifikovanja pitanja, naročito onih koja se odnose na ponašanja koja nisu prisutna u kulturi na koju se upitnik prevodi Kvalitativna (pre-testiranje) i kvantitativna (psihometrijsko testiranje) procena prevoda upitnika su sledeći koraci, kojim se dodatno obezbeđuje da je HRQOL koncept koji se meri određenim upitnikom zaista prenet na novi jezik. Operativni model kulturološke adaptacije upitnika koji je usvojen u ovoj tezi pruža nove mogućnosti i izazove za istraživanje HRQOL u pedijatrijskim populacijama. Međutim, na prvom mestu, važno je da se proceni važnost predloženih koraka u fazama kulturološke adaptacije kroz studije praćenja, koje bi procenjivale navedene korake na različite načine. Takođe, važno bi bilo razviti smernice koje bi bile konsenzualno donete o kulturološkoj adaptaciji upitnika o kavalitetu života, ali i drugih koji se odnose na procenu ishoda lečenja.

Ključne reči – kvalitet života; upitnik; prevod; adaptacija; psihometrija; deca; adolescenti.

Naučna oblast – Medicina

Uža naučna oblast – Dečija Psihijatrija

Applying the principles of good research practice for cross-cultural adaptation to pediatric health-related quality of life questionnaires

Table of Contents

APPLYING THE PRINCIPLES OF GOOD RESEARCH PRACTICE FOR CROSS-CULTURAL ADAPTATION TO PEDIATRIC HEALTH-RELATED QUALITY OF LIFE QUESTIONNAIRES 1

LIST OF ARTICLES UPON WHICH THIS PHD THESIS IS BASED 3

INTRODUCTION 4

CROSS-CULTURAL ADAPTATION OF HRQOL QUESTIONNAIRES 6
RATIONAL FOR THE CROSS-CULTURAL ADAPTATIONS OF HRQOL QUESTIONNAIRES 7
MAKING HRQOL QUESTIONNAIRES AVAILABLE FOR DIFFERENT CULTURES 9
APPROACHING THE CROSS-CULTURAL ADAPTATION OF HRQOL QUESTIONNAIRES 11
PRINCIPALS OF THE CROSS-CULTURAL ADAPTATION PROCESS 12
Equivalence 14
Bias 18
Essence of the translation and cultural adaptation process 20
Guidelines for the cross-cultural adaptation of PRO/HRQOL questionnaires 22
Qualitative Approaches to translation and cultural adaptation 26
Psychometric validation 27
Finalization of the cross-cultural adaptation process 33
USING HRQOL QUESTIONNAIRES WITH CHILDREN AND ADOLESCENTS 34

RATIONALE AND AIMS 36

Aims 39

METHODS 41

OPERATIONAL MODEL OF THE CROSS-CULTURAL ADAPTATION PROCESS 41
PRE-TRANSLATION PHASE 44
Investigating conceptual equivalence 44
Preparation 46
TRANSLATION PHASE 47
Forward translation 48
Reconciliation 49
Backward translation 49
Pre-final version development 50
PRE-TESTING PHASE 52
Cognitive debriefing 53
Questionnaire completion 54
PSYCHOMETRIC PHASE 55
Sample 55

Statistical procedures *56*
FINALIZATION PHASE 58
ARTICLE 1 & 2 *KINDL QUESTIONNAIRE* 59
ARTICLE 3 *KIDSCREEN QUESTIONNAIRE* 64
ARTICLE 4 *PEDSQL QUESTIONNAIRE* 68
ARTICLE 5 *CHEQOL-25 QUESTIONNAIRE* 70
ARTICLE 6 *QOLIE-AD 48 QUESTIONNAIRE* 72
ARTICLE 7 *Q-LES-Q-SF QUESTIONNAIRE* 76
ARTICLE 8 80

SUMMARY OF RESULTS 82

ARTICLE 1 & 2 *KINDL QUESTIONNAIRE* 82
ARTICLE 3 *KIDSCREEN QUESTIONNAIRE* 90
ARTICLE 4 *PEDSQL QUESTIONNAIRE* 94
ARTICLE 5 *CHEQOL-25 QUESTIONNAIRE* 95
ARTICLE 6 *QOLIE-AD 48 QUESTIONNAIRE* 98
ARTICLE 7 *Q-LES-Q- SF QUESTIONNAIRE* 103
ARTICLE 8 105

DISCUSSION 108

DISCUSSION OF RESULTS 111
Article 1 & 2 KINDL questionnaire *111*
Article 3 KIDSCRENN questionnaire *116*
Article 4 PedsQL questionnaire *119*
Article 5 CHEQOL-25 questionnaire *121*
Article 6 QOLIE-AD-48 questionnaire *123*
Article 7 Q-LEQ-Q-SF questionnaire *125*
Article 8 ..... *128*

CONCLUSIONS AND RECOMMENDATIONS 131

REFERENCES 135

APPENDIX LIST 157

BIOGRAPHY 160

List of articles upon which this PhD thesis is based

1. Stevanovic D, Lozanovic-Miladinovic D, Jovic N, Sarenac M. The Serbian QOLIE-AD-48: translation, cultural adaptation, and preliminary psychometric evaluation. Epilepsy Behav. 2005; 7:240-245,
2. Stevanovic D. Health-related quality of life in adolescents with well-controlled epilepsy. Epilepsy Behav. 2007; 10: 571-575.
3. Stevanovic D, Lakic A, Vilotic J. The psychometric study of the Serbian KINDL questionnaire for health-related quality of life assessment in children and adolescents. Scand J Caring Sci. 2009; 23; 361–368,
4. Stevanovic D. Serbian KINDL questionnaire for quality of life assessments in healthy children and adolescents: reproducibility and construct validity. Health Qual Life Outcomes. 2009; 7:79,
5. Stevanovic D, Tepavcevic DK, Jocic-Jakubi B, Jovanovic M, Pekmezovic T, Lakic A, Ronen GM. Health-Related Quality of Life Measure for Children with Epilepsy (CHEQOL-25): preliminary data for the Serbian version. Epilepsy Behav. 2009, 16: 599-602.
6. Stevanovic D, Lakic A, Damnjanovic M. Some psychometric properties of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales (PedsQLTM) in the general Serbian population. Qual Life Res. 2011; 20:945-949,
7. Stevanovic D. Quality of Life Enjoyment and Satisfaction Questionnaire-short form for quality of life assessments in clinical practice: a psychometric study. J Psychiatr Ment Health Nurs 2011, 18: 744-750, and
8. Stevanovic D, Tadic I, Novakovic T, Kisic-Tepavcevic D, Ravens-Sieberer U. Evaluating the Serbian version of the KIDSCREEN quality of life questionnaires: reliability, validity and agreement between children's and parent's ratings. Qual Life Res. 2013; 22: 1729-1737.

Introduction

During the 1990s, it has become evident that measures of a population's well-being, such as mortality and morbidity rates, provide an incomplete picture of public health care needs and are rarely sufficient in prevention and intervention research. This has highlighted the need to develop measures of well-being that are more "patient-centered", specifically developed on the patient's preferences for his/her treatment. With the patients' preferences in the centre of contemporary medicine, the concept of *patient-reported outcome (PRO)* was developed to represent the patient's report of a health condition and its treatment regimen (1). This general term includes different sources of information coming directly from patients about their health status; each providing a unique and valuable perspective like *well-being*, *functional status*, *health-related quality of life (HRQOL),* and many others. These areas of PRO are considered as the primary indicators of the impact of disease, essential parts for evaluating treatment efficacy and interpreting clinical outcomes, and key elements in decision-making (1). Coming directly from patients about their health, PROs are equally valuable to other observational reports such as physiological or clinical data, and together with these outcomes provide a more comprehensive picture of one's health status. Nowadays, the HRQOL concept has become the most significant PRO.

Although there are ongoing debates about its definition and conceptualization, the HRQOL concept is simply considered as the patient's evaluation of the impact of a health condition and its treatment on daily life (1). The central aspect of this definition is *the impact of a health condition*, what separates HRQOL from its "parent" term, *quality of life (QOL),* which implies on an evaluation of the impact of all non-health-related aspects of life on general well-being (2). HRQOL is a multidimensional, changing construct that covers physical, emotional, mental, social, and behavioral components of well-being and functioning as perceived by patients (3). As well as being a subjective and qualitative construct, HRQOL can also be conceptualized through objective and

quantitative measures. (4). A great deal of attention has been paid to ensure reliable and valid HRQOL measurements through the development of HRQOL questionnaires with sound measurement properties.

It has been early recognized that when pharmaceutical companies evaluate new drugs in clinical trials across more countries, they need assurance that patients' self-assessment of their condition and the effects of treatment, including HRQOL data, are reliable and valid across different countries, irrespective of the patients' language and cultural background (2). This requirement has been extended to other research fields that might consider HRQOL data across countries, such as health promotion, public health reporting, or decision-making, which further highlights the need for reliable and valid measures across cultural settings (5-8). The European Agency for the Evaluation of Medicinal Products (EMEA) and the United States Food and Drug Administration (FDA) provided guidelines for the use of patients' self-assessed outcomes in the evaluation of medicinal products in order to maximize validity of clinical trials with PRO/HRQOL as outcomes (2, 6). The central part in these guidelines is the measurement of PRO/HRQOL, and specifically the development and use of specific questionnaires. The importance of these guidelines is the consideration of the cross-cultural settings in PRO/HRQOL measurements.

Over the past decade, HRQOL is gaining significant attention in prevention, treatment, and rehabilitation, both nationally and internationally. With this in mind, HRQOL measures with good psychometric properties need to be simultaneously developed across different nations and cultures. Alternatively, the translation, subsequent analysis, and adaptation of existing and accepted measures into other languages are performed while considering aspects of the cultural setting and cultural diversity. So far, cross-cultural adaptation of questionnaires has been recognized as one of the priorities in HRQOL research.

Cross-cultural adaptation of HRQOL questionnaires

Several terms need clarifications before introducing the cross-cultural adaptation process for HRQOL questionnaires.

Among the first terms are *language, culture, nation*, and *country*. Without going into details about their meaning and relatedness, the following simplistic approach was adopted in this thesis. A specific *language* is an acquired method of communication and as such, it is a component of one *culture* defined by the specific society employing it (7). Actually, one culture has the feature of shared experiences and meanings that result in values, beliefs and practices that are distinctively different from those found in other cultures (162). One or more closely related cultures constitute one *nation* (mainly speaking the same language with different dialects), interchangeably used with *country*, although one country is a politically governed entity. The term *international* is used to refer to various phenomena concerning more than one nation or culture with a possible extension to cultural groups within one nation (7).

Close to the terms above, we can think of *cultural*, *international* and *cross-cultural* as specifically related terms to HRQOL research (7, 8). The term *cultural* would refer to various activities in HRQOL research in a specific culture, such as studies to investigate the aspects of HRQOL in a specific condition in one culture, using one language. The term *international* would refer on various activities of different nations/countries in the HRQOL field, such as studies from different countries (using different languages) aiming to investigate the risk factors of HRQOL in a specific condition. Finally, the term *cross-cultural* would refer to an additional collaborative and comparative effort to investigate cultural differences in the aspects of or risk factors of HRQOL in a specific condition, but mainly concerning the same assessment method in that condition across nations/cultures, understandably with the same language or different languages (8).

The final group of terms relates specifically to HRQOL questionnaires used internationally and cross-culturally. There are no clear definitions proposed for specific terms when reporting levels of validation of HRQOL questionnaires for another language and/or culture. Simply stated, the term *translation* would refer to the process of translating a questionnaire from one language into another with the aim of ensuring that the content of the questionnaires remains unchanged despite changes at the level of expression (9). This implies that the language versions obtained through the translations are conceptually equivalent to the original questionnaire, what is often called *linguistic validation* (9, 10). *Cultural adaptation* relates to ensuring that a well-translated HRQOL questionnaire maintains its content validity at a conceptual level across different cultures (9). Both qualitative (*pre-test*) and quantitative evaluations (*psychometric validation*) of the target language versions demonstrate whether the relevance and measurement properties of the original questionnaire are preserved during the translation and cultural adaptation (9-11). Some authors use linguistic validation and cultural adaptation interchangeably, while others consider them separately as parts of *cross-cultural adaptation* (10). Furthermore, some authors consider that pre-test is integrated into the cross-cultural adaptation followed by psychometric validation, while others include and psychometric validation as a part of cross-cultural adaptation (10-14). To avoid ambiguities, this thesis considers the cross-cultural adaptation process to include translation or linguistic validation, cultural adaptation, pre-test, and psychometric validation, if not stated otherwise. Additionally, the originally developed questionnaire is the *source language questionnaire* and the translated and culturally adapted version is the *target language questionnaire*.

Rational for the cross-cultural adaptations of HRQOL questionnaires

As anthropological and health psychology research suggest, health is "culture-bounded" implying that culture prescribes views on what constitutes the current health of an individual (7, 8, 15, 169). Values, traditions, and beliefs within

communities of one culture interact with environmental conditions and availability of opportunities to influence the health status of individuals at both the group and individual levels (16-18). This observation is particularly relevant when a health problem occurs, because it has been generally shown that rates and patterns of disorders are broadly similar across various culture groups, but still there are differences when the culture itself is taken into account (163). Nikapota and Rutter emphasized that when considering mental health there are variations in rates of disorders across socio-cultural/ethnic groups, with culture-specific mental disorders suspected, different manifestations of disorders, and levels of similarities or differences in risk factors across groups (163). In an extreme view, the meanings of disease for particular culture are most clearly noticeable in so-called "folk-illnesses" – illnesses for which the culture provides an etiology, diagnosis, preventative treatment, and curing process (17).

Inherently, health-related behaviors, functioning, wellbeing, and HRQOL in the final instance cannot be "culture-free" and they need to be evaluated in the context of the culture where the assessments are organized (12, 15, 40). This is particularly relevant for HRQOL, because the perceptions of the impacts of a health condition and its treatment on daily life are strongly influenced by values, traditions, and beliefs about health in one culture, hence the term "cultural-bounded" HRQOL assessments. The World Health Organization (WHO) is the first to recognize the importance of culture stating that HRQOL is "an individual's perceptions of their position in life in the context of the culture and value systems in which they live, and in relation to their goals, expectations, standards and concerns" (19). Therefore, the measurements of HRQOL are sensitive to the language and dialect, customs, beliefs and traditions of one culture (18). In this light, HRQOL questionnaires developed in one language are inherently sensitive to measuring HRQOL only in one or a few, very similar cultures using that language. Administering a HRQOL questionnaire in a new language, culture, and/or country requires unavoidable cross-cultural adaptaion to reach equivalence between the source questionnaire and the target version (14). Reaching equivalence generally

implies obtaining strong levels of comparability in HRQOL measurements across different languages/cultures.

Making HRQOL questionnaires available for different cultures

During the 1990s, great effort has been put on the development of HRQOL questionnaires sensitive to a specific culture. Since then, research on the translation and cultural adaption of HRQOL questionnaires is proliferating rapidly. Considering that HRQOL assessments are "culturally-bounded", several key issues are recognized and need to be resolved prior to considering one HRQOL questionnaire for use in another language/culture (7, 20).

The first issue concerns the relevance of the HRQOL concept. In other words, to what extent the HRQOL concept researched and developed in one culture is consistent to its conceptualization in another (25). During the past 20 years, studies from the Anglo-Saxon cultures have provided various data about HRQOL. However, studies and data about HRQOL from other cultures are scarce; especially data about the adequacy of HRQOL measurements in groups differing in educational levels, ethnicity, or language are insufficient (15).

The second issue concerns the extent to which cultural groups share an identical set of HRQOL domains, what is closely related to the first issue (15). This issue is central to achieving conceptual equivalence between the source and the target language questionnaire, because domains included in one questionnaire are representative of the HRQOL construct.

The third issue concerns the extent to which HRQOL could be assessed with a given questionnaire and the appropriateness of using the same questionnaire to measure HRQOL across cultures (15). This issue concerns whether a given questionnaire is reliable and valid to measure HRQOL considering the relevance of and acceptability of the concept and its possible domains across cultures (15).

Finally, the fourth issue concerns the extent to which HRQOL could be compared across cultures and the extent to which cross-cultural HRQOL results provide a sound basis for decision making in the health care system (20). There are great risks of interpreting results of medical effectiveness research without assurances of equivalence of concepts and measurement properties in light of potential applications of such findings (15, 20).

Taking into account these issues, research on adapting HRQOL questionnaires for use in another language has generally followed two approaches. The first and most widely represented approach is the cross-cultural adaptation process. Questionnaires for HRQOL assessments are developed in one language (source language), mostly in one culture, and they are altered through rigorous translation and cultural adaptation for use in another language (target language) (13). Guillemin and his coworkers were among the first to apply this approach for HRQOL research systematically (13, 14). They provided five scenarios that highlight the importance of the cross-cultural adaptation process when comparing the target and source language/culture. In the first scenario, the source language questionnaire is to be used in the same language/culture where it was developed and no adaptations are required. The second and third scenarios involve a situation when a questionnaire is to be used in another culture within one or different country using the same language, what requires only the cultural adaptation to the questionnaire. The fourth scenario concerns the application of a questionnaire in a different language/culture and within one country, which requires both translation and cultural adaption. The last scenario is the application of a questionnaire in a different culture, language and country than the one it was developed, which again requires translation and cultural adaption. However, Guillemin et al. failed to include psychometric validation into these scenarios, although they recognized its importance (13). Psychometric validation, as an important step, was formally added by the International Quality of Life Assessment Group (IQOLA) (21-23).

The second approach is concerned with developing cross-nationally usable questionnaires for HRQOL that are not sensitive to one particular culture (7, 20). Within this approach, three goals could be distinguished. The first one would be to develop a HRQOL questionnaire universally applicable across all cultures, but using different languages. The more modest second goal would include the development of a core questionnaire, which might be universally applicable, but which contains specific add-on national modules. The last option pertains to the development of a series of national questionnaires, which are specific to each culture. The last approach is the most relevant for HRQOL, but the current cross-cultural research focuses on the first one. To date, two sets of pediatric questionnaires have been developed in this way – the Quality of life questionnaires for children and adolescents - KIDSCREEN Questionnaires (24) and DISABKIDS condition-specific HRQOL modules for asthma, juvenile idiopathic arthritis, atopic dermatitis, cerebral palsy, cystic fibrosis, diabetes and epilepsy (188).

Regardless of the approach, the overarching basic principal is that HRQOL questionnaires need to be equivalent and available across cultures/languages in order for them to be used in cross-cultural research (7, 12, 14, 15).

Approaching the cross-cultural adaptation of HRQOL questionnaires

When considering HRQOL questionnaires for cross-cultural adaptation, two approaches could be followed – absolutist and universalist (25).

The "absolutist" approach makes an initial assumption that there are a nil or negligible differences in the HRQOL concept across cultures, and careful attention to linguistic elements will make a questionnaire developed for use in one culture acceptable for use in another culture (25).

The "universalist" approach does not make the prior assumption that the HRQOL concept is the same across cultures, but it implies on a need to

demonstrate whether the concept exists and it is interpreted similarly in the two cultures and, if so, the degree to which it is interpreted similarly (15, 25). The "universalist" approach aims to elicit those aspects of the HRQOL concept genuinely universal across cultures and to use them in developing questionnaires. Considering the evidence that the definitions and domains of HRQOL are different across cultures (15), the assumptions of the absolutist approach are not well supported. As such, the universalist approach is regarded as more appropriate for cross-cultural research, and it should be adopted in the process of validating HRQOL questionnaires for use in another language (25).

Related to the "universalist" approach is the identity-equivalence method introduced in 1970s in cross-cultural psychology (25-29). A questionnaire with the identity-equivalence method would include items that are likely to transcend the cultures of interest (i.e. *etics*) as well as items that are specific to one or some of the cultures under the interest (i.e., *emics*) (29). Subsequent quantitative analyses would be used to verify empirically which measures were representing the same construct cross-culturally. However, items not identified as "etic" may nonetheless be valid "emic" indicators of the construct of interest if they correlate with the "etic" items within a given cultural group (25). Following this, a HRQOL questionnaire as an "etic" construct might be developed with a common set of "etic" and group-specific sets of "emic" items.

Principals of the cross-cultural adaptation process

As mentioned earlier, great effort has been put to develop methods that ensure the basic principal that questionnaires available for more languages/cultures are equivalent in their measurements. However, the main concern in cross-cultural research is not only to maximize the validity of inferences made through the methodological rigor in terms of establishing cross-cultural equivalence, but also suppressing bias (30). Generally, *equivalence* refers to the level of comparability of measurement outcomes among languages/cultures, while *bias* refers to differences

in a measurement questionnaire that do not have exactly the same meaning within and across languages/cultures (30). Bias and equivalence are two sides of the same coin and they are inseparable, because cross-cultural equivalence requires the absence of biases and the presence of cross-cultural bias always results in some form of non-equivalence.

The general principal of equivalence is ensured throughout the following validation steps (12). The linguistic validation process is required to ensure that the target language questionnaire contains conceptual, semantic and operational equivalence to the source language questionnaire. Cultural adaptation, as an integral part of the translation, ensures that the target language questionnaire is culturally appropriate, relevant and meaningful for the new culture. Psychometric validation is required to ensure that the target language questionnaire contains measurement properties equivalent to the source language questionnaire. The IQOLA group suggests that the final step in validation is establishing normative values for the new version (20-21). Geisinger elaborated several steps that should be involved in any test-adaptation process (Figure 1) (177).

Considering their importance for the cross-cultural adaptations of HRQOL questionnaires, equivalence and bias will be introduced first. In the remainder of the section, the translation cultural adaption, qualitative approaches, and psychometric validation of HRQOL questionnaires are separately introduced for the matter of clarification. Nevertheless, cross-cultural validation is a consecutive, iterative process with all steps closely interrelated to each other, and many methods are devised to address multiple steps simultaneously.

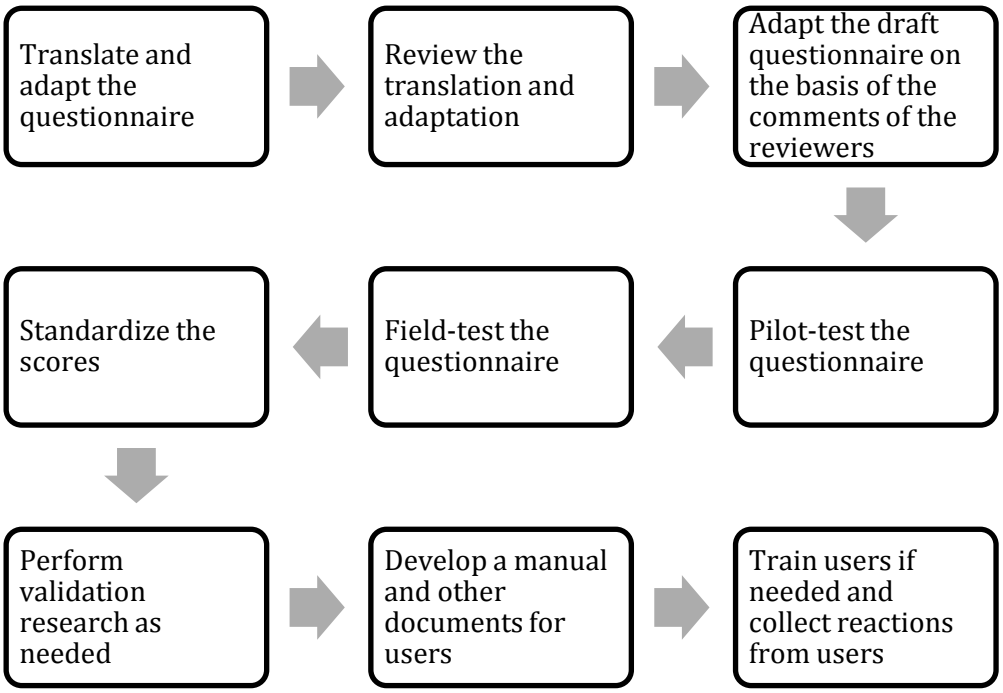

Figure 1 Translation and adaptation process (adapted from ref. 177)

Equivalence

Equivalence refers to the level of comparability of measurement outcomes among languages/cultures (30, 31). In psychological literature, more than 50 definitions of equivalence were identified (32). However, all definitions could be reduced to two main types: *interpretive equivalence* definitions focusing on "equivalence of meaning" and *procedural equivalence* definitions focusing on "measures and procedures used to make cross-cultural comparisons" (30). If applied to cross-cultural HRQOL research in this way, equivalence implies whether the HRQOL concepts have the same meaning across languages/cultures and whether a HRQOL questionnaire developed in one language/culture measures these concepts in the same way as in another.

For psychological tests, the hierarchy of equivalence was suggested including *construct, structural, measurement,* and *scalar equivalence* (30, 33, 186, 187). *Construct equivalence* implies that equivalent constructs have a shared meaning, which is a prerequisite for a cross-cultural comparison. A questionnaire administered across different language/cultural groups shows structural or functional equivalence if it measures the same construct(s) in all these groups. Questionnaires show *metric* (also called *measurement unit*) *equivalence* if their measurement scales have the same units of measurement. Finally, in the case where *scalar* (or *full score*) *equivalence* is achieved, the average scores of two language/culture groups can be directly compared, and conclusions regarding whether these scores are different or equal can be made.

Although borrowed from cross-cultural psychology and educational research, taxonomies of equivalence in cross-cultural HRQOL research are slightly different. Up to 1997, about 19 types of equivalence were reported in HRQOL cross-cultural studies (34). However, due to substantial variations and inconsistencies in defining and assessing various types, five dimensions of equivalence could be considered as the basic ones: *conceptual*, *item*, *semantic*, *operational*, and *measurement* (15, 25, 34, 187).

Conceptual equivalence between two HRQOL questionnaires will be achieved when the questionnaires have the same relationship to a proposed HRQOL concept in both languages/cultures in terms of the domains included (25). This definition implies that HRQOL exists and is relevant and acceptable in both cultures, the questionnaire measures the same HRQOL aspects in each language/culture, and values or emphases placed on different domains are equivalent (13). For example, a HRQOL questionnaire with physical, emotional, and school functioning domains is conceptually equivalent if the target language/culture also considers these domains as relevant to HRQOL. Considering this, it is preferably to establish conceptual equivalence before the translation of a HRQOL questionnaire takes place (25). Conceptual equivalence is hard to achieve,

in part due to the lack of consensus around its definition among HRQOL research. Further, this step is largely neglected as researchers are more focused on claiming that a questionnaire is good measure of HRQOL, without considering the relevance of the underlying concept in the target language/culture (35, 189).

Item equivalence simply concerns the way HRQOL domains are sampled (25). This equivalence is achieved when items estimate the same parameters on the latent trait being measured and when they are equally relevant and acceptable in two different languages/cultures (187). Naturally, the relevance of items varies across languages/cultures, but items are not more difficult to read or understand in target languages/cultures than in the original. Additionally, items vary in acceptability across languages/cultures. Considering that items compose domains, and considering that we still lack clear definitions on HRQOL domains, item relevance and acceptability have substantial implications when testing equivalence. For example, a physical domain might contain an item like running in a park or a social domain might contain an item on dating, and both domains could be considered universally acceptable. When translated, a population in the target languages/cultures might consider running in a park irrelevant as it is not in the source languages/cultures or dating might be an inacceptable item (i.e. taboo).

Semantic equivalence concerns the transfer of HRQOL meanings across languages (25). This implies that items mean the same to people from both languages/cultures, the same expression exists in the target languages/cultures, situations or examples given fit the target languages/cultures, equivalent expressions are found for idioms and colloquialisms, level of language used is appropriate to target population, and technical features of language are equivalent (i.e. complexity, syntax, grammar, and level of abstraction) (15). Research on linguistics and semantics showed a number of different aspects of meaning (referential, stylistic, affective, esc.) (25). Transferring HRQOL meaning could be very challenging and special attention is paid to this process in terms of establishing clear lexical relationships between words. For example, a phrase

"leisure activity" and "leisure time" could be considered to mean the same by some cultures. However, they could imply on different things across different populations, which inherently influences to the interpretation and responses to the HRQOL questionnaire.

Operational equivalence refers to the possibility of using the same questionnaire format, instructions, mode of administration, and measurement in the target languages/cultures as in the source languages/cultures (25). This type of equivalence is far less complex to establish than the other ones. However, if some aspects of operational equivalence are not achieved, cross-cultural compilations might be jeopardized. For example, if the source language questionnaire has pictures for a Likert-like format responses and the target language version omitted the pictures, giving responses for younger children could be problematic and the measurement biased accordingly.

Measurement or *psychometric equivalence* and refers to what degree the target language questionnaire produces reliable and valid HRQOL assessments and to what extent are they consistent with the source language questionnaire (25). Various psychometric properties are evaluated: variability (floor and ceiling effects), missing data, internal consistency and test-retest reliability, factor structure, and construct validity (15). Specific section below deals with this aspect of psychometric equivalence.

Some authors considered also *functional* and *criterion equivalence* as the basic ones (15, 25, 34). *Functional equivalence* is defined to what extent an HRQOL questionnaire does what it is supposed to do equally well in two or more languages/cultures (25). Generally, assessing the degree of functional equivalence actually imply to what extent the other types of equivalence are achieved. *Criterion equivalence* could be achieved only after repeated applications of the target language questionnaire and implies that the interpretation of the measurement of HRQOL remains the same when compared with the norm for each culture (15, 25).

It also refers to its relationship to previously established and independent criteria which does not necessarily mean another questionnaire, which is basically criterion validity (15). Other types of equivalence frequently reported for HRQOL are *technical*, *cultural*, *language*, *experimental*, and others (15, 34).

Bias

Bias refers to differences in a measurement that do not have exactly the same meaning within and across languages/cultures (35). In this way, a cross-cultural study shows bias if differences in a HRQOL measurement do not correspond to cross-cultural differences in HRQOL purportedly measured by the questionnaire. If scores are biased, individual differences within a culture (*within-culture differences*) do not have the same meaning as cultural differences (*between-culture differences*) (30). For example, scores of a HRQOL questionnaire that show bias may be valid measures of HRQOL if they are compared within a single cultural group, whereas cross-cultural differences based on this questionnaire may be influenced by other factors, such as translation issues, item inappropriateness, or differential response styles. Differences due to bias are not random, but systematic, as bias is an inherent characteristic of a questionnaire being applied to at least two cultural groups (30).

Compared to research on equivalence in cross-cultural HRQOL studies, the issue of bias is not investigated systemically, and it is rarely addressed as related to equivalence (15). Therefore, this thesis adopts three main types of bias recognized in cross-cultural psychology (15, 30, 36). These types of bias are distinguished in the questionnaire's translation depending on whether they are brought about by anomalies in the theoretical construct (*construct bias*), questionnaire administration (*method bias*), or specific items (*item bias*) (30).

Construct bias is said to occur when the construct that is measured by a questionnaire shows non-negligible differences across languages/cultures (37).

Either difference in conceptualization or difference in behaviors associated with the construct can underlie construct bias, leading on the final instance to *construct inequivalence*. For example, although the concept of epilepsy as a neurological illness has the same meaning across various languages/cultures, coping with epilepsy is largely culture specific due to different behaviors related to it. If an epilepsy coping questionnaire failed to consider cultural specific behaviors in the target language, but recognize coping as an epilepsy-universal phenomenon, it could be a source for a construct bias. In HRQOL research, the risk of construct bias is very high when developing and using HRQOL questionnaires cross-culturally. The main reason for this is the lack of a clear theoretical basis and definition of the HRQOL construct, as well as a consensus on which domains underlie the construct and how these are best represented by particular items.

Method bias is a generic term for validity-threatening factors that are related to a questionnaire administration (37). Namely, the sample in which the questionnaire was administered, the process of administration, or the questionnaire itself, can be considered as sources for this bias. Method bias is a significant threat to various aspects of equivalence, especially to item and operational equivalence. Examples of method bias include when a translation of a HRQOL questionnaire developed for adults that is used for children with epilepsy, or administering a HRQOL questionnaire to people with severe dementia. Method bias could be also present if only selected scales from a questionnaire are used.

Item bias refers to anomalies at the item level, also known as *differential item functioning* (15, 38). According to a widely used definition in psychology, an item is biased if respondents with the same standing on the underlying construct do not have the same mean score on the item because of different cultural origins (37). Item bias mostly arises from poor item translation, ambiguities in the original items, low familiarity or appropriateness of the item content in certain cultures, or the influence of culture-specific nuisance factors and connotations associated with the item wording (38). Of all types, item bias has been the most extensively studied

with various psychometric techniques available to identify it (30, 31). Item bias is a significant threat to semantic and less to item equivalence in HRQOL research. For example, items on skiing or skating for children who live in a country where snow never falls or an item about pain severity for a condition where pain is not likely to occur.

Numerous procedures have been developed to deal with bias and to examine equivalence to assess the impact of bias on the comparability of questionnaire scores. Procedures to deal with bias can be classified depending on whether they affect how a study is conducted (such as study design, adaptation of stimulus materials, and administration procedure) or how data are analyzed (7, 38-46). The first type is constituted by a priori procedures, because they affect a study before the actual data collection; the second type, a posteriori procedures, affects data analysis. These two kinds of procedures are complementary and not compensatory. For details on the methods, see Appendix I.

Essence of the translation and cultural adaptation process

Broadly speaking, all translations of any questionnaire can be direct and indirect (178, 179). Acquadro et al. define *direct translation* as translation including "…borrowings, calques (loan translations) and word-for-word translation", while *indirect translation* involves "transposition, modulation, equivalence and adaptation" (178). The basic difference is in the notion that in indirect translation there are some "manipulations" that goes beyond simple, literal rendering of the words from the source to the target language to ensure the same purpose and meaning of the questionnaire in both cultures in question (i.e. functional and conceptual equivalence). Considering this, the cross-cultural adaptation process of the source questionnaire to a new language/culture is in its essence an indirect translation. Those who are involved in the cross-cultural adaptation process have to consider the choice between using direct translations, where words are translated literally without conveying their original meaning, or indirect

adaptations where items that do not generalize equivalently in the target culture are modified (177).

Many researchers and clinicians use a simple process of direct translation from one language to another because it is more convenient and less expensive. However, there are some risks associated with it. The simple direct translation is mostly performed by one bilingual person and there is a risk that the translation is biased by the idiosyncratic writing style of that person (50). If the person performing the translation is not familiar with the measuring concept of the questionnaire, it could result in construct bias and compromise conceptual equivalence. Finally, there are no means of the quality control of the direct translation. Therefore, the direct translation is considered as a malpractice, and different approaches were developed as improvements over the direct translation of questionnaires, such as back translation, the committee approach, or decentering (180, 181), which actually represent indirect translation procedures.

Since 1990s, two approaches have been used for the translation and cultural adaptation of PRO questionnaires (12). The first approach is forward-backward translation, also known as *translation/back translation* (173). There are many variations within this approach, including the methods of translation to ensure equivalence, number of translations made, or qualifications of translators (171, 172). In general, the approach includes the following. The first step is the production of one or more translations of a questionnaire from the source language into the target language known as forward translation. All translations were than synthesized in one version, the target language questionnaire. Afterwards, backward translation is carried out where the target language questionnaire is back translated into the source language. The aim of this process is validity checking to make sure that the target questionnaire reflects the same item content as the source questionnaire (i.e. minimizing construct bias) (170).

The second approach is the dual *translation panel*, which is a modification of the committee approach (47, 48, 173, 174, 175). In this approach, a consensus translation is produced by a panel of bilingual people native to the target language together with a representative of the developers of the questionnaire. This is followed by review of the first translation by a second panel consisting of monolingual people of average or below average educational levels to ensure acceptability of wording and ease of completion. The main emphasis accent in this approach is to ensure the quality of the target questionnaire through rigorous forward translation, without considering backward translation at all (47).

In cross-cultural psychology, *decentering* is also recognized as an important approach (173, 175). In the decentering method of translation, items and concepts are paraphrased and translations are made separately for each language based on the paraphrased items. This allows for a better translation of the meaning of the item, instead of a close literal translation of a finalized item, which can result in an un-natural sounding item. This method was far less frequently used with PRO questionnaires than the two presented above (172).

Whichever approach taken, the next main steps are the target questionnaire synthesis, with its pretesting in a population for which the questionnaire was developed to test its content validity (see below). This is important, because a recent study indicated that the initial processes outlined above are insufficient for establishing equivalence and acceptability (182). However, there are a number of modifications in each approach, with many inter-steps, modifications and combinations.

Guidelines for the cross-cultural adaptation of PRO/HRQOL questionnaires

To date, there are 17 proposed guidelines for the cross-cultural adaptation of PRO/HRQOL questionnaires, and almost every HRQOL questionnaire developed during the 2000s had adopted an existing guideline in their own method of cross-

cultural adaptation (12). This practice introduced more inconsistencies in terminologies and methodologies surrounding the cross-cultural adaptation process. Although the development of these guidelines were informed by empirical cross-cultural psychology research, researchers often fail to fully abide by the guidelines and carry out all the steps required in the translation and adaptation process (176, 177). .

Several guidelines provided new insights in the cross-cultural adaptation of HRQOL questionnaires moving the field forward.

The American Association of Orthopedic Surgeons (AAOS) was among the first to suggest strict guidelines for questionnaires' translation (13, 14). According to this guideline, published in 1993 and revised in 2000, the process of translation and cultural adaptation involves six consecutive steps:

- Forward translation performed by at least two independent translators,
- Synthesis of the translations into one consensus version,
- Back-translation,
- Expert committee,
- Test of the pre-final version on 30-40 persons, and
- Submission of the version and all documents for the developers' approval.

Besides the importance of the other parts, the substantial benefit of this approach is the *expert committee* step. The expert committee should be composed

of methodologists, health professionals, language professionals, and translators involved in the process. Its role is to develop the pre-final version for psychometric testing. The committee should review all translations, reaching consensus on any discrepancy. The main goal of this step is to achieve semantic, idiomatic, experiential, and conceptual equivalence.

In 1998, the International Quality of Life Assessment project group (IQOLA) reported on specific guidelines for the translation of SF-36 questionnaire (22). This guideline has very similar steps as the previous one. However, this guideline introduced procedures to be employed during the forward translation and synthesis steps to evaluate the difficulty of the translated items and response choices, using a rating scale to assess clarity of the translation, common language use, and conceptual equivalence. These procedures supplemented the strictly qualitative approach of the cross-cultural adaptation process with a quantitative approach. Unfortunately, this approach has not been accepted for questionnaires other than the SF-36.

During the period 1995-2004, Mapi Research Institute proposed its own guidelines for translation and cultural adaptation (9). The translation methods are similar to those described by Guillemin et al. (13), but the entire process is labeled as *linguistic validation*. This guideline introduced *international harmonization step,* which takes place when the original questionnaire is translated into several languages simultaneously. The aim is to perform further quality control and to ensure greater comparability between source and target versions. In contrast to the other steps of the linguistic validation, this takes place in one country and in the presence of professional translators representing each target language. The harmonization step is achieved at a meeting between the translators, coordinating center and the author. This process highlighted the need to overcome gross errors at the linguistic and conceptual level.

In 2005, the Task Force for Translation and Cultural Adaptation of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) published the *Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures* (50). The guideline proposes the following 10 steps:

- Initial work carried out before the translation work begins,

- Forward translation,

- Reconciliation – comparing and merging more than one forward translation into one single forward translation,

- Back translation,

- Back translation review – comparison of the back-translated versions with the original to highlight and investigate discrepancies between the original and the reconciled translation,

- Harmonization – comparison of back translations of multiple language versions with each other and the original to highlight discrepancies,

- Cognitive debriefing – testing the translation on a small group of relevant population or lay people in order to test alternative wording and to check understandability, interpretation, and cultural relevance,

- Review of cognitive debriefing results and finalization,

- Proofreading – final review of the translation to highlight and correct any typographic, grammatical or other errors, and

- Final report – a report written at the end of the process documenting the development of each translation.

The ISPOR guideline was developed in order to overcome various terminological and methodological inconsistencies arising from the previous guidelines and to standardize the ways of the translation and cultural adaptation process of various PRO questionnaires. This guideline is the most comprehensive.

The four guidelines presented above have been frequently used over the past two decades. Nevertheless, even with these recent activities in developing strict guidelines, we are still lacking a consensus on which one really represents "*the principles of good research practice (GRP)*". This is mainly due to very limited research that compared different methods and generally failing to report substantial differences among them (49).

Qualitative Approaches to translation and cultural adaptation

At some point during the translation and cultural adaptation process, the relevance of the construct represented by a HRQOL questionnaire is evaluated (25). This is known as testing content validity or conceptual equivalence achievement (see above) (15). The goals of this process are to explore how individuals from diverse backgrounds describe the concept, whether some elements are missing, and identify reasons why items may have been problematic during psychometric testing (51). There are three commonly used qualitative methods in the cross-cultural adaptation of HRQOL questionnaires: *cognitive testing, focus groups*, and *expert panels* (15). The first two methods are essentially the pre-testing process, mostly organized with the pre-final target questionnaire before its psychometric evaluation.

Cognitive testing uses theories and methods of cognitive psychology to understand processes used by respondents to understand and answer questions,

and to design questions to increase comprehension (52, 53). These processes could be explored through behavioral coding, probe techniques, and think-aloud interviews. These techniques aim to identify questions that pose problems for either interviewers or respondents and determine the nature and source of the problem to find solutions.

Focus groups are in-depth interviews of small, homogeneous groups (54). They provide researchers with access to the language and concepts used by participants to think and talk about particular topics (55). Hearing participants use their own vocabulary, language, and communication patterns facilitates development and evaluation of optimal item wording for different groups.

An *expert panel* includes consultation with "experts" as a way to learn efficiently about the concept of interest (55). Presumably, the experts (i.e. on the cultural issues and concepts of the group being studied) would have a range of experience, and may include individuals who represent the target group.

Psychometric validation

Following the translation, cultural adaptation, and pre-testing, the next step in the cross-cultural adaptation process is psychometric validation. The paramount aim of this step is to achieve some sort of *measurement equivalence* (12, 25), namely to ensure similar levels of measurement properties between the target language version and the original in terms of reliability, validity, sensitivity to change, and responsiveness. From a psychological point of view, psychometric or measurement equivalence (i.e. invariance) is defined as "…whether or not, under different conditions of observing and studying phenomena, measurements yield measures of the same attributes" (186, 187). In HRQOL literature, measurement equivalence is frequently called *scale* or *construct equivalence*, what is practically *measurement invariance* (34). To avoid confusions, psychometric equivalence in this thesis refers to the levels of comparability in psychometric properties between the source and

the target language questionnaire, while measurement invariance refers to the sameness of the HRQOL construct in the two languages and it is considered only as an aspect of the psychometric equivalence important for questionnaires used for cross-cultural comparison purposes. Therefore, psychometric equivalence as such can be only considered within the operational model proposed here.

Generally, psychometric validation includes the testing of a HRQOL questionnaire within a particular subgroup or across multiple groups and subsequently evaluating its measurement properties through quantitative methods. As for the methods of the translation, cultural adaptation, and pre-testing, the methods of psychometric evaluations for HRQOL questionnaires were adapted from psychology and education following classical and modern test theory (58).

As recognized by regulatory bodies, questionnaires should have sound psychometric properties when used for HRQOL assessments (2, 6, 56). Significant progress has been made over the past decades to establish criteria for methods used in psychometric evaluation when developing HRQOL questionnaires, with various guidelines proposed (22, 57-61). However, firm criteria for psychometric validation during cross-cultural adaptation are lacking. The following paragraphs briefly introduce the most frequently employed psychometric methods used with HRQOL questionnaires in the cross-cultural adaptation process. The reader is referred to classical textbooks for details (58, 59, 62).

All psychometric methods could be simplistically considered as pertaining to the items, scales, or the entire questionnaire (22). At the item level, it is aimed to assess the general “behaviors” of each item, while at the scale level it is aimed to assess how items “behave” together in a scale when measuring a particular domain. At the entire questionnaire levels, it is aimed to assess the behaviors of all scales.

Descriptive statistics. At the item level, the main issues being evaluated are the levels of missing data, distribution of items' scores, and distribution of responses (15, 22, 63). Missing data tends to indicate on problems with particular items, such as irrelevance, ambiguity, difficulty, and inappropriateness (22). Greater number of missing items in a scale might indicate possible problems where the concept represented by the scale is not adequately captured. Further, the distribution of items' scores (mean and standard deviation values) should be roughly equivalent among the items in a scale in order to judge the variability of the specific concept measured by the scale (22). The frequency of responses should roughly indicate whether responders had considered all item responses (15).

In a scale of a questionnaire, it is also important to demonstrate the relationships between items using the correlation matrix (22). This examines to what extent items supposed to correlate are indeed correlated (convergence) and to what extent items not supposed to correlate are indeed not correlated (discriminations). Problems in the correlation matrix are indicative of item's irrelevance, ambiguity, difficulty, inappropriateness, and so on (22).

The scale level descriptive statistics concerns scale variability. Scale scores should represent approximately the full range (or comparable ranges) with comparable (and minimal) floor and ceiling effects to enable worsening or improvement to be detected within each group. Extensively low scores (floor effects) and high scores (ceiling effect) can attenuate reliability and validity, reducing variation in a questionnaire (22, 63).

Reliability. Reliability is the extent to which a questionnaire is free of measurement error (59). Two basic forms of reliability exist: (1) internal-consistency and (2) reproducibility (test-retest reliability). Internal consistency is the extent to which items in a questionnaire scale are correlated (homogeneous), thus measuring the same single underlying concept (57). An internally consistent scale is achieved through sound construct definitions, appropriate items, following

principal component analysis or exploratory factor analysis, and confirmatory factor analysis (57). Reproducibility concerns the degree to which repeated measurements in stable persons provide similar data. This form of reliability is achieved in repeated applications of a HRQOL questionnaire and it is important for HRQOL questionnaires to be used in evaluative purposes (58).

Validity. Validity is the extent to which a questionnaire is measuring the concept it is supposed to measure (59). Three types of validity are important for health status questionnaires: content, criterion, and construct.

Content validity is the extent to which a questionnaire represents all aspects of the defined concept, in this case HRQOL. Content validity is concerned with the inferences that we could come to when assessing HRQOL with a particular questionnaire (64). It is relevant to all elements of a questionnaire that affect the obtained data, including item content, presentation of stimuli, instructions, behavior codes, time-sampling parameters, and scoring (58). Content validity is best achieved during the initial phases of the original questionnaire development (15). However, in cross-cultural adaptation it is mostly achieved before the translation of a questionnaire and in the pre-testing phase to evaluate relative importance, equivalence, relevance, appropriateness, and other aspects of items, responses, and instructions.

Criterion validity is the extent to which a questionnaire corresponds to an accurate or previously validated questionnaire of the same concept or to an external criterion established by the investigators – "gold standard" (59). It is divided into *concurrent* and *predictive validity*. The first type concerns the ability of a questionnaire to provide the same or similar results in HRQOL measuring as the "gold standard" does, which is achieved through comparing a new questionnaire against one or more well-established ones (59). The second type concerns the ability of a questionnaire to predict future HRQOL discovered by an external

criterion, which is achieved through comparing a new questionnaire against criteria available in the near future (59).

Construct validity refers to the extent to which a questionnaire measures some theoretical construct in this case HRQOL. Specifically, it concerns to what extent scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured (58).Construct validity can be divided into *convergent* and *discriminant validity*, where both concern the extent to which a questionnaire measures theoretically similar or different domain respectively then the predicted one. This type of validity is achieved through the "extreme groups" validation, correlations with other questionnaires, and multitrait-multimethod matrix (58, 59). In the "extreme groups" validation, it is assumed that groups differing in a health status would have different levels of HRQOL. The second and third methods are similar; two or more different questionnaires or methods of assessments are considered to evaluate levels of HRQOL and they are correlated to estimate the convergent and discriminant validity of the questionnaire of interest.

Similar but different to construct validity is structural validity, which generally concerns how many "things" in one specific concept does a particular questionnaire measure by its items. This is tested using factor analysis (FA) and it is known as factorial validity (69, 186). Factor analysis (FA) is a statistical technique designed to reveal whether or not the pattern of responses on a number of items can be explained by a smaller number of underlying factors (60, 65-69). In cross-cultural adaptation, FA has become one of the most suitable methods to evaluate dimensions being measured by one questionnaire and to investigate whether the questionnaire shows the same dimensions across different groups (68). The factor structure of a questionnaire can be investigated by means of *exploratory factor analysis (EFA)* or *confirmatory factor analysis (CFA)*. The first aims to determine the number of dimensions and their mutual associations based on responses on particular items, while the other concerns testing an applied

hypothesis on particular dimensions from items. In this way, EFA proceeds with the development of one or more hypothetical models that are later tested in CFA. For cross-cultural adaptation, of particular importance is multiple-group CFA using structural equation modeling (SEM), when the adapted HRQOL questionnaire will be used for cross-cultural compressions (15). Multi-group SEM enables test of significance of measurement invariance between two or more cultural groups (69), and it is defined as "…whether or not, under different conditions of observing and studying phenomena, measurements yield measures of the same attributes" (186). In order to compare estimates by the questionnaire across various nations/countries, an important aspect that needs to be demonstrated is that reproducible factorial structure across different ethnic/cultural groups is also invariant (i.e. that the factors are measured in the same way). Besides multi-group SEM, there are also different item-response theory methods to test measurement invariance between two or more groups (58).

Sensitivity and responsiveness. Two closely related properties to reproducibility are sensitivity and responsiveness. Sensitivity is the ability of a questionnaire to detect differences between groups in levels of HRQOL, which is analog to testing construct validity using the "extreme groups", except that the groups differ in severity or types of one condition and not in two or more different conditions (58, 59). Responsiveness is the ability of a questionnaire to detect changes in HRQOL in an individual over time. An evaluative questionnaire should not only be reliable, yielding reproducible results when a person's condition is stable and unchanged, but in addition it should respond to relevant clinical changes in a person's condition (58). Sensitivity can be assessed by cross-sectional studies, but responsiveness is evaluated by longitudinal assessment of persons in whom a change is expected to occur. Two of the most widely used measures of sensitivity and responsiveness are the standardized response mean (SRM) and the effect size (ES), which are also used for indicating clinical significance (58).

Differential item functioning. During the past decade it has become important to evaluate how items function regardless of groups' characteristics, which is especially important for the cross-cultural adaptation of HRQOL questionnaire (15). Differential item functioning (DIF) arises when one or more items in a scale perform differently in various subgroups of persons (58). This is drawn from the item-response theory, where it is assumed that persons with a particular level of HRQOL will have a certain probability of responding positively to each question (30, 31, 58, 70-72). This probability will depend upon the 'difficulty' of the item in question. One particularly important application of DIF analysis is the detection of linguistic and cultural differences. It provides a powerful tool for detecting whether patients in one language group respond to an item differently from other patients; if an item shows DIF, it may be indicative of cultural differences or, more probably, a translation inadequacy (58). Various methods exist for testing DIF (20, 58, 73).

Finalization of the cross-cultural adaptation process

Two steps at least should be taken to finalize the cross-cultural adaptation process; to standardize the scores of a questionnaire by developing its norms and to develop manual and other documents for test-takers (11, 21, 177, 180). The standardization of the scores of the questionnaire is important for the development of norms that could be used for within-culture and cross-cultural research when equated back to the score scale of the original (22). A manual that describes administration, scoring, and interpretation should be developed to provide information that relates to interpretation, summarization of norms, equating (if any), reliability analyses, validity analyses, and investigations of bias. Additionally, training of users might be needed and collection of reactions of a test takers (177).

Using HRQOL questionnaires with children and adolescents

It is early recognized that HRQOL assessments have some specific characteristics when children and adolescents are considered. So far, the development of pediatric HRQOL measurement followed specific pathways with several important issues recognized; specific HRQOL domains, age and developmental characteristics, self- and proxy-rating, generic and disease specific approaches in assessments, and psychometric considerations.

First, synthesizing the pediatric literature on HRQOL, Davis et al. recognized that HRQOL measurements mostly deal with evaluating levels of functioning, feelings about functioning, health, and value assigned to duration of life (74). The main identified HRQOL domains are physical and psychological well-being, energy and vitality, self-perception, cognitive functioning, social functioning and support, autonomy and independence, psychological relations to the material environment, and general health perception and life quality (3). Therefore, assessment of children and adolescents' HRQOL should be based on a questionnaire including as many of these domains as possible.

Second, it was recognized that HRQOL changes substantially, mostly deteriorating, throughout a child's growth and development (3, 75, 76). Although sound research data are lacking, the deteriorations in HRQOL are probably inherent to physical, emotional, cognitive, and social development. It is suspected that more risk factors for poorer HRQOL emerge at adolescence than in childhood, such as risk-taking behaviors and increased school demands. Additionally, the preferences in HRQOL are changing; adolescents' HRQOL has specific domains such as maturation, intimacy, and sexuality as important components added to classically included domains in assessments (77). Considering these observations, HRQOL assessment should consider relevant age group characteristics and use questionnaires developed for specific groups, mainly for children up to 3, 4-7, 8-12, and adolescents for 13-16 (18) years (3, 74-76, 78). Finally, developmental

characteristics should be specified according to HRQOL domains such as whether a child with cerebral palsy can write (physical domain) or whether a child suffering from parental neglect attend school regularly (social domain) (3, 78).

Third, close to the previous issue is whether the child can rate their own HRQOL or whether a proxy should be considered (3, 76, 79, 80). The proxy is usually the most closely related person or caregiver to the child such as his/her parents, grandparents, or others who can rate the child's HRQOL. It is now widely accepted that using both self- and proxy-ratings provides the most comprehensive picture of the child's HRQOL (80).

Finally, based on the aims of assessment, appropriate HRQOL questionnaires should be selected based on its type and psychometrical characteristics (3, 75, 82). More than 30 generic and 64 disease-specific instruments were developed (83), most of which possess sound psychometric properties in terms of reliability, validity, responsiveness, and interpretability.

These issues need to be resolved prior any HRQOL assessment, because they determine the selection of appropriate questionnaires (167).

Rationale and aims

It is universally accepted that the evaluation of HRQOL in a group of children or adolescents should include all relevant domains to that group and it should be performed using appropriate methodology, taking into account sophisticated measures developed and regulations asserted. Additionally, appropriate HRQOL assessments in children and adolescents should also consider the cross-cultural settings. Specifically, cross-culturally available HRQOL questionnaires are needed for multinational assessments and cross-cultural comparisons of HRQOL data (3). Considering this requirement, many contemporary pediatric HRQOL questionnaires are cross-culturally developed or cross-culturally adapted (3, 83). For example, the KIDSCREEN questionnaire has been cross-culturally developed for 10 languages in Europe, and is cross-culturally adapted for 25 others (24; for details see www.kidscreen.org/english/language-versions/existing-language-versions/).

To date, no clear guidelines for the cross-cultural adaptation of pediatric HRQOL questionnaires have been developed. The adaptations of pediatric HRQOL questionnaires like the KIDSCREEN primarily followed guidelines that were established for the adult population, and to date, no clear guidelines for the cross-cultural adaptation of pediatric HRQOL questionnaires have been developed (183). This is mostly due to widely held opinion that the same principals for the cross-cultural adaptation briefly described in the Introduction apply to children and adolescents as well. Therefore, it could be questionable whether we actually need guidelines for the cross-cultural adaptation of HRQOL questionnaires in pediatric population, when we can use available ones. Thus, the development of guidelines for the cross-cultural adaptation of pediatric HRQOL questionnaires might be seen as “reinventing a wheel”.

Nevertheless, there are several strong arguments for developing the regulations of the cross-cultural adaptation process of pediatric HRQOL questionnaires.

First, the main arguments is in the notion that pediatric HRQOL research is relatively new and still developing, with a majority of available studies being based on Western societies and concepts of well-being and functioning. Cross-cultural studies on the concept in general, but also considering specific conditions, are lacking. Furthermore, there are no clear understandings of the HRQOL concept cross-culturally, either in its definition, measurements, or applications, despite available recommendations for measuring pediatric HRQOL (3, 74-83) or recent efforts to develop HRQOL questionnaires cross-nationally (24, 168, 169). Therefore, one of the focuses of the cross-cultural adaptation of HRQOL questionnaires is answering the question to which extent a HRQOL questionnaire developed for children and adolescents in one language/culture measures the aspects of the HRQOL construct in another, where HRQOL was not previously evaluated or even conceived. This aspect of the cross-cultural adaptation of HRQOL questionnaires is frequently skipped and authors simply engage in the translation and cultural adaptation of one questionnaire assuming that it inherently measures the same underlying concept of HRQOL in two different languages/cultures.

Second, the source language questionnaires are not cross-culturally adapted in an appropriate way. This implies that not all steps of the process are taken or even considered appropriate to ensure levels of equivalence - translation, cultural adaptation, pre-testing, or psychometric validation. For example, Rajmil with his colleagues recently reviewed HRQOL measurements in children and adolescents in Ibero-American countries for the period 2000 – 2010 (84). Adaptation of instruments in Spanish-speaking countries generally followed international guidelines developed by Guillemin et al., (13), but they concluded that the available information about the translation and adaptation process was insufficient. Forward and back translations were most often used, although a simple direct translation was also present. Some HRQOL questionnaires were adapted only through exploratory factor analysis to identify the relationships between the concepts of life satisfaction, without fully following all other steps required to develop a standardized questionnaire (84). The practice that the source language questionnaires are not cross-culturally adapted appropriately is also evident in our

community. There are about ten pediatric HRQOL questionnaires cross-culturally adapted for Serbian children from 2000 to 2013 (Stevanovic 2013, unpublished data). All of them mentioned the cross-cultural adaptation process, but not all aspects of the process were considered in some studies; such as which steps were followed during the process (164), how pre-testing was organized (165), or not organizing the questionnaire's pre-testing in the target group of children and adolescents (85). More importantly, these authors failed to report any aspects of equivalence testing between the original and Serbian version.

Close to the previous argument, it is not infrequent that the source language questionnaire is not cross-culturally adapted in the same way or as proposed by the developer to ensure satisfactory equivalence with more the target language versions. For example, a HRQOL questionnaire for children with systemic lupus erythematosus (SLE) was developed in the United States; called Simple Measure of Impact of Lupus Erythematosus in Youngsters (SMILEY). The SMILEY was recently cross-culturally adapted for 13 other languages. However, although the authors claimed that rigorous translation method was applied, across the cultures one to three forward/backward translations were considered, with two to three experts as reviewers of the versions (166). Therefore, the regulations of the cross-cultural adaptation of HRQOL questionnaires could overcome this practice in the way of standardizing the process.

Third, the methodological frameworks of HRQOL questionnaire are frequently not respected during the cross-cultural adaptation process. Specifically, the source language questionnaire's HRQOL domains, target age group, and response styles, are often neglected when the questionnaire is adapted or used for another culture (86-89). Additionally, the decision to use one HRQOL questionnaire or another for children and adolescents depends on many factors such as availability of a questionnaire, type of measurement, domains a questionnaire assesses, and others (3, 75, 76, 82). Today, the decision to use one HRQOL is almost exclusively based on the notion that one specific measure was used previously for a specific purpose in one culture, without explorations of its measuring concept, operational characteristics, or psychometric properties for

another culture. Therefore, the selection of questionnaires has mostly been guided by the purpose of HRQOL assessment, rather than their psychometric qualities or the other aforementioned aspects (167, 189). Thus, regulations of the cross-cultural adaptation of HRQOL questionnaires would assist the questionnaire's selection and standardize the methodological frameworks of HRQOL measurements.

Finally, the cross-cultural adaptation is a complex process involving different steps, thus the regulations should follow one clear framework of the cross-cultural adaptation process that will include all important aspects and guide translation, cultural adaptation, pre-testing, or psychometric validation. This is important because it is not always possible to follow suggested regulations with pediatric HRQOL questionnaires. For example, even when reasonable steps are employed to evaluate the levels of measurement equivalence in a target language questionnaire, there might not be enough subjects with a particular child disorder to organize the field-testing with a sufficient number of participants. Another example would be that a previously adapted HRQOL questionnaire is also suggested for younger age group by the developers. Thus, the translation needs to be further evaluated in that particular group.

Aims

In the light of the rationales above, the purpose of this thesis was to operationalize a model of the cross-cultural adaptation of pediatric HRQOL questionnaires. Taking a systematic approach, the cross-cultural adaptation process of pediatric HRQOL questionnaires has been operationalized including the procedures of translation and cultural adaptation, pretesting, psychometric evaluation, and assessing equivalence with bias elimination. Additionally, it aimed to present the importance of risk factors for the cross-cultural assessments of HRQOL. Although not part of the cross-cultural adaptation process itself, identifying the risk factors associated with an assessment with a specific HRQOL questionnaire is important for cross-cultural applications of that questionnaire. The operational model was

presented considering four generic and two epilepsy specific questionnaires to include questionnaires for different age groups, self- and proxy-ratings, and generic and desease specific approach.

The specific aims were:

- To translate, culturally adapt, and psychometrically analyze the Serbian version of the KINDL questionnaire for HRQOL assessment in children and adolescents (90),
- To translate, culturally adapt, and psychometrically analyze the psychometric properties of the Serbian set of the KIDSCREEN questionnaires: KIDSCREEN-52, KIDSCREEN-27, and KIDSCREEN-10 questionnaire (24),
- To psychometrically analyze the Serbian version of the Pediatric Quality of Life Inventory Version 4.0 Generic Core Scales – PedsQL (91),
- To translate, culturally adapt, and psychometrically analyze the Serbian version of the Quality of Life Enjoyment and Satisfaction Questionnaire Short form – Q-LES-Q – SF (92);
- To translate, culturally adapt, and psychometrically analyze the Serbian version of the Health-Related Quality of Life Measure for Children with Epilepsy - CEQOL-25 (93),
- To translate, culturally adapt, and psychometrically analyze the Serbian version of the Quality of Life in Epilepsy Inventory for Adolescents - QOLIE-AD-48 (94), and
- To evaluate determinants of HRQOL in adolescents with well-controlled epilepsy.

Methods

This thesis comprises multiple individual studies and the methodology of each study is provided in their original published articles. However, the procedures of the translation and cultural adaptation in some articles were described to a lesser extent than the psychometric procedures were, due to the limitations for presentation set by particular journals. Therefore, the translation/cultural adaptation and pretesting, as well as the assessment and reporting of equivalence and bias, are described in details here.

Operational model of the cross-cultural adaptation process

The operational model of the cross-cultural adaptation process for pediatric HRQOL questionnaires was developed following the previously established guidelines for translation and cultural adaptation, psychometric evaluation, and equivalence testing with bias elimination. Specifically, the operational model presented here is an adaptation for pediatric HRQOL questionnaires of the ISPOR's principles (50), the AAOS guidelines (13, 14), the IQOLA guidelines (22), the model of equivalence modified according to Herman et al. (25), and standard statistical methods for health measurement scales (21, 56-61).

The operational model has five consecutive phases (Figure 2): pre-translation, translation, pre-testing, psychometric, and finalization phase. Within the cross-cultural adaptation model, five aspects of equivalence are recognized namely conceptual, semantic, items, operational, and psychometric/measurement and three types of bias namely construct, method, and item. As outlined in (Table 1), the suitability of the questionnaire to be used cross-culturally is determined by the level of equivalence achieved.

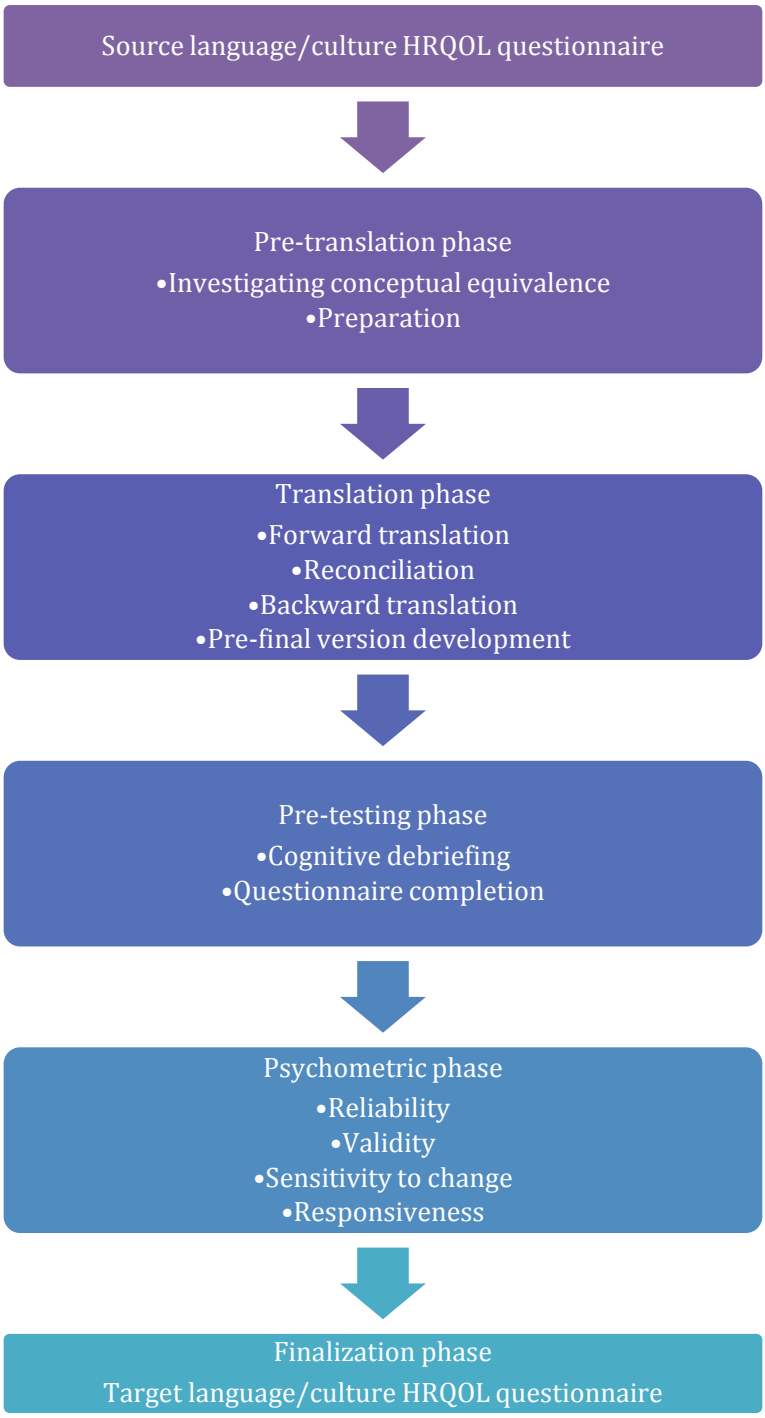

Figure 2 Operational model of the cross-cultural adaptation process

Table 1 Operational model of investigating equivalence

Equivalence	**Domains/items for the target language/culture vs. the source language/culture**	**Cross-cultural adaptation**	**The target vs. the source questionnaire equivalence**	**Cross-cultural comparison with both questionnaires**
Conceptual	Relevant and important	Suitable	Satisfactory	Suitable
	Relevant and important, but the importance of some items might be differently perceived	Suitable only with domains weighting	Achieved with weighted domains	
	Some domains not important or relevant	Limited*	Limited	Not suitable
	All domains not important or irrelevant	Not suitable*	/	/
Item	Same items could be used		Satisfactory	Suitable
	Minor cultural adaptations to some items			
	Some items had to be replaced		Limited	Not suitable
	Some items had to be omitted		Limited	Not suitable
Semantic	Easy to translate all items		Satisfactory	Suitable
	Difficult to translate some items			
	Impossible to translate some items		Limited	Not suitable
Operational	Same aspects of operationalisation		Satisfactory	Suitable
	Some aspects of operationalisation different			
	Impossible to use the same operationalisation for the entire questionnaire		Not achieved	Not suitable
Psychometric	Reliable, valid, sensitive to change and responsive assessments Reproducible factorial structure across different ethnic/cultural groups is invariant (i.e. measurement invariance)		Satisfactory psychometric characteristics and measurement invariance achieved	Suitable
			Satisfactory psychometric characteristics, but measurement invariance not achieved	Not suitable
			Not satisfactory psychometric characteristics	Not suitable

* The cross-cultural adaptation should not be performed for a selected questionnaire, because it could be considered as conceptually inappropriate for the target language. However, it can be used for further HRQOL conceptualizations in our community.

Pre-translation phase

Inevitable steps before engaging in the cross-cultural adaptation process are to investigate *conceptual equivalence,* to minimize *construct bias,* and to organize *preparation* activities (25, 50).

Investigating conceptual equivalence

Investigating conceptual equivalence involves assessing of the nature of the HRQOL concepts that are represented by a questionnaire, and to what extent the source and target language questionnaire could measure the same HRQOL aspects (25). The essence of investigating conceptual equivalence is actually evaluating how children and adolescents from two languages/cultures perceive the HRQOL concept. The aspects of conceptual equivalence were tested in the subsequent steps during the cross-cultural adaptation process of a specific questionnaire. However, its evaluation in the initial phase is important in order to determine whether the HRQOL construct could be considered for Serbian children and adolescents as it was for other cultures and to minimize construct bias at an early stage. There are different methods proposed for investigating conceptual equivalence and the HRQOL construct itself (15, 25), but in this thesis three methods were considered. These methods are also implemented to eliminate construct bias.

The first method for investigating the HRQOL concept was the evaluation of previous HRQOL assessments among children and adolescents in Serbia. This method included performing a mini-literature review of studies dealing with HRQOL in pediatric populations and assessing the previously adapted questionnaires for pediatric HRQOL or similar constructs such as health status or functioning. It was aimed to elicit information about how the previously used health questionnaires "behaved", what difficulties authors encountered during the measurements, and what were the measurement outcomes.

The second method for investigating the HRQOL concept involved consulting pediatricians and child psychologists about the HRQOL concept among children and adolescents. With this method, it was aimed to obtain relevant information for HRQOL assessments from care providers about specific health care needs of children and adolescents with chronic disorders. Specifically, this method elicited information on relevant HRQOL domains and their risk factors such as levels of physical and psychosocial functioning or disease characteristics.

The third method for investigating the HRQOL concept included semi-structured interviews with children and adolescents. In brief, the domains of HRQOL and items representing them in available generic questionnaires were discussed in a small focus group including 3-5 children and adolescents. Epilepsy specific questionnaires were considered with children and adolescents with epilepsy. Participants were asked to comment on the items meanings and how relevant they were for eliciting information on one's well-being and functioning. With this method, it was aimed to obtain relevant information for the HRQOL concept directly from children and adolescents.

The results of these outcomes largely guided the selection of the questionnaires used.

Four possible outcomes of investigating conceptual equivalence (modified from ref. 25) and levels of construct bias were considered in determining whether one or more HRQOL questionnaires could be used for Serbian children and adolescents:

1. The HRQOL domains constituting the source language questionnaire were relevant and important for Serbian children and adolescents, with minimal levels of construct bias if the questionnaire is translated and culturally adapted,

2. The HRQOL domains constituting the source language questionnaire were relevant and important for Serbian children and adolescents, but the importance of the domains varied between the two cultures, with possible levels of construct bias if the questionnaire is translated and culturally adapted,

3. One or more of the HRQOL domains constituting the source language questionnaire were not important or relevant for Serbian children and adolescents, with certain levels of construct bias if the questionnaire is translated and culturally adapted, or
4. The HRQOL domains constituting the source language questionnaire were not important or relevant for Serbian children and adolescents.

As Herdman et al. suggested, in the case of the first and second outcome, it is safe to proceed with the cross-cultural adaptation process, although the second outcome would require the domains to be weighed in order to reflect differing importance (25) in order to minimize the occurrence of construct bias. In the case of the third outcome, it might be possible to use the relevant HRQOL domains from the source questionnaires, though careful attention should be paid to the effects on its psychometric properties, and this will only allow a partial comparison across cultures due to certain levels of construct bias. Further, the measurement invariance of the questionnaire needs to be demonstrated to allow comparisons across cultures (187). In the case of the fourth outcome, the HRQOL questionnaire could not be considered for the cross-cultural adaptation. Nevertheless, the questionnaire can still be used for HRQOL conceptualizations in our culture or in the development of a national HRQOL questionnaire.

Preparation

The next step of the pre-translation phase is the preparation of the translation and cultural adaptation of the selected HRQOL questionnaire (50).

This step involved obtaining permission to use the questionnaire in our population, which served to resolve copyright issues. All pediatric HRQOL questionnaires are copyrighted and the use without permission of the copyright holder could be a subject of lawsuit. Moreover, obtaining permission is important, because it is the only way to produce the official translation/adaptation of that particular questionnaire. Today, it is possible to locate one HRQOL questionnaire

on the Internet and translate/adapt it without permission, what might result with unofficial translations of questionable quality.

The next key aspect of the preparation step was to invite the questionnaire's developers to be involved in the cross-cultural adaptation process. The developers might help to clarify the meanings of items and domains and help overcome ambiguities while transferring the meanings into the target language. Additionally, each pediatric HRQOL questionnaire is accompanied by manuals describing the measurement concept in detail. Sometimes, the developers provided reports on the translations of the questionnaire into other languages as in the case of the KINDL and KIDSCREEN (*Article 1, 2 and 3*). Considering the contacts with the developers and the manuals, the explanations were developed for the concepts that will be translated in the questionnaires used. These steps were taken to provide more data about conceptual equivalence and to minimize construct bias early.

Finally, a team for the cross-cultural adaptation process was recruited during the preparation step (50), and the team members were allocated the tasks of conducting forward/backward translations or attending panel meetings.

Translation phase

The translation phase of the operational model for pediatric HRQOL was considered to include the following:

1. Forward translation,
2. Reconciliation,
3. Backward translation, and
4. Pre-final version development.

Forward translation

Initially, two independent forward translations were performed from English into Serbian for all questionnaires. Bilingual translators, with the Serbian mother tongue, produced two independent translations. This is important because the translations into the mother tongue more accurately reflect the nuances of the language (14). At least one translator was familiar with the HRQOL concept, while both were experienced in working with children and adolescents.

Each translator was provided with details on the translation/adaptation procedures briefly presented in Appendix II. Additionally, the translators were also provided with short descriptions of the meanings of items and domains in general as provided by the developers of the questionnaires, as well as the results of the conceptual equivalence testing (see above). During the translation/adaptation process, maintaining the meaning of an item/response/instruction was the paramount aim (i.e. achieving conceptual and semantic equivalence, but eliminating construct and item bias) (25, 30). In this regard, the translators were requested to choose between altering the words in the items/responses/instructions literally without conveying the sense of the original and removing, changing, adding, supplementing and/or modifying those items/responses/instructions that deal with behavior that does not generalize equivalently in our language/culture.

Items, response options, and instructions were translated/adapted in the same way. A "*pool of possible translations*" was provided by each translator including all possible translation/adaptation options for one item/response/instruction. Additionally, each translation of items/responses/instructions was graded as easy, difficult, or impossible to translate (i.e. item equivalence). The translators produced a written report outlining the challenging phrases or uncertainties and rationales for their choices.

Reconciliation

Two forward translations were synthesized into single translation during a small panel meeting. This panel was run by the principal author and included translators, one child psychologist and one layperson. When two epilepsy specific questionnaires were reconciled, three to five children and adolescents with epilepsy were included to overcome problems that arose during the translation related to this specific disorder. During the panel section, each item/response/instruction was read one-by-one, and the panel decided on the best option from the "*pool of possible translations*". The translators' reports were also considered carefully during this process. The reconciliation resolved discrepancies between the original independent translations (50), with single forward translation developed. In the case that the discrepancies could not be resolved, that item/response/instruction was left as translated with all possibilities. A summary report with comments was created for the translation.

Backward translation

The backward translation process represents the quality control step, as it demonstrates the quality of the Serbian translation such that the same meaning is derived when the translation is moved back into the English language (50). The first Serbian version developed during the reconciliation phase was back translated into English by two independent bilingual translators not included in the forward translation and not familiar with the HRQOL concept, but with a medical/clinical background and/or experience in working with children and adolescents. The translators were instructed to weigh between the literal and conceptual back-translation of items/response options/instructions. Again, a "*pool of possible translations*" was provided including all possible back-translation options for a particular item/response/instruction, as well as each was graded as easy, difficult, or impossible to back-translate.

Pre-final version development

Panel meetings were organized in order to review the backward translation process, run harmonization, and develop the pre-final Serbian version (sometimes numbered as 1.0 or named Serb V) that would be pre-tested. This panel was again run by the principal author and included at least one forward translator and both backward translators, a child psychologist, pediatrician, and three to five children with at least one parent. In the case of two epilepsy specific questionnaires, children and adolescents with epilepsy were included.

The first part of the panel was devoted to the backward translation review and to the evaluation of conceptual equivalence and construct bias. One-by-one, items/responses/instructions were analyzed considering levels of agreement in the concepts among the two back-translations and the source questionnaire. Special attention had been paid to the difficulty of the translation. These procedures aimed to eliminate any discrepancies by consensus.

The second part of the panel was devoted to the harmonization process that aimed to detect and deal with any translation discrepancies that arise between different language versions, thus ensuring conceptual equivalence between the source and Serbian language version and minimizing potential construct bias (15, 25). On a one-by-one basis, items/responses/instructions were compared with the translation/adaptation solutions and discrepancies for the previous language versions of the questionnaires were considered. Two approaches were taken. First, considering the cultural and language similarities, Croatian versions were considered for harmonization (the KINDL and QOLIE-AD 48 questionnaire, Article 1 and 2 and Article 6). Second, the translation discrepancies in the Serbian version were compared with ones provided by developers for the source language questionnaire or other versions (KIDSCREEN and CHEQOL-25, Article 1 and 2 and Article 5).

The third part of the panel was devoted to the development of the pre-final Serbian version for the pre-testing phase. The best translation option for each

item/response/instruction was used. However, in the case more than one option could be used or the discrepancies during the translation were not resolved, that item/response/instruction was left with all possible translations to be discussed in details during the pre-testing phase. The pre-final Serbian version was formatted as the original questionnaire.

The final part of the panel was devoted to estimate the degree of the conceptual, item, and semantic equivalence/bias between the source questionnaire and Serbian version. Item equivalence concerns the way in which domains are sampled and it involves both qualitative estimation of the relevance of items, and quantitative psychometric evaluation of the items (25). As proposed in the source questionnaire, each item from all domains was analyzed to what extent it required some modification to be used for valid assessments among Serbian children and adolescents. In this phase, four possible outcomes of investigating item equivalence (25) and levels of bias were considered for Serbian versions:

1. Each item from the source domain could be used without major modifications in the Serbian versions, except necessary translation, with minimal levels of concept and item bias,

2. Item/items from the source domain required minor cultural adaptations in the Serbian versions, in order to represent the underlying measurement concept of the domain more culturally sensitive to Serbian children and adolescents, with possible levels of concept and item bias,

3. Item/items from the source domain had to be replaced, in order to represent the underlying measurement concepts of the domains more appropriately for Serbian children and adolescents, with defined levels of concept and item bias, or

4. Item/items from the source domain should be omitted due to inappropriateness or irrelevance for Serbian children and adolescents.

Herdman et al. suggested the first two outcomes are desirable and items could be further tested (25). In the case of the third and fourth outcome, new versions of the questionnaires are developed depending of a number of items replaced or omitted, which would allow the questionnaire to be available for HRQOL in the target language, but its use in cross-cultural comparisons could be bias, due to the existence of construct and items bias. As for conceptual equivalence, the measurement invariance of the questionnaire needs to be demonstrated to allow comparisons across cultures (187).

Semantic equivalence concerns the transfer of meaning from the source language into Serbian (25). It involved qualitative estimation of the difficulty of translation as provided by the translators during the forward and backward translation processes. Each item from all domains was analyzed separately. Three possible outcomes of investigating semantic equivalence were considered for the Serbian versions (25):

1. Each item from the source domain was easy to translate into Serbian,
2. Item/items from the source domain was/were difficult to translate into Serbian, and
3. Item/items from the source domain was/were impossible to translate into Serbian.

Pre-testing phase

The pre-testing process serves to assess the level of comprehensibility of items/responses/instructions, to test any translation alternatives that might not be resolved in the previous steps, to highlight any items that might be inappropriate, and to test the operational properties of the translation such as its format, layout, instructions, and others (15, 25, 50). The pre-testing of each questionnaire was organized using cognitive debriefing and the questionnaires completion with a group of children/adolescents.

Cognitive debriefing

Cognitive debriefing is a qualitative method used to determine whether items/responses/instructions in the translation are understandable, interpretable, and relevant for children and adolescents as it intended by the original questionnaire. This is a direct examination of construct, method, and item bias.

Five to twenty children and adolescents with their parents participated in independent semi-structure interviews. The parents were included in order to provide inputs for the proxy versions of the translations. Children and adolescents aged 8-18 (20) years of both genders, were healthy or with chronic disorders (i.e. asthma, diabetes, juvenile idiopathic arthritis, and anxiety disorders). In the case of two epilepsy specific questionnaires, children and adolescents with various types of seizures and epilepsy participated. One-by-one, items/responses/instructions were read to the child/adolescent and they were asked the following:

- To explain what exactly the item was measuring or what the instruction was about,
- To respond to the item in the way it applied to him/her,
- How he/she formulated the response,
- Whether the item is relevant,
- Whether the item is appropriate, difficult, or confusing,
- Whether the response format is appropriate, and
- Whether format, layout, and mode of administration is convenient.

This process explored the precision, clarity, effectiveness, relevance, and appropriateness of the item/response/instruction. Special attention was paid to those that were left with more translation options or unresolved discrepancies in the meanings during the previous phases.

Questionnaire completion

A questionnaire completion was organized with another group of children and adolescents after the he questionnaire was formatted as the original one. The missing-data, responding distribution and the period to complete the questionnaire were recorded. Participants were also asked to comment on the questionnaire overall presentation and readability, as well as on any other aspect they found relevant.

After the pretesting, the final panel of the main investigators was met to review the results of the pre-testing together with the reports of the previous phases. It was aimed to incorporate the findings of the pre-testing process into the translation, to check for minor errors, which have been missed during the translation process, and to finalize the version that would be used for psychometric evaluation (50). The final report for the Serbian translation was created and submitted to the developers for approval. The results of the cross-cultutal adaptation process were sometimes discussed through telephone or Skype with the developers, as well as through written reports (KINDL, KIDSCREEN, and CHEQOL-25; Article 1, 2, 3, and 5). This practice was important, because the item/response/instruction with more translation options or unresolved discrepancies in the meanings were discussed with the developers.

The panel meetings also focused on discussing operational equivalence, which refers to the possibility of using a similar questionnaire format, instructions, mode of administration, measurement methods, and scoring (25). Three possible outcomes of assessing operational equivalence were adopted:

1. For the Serbian versions, the same methods (mode of administration, measurement methods, format, time frame, etc.) could be used as for the source questionnaire, with minimal levels of method bias in measurements with the questionnaire,

2. Some aspects of operationalisation needed to be different for the Serbian versions, with defined levels of method bias in measurements with the questionnaire, and

3. It was impossible to achieve operational equivalence.

For all questionnaires in the thesis, the first outcome was found when assessing operational equivalence. It was decided that in the case of second and third outcome the developer of a questionnaire should be contacted to resolve the issue. This is important because the format, instructions, and/or mode of administration of a questionnaire, can significantly influence the measurements of the questionnaire (59), what threatens the equivalence and introduces method bias.

Psychometric phase

Once the final Serbian versions were approved by the developers, the psychometric phase was organized. The aim of psychometric phase was to investigate psychometric equivalence or measurement equivalence as proposed according to Herdman (25). Specifically, it aimed to ensure that the Serbian language versions achieved acceptable levels in terms of their measurement properties. All final Serbian versions developed in the previous phases were field tested as the final phase of the operational model of the cross-cultural adaptation. The field-testing included the surveying of children and/or adolescents and/or parents with the questionnaire/s and the evaluation of the measurement properties using various statistical procedures. The field-testing of each questionnaire was mostly organized following the developers' recommendations or following the procedures considered for the original questionnaires.

Sample

Each study recruited samples that closely resembled those in the psychometric studies of the source questionnaire. Children and/or adolescents for the field-testing of three generic questionnaires, namely the KINDL, KIDSCREEN, and

PedsQL (Article 1, 2, 3, and 4) were sampled from the general population. For the two epilepsy specific questionnaires (Article 5 and 6), clinical populations of children and/or adolescents with epilepsy were recruited.

The sample sizes in the studies were determined according to the requirements for specific statistical procedures or following general recommendations for a sample size in psychometric studies with health measurement scales. Considering specific statistical procedures such as factor analysis, rules-of-thumb vary from four to 10 subjects per variable, with a minimum number of 100 subjects to ensure stability of the variance covariance matrix (95-97). In psychometric studies with health measurement scales, a sample size of at least 50 participants was suggested (57).

For details on the samples and sampling procedures, see the method sections of the original articles below.

Statistical procedures

The paramount aim of the psychometric phase is to demonstrate that measurement properties of the translation are not only acceptable, but also similar to the ones reported for the source language questionnaire (15, 25, 187). Only in this way, it is possible to use questionnaires for cross-cultural comparisons of HRQOL. Although it is undoubtedly important to achieve very similar or equivalent results, considering that the HRQOL construct is culturally sensitive, it might be unrealistic to expect similar results, and similar results might even indicate that the response patterns were confounded (25). Additionally, we are still lacking the clear definitions of the HRQOL construct in pediatric populations, and HRQOL questionnaires are only claiming to measure the construct. Therefore, the primarily interests of the psychometric phase of the Serbian versions were to demonstrate appropriate measurement properties analyzing how the target items and scales in the Serbian version "behave" together and to explore the underlying construct. Of the second interest was ensuring measurement equivalence and whether the questionnaire yield comparable results in cross-cultural studies (i.e. measurement invariance), which can only be explored if the questionnaire is tested

simultaneously across two or more languages/cultures (186), ideally involving samples from the original language/culture and the target language/culture, namely Serbian. Therefore, three possible outcomes of assessing psychometric equivalence and levels of bias were adopted:

1. Sound psychometric properties for the source and target language, as well as measurement invariance, were demonstrated for the HRQOL questionnaire, with minimal levels of construct and item bias in measurements with the questionnaire,
2. Sound psychometric properties for the source and target language were demonstrated, but the measurement invariance of the HRQOL questionnaire was not tested or achieved, possible levels of construct and item bias in measurements with the questionnaire, and
3. The target language questionnaire does not possess sound psychometric properties.

If one HRQOL questionnaire has sound psychometric properties in both languages, and the measurement invariance was demonstrated, then we can infer satisfactory equivalence, with minimal bias in its measurements, and the questionnaire can be used for cross-cultural comparisons. In the case of the second outcome, if the source and target language questionnaire have sound psychometric properties, but its measurement invariance was not tested or demonstrated, the questionnaire could be used only for in-culture HRQOL assessments and comparisons, but cross-cultural comparisons are biased. In the third situation, the target language questionnaire does not possess sound psychometric properties. Thus, it needs to be revised before use or to be discarded as culturally inappropriate, although this needs to be shown in several validation studies.

The measurement properties of the questionnaires were evaluated using various statistical procedures. The selection of statistical procedures that were used in the following articles had been mostly determined by the procedures used during the field testing of the original questionnaire and following general recommendations (22, 57-61). Considering that the psychometric evaluation is an

iterative process, the measurement properties are examined in two or more different samples using various psychometric procedures. Most of these procedures representing those that are based on *classical test theory*, while factor analysis and structural equation modeling were also considered as modern psychometric procedures.

All statistical procedures were considered as item level analyses, scale level analyses, or the entire questionnaire analyses (22), including descriptive item and scale statistics, reliability, validity, and sensitivity to change and responsiveness.

The statistical procedures used for the Serbian samples were described in details in the method sections of the original articles below. All analyses were performed using SPSS software (193).

Finalization phase

The finalization phase is the last step in the cross-cultural adaption process. In this phase, the target questionnaire with all possible versions (i.e. self- and proxy-versions for various ages if appropriate.) was issued, along with a manual that includes instructions for scoring and administration, as well as details of the field-testing. Nevertheless, the finalization phase does not imply that the cross-cultural adaptation process ends. The final remarks on the equivalence achievement according to Table 1 were also considered, as well as future directions on the use with further development of the Serbian version. This is important because of the following. First, the cross-cultural adaptation process can last for years to obtain enough information about equivalence, mostly due to longitudinal psychometric testing for measurement equivalence in repeated psychometric studies. Second, the target questionnaire might need revisions in order to improve its measurement properties. Third, the measurement data, structure, or even concept of the source language questionnaire might have changed over a period of its use, thus the target questionnaire might also need reworking.

Article 1 & 2 *KINDL questionnaire*

Article 1

Questionnaire

The KINDL is developed in three self-report versions (90): KINDL-Kiddy (4–7 year-olds), KINDL-Kid (8–12 year-olds), and KINDL-Kiddo (13–16 year-olds). Twenty-four items are classified into six sub-scales: Physical well-being (items no. 1–4), Emotional well-being (items no. 5–8), Self-esteem (items no. 9–12), Family (items no. 13–16), Friends (items no. 17–20), and School (items no. 20–24). Six additional items are in the 'Disease' module. The parent version represents an equivalent to the self-report. The items are five-Likert-scaled, 1 = never to 5 = all the time, while the sub-scales and total scores are formulated from items' means. The scores are transformed into a 0–100 scale, with the higher the value, the better the HRQOL is. The KINDL possesses satisfactory psychometric characteristics. Cronbach's alpha coefficients for the scales are from 0.63 to 0.84, the underlying theoretical model of the seven domains of importance was confirmed by factor analysis, and it possesses good discriminative validity as well. The questionnaires and the scoring procedures are provided in the KINDL Manual (98). The questionnaires were translated into the Serbian language and named as the KINDL-Kiddy-S, KINDL-Kid-S and KINDL-Kiddo-S. The translation was performed in several steps; forward and backward translations, cultural adaptation and harmonization. During the translation and cultural adaptation, none of the items was significantly changed; no items were deleted, no new items added. The entire process ensured a good feasibility and content validity to the versions.

Sample

The children and adolescents for the study were recruited from a public school in Belgrade. A school psychologist randomly contacted 300 subjects, asking them to participate and sending an informed consent to their parents along with the questionnaires. The subjects were informed that the purpose of the study was to evaluate the characteristics of the questionnaire offered, but not their health

status, and they were instructed how to complete it. Exclusion criteria were any reading or writing disability and the presence of some illness. Those who had accepted to participate and had returned the consent were included. Of all contacted, 94 returned the questionnaires without the consent or refused to participate later, 10 returned the questionnaires with many missing and inappropriately answered items, while 14 children were diseased at that time. The final samples consisted of 100 children (46 boys and 54 girls, 9.5–12.2 years), 92 adolescents (35 boys and 57 girls, 13–16 years), and 189 parents.

Statistical procedures

The analysis included descriptive and multivariable statistics (22, 62).

The distribution of missing data, mean (M), standard deviation (SD), skewness and kurtosis and floor and ceiling effects was calculated for each item, subscale and total. Based on these parameters, the following assumptions were assigned. The overall amount of missing data should indicate on the feasibility of the version and a significant amount of missing data for an item on a problem with translation, difficulty or irrelevance. Items means and SDs should be roughly equivalent within a sub-scale, considering that the originally hypothesized items in its sub-scales measure the same concept. If items' variances differ greatly within a sub-scale, the items would be standardized. Together with SD, skewness and kurtosis, means should report on the distributional characteristics. Finally, floor and ceiling effects should be below 10%, and those above 15% are considered importantly high.

The relationships between the items, sub-scales and the total score were analyzed by Spearman's rho coefficient, while Cronbach's alpha coefficient (a coefficient) was used to explore the reliability. The following assumptions were made. First, item internal consistency (convergent validity) assumed that an item should be substantially correlated to the underlying concept measured, with the correlations between the score for each item and its sub-scale total above 0.4.

Second, equality of item–sub-scale correlations assumed that items in a sub-scale should contribute roughly equal proportions of information to the total sub-scale score, and the correlations between 0.4 and 0.7 should be considered important.

Third, item discriminant validity assumed that an item measured other concepts that it was not supposed to measure, with the score more correlated with the other scores than the score of its sub-scale.

Forth, internal consistency reliability tested the extent to which the sub-scales and the total score were consistent and free of measurement error, and a coefficients should be at least 0.6.

Fifth, inter-sub-scale correlations were evaluated in relations to the reliability to explore the variance measured by each sub-scale as a parameter of measuring a distinct concept. The KINDL-S parent versions were analyzed using descriptive statistics, Spearman's rho correlation (ρ) coefficient for analyzing child–parent relationships and Mann–Whitney U-test to compare the differences in the HRQOL valuing between the children and parents. The analyses were performed on child–parent pairs.

Article 2

Sample

School psychologists contacted 800 pupils (aged 8-16 years and equally boys and girls) from nine elementary schools in Western Vojvodina to participate in the study. They informed all children and adolescents about the purpose of the study, as well as their parents and teachers. Those agreed to participate and returned the written consent from the parents completed the questionnaire in the schools to prevent a low responding rate. The participants were instructed carefully how to fill the KINDL out. One hundred and twenty randomly selected pupils completed the questionnaires after a seven-day period. The data from healthy subjects were

used for the present analysis and those with major psychological or physical chronic diseases or acutely diseased were not considered relevant. As in the previous article, only health subjects were included, assuming to develop a questionnaire with appropriate measurement properties for HRQOL assessments in healthy populations. The data about the subjects' health were taken from medical records available in schools.

Statistical procedures

The distribution of missing data was calculated as the percentage of missing responses on all possible responses. Only subscales with less than 30% of missing items were considered, whereby mean value replacement dealt with such missing values. For each item, subscale, and total M and SD were calculated. Reproducibility, test-retest reliability, concerns the degree to which repeated assessments in stable persons produce similar responses (58). It was evaluated using the intarclass correlation coefficient - ICC, the two-way random method of absolute agreement (58). Assuming reliability is the degree to which people can be distinguished from each other, the KINDL's ICCs should be 0.6 or higher for healthy group comparisons. The retest took place seven days later. Construct validity was assessed using factor analysis that combines observable variables into unobservable, latent variables, giving insights into the theoretical model of some construct (58, 99). This is known as factorial validity that is assessed using EFA and/or CFA. The present study gave priority to CFA, whereas we already have the hypothesized theoretical model of the KINDL assuming to be confirmed as valid for HRQOL assessments and it is not necessary to re-explore the latent variables using EFA. Moreover, the current perspectives are to use CFA in HRQOL models, whereas EFA could produce strange combinations of HRQOL items with unexpected latent constructs (58). This is mainly because HRQOL questionnaires often combine items with a causal relationship with the latent variables, causal variables, and items dependent upon the latent variables, indicator variables, while EFA requires only the later (58, 59, 100). Finally, CFA provides some data on convergent (the extent to which similar theoretical constructs are related) and discriminant validity (the extent to which different theoretical constructs are relatively

unrelated) as the aspects of construct validity (99). A CFA was conducted using Analysis of Moment Structures Version 7 (AMOS-7) on a model representing the items and the corresponding factors as originally assumed. Therefore, the tested model, as a second order CFA model, had three levels: items (24), primary factors (six subscales), and one secondary factor (HRQOL). The primary goal is to determine the goodness of fit between the hypothesized model and the sample data. To test the hypothesized model the variance-covariance matrix was used and maximum likelihood (ML) estimation was employed. ML is robust in terms of using non-continuous data and there is evidence of robustness in the terms of the violation of multivariate normality assumption (101, 102). However, Bollen-Stine bootstrap and associated test of overall model fit were used to study and manage the effects non-normality in the underlying database since research has also demonstrated that ML test statistic (TML) and ML parameter standard errors may be affected when data deviate from normal (101, 102). Bollen-Stine bootstrap provides more realistic standard errors if there is serious departure from multivariate normality. Based on the recommendations, 2,000 bootstrap samples were drawn to obtain overall model fit and 250 bootstrap samples to obtain parameter estimates and associated standard errors (101). Model identification was established by estimating the factor variances and fixing one factor loading to 1.0 for each factor. The following statistics assessed the adequacy of the model, indirectly construct validity, as the degree of fit between estimated and observed variance: chi square, Tucker Lewis Index (TLI) (>0.90 acceptable, >0.95 excellent), the Comparative Fit Index (CFI) (>0.90 acceptable, >0.95 excellent), and root mean square error of approximation (RMSEA) (<0.08 acceptable, <0.05 excellent) (58, 99, 101-103). It was assumed the factor loadings of the items within the subscale and the standardized coefficient of the subscales should be at least moderate to support convergent validity, while the correlations between the estimated parameters of the latent factors should be low to support discriminant validity (58, 99, 102).

Article 3 *KIDSCREEN questionnaire*

Questionnaires

The content of the KIDSCREEN questionnaire was generated through literature reviews, expert consultations using the Delphi method and focus groups with children (24). Three KIDSCREEN instruments are available in child/adolescent and parent/proxy versions: KIDSCREEN-52, KIDSCREEN-27, and the KIDSCREEN-10 Index. The KIDSCREEN-52 (long version) and KIDSCREEN-27 (short version) allows detailed profile information for ten and five HRQOL dimensions respectfully, while the KIDSCREEN-10 Index gives a global HRQOL score for monitoring and screening purposes.

The KIDSCREEN-52 has ten dimensions (scales): Physical Well-Being (5 items), Psychological Well-Being (6 items), Moods & Emotions (7 items), Self-Perception (5 items), Autonomy (5 items), Relations with Parents & Home Life (6 items), Social Support & Peers (6 items), School Environment (6 items), Social Acceptance (Bullying) (3 items), and Financial Resources (3 items). The KIDSCREEN-27 is embedded within the KIDSCREEN-52 and consists of 27 items in five scales: Physical Well-Being (5 same items as in the longer version), Psychological Well-Being (7 items from the Psychological Well-Being, Moods & Emotions, and Self-Perception original scale), Parent Relations & Autonomy (7 items from the Autonomy, Relations with Parents & Home Life, and Financial Resources original scale), Social Support & Peers (4 items from the longer version), and School Environment (4 items from the longer version). The KIDSCREEN - 10 Index represents a global HRQOL score including 10 items from the short version that addresses affective symptoms of depressed mood, cognitive symptoms of disturbed concentration, psycho-vegetative aspects of vitality, energy and feeling well, and psychosocial aspects correlated with mental health, such as the ability to experience fun with friends or getting along well at school. All versions have sound psychometric qualities (3, 24, 104-110).

All KIDSCREEN items use the same 5-point Likert-type response scales to assess the frequency (never, seldom, sometimes, often and always) of certain behaviors and feelings and the intensity of an attitude (not at all, slightly, moderately, very, and extremely). Rasch scores were computed for each dimension and transformed into T-values with a mean of 50 and a standard deviation of 10. Higher scores indicate better HRQOL. Details how the scores were computed were given in the manual (24).

The Serbian KIDSCREEN set was developed in close collaboration with the KIDSCREEN Group following their procedures for translation and cross-cultural adaptation (24). Briefly, the process included two forward translations (English into Serbian), reconciliation of the translations and development of one version, backward translation, followed by the review of the forward and backward translation, the assessment of conceptual equivalence, and a pre-test (cognitive interviews). During the translation process, the Serbian versions were developed with no items added, replaced or omitted, while important translational strategies applied were semantic rearrangements, supplementations to items, and substitutions of certain words with synonyms. All items were considered short, precise, clear and equivalent to the originals (conceptual, item, semantic, and operational equivalence). The KIDSCREEN Group approved the Serbian version. In the rest of the text, the term "KIDSCREEN" implies on the Serbian version, if not specified differently.

The KINDL is a self- and parent-rated HRQOL instrument developed for children aged 4–18 years (90). Twenty-four items are assigned to six different scales: Physical well-being (4 items), Emotional well-being (4 items), Self-esteem (4 items), Family (4 items), Friends (4 items), and School (4 items). The parent version is conceptually equivalent to the self-report. All items use 5-point Likert-type scales to assess the frequency (never to all the time), while the scales and total scores are calculated are mean scores of the relevant items. The scores are transformed into a 0–100 scale (higher value represents better HRQOL).

Sample

Children and adolescents (hereinafter referred to as children) from five elementary and three high public schools in Serbia and their parents participated. The inability to read and write Serbian served as the only exclusion criterion. School psychologists contacted 500 pupils (aged 8-18 years) informing them about the purpose of the study. The same psychologists contacted the parents for participation on parent's meetings, via e-mails, phone, or fax. After the parents sent the signed informed consent forms, the children completed the KIDSCREEN-52 and KINDL questionnaire assembled in a random order during scheduled time in classrooms. All children were instructed how to complete the questionnaires and the completion was supervised to ensure that all questionnaires were completed and returned. Afterwards, all children were given a sealed envelope with a similar set of questionnaires to be filled up by the parents or significant others (proxies) who gave their consents at home without consulting their children and afterward returned them to the psychologists. The Ethics Committee of the Clinic for Neurology and Psychiatry for Children and Youth Belgrade approved the study.

In total, 330 children (66% reporting rate) and 314 parents (62.8% reporting rate) participated in the study. All participants were Caucasians. In the children sample, 153 (46.4%) were males and 177 (53.6%) females. Of these, 129 (39.1%) were children aged 8 –12 years (M = 9.86, SD = 0.91) and 201 (60.9%) adolescents aged 13 – 18 years (M = 14.34, SD = 1.92). According to the answers provided by the children and parents on the questionnaires, sixteen participants (4.8%) had a long-term disability, illness or medical condition (asthma, arrhythmia, atopic dermatitis, migraine, physical disability of the knee). Among the proxies, 262 (83.4%) were mothers, 47 (15%) fathers, and five (1.6%) significant-others.

Statistical procedures

The same set of analyses was performed to the child and parent versions following the procedures from the previous KIDSCREEN studies in order to facilitate the interpretation of the results.

Scale descriptions. The distribution of mean T-value, SD, percentage of missing cases, and floor and ceiling effects were calculated for all scales. A floor and ceiling effect was defined as the percentage of individuals with the best and worst results respectfully. More than 15% of participants with the highest or lowest score on one particular scale were considered as a relevant effect (62).

Internal consistency reliability was evaluated using Cronbach's α coefficient and the values of $\alpha \geq 0.7$ were considered acceptable (62).

Criterion validity was assessed by determination of the degree of correlation between the KIDSCREEN-27 scales and the KIDSCREEN-52 scales assessing similar dimensions and between the KIDSCREEN-10 index and scales and the KIDSCREEN-27. Coefficients exceeding $r = 0.7$ were considered satisfactory (106). How well the short versions replicate the longer ones were evaluated by examining the proportion of variance in each score explained by the corresponding dimension using regression coefficients.

Convergent and discriminant validity was evaluated in two ways. The first one, Pearson correlation coefficients between the KIDSCREEN and KINDL scales were calculated. Convergent validity was considered to be demonstrated when correlations between comparable dimensions exceeded correlations between theoretically different dimensions. Correlation coefficients ranging 0.1–0.3 were considered low, those 0.31–0.5 moderate, and those exceeding 0.5 high (107). Considering that at the time of the study initiation no other HRQOL questionnaire was available, the second method to evaluate convergent/discriminant validity included the multitrait-multimethod matrix (MTMM) analysis of children-parent correlations (Pearson correlation coefficients) between the KIDSCREEN scales (18). It was assumed that a correlation between the same dimensions needed to be higher (monotrait-heteromethod) than a correlation between different dimensions (heterotrait-hetero method) (24).

Levels of agreement between children and parents in reporting HRQOL was assessed using the ICC. The agreement assessed with the ICC was interpreted as follows: <0.4, poor to fair; 0.41– 0.6, moderate; and >0.61, good to excellent (109).

To evaluate the magnitude of discrepancies between children and proxy ratings paired t-test of the mean differences (Δ) and unbiased Cohen's effect size corrected for sample size (d) were computed (111). A positive Δ or d indicated a higher level of HRQOL reported by proxies. Effect size should be interpreted as: 0.2 small, 0.5 moderate, and 0.8 large under- or overestimation by parents (111). A domain is overestimated (respectively, underestimated) by proxies, if d is greater than +0.2 (respectively, smaller than - 0.2) (111).

Article 4 *PedsQL questionnaire*

Questionnaires

The PedsQL includes (1) Physical Functioning – PF (8 items), (2) Emotional Functioning – EF (5 items), (3) Social Functioning – SoF (5 items), and (4) School Functioning – ScF (5 items) Scale (91). Within the scales, all items are in a 5-point response scale (0 = never a problem; 1 = almost never a problem; 2 = sometimes a problem; 3 = often a problem; 4 = almost always a problem), all are reverse-scored, and linearly transformed to a 0–100 scale. Higher scores indicate better QOL. Three scale scores are computed. The Physical Health Summary Score is computed as the sum of the items divided by the number of items answered in the Physical Functioning Scale. To create the Psychosocial Health Summary Score (15 items), the mean is computed as the sum of the items divided by the number of items answered in the Emotional, Social, and School Functioning Scales. Finally, the total PedsQL score was computed as the sum of all items divided by the number of items answered. If more than 50% of the items in the scale are missing, a score is not computed. The Serbian version was provided by MAPI Research Trust that translated and cross-cultural validated the version according to their standardized procedures. The permission was also obtained from its developer, James W. Varni, PhD.

The Strengths and Difficulties Questionnaire (SDQ) was administered to assess general mental health problems (112). The SDQ possesses 25 items

comprising the following five-item scales: emotional and conduct problems, hyperactivity/inattention, peer relationship problems, and pro-social behavior. Each item is scored on a 3-point scale and the sum of all answered items in a scale creates its total score. Higher scores indicate larger difficulties.

Sample

School psychologists contacted nearly 300 pupils (aged 8-18 years, equally boys and girls) from four elementary schools in Serbia to participate in the study. They informed all children and adolescents about the purpose of the study, as well as their parents. The only exclusion criterion was inability to read and write the Serbian language. Of all contacted, 238 agreed to participate and returned the written consent from their parents. The children completed the questionnaire at school in order to prevent a low responding rate.

Of 238 subjects, 107 (45%) were boys and 131 (55%) girls. The mean age of the sample was 12.76 (2.17), with 118 (49.6%) children aged 8-12 years and 120 (50.4%) adolescents aged 13-18 years.

Statistical procedures

The distribution of missing data, M, SD, skewness and kurtosis, and floor and ceiling effects were calculated.

Scale internal consistency reliability was determined by calculating Cronbach's coefficient (α) (22, 58). Scales with reliabilities of 0.70 or greater are recommended for comparing groups, while a reliability criterion of 0.90 is recommended for analyzing individual patient scale scores (57).

Confirmative factor analysis was used to analyze the factor structure of the PedsQL using AMOS-7. Four models were tested: a model representing 23 items and the corresponding five factors as originally suggested (113), a second-order four-factor model representing four scales (114), the PedsQL Physical Health model (one factor represented by 8 items), and PedsQL Psychosocial Health model (a second-order model derived from the emotional, social, and school scale). The

following statistics assessed the adequacy of the models, as the degree of fit between estimated and observed variance: chi square test, Tucker Lewis Index (TLI) (>0.90 acceptable, >0.95 excellent), the Comparative Fit Index (CFI) (>0.90 acceptable, >0.95 excellent), and Root Mean Square Error of Approximation (RMSEA) (<0.08 acceptable, <0.05 excellent) (101, 103).

Convergent validity was assessed through correlations between the PedsQL and SDQ scales. It was hypothesized the PedsQL total as well as the PedsQL Psychosocial Health scales would show negative correlations with the SDQ emotional and conduct problems, hyperactivity/inattention, and peer relationship problems scale and positive correlations with the pro-social scale. Furthermore, the PedsQL Physical Health scale was expected to show a comparable weaker correlation with all SDQ scales. Correlations were estimated by use of Spearman's rho statistics.

Article 5 *CHEQOL-25 questionnaire*

Questionnaire

The CHEQOL-25 was developed in two forms: a self-report for children and a proxy measure for parents or caregivers (93). It contains five subscales with five items in each: (1) Interpersonal/Social, (2) Intrapersonal/Emotional, (3) Present Worries/Concerns, (4) Secrecy, (5) Quest for Normality (child only) and Future Worries and Concerns (parent only). This measure uses the alternative format of paired options of forced responses to avoid neutral or extreme scoring, avoids overt positive and negative phrasing of the items and eliminates emotional burden that may result from negative formulation (115). Each item is scored on a scale of 1-4 and the sum of all items of the subscale forms its total score (scores ranging between 5-20). A higher score reflects more positive perceptions in that domain. The psychometric study showed that the CHEQOL-25 produces scores that are reliable (the reliability coefficients ranging 0.64-0.86 for internal consistency and

0.60-0.81 for reproducibility) and valid (with appropriate face, content, and construct validity) (93).

Translation and cultural adaptation. The Serbian translation and cultural adaptation of the CHEQOL-25 were performed according to the ISPOR guidelines that included preparation, forward translation and reconciliation, back translation, harmonization, cognitive debriefing, and finalization (50). Using these principals, it was aimed to develop a version that is equivalent to the original (item, semantic, operational, and measurement equivalence) and culturally sensitive, too (25). Researchers who are working with HRQOL performed two independent forward translations. From these versions, a single form was developed (reconciliation) that was back translated into English by two independent English-Serbian translators. One back translation was developed and it was compared with the original (harmonization). The purpose was to highlight all differences that arose during the translation and to overcome the discrepancies in the concepts between the original and the translation. The entire processes resulted in a version named the CHEQOL-Srb that was pretested in semi-structure interviews with a group of ten children with epilepsy. It was aimed at exploring comprehensibility, judgment, and response process of each item, as well as clarity and appropriateness (cognitive debriefing). The children were requested to "think aloud" about the items and how they formed an answer, to try to explain the meaning of each item, and to give an opinion whether all items are clear and appropriate (unambiguous and free of wording that might adversely affect a child) (50). Finally, an expert panel meeting was organized to evaluate the content and face validity of the version, the results of the pretesting, and the equivalence with the original (conceptual, item, semantic, etc.) (25).

Sample

Ten children with epilepsy were included in the translation and cultural adaptation process. For psychometric validation, 50 children with epilepsy aged 8-12 years were recruited as the minimal number of subjects required for this psychometric analysis (57). Inclusion criteria were: (1) active epilepsy (i.e., at least

one seizure in the last 2 years); (2) epilepsy duration greater than 2 years; (3) regular school attendance; and (4) ability to read and write Serbian. Children who had comorbid psychiatric or other chronic disorders, who had additional neurological impairments or disabilities, or who were hospitalized during the last year were not eligible. All subjects participated during regular clinic visits and all provided their informed consent prior to the inclusion. The ethics committee of the principal author's institution (DS) approved the study.

Statistical procedures

Descriptive and inferential analyses were performed to explore the distributional characteristics of the items and the relationship between them (22, 57, 59).

Descriptive statistics included the distribution of missing data, M, SD, skewness and kurtosis, and floor and ceiling effects calculated for each item and subscale.

The relationship between the items was explored using Spearman's rho coefficient and it was assumed that an item should be substantially correlated to the underlying concept measured (i.e., an item should more significantly correlate with the total of the subscale from which it originated than with the total of the other subscales). The internal consistency of the subscales was assessed with Cronbach's alpha coefficient. Finally, the level of agreement between parent reports and child self-reports was determined using the ICC as a more appropriate parameter than the statistical significance of correlation between the raters (116).

Article 6 *QOLIE-AD 48 questionnaire*

Questionnaires

The QOLIE-AD-48 is a disease-specific instrument for the assessment of HRQOL in adolescents with epilepsy (94). It contains 48 (and 3 optional) items on eight subscales: Epilepsy Impact (12 items), Memory/Concentration (10 items), Attitudes toward Epilepsy (4 items), Physical Functioning (5 items), Stigma (6

items), Social Support (4 items), School Behavior (4 items), and Health Perceptions (3 items). The subscale scores, as well as an overall score, range from 0 to 100; the higher the score, the better HRQOL is. It is a sensitive, validated, and test–retested (r = 0.83) instrument, easy to administer and to score. The average time for completion is 15–20 minutes. The QOLIE-AD-48 could provide information about a variety of issues pertinent to adolescents with epilepsy, foremost HRQOL.

Translation, cultural adaptation, and pre-testing. The paramount aim was to confirm that the Serbian version of the QOLIE-AD-48 refers to a concept of HRQOL in adolescents with epilepsy comparable to that of the original version, so the cultural adaptation consisted of gaining a full understanding of the conceptual structure underlying the original and its complete transposition into Serbian (23, 25, 76, 117, 118). With respect to the postulate that the translation and cultural adaptation of the QOLIE-AD-48 must ensure item (''the way in which domains are sampled), semantic (''the transfer of meaning across language''), and operational (''a similar questionnaire format, instructions, mode of administration, and measurement methods'') (25) equivalence, the entire process required systematic work, briefly discussed here.

Step 1: forward translation. Initially two English–Serbian translations were made independently. Then, by their combination, test version 1.0 of the Serbian QOLIE-AD-48 was created.

Step 2: face validation. Five adolescents with epilepsy were asked to complete this version.

Step 3: panel session. This step aimed to identify inadequate expressions/concepts of the translation, to check item compatibility, to complete the acceptability, and to denote the ''stumbling blocks'' to further adaptation.

Step 4: back translation. Version 1.0 was back translated into English. This process of validity checking ensured that the translation reflected the same item content as the original version. In addition, harmonization with the Croatian

translation was performed, for translation of culturally sensitive items (23). An intermediate version was completed by another five adolescents with epilepsy.

Step 5: pre- testing. Sixteen adolescents with epilepsy, ranging widely in age from 10 to 20 years and with different seizure types, were randomly divided into two equal groups. For the groups, the combination of a semi-structured interview and questionnaire was conducted in a single session. During the interview, for every item and response, subjects were asked to explain their opinion on item precision, clarity, effectiveness, and appropriateness. During completion of the questionnaire, subjects were asked to identify confusing, unclear, imprecise, non-relevant, or inappropriate items. Finally, all subjects subjectively assessed presentation and relevance of the questionnaire on a scale from 1 (negative) to 5 (positive), and were asked to write comments on the questionnaire.

Step 6: panel session II. Considering the results of the pilot testing, version 2.0 was developed for psychometric evaluation.

Illness Severity Index (ISI) was used for the evaluation of seizure severity, and it comprised seizure frequency, seizure type, and antiepileptic drugs (AEDs) (119). For formulation of the ISI, the same method was used as in the study of Raty et al. (120).

Sample

Twenty-six adolescents with epilepsy were included in the process of translation and adaptation. For psychometric validation, 67 youths with epilepsy aged 10–19 years (36 males, 31 females) were recruited. Inclusion criteria for study participation were: (1) active epilepsy (i.e., at least one seizure in the last 2 years); (2) epilepsy duration greater than 2 years; (3) regular school attendance; and (4) ability to read Serbian. Young people who had co-morbid psychiatric or other chronic illnesses, who had additional neurological impairments or handicaps, and who were hospitalized during the last year were not included in the study. All subjects were asked to participate in the study during regular clinic visits.

Statistical analysis

Serbian version 2.0 comprised eight Likert-type subscales, as well as the original, and the summary and subscale scores were calculated following the suggestions in the scoring manual (94). Psychometric evaluation included descriptive statistics and multitrait analysis (25, 76, 94, 117, 121). The distribution of missing data, M and SD for responses to each question, and means and SDs for subscale scores were calculated. The percentage of responses on extreme points was examined for each subscale to detect floor or ceiling effects.

It was hypothesized that the response to each item should be linearly correlated with the score of the subscale to which it belongs - convergent validity (item internal consistency). For this purpose, the correlation (Pearson's correlation) between the score for each item and the total score for the subscale to which it belonged should be moderate to large ($p \geq 0.4$).

The score for each item should be significantly more correlated ($P < 0.05$) with the score of its hypothesized subscale than with the scores of other subscales - discriminant validity (item discriminant validity).

Construct validity. First, all subscale scores should be correlated to the total summary score (Pearson's correlation); second, there should be an inverse relation between HRQOL and the Illness Severity Index (ANOVA). The first hypothesis testing included a principal-axis factor analysis using an oblique rotation.

Cronbach's α was calculated to estimate internal consistency. Satisfactory a coefficient should exceed 0.7 for all subscales. We performed a split-half method to obtain a reliability coefficient for version 2.0. The split-half technique involves testing for correlation between two halves of the scale (121). In the first half were included items 1–6 and 8–25; item 7 was added to the second.

A paired t test was used to evaluate the relation of the summary score to the Illness Severity Index - sensitivity.

Article 7 *Q-LES-Q-SF questionnaire*

Questionnaire

The Q-LES-Q – SF is a self-report questionnaire, with 16 items, derived from the general activities scale of the original 93-item form (92). It evaluates overall enjoyment and satisfaction with physical health, mood, work, household and leisure activities, social and family relationships, daily functioning, sexual life, economic status, overall well-being, and medications. Responses are scored on a 5-point scale ('not at all or never' to 'frequently or all the time'), where higher scores indicate better enjoyment and satisfaction with life (possible range 14-70). Fourteen summated items create the total Q-LES-Q – SF score. Two last items, about medications and overall life satisfaction, are considered independently.

The MAPI Research Trust Institute developed the Serbian version provided to the author by the developer J. Endicott, PhD. The version was pre-tested in a small group of psychiatric patients prior this study to test the cultural appropriateness and content validity (58).

Sample

All adults admitted to daily hospital for some psychiatric treatment between December 2008 and January 2009 were eligible. Exclusion criteria were the presence of any other major medical problem (e.g. chronic illness, impairment), inability to read or write, and living in an institution. All patients were diagnosed according to the International Classification of Diseases (ICD – 10) (122) and to all was initiated some kind of treatment, medications, social therapy and/or psychotherapy.

Fifty-seven subjects who accepted to participate and provided a written consent were assessed independently (Table 2). However, each analysis below included only subjects who provided all data.

On the admission, each participant was evaluated on the Clinical Global Impression severity scale (CGIs) rated by a clinician and the Patient-reported Global

Impression severity scale (PGIs) rated by a participant, both as a 7-point-scale, from1 – 'extremely ill' to 7 – 'not ill at all' (92). The Q-LES-Q – SF was completed afterwards.

On the second appointment, seven days later, the subjects were evaluated again and those who remained stable over this period completed the Q-LES-Q – SF. A "stable" subject is one whose health status has not changed in any domain since the previous assessment.

Table 2 Demographic characteristics of 57 subjects

Age, years (M, SD)	47.16, 9.22
Gender, N (%)	38 (66.7) M, 19 (33.3) F
Marital status, N (%)	25 (43.87) married/live with a partner, 22 (38.59) never married, 10 (17.54) separated/divorced/ a partner died
Educational level, N (%)	11 (19.29) primary school, 39 (68.43) secondary school, 7 (12.28) high school/university degree
Occupational status, N (%)	21 (36.84) full or part-time work, 16 (28.07) retired, 19 (33.33) not employed, 1 (1.75) college student
Psychiatric diagnosis according to the ICD-10 categories (WHO, 1993), N (%)	20 (35.08) Schizophrenia, schizotypal, and delusional disorders (F20-F29), 16 (28.07) Mood disorders (F30 - F39), 11 (19.29) Anxiety, stress-related, and somatoform disorders F40-F48, 10 (17.56) Disorders of adult personality and behavior (F60-F62)

At the follow-up assessment, four weeks later, beside the CGIs, PGIs and the Q-LES-Q – SF, the Clinical Global Impression improvement scale (GCIi) was completed by a clinician, from 0 – state unchanged to 6 – ideal improvement (92). The three indexes, CGIs, PGIs, and GCIi, were the external criteria of change (anchors) and each should significantly correlate with the Q-LES-Q – SF (58). Thus, any change in the Q-LES-Q – SF would be observed in the external criteria of change.

After the final assessment, all participants were assigned to two groups. The "unchanged" group (n = 22) included subjects without changes in mental health status from the first assessment and the "changed" group (n = 14) included those subjects who improved over the study period (the improvement was measured as change in scores in two or more of the anchors used). During the follow-up period, two subjects worsened significantly, four were acutely diseased (a viral infection), five were discharged earlier (due to financial problems), and 10 refused to participate this time.

After the approval from the Psychiatric Board about the involvement of all patients in the study activities, the Ethics Comity of the author's institution (General Hospital Sombor) approved the study.

Statistical procedures

The number and distribution of missing data were examined to assess the acceptability of the questionnaire assuming that this value should not be more than one third of unanswered items for an individual (57, 58). The distribution of responses for each item was assessed visually. Further, the mean, standard deviation, skewness, and kurtosis for responses to every item were calculated (22). Finally, the percentage of responses on anchor points was examined to detect floor or ceiling effects, which should not exceed 15% (57).

Three assumptions were considered for validity. All items should be correlated minimally 0.4 to the total score corrected for overlap using Spearman's correlation. These correlations should not be substantially different between the items, indicating that the amount of information to the total concept being measured is similar for each item (22). Finally, the Q-LES-Q – SF should be significantly correlated with the CGIs, PGIs, and GCIi ($p < 0.05$, Spearman's correlation) as the means of criterion validity.

It was assumed that the questionnaire should have appropriate internally consistency reliability and it should be stable in stable subjects (test-retest reliability, reproducibility), it should be sensitive to detect differences in HRQOL

between different people (sensitivity), and it should be able to detect HRQOL changes in unstable subjects (responsiveness) (58).

Internal consistency reliability was assessed by Cronbach's alpha and it should exceed 0.7. Test-retest reliability was analyzed using the intraclass correlation coefficient (ICC, the two-way random method of absolute agreement) to derive the reliability coefficient (59). The re-test took place seven days later considering the 1-week timeframe of the Q-LES-Q – SF. Test-retest reliability should be at least 0.9 in order to evaluate the ability of the questionnaire in detecting changes important for individual comparisons (58). Eventually, the responsiveness was evaluated as the following. First, the standard error of measurement (SEM) was derived from the standard deviation of the sample, from which was assessed the reliability, multiplied by square root of (1 – the reliability) (123, 124). For the SEM, 90% confidence intervals (90% CI) were calculated around individual patient scores, reflecting the questionnaire's accuracy for individual assessments and clinical decision-making (125). Further, the SEM was converted into the smallest detectable change (SDC) reflecting the smallest within-individual change in score that could be interpreted as a "real" change, above measurement error in one individual (57). The SDC was calculated by multiplying the z-score corresponding to the level of significance, the square root of 2, and the SEM. A z-score of 1.64 was chosen to reflect an acceptable 90% CI for clinical application to individual patients (Schmitt and Di Fabio, 2004). Finally, the minimal important difference (MID; also known as minimal clinically important difference), defined as the smallest difference in score perceived as beneficial, was determined using mean change scores for patients with small but meaningful change according to some external criteria (126).

Change over the 4-week interval was analyzed using three responsiveness indices at the individual patient level. Only data from all followed-up patients were used (126, 127). Reliable change proportion (RCP) – the proportion of the sample with change scores exceeding the SDC, which is also sensitivity to change; MID proportion – the proportion of the sample with change scores exceeding the MID,

and specificity to change – the proportion of those who claimed perceiving no change and their scores did not exceed the SDC.

Article 8

Questionnaire

HRQOL was measured with the Serbian version of the Quality of Life in Epilepsy Inventory for Adolescents 48 (QOLIE-AD-48) (reported in Article 6.), which incorporates a majority of domains important to the health of adolescents with epilepsy and is the most suitable measure for youth with uncomplicated epilepsy (128). The QOLIE-AD-48 is a disease-specific instrument assessing 48 items in eight domains of HRQOL: Health Perception, Epilepsy Impact, Memory/Concentration, Physical Functioning, Stigma, Social Support, School Behavior, and Attitudes toward Epilepsy. The raw (5-point Likert scale) total and domain scores were translated into scores on a 0–100 response scale, with higher scores indicating better HRQOL. In addition, the level of concern over seizures recurring, epilepsy concern (''Are you worried about having seizures again?''), was measured on a 0–100 response scale, where higher scores indicated less concern.

Sample

Participants were recruited from the Mother and Child Health Care Institute of Serbia and the Clinic for Neurology and Psychiatry for Children and Youth, Belgrade. To be included, participants had to have had active, uncomplicated epilepsy for more than 5 years, with well-controlled seizures, that is, no seizures for more than 1 year. Participants who had other neurological and psychiatric disorders, major neuropsychological deficits, or any other chronic condition; who had undergone significant therapy or EEG changes; who had been hospitalized or institutionalized for any reason or lived away from their parents; and who had had a failing grade or dropped out of school altogether were excluded. During the last trimester of 2005, 150 adolescents met the criteria. Only those who agreed to

participate and who obtained parental consent and completed the instrument properly were included. They numbered 71 (Table 3).

Table 3 Major demographics of the sample (N = 71)

Gender	54.9% boys (39), 45.1% girls (32)
Age (yr)	M = 15.10, SD = 2.15, range 11.5–18
School grade	M = 8.12, SD = 2.47, range 5–12
School achievement (aver. mark)	M = 3.71, SD = 0.94
Seizure type, % (n)	
Simple partial	17.91% (13)
Complex partial	20.89% (15)
Generalized tonic–clonic, secondary generalized	13.43% (9)
Absence	17.91% (13)
Myoclonic	29.86% (21)
Duration of epilepsy (yr)	M = 11.5 SD = 5.12, range 5–18
Age at onset	M = 5.16 SD = 3.85, range 0–13
Antiepileptic drugs (AEDs)	
None	40.8% (29)
Monotherapy	46.5% (33)
Di-or polytherapy	12.7% (9)

Statistical procedures

First, HRQOL total and domain scores were analyzed descriptively. Ms and SDs with confidence intervals (CIs) and percentiles for both genders were calculated. Second, multivariable analysis involved t tests, Pearson's correlation coefficients, and linear regression analysis. The t tests were used to analyze the differences between the HRQOL scores in this study and compare them with those reported for adolescents without seizures during the preceding year in Cramer et al. (94). Pearson correlations evaluated relationships between the total and domain scores. Stepwise linear regression analysis was performed to test the influence of gender, age, school achievement, number of AEDs taken, and epilepsy concern on HRQOL. In the analysis, total and domain scores on the QOLIE-AD-48 were dependent variables, and every variable that significantly deviated from the normal distribution was logarithmically transformed.

Summary of Results

Article 1 & 2 *KINDL questionnaire*

The KINDL-Kid-S

The amount of missing data was 0.001% (Table 4). The M and SD values of items in the sub-scales were roughly equivalent, ranging 2.56–4.77 and 0.48–1.51, respectfully. However, items no 9, 18 and 24 in the Friends and School sub-scale differed greatly from the others in its sub-scale, so the rest of the analysis was performed on the standardized items and the scores transformed to a 0–100 scale. The sub-scale means were from 74.94 for the Self-esteem to 88.19 for the Family sub-scale, with the total 78.84. The scores were negatively skewed and more leptokurtic. Except for the total score, the SD of each sub-scale was above 10. Ceiling effects above 15% were observed for the Physical and Emotional well-being, Self-esteem, and Family sub-scale, while no floor effects were reported.

Table 4 Descriptive statistics of the KINDL-Kid-S (N = 100)

Sub-scale	**Row score range (M)**	**Transformed score M (SD)**	**Skewness**	**Kurtosis**	**Floor (%)**	**Ceiling (%)**
Physical well-being	3.86 – 4.48	80.44 (15.96)	- 0.9	0.5	0	16
Emotional well-being	4.17 – 4.76	87 (12.43)	-1.61	3.86	0	22
Self-esteem	3.52 – 4.33	74.94 (16.35)	- 0.42	0.3	0	12
Family	4.09 – 4.77	88.19 (11.34)	-1.18	1.71	0	27
Friends	3.43 – 4.6	79.31 (13.78)	-1.38	3.37	0	5
School	2.56 – 4.09	63.18 (19.29)	- 0.01	- 0.69	0	5
Total	/	78.84 (9.48)	- 0.52	0.19	0	1

Item internal consistency. All correlations between an item in a sub-scale and the total sub-scale score of the others in the same sub-scale were above 0.4. However, item no. 17 was also substantially correlated with the Self-esteem sub-scale (Table 5).

Equality of item – sub-scale correlations. The item–sub-scale correlations differed substantially between the items within the proposed sub-scales (Table 5).

Item discriminant validity. The score for each item was significantly more correlated with its sub-scale score than with the others (Table 5). However, the majority of the items showed low to moderate correlation with the other sub-scales.

Internal consistency reliability. The a coefficient for the summary was 0.81 (Table 5). The Physical and Emotional well-being and Self-esteem sub-scale had the a coefficients above 0.6.

Inter-sub-scale correlations. The sub-scales showed significant correlations in between and moderate to high with the total score. The correlation coefficient between two sub-scales was less than their reliability coefficients were, for all possible correlations (details not given).

Table 5 Scaling properties of the KINDL-Kid-S (N = 100)

Sub-scale	**α**	**Items correlations with own sub-scale**	**Items correlations with other sub-scales**	**Convergent validity**	**Discriminant validity**
Physical well-being	0.68	0.66 – 0.72	0.13 – 0.34	4/4 (100%)	24/24 (100%)
Emotional well-being	0.61	0.46 – 0.78	0.18 – 0.39	4/4 (100%)	24/24 (100%)
Self-esteem	0.62	0.63 – 0.82	0.07 – 0.55	4/4 (100%)	24/24 (100%)
Family	0.55	0.49 – 0.74	- 0.01 – 0.27	4/4 (100%)	24/24 (100%)
Friends	0.46	0.42 – 0.67	0.09 – 0.42	3/4 (75%)	21/24 (87.5%)
School	0.47	0.44 – 0.81	0.09 – 0.4	4/4 (100%)	24/24 (100%)
Total	0.81	/	/	/	/

The KINDL-Kiddo-S

The amount of missing data was 0.02% (Table 6). The M and SD values of items in the subscales were roughly equivalent, ranging 2.55–4.66 and 0.58–1.47, respectfully. Items No. 3 and 24 in the Physical well-being and School sub-scale differed greatly from the others within the sub-scale, and here the rest of the

analysis was performed on the standardized items and the scores transformed to a 0–100 scale. The sub-scale means were from 54.35 for the Friends to 87.02 for the Self-esteem sub-scale, while the total was 75.51. The scores were negatively skewed and less leptokurtic. The SD for each score was above 10 and ceiling effects above 15% were observed for the Family and Friends sub-scale. No floor effects were reported.

Table 6 Descriptive statistics of the KINDL-Kiddo-S (N = 92)

Sub-scale	**Row score range (M)**	**Transformed score M (SD)**	**Skewness**	**Kurtosis**	**Floor (%)**	**Ceiling (%)**
Physical well-being	3.39 – 4.3	73.09 (16.8)	- 0.67	- 0.09	0	4.3
Emotional well-being	3.79 – 4.61	82.74 (12.98)	-0.92	0.67	0	9.8
Self-esteem	3.45 – 4.09	70.52 (16.11)	- 0.21	- 0.41	0	4.3
Family	4.40 – 4.56	87.02 (13.4)	- 1.14	0.72	0	27.2
Friends	3.97 – 4.66	85.36 (12.44)	- 0.84	0.69	0	22.8
School	2.55 – 3.66	54.35 (19.36)	- 0.26	- 0.13	0	2.2
Total	/	75.51 (10.16)	- 0.51	- 0.06	0	0

Item internal consistency. All correlations between an item in a sub-scale and the total sub-scale score of other items in the same sub-scale were above 0.4, except item no. 21 more significantly correlated with the Self-esteem subscale (Table 7).

Equality of item–sub-scale correlations. The item–sub-scale correlations differed substantially between the items within the proposed sub-scales (Table 7).

Item discriminant validity. Item no. 21 is significantly correlated with others than with its sub-scale score (Table 7). However, the majority of the items showed low to moderate correlation with the other sub-scales.

Internal consistency reliability. The α coefficient for the summary was 0.83 and for the Physical well-being, Self-esteem, and Family sub-scale exceeded 0.6 (Table 7).

Inter-sub-scale correlations. All sub-scales showed significant correlations in-between and moderate to high with the total score. The correlation coefficient between the Emotional well-being sub-scale and the Friends was0.59, what exceeded the reliability coefficients in both. The others correlations were lower than the reliability ones (details not given).

Table 7 Scaling properties of the KINDL-Kiddo-S (N = 92)

Sub-scale	**α**	**Items correlations with own sub-scale**	**Items correlations with other sub-scales**	**Convergent validity**	**Discriminant validity**
Physical well-being	0.72	0.67 – 0.8	0.12 – 0.47	4/4 (100%)	24/24 (100%)
Emotional well-being	0.47	0.54 – 0.64	0.15 – 0.48	4/4 (100%)	24/24 (100%)
Self-esteem	0.67	0.61 – 0.77	0.06 – 0.45	4/4 (100%)	24/24 (100%)
Family	0.69	0.66 – 0.75	- 0.17 – 0.35	4/4 (100%)	24/24 (100%)
Friends	0.49	0.49 – 0.69	- 0.05 – 0.33	4/4 (100%)	24/24 (100%)
School	0.45	0.39 – 0.75	0.04 – 0.4	3/4 (75%)	21/24 (87.5%)
Total	0.83	/	/	/	/

Table 8 KINDL-S parent form

Sub-scale	**KINDL-Kid-S, N = 99 child-parent pairs**				**KINDL-Kiddo-S, N = 90 child-parent pairs**			
	M (SD)	α	ρ	Z value	M (SD)	α	ρ	Z value
Physical well-being	79.23 (14.47)	0.69	0.66**	- 1.02	73.68 (15.74)	0.72	0.71**	- 0.13
Emotional well-being	86.93 (12.31)	0.67	0.4**	- 0.04	81.94 (14.46)	0.62	0.48**	- 0.15
Self-esteem	73.8 (15.9)	0.79	0.67**	- 0.3	72.43 (13.31)	0.7	0.5**	- 0.71
Family	85.55 (11.86)	0.63	0.45**	- 1.47	85.35 (13.93)	0.78	0.61**	- 0.81
Friends	82.89 (12.29)	0.6	0.38**	- 2.12*	77.71 (15.48)	0.74	0.59**	- 3.54**
School	71 (16.3)	0.5	0.7**	- 2.91**	62.92 (16.62)	0.55	0.59*	- 3**
Total	79.9 (9.43)	0.85	0.66**	- 0.46	75.67 (10.67)	0.88	0.7*	- 0.21

*p < 0.05; **p < 0.01

Parents reporting

There were significant correlations between the children and parents' scores in both KINDL-S versions, but significant mean differences were observed in the Friends and School sub-scale (Table 8). The α coefficients were above 0.6 for all, except for the School sub-scale.

Article 2

The overall responding rate was 80% for the children and 77% for the adolescents, while the amounts of missing data were 0.17% and 0.32%, respectfully. The Kid completed 303 subjects (160 males and 143 females, mean age 10.77 ± 1.25 years) and the Kiddo 261 (114 males and 147 females, mean age 14.02 ± 0.84). The reproducibility of majority of the subscales was above 0.6 and appropriate (Table 9). For the total score, the ICC was above 0.8. However, some subscales, like the School Kiddo with the ICC of 0.03, possess very low levels of reproducibility.

Table 9 Means (M), standard deviations (SD), and the intraclass correlation coefficients (ICC) of the KINDL questionnaires

KINDL	**Kid**		**Kiddo**	
Subscale	M (SD)	ICC, n = 63	M (SD)	ICC, n = 33
Physical well-being	4.07 (0.66)	0.55	4.03 (0.65)	0.63
Emotional well-being	4.29 (0.58)	0.64	4.14 (0.55)	0.51
Self-esteem	3.87 (0.75)	0.6	3.87 (0.74)	0.75
Family	4.41 (0.55)	0.57	4.52 (0.57)	0.66
Friends	4.07 (0.66)	0.7	4.18 (0.68)	0.54
School	3.61 (0.81)	0.62	3.13 (0.79)	0.03
Total score	4.05 (0.45)	0.84	4.02 (0.43)	0.8

The final second-order CFA models for both versions are presented in Figure 3 and 4. Above the arrows pointed at the observable variables (rectangles) are given their factor loadings (standardized parameters) and the standardized regression weights of the subscales on the total score are given on the left side of the figures. The fit indices indicated a bad fit of the data to the hypothesized

structure. For the Kid-KINDL, the average chi-square from the 2000 bootstrap samples was 316.38 (SE = 1.05), with Bollen-Stine bootstrap $p < 0.001$, while TLI = 0.67, CFI = 0.706, and RMSEA = 0.077. For the Kiddo-KINDL, the average chi-square from the 2000 bootstrap samples was 325.21 (SE = 1.17), with Bollen-Stine bootstrap p = .000, while TLI = 0.618, CFI = 0.66, and RMSEA = 0.092. The factor loadings varied within each subscale of both versions from low (0.18) to moderate/high (0.79) indicating different level of associations between the latent factors and the respective items (Figure 3 and 4). On the other hand, the correlations between the factors were very low ranging 0.050.09 (details not given). Finally, the standardized coefficient values are moderate (0.64) to high (0.92) for the subscales.

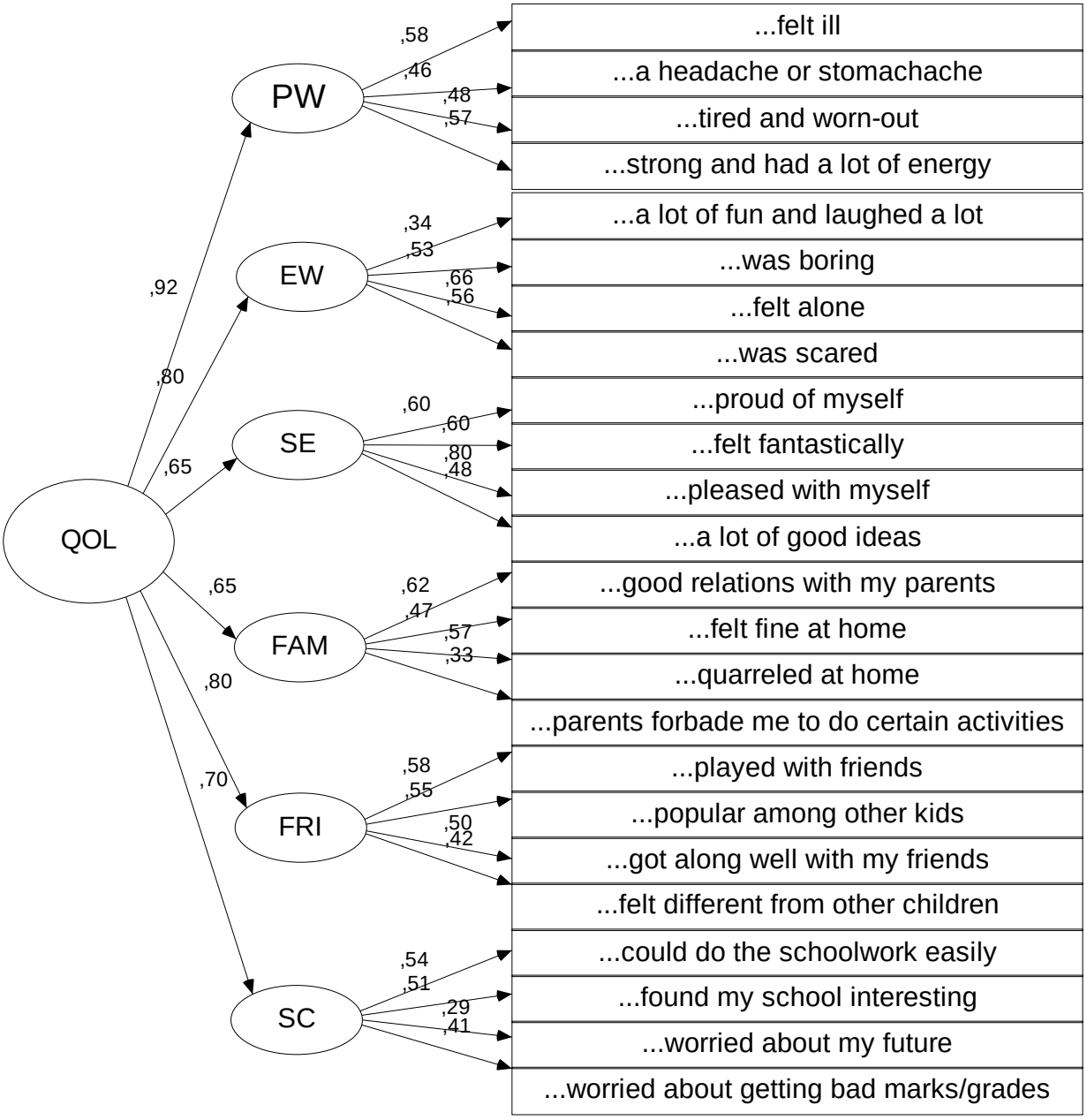

Figure 3 Final second-ordered CFA model for the Kid-KINDL

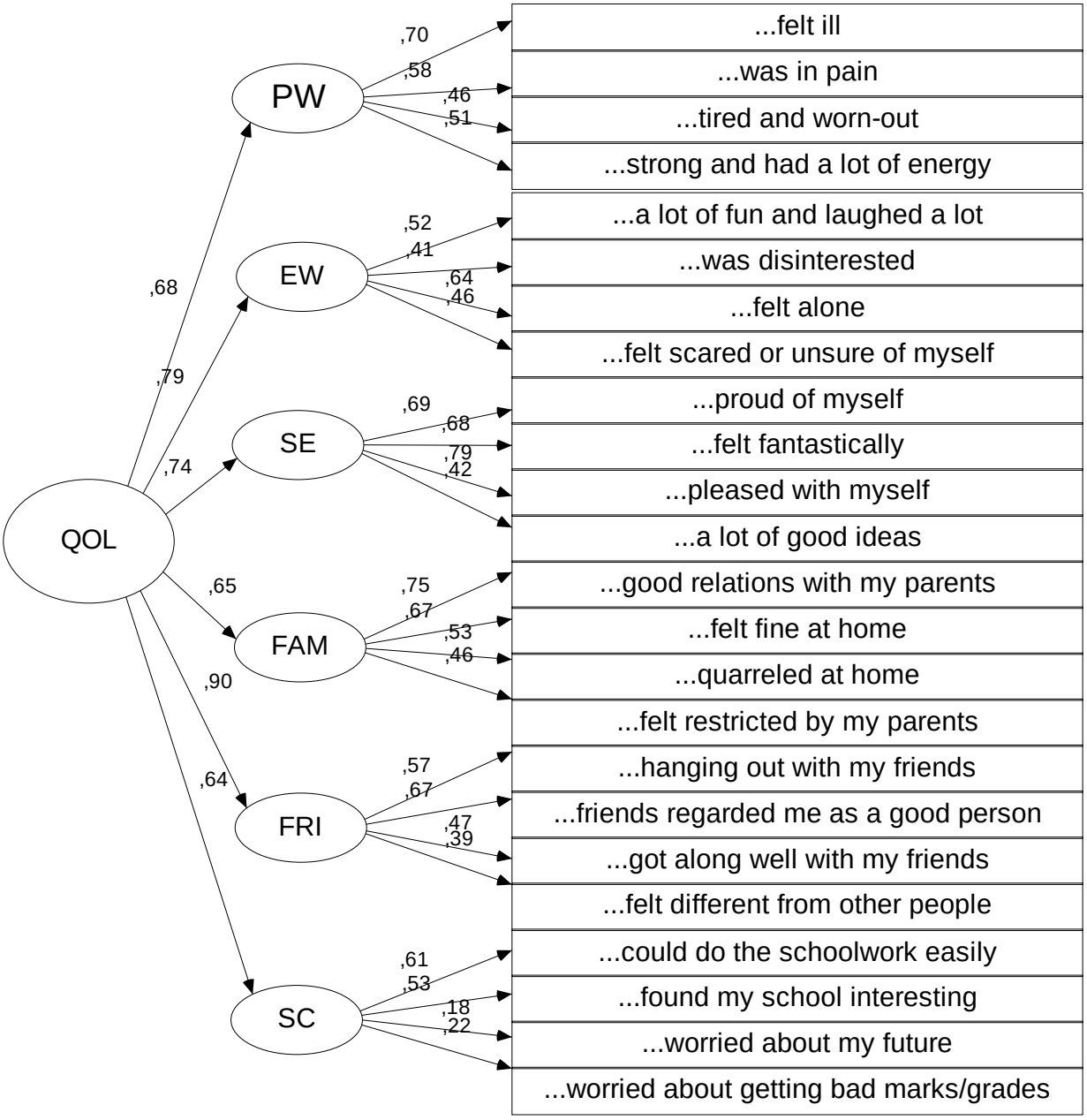

Figure 4 Final second-ordered CFA model for the Kiddo-KINDL

Article 3 *KIDSCREEN questionnaire*

The distribution of the mean T-value (SD), percentage of missing cases, and floor and ceiling effects for of all KIDSCREEN scales were given in Table 10 and Table 11.

Cronbach's α of 0.7 and above was found for all except for the Self-Perception scale of the KIDSCREEN-52 in the child and parent version, 0.58 and 0.63, respectively (Table 10 and Table 11).

Table 10 Scale description and internal consistency of the KIDSCREEN child version

	n	M	SD	Missing (%)	Floor (%)	Ceiling (%)	α	α*
KIDSCREEN-52 Scales								
Physical Well-being	323	55.96	11.9	2.11	0.33	20.33	0.83	0.80 (0.75-0.86)
Psychological Well-being	322	54.20	9.93	2.39	0.33	20.33	0.85	0.89 (0.85-0.91)
Moods & Emotions	321	49.92	11.1	2.68	0.66	10.00	0.87	0.86 (0.80-0.89)
Self-Perception	321	50.92	8.81	2.28	0.33	10.90	0.58	0.79 (0.71-0.84)
Autonomy	320	54.04	10.87	3.00	0.33	23.00	0.88	0.84 (0.79-0.86)
Parent relation & Home Life	318	53.39	9.08	3.61	0.33	21.84	0.85	0.89 (0.85-0.90)
Financial Resources	321	55.43	10.96	2.68	0.33	20.90	0.87	0.89 (0.82-0.91)
Social Support & Peers	317	55.20	11.51	3.86	0.33	16.12	0.85	0.85 (0.81-0.87)
School Environment	319	52.67	8.80	3.33	0.89	30.87	0.87	0.87 (0.81-0.88)
Social Acceptance (Bullying)	325	50.68	10.21	1.51	0.66	53.33	0.78	0.77 (0.61-0.83)
KIDSCREEN-27 Scales								
Physical Well-being	323	55.96	11.93	2.11	0.33	20.33	0.83	0.80 (0.75-0.86)
Psychological Well-being	321	51.85	10.46	2.68	0.33	9.74	0.79	0.84 (0.79-0.88)
Autonomy & Parent relation	313	54.89	10.84	5.22	0.33	14.55	0.80	0.81 (0.74-0.82)
Social Support & Peers	324	53.81	9.95	1.79	0.33	27.78	0.78	0.81 (0.76-0.84)
School Environment	319	54.94	10.98	3.33	0.33	20.90	0.79	0.81 (0.76-0.82)
KIDSCREEN-10 Index								
General HRQOL index	311	54.88	11.48	5.81	0.33	5.81	0.80	0.82

*Cronbach's alpha values from international studies (see text)

Table 11 Scale description and internal consistency of the KIDSCREEN parent version

	n	M	SD	Missing (%)	Floor (%)	Ceiling (%)	α	α*
KIDSCREEN-52 Scales								
Physical Well-being	307	54.26	10.78	2.78	0.33	13.91	0.83	0.82
Psychological Well-being	299	56.08	9.08	5.41	0.33	14.90	0.84	0.90
Moods & Emotions	302	50.16	11.50	4.44	0.33	9.50	0.86	0.84
Self-Perception	308	50.31	10.06	2.48	0.33	10.85	0.63	0.76
Autonomy	311	55.30	9.72	1.61	0.33	24.71	0.85	0.86
Parent relation & Home Life	308	55.53	9.25	2.55	0.33	19.66	0.81	0.87
Financial Resources	299	52.76	8.81	5.42	0.33	18.44	0.89	0.89
Social Support & Peers	311	56.90	9.13	1.63	0.33	12.33	0.84	0.87
School Environment	308	55.29	10.51	2.51	0.33	14.21	0.88	0.88
Social Acceptance (Bullying)	312	50.40	10.46	1.33	0.33	50.00	0.84	0.82
KIDSCREEN-27 Scales								
Physical Well-being	307	54.26	10.78	2.78	0.33	13.91	0.83	0.80
Psychological Well-being	300	53.59	10.63	5.12	0.33	7.00	0.77	0.82
Autonomy & Parent relation	294	55.21	10.11	7.00	0.66	7.67	0.70	0.78
Social Support & Peers	311	56.38	8.68	1.66	1.33	15.55	0.79	0.84
School Environment	308	54.83	10.27	2.51	0.33	16.13	0.83	0.83
KIDSCREEN-10 Index								
General HRQOL index	298	55.89	11.62	5.71	0.33	5.13	0.76	0.78

*Cronbach's alpha values from international studies (see text)

Criterion validity. In the child version, correlations between the KIDSCREEN-27 scales and scales of the KIDSCREEN-52 that measure similar dimensions ranged 0.71 – 0.96. The lowest value of correlation was for the Self-Perception scale (0.58) (Table 12). All KIDSCREEN-27 scales explained between 33% and 92% of the variance in the corresponding KIDSCREEN-52 scales. For the KIDSCREEN-10 Index, correlations with the corresponding scales were 0.68 for the Physical Well-being and 0.55 for the Social Support & Peers, while they ranged 0.71 – 0.8 for the other three scales. The KIDSCREEN-10 Index explained between 30% and 65% of the variance in the corresponding KIDSCREEN-27 scales.

Correlations between the KIDSCREEN-27 scales parent version and scales of the KIDSCREEN-52 that measure similar dimensions were 0.59 for the Self-perception and Autonomy scale, while they ranged 0.72 – 0.98 for the other scales (Table 12). The KIDSCREEN-27 scales explained between 35% and 96% of the

variance in the corresponding KIDSCREEN-52. Considering the KIDSCREEN-10 Index, correlations with corresponding scales were 0.65 for the Physical Well-being and 0.61 for the Social Support & Peers, while correlations were ranged 0.71 – 0.79 for the other three scales. The variance explained in the corresponding dimensions in the KIDSCREEN-27 ranged between 37% and 62%.

Table 12 Correlations of the KIDSCREEN-27 with the KIDSCREEN-52 and the KIDSCREEN-10 index with the KIDSCREEN-27 – child and parent versions

KIDSCREEN-27 Scales	**Correlation and regression analyses with the corresponding KIDSCREEN-52 Scales r_{child} (r^2)/r_{parent} (r^2)***				
Physical Well-being	Identical scale				
Psychological Well-being	0.81 (0.65)/0.79 (0.62) Psychological Well-being	0.82 (0.69)/0.82 (0.67) Moods & Emotions	0.58 (0.33)/0.59 (0.35) Self-perception		
Autonomy & Parent relation	0.81 (0.65)/0.81 (0.65) Parent relation & Home Life	0.75 (0.56)/0.59 (0.35) Autonomy	0.71 (0.5)/0.72 (0.52) Financial Resources		
Social Support & Peers	0.96 (0.92)/0.98/ (0.96) Social Support & Peers				
School Environment	0.96 (0.92)/0.97 (0.94) School Environment				
KIDSCREN-10 Index	*Correlation and regression analyses with the corresponding KIDSCREEN-27 Scales r_{child} (r^2)/r_{parent} (r^2)**				
General HRQOL index	0.68 (0.65)/0.65 (0.42) Physical Well-being	0.8 (0.64)/0.79 (0.62) Psychological Well-being	0.73 (0.53)/0.72 (0.52) Autonomy & Parent relation	0.55 (0.30)/0.61 (0.37) Social Support & Peers	0.71 (0.50)/0.71 (0.50) School Environment

* r_{child} (r^2)/r_{parent} (r^2) Pearson correlation coefficient (squared Pearson correlation)

Convergent and discriminant validity. For children and parents, correlations between the KIDSCREEN and KINDL comparable scales (ranging 0.45 – 65) exceeded correlations between theoretically different scales (see Electronic supplementary material). However, correlations between theoretically different scales were low to moderate (ranging 0.19 – 0.58). Additionally, the MTMM analysis of the child-parent correlations between the KIDSCREEN-52 scales indicated that correlations between the same scales were higher than those

between different scales, except the Moods & Emotions of the parent version that also correlated substantially to the Self-Perception and Psychological Well-being of the child version (see Electronic supplementary material). For the KIDSCREEN-27, all correlations between the same scales were higher than correlations between different ones.

Table 13 Agreement and magnitude of discrepancies between children and parents

	n	Δ, d	ICC	95% Confidence Interval		ICC**
				Lower Bound	Upper Bound	
KIDSCREEN-52 Scales						
Physical Well-being	277	1.74, 0.15	0.59	0.51	0.66	0.62
Psychological Well-being	270	-1.78*, -0.19	0.49	0.39	0.58	0.51
Moods & Emotions	272	-0.13, 0.01	0.34	0.22	0.44	0.45
Self-Perception	277	0.91, 0.10	0.57	0.49	0.65	0.53
Autonomy	277	-0.98, -0.10	0.49	0.41	0.58	0.48
Parent relation & Home Life	275	-1.79*, 0.19	0.51	0.41	0.59	0.50
Financial Resources	277	-1.74*, 0.17	0.65	0.58	0.72	0.53
Social Support & Peers	269	-2.48*, -0.26	0.69	0.63	0.75	0.48
School Environment	278	2.79*, 0.28	0.42	0.32	0.52	0.62
Social Acceptance (Bullying)	284	0.31, 0.03	0.44	0.34	0.53	0.48
KIDSCREEN-27 Scales						
Physical Well-being	277	1.74, 0.15	0.59	0.51	0.66	0.61
Psychological Well-being	271	-1.16, -0.11	0.44	0.33	0.53	0.52
Autonomy & Parent relation	260	0.55, 0.01	0.54	0.44	0.62	0.51
Social Support & Peers	281	-2.44*, -0.26	0.38	0.27	0.48	0.44
School Environment	276	0.43, 0.04	0.63	0.56	0.70	0.60
KIDSCREEN-10 Index						
General HRQOL index	264	0.19, 0.02	0.36	0.25	0.46	0.56

*p < 0.05, **ICC values from international studies (see text), In bold, ICC > 0.4 and d > 0.20.

Children-proxy score comparison: level of agreement and magnitude of discrepancies. In the KIDSCREEN-52, the ICCs were moderate to excellent (ranging 0.42 – 0.69) for all scales, except for the Moods & Emotions that was fair (0.34) (Table 13). In the KIDSCREEN-27, ICCs were also moderate to excellent (ranging 0.44 – 0.63) for all scales, except for the Social Support & Peers that was fair (0.38). For the KIDSCREEN-10 Index, the ICC was fair (0.36). Finally, the largest discrepancies between children and parents were in the scores of the Social Support & Peers and School Environment scale. The Social Support & Peers scale

was underestimated (-0.26), while the School Environment (0.28) overestimated by the parents.

Article 4 *PedsQL questionnaire*

The overall amount of missing data was 0.27%, while for the items it ranged 0.4-1.3%. The scale means were from 70.65 to 88.34, with the total score of 80.74 (Table 14). The scores were negatively skewed. Ceiling effects above 15% were observed only for the Social Functioning Scale, while no floor effects were found.

Internal consistency reliability. Scale internal consistency reliability determined by Cronbach's coefficient (α) was above 0.7 for all except the School, 0.65, and Emotional Functioning Scale, 0.69 (Table 14).

Table 14 Descriptive statistics and Cronbach's coefficients (α) of the PedsQL Generic Core Scales (N = 238)

Scale	M (SD)	Floor (%)	Ceiling (%)	Skewness	Kurtosis	α
Emotional Functioning	70.65 (17.34)	0	8.8	- 0.14	- 0.56	0.69
Social Functioning	88.34 (14.61)	0	37	-1.63	2.61	0.75
School Functioning	78.49 (15.49)	0	13	- 0.46	- 0.61	0.65
Physical Health	82.32 (12.75)	0	10.1	- 0.55	- 0.41	0.70
Psychosocial Health	79.16 (12.5)	0	1.7	- 0.62	- 0.16	0.82
Total	80.74 (11.23)	0	1.3	- 0.58	- 0.23	0.86

Construct validity. The statistics assessing the adequacy of the models revealed good model fit only for the PedsQL Psychosocial Health model, while for the other three models, poor model fit (Table 15). However, RMSEA for the original five-factor model was acceptable (0.075) and this model was better explained with the data than the four-factor model.

Table 15 Confirmatory factor analyses of the PedsQL Generic Core Scales

Statistic	Model			
	Five-factor	**Four-Factor**	**Physical Health**	**Psychosocial Health**
Root Mean Square Error of Approximation (RMSEA)	0.075	0.085	0.114	0.061
Comparative Fit Index (CFI)	0.779	0.721	0.776	0.911
Tucker Lewis Index (TLI)	0.75	0.688	0.686	0.891
Chi-square value (df)	528.8 (225)	609 (226)	81.4 (20)	161.86 (86)

Convergent validity. As hypothesized, the PedsQL total and psychosocial health showed convincing negative correlations with the SDQ scales measuring emotional and conduct problems, hyperactivity/inattention and peer relationship problems, and positive correlations with the SDQ pro-social scale, whereas the PedsQL physical health showed weaker correlations (Table 16).

Table 16 Correlations between the PedsQL Generic Core Scales and the scales of the SDQ

SDQ Scale	PedsQL					
	Emotional Functioning	**Social Functioning**	**School Functioning**	**Psychosocial Health**	**Physical Health**	**Total**
Emotional symptoms	- 0.42*	- 0.37*	- 0.36*	- 0.49*	- 0.41**	- 0.49*
Conduct problems	- 0.25*	- 0.12	- 0.18*	- 0.25*	- 0.09	- 0.18*
Hyperactivity - inattention problems	- 0.25*	- 0.19*	- 0.39*	- 0.33*	- 0.2**	- 0.28*
Peers problems	- 0.14*	- 0.39*	- 0.2*	- 0.29*	- 0.15*	- 0.25*
Pro-social behavior	0.08	0.25**	0.23*	0.21**	0.09	0.16*

*correlations are significant at the p < 0.001 level. ** Correlation is significant at the 0.01 level (2-tailed).

Article 5 *CHEQOL-25 questionnaire*

The main results of the translation and adaptation process were that the Canadian HRQOL questionnaire for children with epilepsy was translated and culturally adapted to be appropriate for HRQOL assessment in children with epilepsy in Serbia. The most important translational strategies applied were semantic

rearrangements, supplementations to items, and substitutions of certain words with synonyms. No item was added, but in some, the word "epilepsy" was replaced with a phrase "seizures", and in four items the wording was slightly changed, but the original concept preserved (e.g., item No 15 "…some kids find they have to take medications for the rest of their life", while the original phrase reads: " Some kids with epilepsy think they will have to take seizure medicine for the rest of their life"). Finally, from the children's point of view, all items in the version were felt to be comprehensive, precise, and relevant for HRQOL assessment.

Table 17 Sociodemographic characteristics of the sample (N = 50)

Gender	52% male (26), 48% female (24)
Age (years)	M = 10.56, SD = 1.45, range 8–12
Class level	100% regular
School grade	M = 3.56, SD = 1.49, range 2–5
School grade repeated	None
Seizure type, % (n) Partial Generalized tonic-clonic, secondary generalized, drop attacks Absence	 72% (36) 4% (2) 24% (12)
Age of seizures onset (years)	M = 6.1 SD = 2.63, range 0.5–11
Antiepileptic drugs (AEDs) Mono-therapy Duo-therapy Tri-therapy	 68% (34) 28% (14) 4% (2)

The questionnaire was administered to 50 children with epilepsy and their parents (38 mothers, 12 fathers) (Table 17). The children and parents completed the questionnaire independently. Only five items (0.4% of all items) were unanswered. The mean scores of the subscales ranged 12.2-14.42 for the children and 12.42-15 for the parents (Table 18). Important ceiling effects were observed for the Interpersonal/social subscale of both versions, and the Intrapersonal/emotional of the parent.

Table 18 Descriptive statistics of the CHEQOL-25 subscales for children (N = 50) and parents (N = 50)

Subscale	M (SD)		Skewness		Kurtosis		Floor (%)		Ceiling (%)	
	Child	Parent	Child	Parent	Child	Parent	Child	Parent	Child	Parent
Interpersonal/ social	14.42 (4.56)	15 (4.05)	-0.33	-0.35	-0.91	-0.85	4	2	24	20
Worries/concerns (present)	12.2 (3.07)	13.12 (2.74)	-0.01	-0.28	-0.47	0.32	10	2	2	4
Intrapersonal/ emotional	12.9 (4.26)	13.76 (4.74)	-0.08	-0.23	-0.78	-1.22	8	4	8	16
Secrecy	12.58 (4.29)	12.44 (3.56)	-0.14	0.04	-1.07	-0.52	6	2	2	4
Normality	14.14 (2.68)	12.42 (3.25)	-0.38	0.45	-0.47	-0.78	2	10	2	2

In general, items correlated better with the score of the subscale from which they originated than with the other subscales (Table 19). For the Present Worries/Concerns and Quest for Normality subscales the internal consistency coefficient were low in subscales, 0.41 and 0.11 respectfully, while for the others they exceeded 0.7 (Table 19).

Table 19 Scaling properties of the CHEQOL-25 subscales for children, N = 50

Subscale	Items correlations with own subscale	Items correlations with other subscales	α
Interpersonal/social	0.49 – 0.77 (100%)	0.03 – 0.46 (45%)	0.85
Worries/concerns (present)	-0.04 – 0.48 (20%)	0.03 – 0.43 (10%)	0.44
Intrapersonal/emotional	0.24 – 0.75 (80%)	-0.01 – 0.56 (35%)	0.76
Secrecy	0.34 – 0.68 (60%)	0.1 – 0.49 (35%)	0.73
Normality	-0.39 – 0.46 (40%)	-0.02 – 0.52 (25%)	0.11

The parent version demonstrated almost the same levels of correlations between the items and the subscales, with the internal consistency of the subscales ranging from 0.44-0.87 (Table 20).

Table 20 Scaling properties of the CHEQOL-25 subscales for parents, N = 50

Subscale	**Items correlations with own subscale**	**Items correlations with other subscales**	**α**
Interpersonal/social	0.37 – 0.75 (80%)	0.03 – 0.46 (55%)	0.79
Worries/concerns (present)	0.03 – 0.41 (20%)	-0.13 – 0.41 (10%)	0.44
Intrapersonal/emotional	0.54 – 0.84 (80%)	-0.01 – 0.56 (55%)	0.87
Secrecy	0.1 – 0.62 (80%)	0.15 – 0.54 (30%)	0.71
Normality	0.13 – 0.7 (20%)	-0.05 – 0.59 (30%)	0.5

Moreover, although the parents rated HRQOL slightly higher across a majority of the subscales, there was an acceptable level of agreement between the children's and parents' reports, with the ICCs ranging from 0.43-0.57 (Table 21).

Table 21 ICCs between parent and child on the CHEQOL-25 subscales

	ICC
Interpersonal/social	0.43*
Worries/concerns (present)	0.56*
Intrapersonal/emotional	0.57*
Secrecy	0.52*

*p < 0.025

Article 6 *QOLIE-AD 48 questionnaire*

The main translation problems in developing the Serbian version of the QOLIE-AD-48 involved discrepancies in verbs forms, exact word meanings, and expression of phrases. Because of this, we systematically attempted to find appropriate substitutes for problematic words and phrases. To achieve semantic equivalence, but to maintain as far as possible the original construction, solutions were found in the semantic rearrangements, in supplementing the item, and rarely in substituting a specific word with a synonym. The supplementation referred to a single word, not an expression or phrase, and the substitution to a verb or problematic word. Both processes were adequately performed throughout harmonization with the Croatian translation and appropriate alternative renderings made throughout completion of the questionnaire by the 10 adolescents. The most frequent dilemmas in item translation were in Epilepsy

Impact, Attitudes toward Epilepsy, and Stigma subscales. Problems occurred in items 37–47. The least problematic were responses, except for responses on the Attitudes toward Epilepsy subscale.

Pilot testing not only confirmed that ambiguous items appeared in the translation, but also provided ways to avoid them. Rating the subjective assessment of presentation and relevance of the questionnaire was almost uniform in the pretesting groups (with a mean score of 4.25); likewise, the difficulties reported in understanding the items and responses. A great majority of the items were labeled imprecise, irrelevant, or inappropriate. These items were in the stigma (items 38, 39, and 41) and Epilepsy Impact (items 28, 33, and 48) subscales. Furthermore, participants reported difficulties in understanding items 45 and 46 on the Attitudes toward Epilepsy subscale, and, separately, items 7, 12, and 18. The subjects felt that these items were not representative for scoring HRQOL, and it would be better to omit than to change them. Items 28, 37, and 44 were labeled as confusing and required re-adaptation. A particular problem was discovered in answering items 1 and 2. Seven participants could not adequately compare their health. The ambiguity originated when the first response was ‘‘excellent’’ and the second was ‘‘about the same.’’ Generally, the responses in the entire instrument were not problematic for pilot testing groups, except for one boy who said that he had difficulties in choosing an appropriate answer if there were multiple choices. The total number of unanswered items in the first group was 6, and in the second group, 40. Two boys refused to answer items 44, 45, 46, and an optional item (‘‘Fear dying because of seizures?’’). A majority of the other unanswered items were from the Stigma subscale. Comments indicated that refusing to answer the items resulted from the opinion that these items were not representative of total HRQOL scoring. Four subjects offered some suggestions on item changes for the Stigma and Attitudes toward Epilepsy subscales, which clarified dilemmas in adaptation of the final version. The panel session reported that the Serbian translation was developed with minor modifications of the items and responses, which were quite easy to translate and to adapt.

The final revision of the entire instrument (especially items 27, 33, 41–48, and optional items), version 2.0, was administered to 67 adolescents for psychometric analysis (Table 22).

Table 22 Sociodemographic characteristics of the sample (N = 67)

Gender	53.73% male (36), 46.27% female (31)
Age (years)	M = 14.56, SD = 2.83; range 10-19
Living	95.55 (64) % with parents 4.45% (3) cohabit/alone
Class level	100% regular
School grade	M = 7.56, SD = 2.49; range 4-12
School grade repeated	Nobody
Seizure type (%, n)	
Simple partial	17.91% (12)
Complex partial	20.89% (14)
Generalized tonic-clonic, secondary generalized	13.43 % (9)
Absence	17.91% (12)
Myoclonic	29.86% (20)
Duration of epilepsy (yr)	M = 10.01 SD = 5.07; range 2-19
Age at onset (yr)	M = 9.02 SD = 4.81; range 0-16
Antiepileptic therapy (AEDs)	
No AEDs	13.43% (9)
Monotherapy	56.71% (38)
Di- or politherapy	29.86% (20)
Illness severity index (ISI)*	
Low	24 (35.82%)
Moderate	33 (49.25%)
High	10 (14.92%)

*Illness severity index (ISI score) included seizure frequency, seizure type, and antiepileptic drugs (AEDs).

The number of missing data was small; only 2.2% of the items were unanswered (Table 23). These data were on the Attitudes toward Epilepsy, Stigma, and Health Perceptions subscales. Several subjects commented that these items were not representative of their HRQOL. Mean scores for subscales ranged between 75.4 and 91.3, and SDs, from 13.8 to 21. No floor effects, but important ceiling effects (>10% per subscale) were observed for all subscales except Epilepsy Impact and Memory/Concentration. In the summary scores, neither floor nor ceiling effects were observed.

Table 22 Descriptive statistics of the QOLIE-AD-48 Serbian version 2.0 (N=67)

	Epilepsy impact	**Memory/ Concentration**	**Attitudes toward epilepsy**	**Physical functioning**	**Stigma**	**Social support**	**School behavior**	**Health perceptions**	**Total score**
Number of items	12	10	4	5	6	4	4	3	48
Missing data (%)	1.12	0	5.97	0	6.22	0	0	4.45	2.22
M	83.2	78	75.4	90	82.2	75.65	91.3	79.6	81.9
SD	17.3	19.9	20	14.5	19.7	21	13.8	18.3	14.1
Percentiles									
25th	75	70	64.5	87.5	66.7	62.5	87.5	66.7	75.44
50th	87.5	85	75	95	91.6	81.2	100	83.5	85.94
75th	95.8	92.5	91.6	100	100	87.5	100	91.7	92.1
Ceiling (%)	7.46	4.47	13.43	37.31	26.86	16.42	56.71	19.40	0

Table 23 Psychometric properties of the subscales of the QOLIE-AD-48, Serbian version

Subscale	**α**	**Correlations of item with own subscale (range)**	**Correlations of item with other subscale (range)**	**Convergent validity**	**Discriminant validity**	**Structure coefficient**
Epilepsy impact	0.89	0.55-0.79	0.08-0.62	12/12 (100%)	96/96 (100%)	0.98
Memory/ Concentration	0.91	0.69-0.76	0.09-0.79	10/10 (100%)	80/80 (100%)	0.70
Attitudes toward Epilepsy	0.78	0.61-0.87	0.18-0.74	4/4 (100%)	30/32 (93.8%)	0.60
Physical functioning	0.80	0.71-0.85	0.17-0.65	5/5 (100%)	40/40 (100%)	0.74
Stigma	0.75	0.55-0.77	0.12-0.69	6/6 (100%)	48/48 (100%)	0.4
Social support	0.80	0.69-0.84	0.1-0.44	4/4 (100%)	32/32 (100%)	0.34
School behavior	0.71	0.66-0.82	0.18-0.64	4/4 (100%)	32/32 (100%)	0.47
Health perceptions	0.66	0.48-0.87	0.11-0.62	3/3 (100%)	24/24 (100%)	0.45

*It was derived from factor analysis, and presents Pearson's correlations between the subscales and the summary score.

Convergent/divergent validity. All correlations between an item on a subscale and the scores for other items on the same subscale were above 0.4. Within a given subscale, these correlations were relatively close (Table 23). The score for each item, in general, correlated significantly more closely with its

subscale than with other subscales, except one item 47, from the Attitudes toward Epilepsy subscales, more closely correlated with the Epilepsy Impact subscale.

Table 24 Illness severity in relation to the subscales and the summary score

	Illness severity index (ISI)				
Subscales	**Correlation Coefficient**	**Low ISI (M)**	**Moderate ISI (M)**	**High ISI (M)**	**F** (p)**
Epilepsy impact	-0.58*	95.62	82.24	86.04	9.7*
Memory/ Concentration	-0.66*	93.23	75.91	46.65	13.8*
Attitudes toward epilepsy	-0.55*	88.71	72.97	45.62	6.5*
Physical functioning	-0.54*	98.54	89.24	71	7.4*
Stigma	-0.42*	94.70	80.58	59.67	5.9*
Social support	-0.35*	84.9	73.29	60	2.3*
School behavior	-0.42*	98.56	89.58	78.75	3.1*
Health perceptions	-0.35*	86.45	78.91	65.83	2.2*
Total score***	-0.68*	93.58	80.39	58.02	15.3*

*p < 0.05; **F-value, one-way ANOVA test, ***Two tailed t test: LISI versus MISI, p < 0.001; LISI versus HISI, p < 0.001; MISI versus HISI, p = 0.01.

Construct validity. A principal-axis factor analysis suggested two second-order factors, operationally derived from the Social Support and Health Perceptions subscales, which had low rotated loading patterns. However, Pearson correlations (a structure coefficient) between the second-order factor and each of the eight first-order factors from which it was composed demonstrated that all the subscales were positively correlated to the summary score measure. This moderate to strong correlation ranged from r = 0.98 to r = 0.34 (Table 23). In addition, multitrait analysis also revealed that there was significant inverse correlation between the Illness Severity Index and the subscale scores, and that the influence of illness severity on the overall score and the subscales was significant (ANOVA, P 6 0.05), except for the Health Perceptions subscale (P = 0.065) (Table 24).

Reliability analysis. Cronbach's a coefficients exceeded 0.7 for all subscales (range, 0.71–0.91), except for the Health Perceptions (0.66). Alpha for the summary score was 0.92 (Table 23).

Sensitivity. The summary scores of version 2.0 demonstrated a significant tendency to decrease as Illness Severity Index Increased (Table 24).

Article 7 *Q-LES-Q- SF questionnaire*

From the pre-testing, it was accepted that the Q-LES-Q – SF possesses culturally appropriate items with sufficient content validity. Only item that measures satisfaction with economic status possesses insufficient characteristics for evaluative purposes.

The amount of missing data was 5.3% and none of the subjects left more than one third of unanswered items. The data are particularly missing for the third item ("satisfaction with work"), whereas the majority of the subjects were unemployed or retired. On the other hand, the group considered all possible responses and there were no biased patterns in responding. The differences between the distribution measures are roughly equivalent, except between Item No 3 ("satisfaction with work") and the others, while no floor or ceiling effects were observed for the total and they were below 15% for all items. In Table 25 were given the distributional data of the Q-LES-Q – SF items.

Validity. All items, except Item No 3 (r = 0.18), were significantly correlated to the total score and the correlations ranged between 0.41 and 0.81. Two last, uncommitted items about medications and overall life satisfaction, were correlated to the total 0.66 and 0.83, respectfully. Finally, the Q-LES-Q – SF score was significantly correlated with the CGIs, PGIs, and GCIi, 0.89, 0.43, and 0.47 respectfully.

Table 25 Distributional data of the Q-LES-Q – SF items (N = 57)

Items	**M**	**SD**	**Skewness**	**Kurtosis**
1... physical health	3.07	1.24	-0.37	-0.94
2... mood	2.86	1.11	-0.36	-0.98
3... work	2.13	0.51	.45	3.79
4... household activities	3.16	1.01	-.01	-0.63
5... social relationships	2.73	1.16	0.19	-0.73
6... family relationships	3.25	1.18	-0.44	-0.66
7... leisure time activities	2.96	1.05	-0.31	-0.65
8... ability to function in daily life	2.91	1.23	0.11	-1.07
9... sexual drive. interest and/or performance	2.75	1.18	-0.25	-1.2
10... economic status	2.34	1.15	0.51	-0.67
11... living/housing situation	3.17	1.23	-0.51	-0.73
12... ability to get around physically without feeling dizzy or falling	3.18	1.23	-0.23	-0.93
13... your vision in terms of ability to do work or hobbies	2.88	1.17	0.18	-0.91
14... overall sense of well being	2.80	1.02	0.12	-0.7
Row total (1-14 item)	37.27	9.28	-0.67	0.53
15... medication	3.55	0.94	-0.33	-0.63
16 … overall life satisfaction and contentment	3.04	1.08	0.14	-0.82

Internal consistency reliability, test-retest reliability, sensitivity, and responsiveness. The internal consistency reliability of the questionnaire was 0.90, while test-retest reliability was 0.93. Fifty-four subjects were re-tested, as the subjects whose health status has not changed in any domain since the first assessment, while three subjects had not appeared to the testing. Test-retest reliability of Item No 15 was 0.75 and 0.80 of Item No 16.

In Table 26 were given the descriptive data of the "change" and "unchanged" group and in Table27 the responsiveness parameters. The smallest detectable change (SDC) of the measure is almost 6.5, while the minimal important difference (MID) almost nine. The Q-LES-Q – SF is able to detect changes in HRQOL in nearly 80% (sensitivity to change) of those who claimed that there was change during the follow-up, while it detects 100% the absence of HRQOL change when there was no real change (specificity to change).

Table 26 Total Q-LES-Q – SF score at baseline and the 4-week follow-up of the groups

Group	**Baseline score M (SD)**	**Follow-up score M (SD)**
"Changed", n = 14	33.57 (6.24)	42.52 (8.48)
"Unchanged", n = 22	45.02 (10.92)	44.76 (11.5)

Table 27 Responsiveness parameters for the Q-LES-Q – SF

Parameter	**Value**
Standard error of measurement (SEM)	2.74, 90% CI = ± 4.49
Smallest detectable change (SDC)	6.34
Minimal important difference (MID)	8.95
Sensitivity to change	78.57%
MID proportion	71.43%
Specificity to change	100%

Article 8

Mean QOLIE-AD-48 scores ranged from 69.98 for the Attitudes toward Epilepsy domain to 93.57 for the School Behavior domain (Table 28), and the 95% CI was 57.22–95.94. The standard deviations indicated a considerable range of response, particularly higher in the Stigma, Epilepsy Impact, Social Support, and Attitudes toward Epilepsy domains. Compared with the scores reported for the adolescents without seizures in the study of Cramer et al. (mean ± SD = 77.3 ± 12.6) (19), the scores in this sample differed significantly ($t = 4.74$, $P < 0.001$).

As for gender, there were no significant differences in HRQOL between boys and girls, but in the Epilepsy Impact domain, girls reported significantly differently with respect to perception of the effects of epilepsy. In addition, the 50th percentile of the HRQOL reflected greater differences between the sexes in the Social Support domain, and this domain was only significantly correlated with HRQOL total score (Table 28). Epilepsy concern (mean = 82.04, SD = 24.3) was inversely correlated with all HRQOL scores, indicating that the subjects who were less concerned that seizures would recur reported better HRQOL.

Table 28 The QOLIE-AD 48 scores (N = 71)

	M (SD)	95% CI	Percentiles		
			25	50	75
Total HRQOL	83.52 (11.61)				
Boys	83.9 (10.56)	80.47 - 87.33	79.25	86.45	91.92
Girls	83.06 (12.92)	78.4 - 87.72	74.14	89.31	93.22
Health perception	81.42 (14.98)				
Boys	82.1 (14.7)	77.32 - 86.86	75	81.25	93.75
Girls	80.6 (15.5)	75.02 - 86.19	73.2	84.37	91.95
Epilepsy impact*	85.23 (15.61)				
Boys	87.41 (11.5)	83.66 - 91.14	82.3	91.67	95.83
Girls	82.6 (19.34)	75.62 - 89.57	70.83	93.13	95.83
Memory/Concentration	83.49 (14.55)				
Boys	83.78 (14.62)	79.04 - 88.52	75	87	95
Girls	83.12 (14.69)	77.82 - 88.42	75	87	93.75
Physical functioning	91.69 (11.52)				
Boys	91.9 (10.74)	88.44 - 95.4	87.5	95	100
Girls	91.4 (12.59)	86.86 - 95.94	90	95	100
Stigma	82.96 (19.4)				
Boys	81.1 (21.21)	74.21 - 87.96	66.67	88.89	100
Girls	85.24 (16.97)	79.12 - 91.36	69.44	94.44	100
Social support	74.78 (20.18)				
Boys	71.19 (22.27)	63.97 - 78.4	56.25	75	87.5
Girls	79.15 (16.62)	73.16 - 85.14	65.62	84.37	90.62
School behavior	93.57 (11.86)				
Boys	94.39 (10.80)	90.88 - 97.89	93.75	100	100
Girls	92.57 (13.13)	87.84 - 97.31	87.5	100	100
Attitudes towards epilepsy	69.98 (26.59)				
Boys	72.27 (25.84)	63.89 - 80.65	50	81.25	93.75
Girls	67.18 (27.63)	57.22 - 77.15	50	75	87.25

* F = 8.01, p = 0.006

In the regression analysis, number of AEDs taken was revealed as the most significant predictor in all domains studied (Table 30); variance was additionally explained by epilepsy concern in the Epilepsy Impact, Physical Functioning, Stigma, School Behavior, and Attitudes toward Epilepsy domains. Female sex appeared to be a third predictor in the variance model of Epilepsy Impact. None of the other variables significantly contributed to the variance in HRQOL.

Table 29 Correlations between overall HRQOL scores and its domains (N = 71)

	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
1. Total HRQOL	-							
2. Health perception	0.49*	-						
3. Epilepsy impact	0.86*	0.40*	-					
4. Memory/Concentration	0.65*	0.35*	0.44*	-				
5. Physical functioning	0.66*	0.51*	0.62*	0.56*	-			
6. Stigma	0.6*	0.25*	0.57*	0.25*	0.28*	-		
7. Social support	0.32*	0.04	0.16	0.12	0.13	- 0.02	-	
8. School behaviour	0.62*	0.43*	0.57*	0.55*	0.56*	0.24*	0.08	-
9. Attitudes towards epilepsy	0.67*	0.27*	0.6*	0.33*	0.38*	0.42*	0.04	0.23 *

* Correlations significant at the 0.05 or higher level (2-tailed)

Table 30 Overall HRQOL and its domains' variances explained by the most significant predictors

	Standardized Estimate – beta	***F***	***Variance***
Total HRQOL			
AEDs	0.28*	45.78	57%
Epilepsy concern	- 0.006*		
Health perception			
AEDs	0.32*	17.84	20%
Epilepsy impact			
AEDs	0.27*		
Epilepsy concern	- 0.008*	25.25	53%
Gender	0.15*		
Memory/Concentration			
AEDs	0.41*	38.5	60%
Physical functioning			
AEDs	0.38*	17.88	34.5%
Epilepsy concern	- 0.006*		
Stigma			
AEDs	0.233*	15.47	31%
Epilepsy concern	- 0.01*		
Social support			
AEDs	0.23*	5.37	7%
School behaviour			
AEDs	0.32*	20.5	38%
Epilepsy concern	- 0.009*		
Attitudes towards epilepsy			
Epilepsy concern	- 0.01*	9.45	22%
AEDs	0.27*		

*$p < 0.05$

Discussion

The cross-cultural adaptation process of a HRQOL questionnaire to a new language/culture is a complex process involving several consecutive steps in order to ensure that the HRQOL concept represented by the questionnaire is appropriately transferred to that new language/culture. The essence of the process actually involves weighting between the altering of the source questionnaire's items literally (i.e. translation) and removing, changing, adding, supplementing and/or modifying those items that deal with behavior that does not generalize equivalently in the target culture (i.e. cultural adaptation). Qualitative evaluations (i.e. pre-testing) and quantitative evaluations of the target questionnaires (i.e. psychometric evaluations) are added to confirm that the measuring concept represented by the questionnaire is appropriately transferred to the target culture.

In this thesis, it was presented the operational model of the cross-cultural adaptation of HRQOL questionnaires for pediatric population in Serbia. The framework for cross-cultural adaptation presented here allows a systematic approach to translation and cultural adaptation, pre-testing, and psychometric evaluation in order to ensure that the pediatric HRQOL concept represented by one questionnaire is translated appropriately. Using five different HRQOL questionnaires, the proposed cross-cultural adaptation model is fully operationalized in the "step-by-step" approach throughout several phases: pre-translation phase (investigating conceptual equivalence and eliminating bias, with preparation activities), translation phase (forward translation, reconciliation, backward translation, and pre-final version development), pre-testing phase (cognitive debriefing and questionnaire completion), psychometric phase, and final report and the target questionnaire.

Although all are necessary for a reliable and valid cross-cultural adaptation, several aspects are the cornerstones of the model.

First, the investigation of the HRQOL concept before engaging into the translation and cultural adaptation of the questionnaires for children provided

important insights not only on how the children perceived the concept and what domains constitute it, but also for the translation and cultural adaptation process itself. In this way, during the forward-translations the translators were provided with the details of the HRQOL concept evaluation, what helped approaching the adaptation in more child-sensitive ways. Considering the approach to evaluate the HRQOL concept in a particular group of children, it was concluded that Serbian translations were likely to be equally valid for HRQOL assessments as the original questionnaires were, although more work was warranted for some questionnaires.

Second, the systematic approach to equivalence and bias evaluation included five main aspects: conceptual, semantic, item, operational, and psychometric and three types of bias: construct, method, and item. As the sixth aspect of equivalence testing, Herdman and his colleagues suggested functional equivalence. They defined functional equivalence as the extent to which a questionnaire does what it is supposed to do equally well in two or more cultures (25), while cross-cultural psychology suggests that functional equivalence is achieved when the domain of behaviors sampled on a test has the same purpose and meaning in both cultures in question (180). Functional equivalence is also known as structural and it donates on the identity of underlying dimensions (factors) in all groups, what is actually *measurement invariance* (30). Functional equivalence is not recognized in the operational model suggested in this thesis as the separate one. The main reasons for this are the following. First, we found this type of equivalence to be vaguely defined for a routine use in the cross-cultural adaptation alongside with the other five types of equivalence and it is actually in the basis of conceptual and item equivalence (25, 180). Second, the HRQOL concept is still elusive and vaguely defined as well, with a great majority of the studies based on particular questionnaires claiming to measure HRQOL, instead of conceptual studies dealing with the concept itself. Third, the psychometric evaluation of a questionnaire is an iterative process and longitudinal assessments are needed to demonstrate its measurement properties, especially considering its structure and measurement invariance. On the other hand, there appears to be a prevailing notion that the replicability of a factorial structure of one questionnaire

across cultural groups guarantees that the questionnaire will operate equivalently across these groups and it is suitable for cross-cultural comparisons (187). However, a prerequisite for cross-cultural comparisons is that the same theoretical construct is measured in each culture, namely that construct equivalence is achieved for the questionnaire measuring the construct (186). Therefore, in order to compare estimates by the questionnaire across various nations/countries, an important aspect that needs to be demonstrated is that reproducible structure represented by specific items across different ethnic/cultural groups is also invariant, what implies that the items measure in the same way across the groups (186). In this light, if at one point of time the target HRQOL questionnaire has not equivalent structure to the source questionnaire that does not mean that with more data on the concept available or with different samples tested in future, the equivalence will not be achieved. Therefore, claiming functional equivalence could be premature in the cross-cultural adaptation process of pediatric HRQOL questionnaires and with conceptual, semantic, items, operational, and psychometric equivalence are covered the main aspects of equivalence testing and their outcomes are sufficient for future directions on the use and further development of the target questionnaires especially.

Third, during the forward/back translation it was provided a "pool of possible translations" by each translator including all possible translation/adaptation options for one item/response/instruction. This is an adaptation from the IQOLA procedure (22) and it allowed selecting the best possible option for the item/response/instruction during the reconciliation steps. This is actually the consensual decision on the option, with minimal influence of the translators their selves.

Forth, during the process several panel meetings were organized devoted to the review of the forward/back translation, evaluation of equivalence achievement, and harmonization testing with the originals and other versions mostly. As the previous, the panel meeting served to eliminate any discrepancies by consensus (50).

Fifth, it was ensured that the cross-adaptation process was carried out by people working with children and adolescents and that the process included children and adolescents during panel meetings. In this way, it was minimized possibility of method bias.

Sixth, a great effort was put on the qualitative and quantitative evaluations of the Serbian versions. As a qualitative, pre-testing procedure of the questionnaires, it was considered cognitive debriefing. Cognitive debriefing allowed determining whether items/responses/instructions in the translation are understandable, interpretable, and relevant directly from the interviews with children and adolescents (50). On the other hand, various statistical procedures were considered to evaluate the items, scales, or the entire questionnaire measurement properties. At the item level, it is aimed to assess the general "behaviors" of each item, while at the scale level it is aimed to assess how items "behave" together in a scale when measuring a particular domain. At the entire questionnaire levels, it is aimed to assess the behaviors of all scales.

Finally, following different steps in this operational model and various outcomes, it is possible to determine the "fate" of the target language questionnaire. The translated questionnaire can be used for cross-cultural HRQOL comparisons, can be used only for in-culture HRQOL evaluations and comparisons, or it needs to be further developed (i.e. revised).

Discussion of results

Article 1 & 2 *KINDL questionnaire*

The KINDL is the first generic questionnaire for the HRQOL assessment of children and adolescents culturally adapted and psychometrically tested for the Serbian language. The Serbian KINDL is developed in two versions –the KINDL-Kid-S for children, the KINDL-Kiddo-S for adolescents, and a parent form as a proxy measure. Based on this report, the versions have very similar performances and promising metric characteristics. In the entire questionnaire, the response rate was high; the amount of missing data was negligible indicating good feasibility,

acceptability, and relevance of the concept measured, as well as appropriate translation and cultural adaptation performed. The distributional measures revealed that both versions have sufficient basic characteristics. The mean values of the items in the sub-scales were roughly equivalent, as if the SD values, although located on a positive side of the response scale. Only the items in the School sub-scale differed greater in between. However, the normality parameters showed the negative skewness of all sub-scales, with higher scores, and the distribution more peaked, particularly for the Kid-S version. Negligible ceiling effects were reported for the total score, while for the sub-scales, these ranged from very low to high, like in the Family subscale. No floor effects were reported in both. Together, these findings addresses the notion that the items in its assumed sub-scales measure roughly the same level of a proposed concept (22), but with a greater overall variance and the possibility that the positive perceptions of the HRQOL would be rather reported. The important ceiling effects indicate that the KINDL-S is not able to detect possible changes occurring during the time and is unlikely to have evaluative characteristics.

The multivariable analysis showed that the items and sub-scales possess consistent measurement relationships, but the underlying concepts might be overlapping. The majority of the items were significantly more correlated to the sub-scale score where it was hypothesized to belong than to the others, except item no. 17 in the Kid-S ('I did things together with my friends') and no. 21 in the Kiddo-S ('Doing the schoolwork was easy') importantly correlated to the Self-esteem sub-scale. The correlations between the items within the sub-scales differed from each other in a broad range and showed low to moderate correlation with the total score of the other sub-scales. These two findings show that the internal consistency of the items is satisfactory for the KINDL-S, but the proportion of the information contributed to the sub-scale score was not equal for all items and some of them were likely to have superior positions. Additionally, the concepts overlapped, with an insufficient integrity of the assumed items' groups, and convergent and discriminate validity reduced. The overall reliability of the KINDL is adequate, whereas the KINDL-Kid-S has the Cronbach's coefficient 0.81 and the

KINDL-Kiddo-S 0.83, as a sufficient level of reliability for group comparisons. The alpha value exceeded 0.6 for the Physical and Emotional well-being and Self-esteem sub-scale of the Kid-S version and the Physical well-being, Self-esteem and Family sub-scale of the Kiddo-S. This value is an accepted level of reliability for the measures in development and basic research for sorting children and adolescents, but not for any comparisons (62). The other sub-scales do not possess an appropriate reliability and deserve further studying. Nevertheless, the correlation coefficients between two sub-scales was less than their reliability, what is an evidence of unique reliable variance measured by each sub-scale and a parameter of a sound construct validity, where the proposed sub-scales did not measure the same concept (62). An exception is the Emotional well-being and Friends subscale in the Kiddo-S version that might possess a substantial concept overlapping. Therefore, the reliability findings suggest on a considerable inconsistency in giving an accurate response across the sub-scales, a considerable level of common variability of the items, or an important measurement error. On the other hand, the construct validity might be sufficient, but there might be some irrelevant items within the sub-scales biasing the reliability, what should be best-clarified using factor analysis. Overall, the metric performances of the parent form are better than the children's are. The parents tended to value the HRQOL higher across the items, except in the Family and School sub-scale where they reported on significantly lower, but the pairs' correlations between the sub-scales showed significant linear relationships. The parents' reporting showed the internal consistency coefficients above 0.6 in all sub-scales and the total score, except the School sub-scale, what is an accepted level of reliability; and currently, a parent version can be used as a proxy measure, although the further studying is needed.

Comparing the Serbian KINDL with the other validated versions, the following was observed. The mean and SD values are generally higher, with more ceiling effects in the Serbian than in the original (90), Norwegian (129), English version for the Asian population (130), and for some subscales in the Spanish (131). The School sub-scale has the lowest scores in all, but higher in the Serbian translation. Further, the reliability is much better in the original and Spanish

KINDL, while the Serbian has the reliability closely related to the others. The construct validity is not evaluated here and is not possible to compare the characteristics appropriately. Finally, for the Norwegian and English version for the Asian population, it was suggested that the sub-scales possessed the overlapped concepts, like in the Serbian, which deserves further evaluations.

The peculiarities of the KINDL-S can be explained by the translation and cultural adaptation process, different values placed to the HRQOL domains, and overall health of the children (25, 76). The study has several limitations that could also bias the results. The study samples were small, so the factor analysis was not performed to explore construct validity and the offered solutions were only for studying the scaling assumptions (95). Further, although the samples were heterogeneous, they were from a central area of Belgrade city and the children from rural societies were not included. No comparable HRQOL or well-being measure was used and convergent validity was not tested, and finally, test–retest was not organized so the temporal stability of the KINDL-S was not evaluated.

The second study further assessed the measurement properties of the Serbian KINDL questionnaire for HRQOL assessments in healthy children and adolescents. Here, the results indicated the translation has appropriate stability in repeated assessments for general groups' comparisons, but the hypothesized theoretical model of HRQOL is not appropriately represented with the KINDL items. The reproducibility, as test-retest reliability, of the Serbian KINDL is different across the subscales, ranging from very low (0.03) to moderate (0.75) and it is high (0.8 and 0.84) for the total score only. The Kid version is more stable in repeated assessments than the Kiddo. This level of measurement stability for some subscales is possible to explain with assumption the concepts measured by the items of that subscales are possibly more dynamic in nature and sensible to even subtle changes in HRQOL than expected for healthy individuals. Taking into account the results of internal consistency from the previous study, where Cronbach's coefficient ranged 0.42-0.72 for the subscales and 0.8 for the total, the level of reliability indicates the total KINDL could only produce reliable assessments for group comparisons. On the contrary, the sub- scales could produce

reliable measurements only for basic evaluations, like sorting subjects or preliminary decisions, considering that some possess inappropriate reliability as an indicator of low discriminatory ability (58). These data requires more explorations, whereas the recent researches of the Taiwanese version of the Kiddo-KINDL and the Spanish KINDL in healthy populations also reported very similar levels for test-retest reliability (131, 132).

The indices from the CFA analysis show the data failed to fit appropriately the hypothesized model of the KINDL, whereas they were below acceptable ranges (58, 102). This implies the original theoretical model could be discarded for the Serbian version and appropriate construct validity is not possible to support for valid HRQOL assessments. From this analysis, it was observed that the items share common latent construct partially, whereas there are low to moderate associations between the subscales and the respective items (based on the factor loadings) with a high variability of the associations within each subscale of both versions. On the contrary, the correlations between the factors were very low between the subscales, showing the subscales measure different constructs to a substantial degree. Together, these findings suggest that there is a partial level of convergent validity, while the subscales possess even excellent discriminant validity. Placing these observations on the continuum of construct validity, we have on its very left side an excellent distinctiveness of the KINDL subscales, discriminant validity, and somewhere on its middle a moderate possibility of the items to measure common underlying constructs of each domain. Therefore, the above findings show that there are complex associations among the items and their underlying constructs are incompletely represented with the present subscales, although they had strong effects on the total score, suspecting that there might be some third constructs involved in these relations and it needs to be discovered in the future examinations of construct validity (58, 99). The present study is the only one to use CFA for the KINDL in healthy children and adolescent, so it is hard to compare the findings. Nevertheless, the findings from the studies of exploratory factor analysis performed on healthy samples showed the subscales possess unimportant items or some that could be regrouped differently, suggesting revisions for the KINDL (129,

131, 132). For the model studied here, AMOS suggested several modification indices that would let to the model improvement as the means of structural equation modeling (58, 99). However, this is beyond the article's scope and such a revision should be best undertaken applying a cross-cultural simultaneous approach to ensure comparability of different national versions and to avoid running into results due to chance. An important consideration during a revision shall be to study the causal effects of those items that influence HRQOL, causal variables, separately from those indicating a HRQOL level, indicator variables (58, 100).

The study has some limitations that could explain the results as well. First, restricting the sample to healthy subjects leads to restricted distribution of scores and variances, therefore the results of a CFA might be significantly affected. Further, the results might be also affected even Bollen-Stine bootstrap was used to manage the effect of deviation from normality, so the usage of polychoric correlations would be an alternative. Finally, there are no available HRQOL measures in Serbia with appropriate measurements characteristics against witch to confirm the results of construct validity and no studies reported evaluating the KINDL with CFA in healthy subjects.

Article 3 *KIDSCRENN questionnaire*

The KIDSCREEN questionnaire, as a generic HRQOL measure for children and adolescents, was developed simultaneously in 13 European countries and cross-culturally translated and adapted for 15 other languages (24). Within the study, it was found that the T-scores of all Serbian KIDSCREEN scales were only slightly above the scores from international studies (24, 104-110, 133-136). It means that Serbian children similarly perceived levels of well-being and functioning across various HRQOL dimensions to children from other European countries. Additionally, the floor effects for all dimensions were similar to those found in international studies (24, 104-110). However, the Physical Well-being, Autonomy, Parent Relation & Home Life, Financial Resources, Social Support & Peers, School Environment, and Social Acceptance items scale of the children version shown

severe ceiling effects opposite to the international data that reported a ceiling effect only for the Financial Resources and Social Acceptance (24). For the parent version, less severe ceiling scores were observed. The difference between the findings in this sample and international ones which consider the ceiling effects indicate that children in our study more readily endorsed positive values on HRQOL domains and such findings could bring a high inclusion effect of more healthy children rather than those with chronic conditions whom are generally expected to have lower levels of HRQOL.

The internal consistency reliability of the Serbian versions is appropriate for all scales, and the index, except for the KIDSCREEN-52 Self-Perception that is low for both the children/adolescent and parent/proxy version. Cronbach's coefficients were similar between children and parents. All KIDCREEN scales in different international studies had Cronbach's coefficients above 0.7 (24, 104-106, 110, 136). However, in these studies, the Self-Perception of the KIDSCREEN-52 had Cronbach's coefficients slightly lower than found for other scales (24).

Turning to validity, the analyses revealed that the KIDSCREEN-27 scales had sound criterion validity with high percentages (56–92%) of the variance explained in the corresponding dimensions in the KIDSCREEN-52 and only lower for the Self-perception scale. Apart from the Self-perception, the Autonomy scale of the parent version also showed lower levels of criterion validity coefficients. In the original study, the KIDSCREEN-52 Self-Perception scale correlated slightly below the a priori specified threshold with the corresponding dimension of Psychological Well-being, but this is likely to be because few items from this scale were incorporated into the shorter version (106). Convergence between the KIDSCREEN-10 Index child/adolescent and parent version and corresponding dimensions of the KIDSCREEN-27 was acceptable considering the Psychological Well-being, Autonomy & Parent relation, and School Environment dimension with the variance explained above 50%, but lower only for the Social Support & Peers and School Environment. Such findings of criterion validity might imply that the shorter version and index adequately represent the original measurement model of the KINSCREEN-52. Furthermore, the analyses of convergent validity generally

indicated that the KIDSCREEN measurement model displayed a reasonable pattern of associations. Correlations between KINDL and KIDSCREEN scales demonstrated convergent validity with medium to large effects (24). However, correlations between theoretically different scales were low to moderate, what might indicate poor discriminant validity of the scales. Nevertheless, in our previous studies with the KINDL, it was found that this measure possessed questionable convergent/discriminant validity, what might be reflected here. Considering this, we included an additional analysis to demonstrate convergent/discriminant validity, the MTMM analyses of the child-parent correlations between the KIDSCREEN scales. This analysis demonstrated that there were higher correlations between the same KIDSCREEN-52 scales than between different and only the Moods & Emotions scale of the parent version correlated similarly to the Self-Perception and Psychological Well-being of the child version. However, for the KIDSCREEN-27, all correlations were in the predicted fashion. The MTMM results closely resemble those found in international studies (24). The above results, as a whole, indicates that the KIDSCREEN questionnaire possesses appropriate levels of convergent/discriminant validity. However, construct validity and convergent/discriminant validity needs to be further explored using confirmatory factor analysis in order to confirm the KIDSCREEN constructs especially across different groups.

Finally, assessing children's and parents' responding together, moderate to excellent levels of agreement were found for all KIDSCREEN-52 scales, except for the Moods & Emotions, that was fair. Additionally, the magnitude of discrepancies between children and parents scores was relatively small for a majority of the scales, while moderate for the Social Support & Peers and School Environment. The Social Support & Peers domain was underestimated, while the School Environment overestimated by the parents. In the KIDSCREEN-27, moderate to excellent levels of agreement were found for all dimensions, expect for the Social Support & Peers dimension, which was also underestimated by the parents. Finally, for the KIDSCREEN-10 Index, levels of agreement were fair, besides that the magnitude of discrepancies between children and parents scores was relatively

small. Taken together such data indicate that the parent KIDSCREEN-52 and -27 versions could be use as good approximations of child's HRQOL, but it is always favorable to consider both (3, 76). It is advisable to include both ratings for global HRQOL scores obtained from the KIDSCREEN-10 Index. Our findings generally follow the patterns of the agreement observed in other European countries using the KIDSCREEN-52 (10). However, in that study it was indicated that the Physical and School Environment domains in general have the highest agreement between youths and parents, while social and psychological facets presented the main discrepancies (Autonomy and Moods & Emotions) (108).

The main limitations of the study need to be acknowledged. First, the study did not evaluate the psychometric properties of the children and adolescents separately. Second, relatively low numbers of included children with chronic illness or conditions did not allow evaluating measurement properties of the general population or the clinical settings. Third, factorial validity was not evaluated due to small samples and test-retest reliability. Finally, only the general data of the patterns of agreement between children and adolescents which was of any particular importance to evaluate factors influencing the agreement were included.

Article 4 *PedsQL questionnaire*

The PedsQL is one of the most frequently used HRQOL measure around the world. In this study, we report on some psychometrics for the Serbian self-report version for children and adolescents.

In general, the version has sufficient basic measurement characteristics. Negligible ceiling effects were reported for the total score, while for the scales, they were from very low to high, like in the Social Functioning Scale (58). No floor effects were reported in both. However, the normality parameters showed the negative skewness of all scales, with higher scores, and the distribution more peaked than expected.

The overall internal consistency reliability of the PedsQL Serbian version is adequate, but the alpha value did not exceed 0.70 for the School (0.65) and Emotional Functioning (0.69) Scale. Nevertheless, this level of the reliability of the Serbian PedsQL is appropriate for comparing groups, while it is not for analyzing individual patient scale scores, where alpha should exceed 0.90 (58). Finally, in terms of internal consistency measured by Cronbach's coefficients in similar samples as in this study, the Serbian version has slightly lower coefficients for all scales then reported for the original (137, 138), Swedish (114), and Greek (139), but higher then reported for the Chinese version (140).

Furthermore, the fit indices from the CFA analysis showed there are some problems with the current construct of the Serbian version. The data failed to fit appropriately the hypothesized models of four and five factors for the PedsQL self-report, whereas they were below acceptable ranges (58, 103). Nevertheless, the originally hypothesized model of five factors (113) has better-fit indices than the four-factor model confirmed for the Swedish version (114). Further, when considered as independent scores, the PedsQL Physical Health is not appropriately represented by the items in one latent factor, while the PedsQL Psychosocial Health is appropriately represented by the items allocated to four latent factors. Therefore, based on the CFA statistics, construct validity could be only supported for the PedsQL Psychosocial Health, what was previously already reported (114). Nevertheless, this is only preliminary evidence for construct validity, whereas we evaluated only the basic structure, without considering multiple groups CFA to examine factorial invariance across different groups (configural invariance, metric invariance, and scalar invariance) (141). This was not possible in the current study due to a small sample size and not including children and adolescent with chronic conditions.

Finally, several studies of correlations between the PedsQL and measures targeting mental health (CBCL / YSR) gave further empirical support for the theoretically driven PedsQL psychosocial health scales in general populations, a result consistent with prior findings (114, 142). Furthermore, evidence was provided, that the hypothesized PedsQL physical health scale measured a non-

psychosocial construct, but still, additional work is required to fully confirm the convergent validity, considering that this scale strongly correlated with emotional problems reported by the SDQ. It is possible that some of the items in the PedsQL Physical Functioning measure some emotional aspects and not purely physical functioning.

The study has several limitations. No comparable HRQOL or well-being measure was used to evaluated convergent validity; test-retest was not organized so the temporal stability of the version was not reported, as well as responsiveness. Finally, the parent version was not assessed.

Article 5 ***CHEQOL-25 questionnaire***

In this study, we report on the initial development of the Serbian version of the CHEQOL-25 for HRQOL assessments in children with epilepsy. Our report focuses on the translation and cultural adaptation of the measure and gives preliminary data on its reliability.

During the translation process, several strategies were applied that ensured the transference of the original concept of the measure into the Serbian version. The first and most obvious was cultural adaptation of the underlying concept, then readjustment in the form of items, followed by syntactic changes, supplementations, and subtitle substitutions. Particularly, the systemic adjustments adequately transposed the concept underlying every item, so none of the items was excluded or drastically different from the original. The pre-testing revealed that the Serbian version is a comprehensive and feasible measure, equivalent to the original in the means of item, semantic, and conceptual equivalence and could be psychometrically analyzed (115).

The descriptive analyses of the translation revealed that in both versions the amount of missing data was negligible, while the mean values of the subscales were roughly equivalent, but located on a positive side of the response scale. In addition, there is the negative skewness of all subscales and the distribution more peaked, particularly for the child version, although this deviation was not

significant. Finally, no floor effects were reported in both and only two subscales possess important ceiling effects, the Interpersonal/social subscale of both versions, and the Intrapersonal/emotional of the parent. Together, these findings addresses the notion that this is an acceptable measure and the items in its assumed subscales measure roughly the same level of a proposed concept (25), but with a greater overall variance and the possibility that the positive perceptions of the HRQOL would be rather reported. The important ceiling effects indicate that the two subscales are not able to detect possible changes occurring during the time and are unlikely to have evaluative characteristics.

The multivariable analysis revealed that the majority of the items correlated significantly to the subscale score where it was hypothesized to belong. However, within the subscales, not all of the items correlated linearly to the concept being measured and some of them correlated significantly with other subscales scores. In addition, the Present worries/concerns and Quest for Normality subscales possess low internal consistency coefficients (0.41 and 0.11, respectfully) that are below the values for the original (0.64, 0.63 respectfully) (93). For the other subscales, the internal consistency coefficients exceeded 0.7. Together, these findings imply that not all subscale of the Serbian version achieved appropriate internal consistency and there might be a considerable inconsistency in giving an accurate response across the subscales, a considerable level of common variability of the items, an important measurement error, or irrelevant items when compared to others in a subscale (22). However, we believe that the choice of the specific items is more important here than a higher value of coefficient alpha, because omitting an uncorrelated item may not be covered by the remaining items and the specific life experience will be lost.

Finally, Cronbach's alpha demonstrated sufficient internal consistency for a majority of the subscales in a parent form, while the intraclass correlation coefficients show similar level of agreement with the original version (116). These findings imply that the parent form of the Serbian version could be used as proxy measure for HRQOL of children with epilepsy with the understanding that parents' perspectives alone may not be sufficient to measure the child's HRQOL.

The study has several limitations that could bias the results, besides that the peculiarities of the Serbian CHEQOL-25 could be also explained by the translation, different values placed to the HRQOL domains, and overall health of the children (58, 115). The study samples were small, so the factor analysis was not performed to explore construct validity, where the offered solutions were only for studying the scaling assumptions (25), and the maximum age was limited to 12 years and not 15 years as in the original Canadian measure. Finally, no other HRQOL measure was included for a validation purpose, whereas such one is lacking in Serbia.

Article 6 *QOLIE-AD-48 questionnaire*

The QOLIE-AD-48 Serbian version 2.0 is the first translated and culturally adapted questionnaire for the assessment of HRQOL in adolescents with epilepsy in Serbia. It is equivalent to the original; the overall developmental process was straightforward and performed without major difficulties, and the most apparent ‘‘alterations’’ appeared as an integral part of the process of translation and adaptation.

In this translation of the QOLIE-AD-48, several strategies could be distinguished. The first and most obvious was cultural adaptation of the underlying concept of the original, then readjustment in the form of items, followed by syntactic changes, supplementations, and subtitle substitutions. Most of the modifications were made to items 7, 13, 19, 26, 38, 43, 44, 46, and 47. Other items required only minimal changes, hardly observed. The responses were actually easy to translate with minimal changes, except on the Attitudes toward Epilepsy subscale, where it was necessary to provide a separate response for every item. Nevertheless, the systemic adjustments adequately transposed the concept underlying every item, so none of the items were excluded or drastically different from the original, which produced a coherent version for the final adaptation process, the psychometric evaluation. The data from multivariate analysis confirmed that translation and cultural adaptation of the QOLIE-AD-48 into Serbian preserved the measurement properties of the original, with satisfactory validity, reliability, and sensitivity. First, the translation has excellent construct,

with good item convergence and divergence, whereas all of the items were correlated linearly to the concept being measured and these items were more strongly correlated with their own subscale than with other subscales. Then, the items referring to the same concept had approximately the same variance, and within a given subscale the items contained about the same amount of information on the concept measured (the exception was the Health Perceptions Subscale). In addition, the subscale scores moderately to strongly contribute to the summary score, indicating that the items in version 2.0 and eight subscales could be subsumed under a single construct. This is a slight difference from findings reported for the original and Greek versions, where the Social Support subscale did not significantly correlate with the summary score (94, 143). The degree of homogeneity of the items to the attributes being measured was confidentially achieved with an excellent internal consistency score for the translation (alpha was 0.92), as well as for the subscales; the only exception was the Health Perceptions subscale. The nonconforming responses to this subscale originated because of items 1 and 2; a majority of subjects were in a dilemma over whether their health was better or worse in comparison to the previous year. Unfortunately, we could not resolve the problem, which first originated in the pilot testing. Finally, the sensitivity of version 2.0, as the crown of the psychometric evaluation (121), was also good. The data estimated that it was possible to detect differences in quality of life among adolescents with epilepsy, where the more severe the impact of epilepsy, the more HRQOL would be compromised. In consideration of these data, to the extent that equivalence was achieved, four types could be reported: item, semantic, operational, and measurement (25, 94). The comprehensibility, accuracy, and simplicity of the translation confirmed good semantic equivalence, while the item equivalence was achieved by minor modifications of the items. The ability to use a similar questionnaire format, mode of administration, and methods for measurement proved the operational equivalence of the translation. However, the degree of measurement equivalence is also good (due to reported psychometric properties), but there are observable differences. The differences are in the alpha coefficients, means, SDs, and factor analysis. In addition, we did not analyze test–retest validation and responsiveness,

reported for the original. The psychometric properties of the Serbian version differ not only from the original, but also from the other versions (143-145).

Article 7 *Q-LEQ-Q-SF questionnaire*

The Q-LES-Q – SF is not a pediatric HRQOL questionnaire, although it was used before with children in clinical trials (92). Recently, it was developed a modification for children and adolescents that was adapted for Serbian was well (184, 185). However, it was included here to show how responsiveness should be evaluated, because this is the only questionnaire that has responsiveness reported at this time.

The psychometric study of the Q-LES-Q – SF examined its measurement properties in a small sample of people with psychiatric illnesses in Serbia. The study demonstrated that, as a generic HRQOL questionnaire, the Q-LES-Q – SF possesses appropriate measurement properties of an evaluative measure for assessing HRQOL changes in individual patients.

The multivariable analysis showed the questionnaire possesses appropriate convergent and criterion validity, as well as internal consistency reliability, for a measure to be used in the clinical settings, but these parameters deserve more explorations. First, except Item No 3 (“satisfaction with work”), all items correlated significantly to the total score, but the correlations are substantially different among the items, showing the amount of information provided to the total score is different for each item (125). This indicates that some items are more relevant or superior in reporting HRQOL than others are which is best possible to examine using factor analysis. Factor analysis could identify the underlying constructs of the items, to sort them in subscales, and to equalize the importance (22, 58). Second, the correlations with the Global Clinical Index indicate the Q-LES-Q - SF is valuable for evaluating illness severity, but its value shall be explore in depth for people with different psychiatric conditions (92). Finally, the internal consistency reliability of the Q-LES-Q – SF is high, 0.90, what pictures high homogeneity among

the items in measuring the intended concept, consistency in giving a response across the items, but also might indicate redundancy among the items (59). The internal consistency should be also evaluated using factor analysis.

The analysis of test-retest reliability, sensitivity, and responsiveness confirmed the appropriateness of the Q-LES-Q – SF for evaluative purposes (58). First, the Q-LES-Q – SF test-retest reliability is high, 0.93, and this is appropriate for individual comparisons, implying its stability in repeated assessments (59). Second, the approach used here to evaluate the sensitivity and responsiveness to change is suitable for measures used in clinical practice (126, 127, 146). When the standard error of measurement (SEM) is considered for an instrument, clinicians will be aware of the amount of the measurement error, while the smallest detectable change (SDC) will guide them how much an individual has to change to be judged as having really changed. Appling this to the Q-LES-Q – SF, the SEM is 2.74 points of the total (with ± 4.49 in 90% confidence interval), while the SDC is 6.34, meaning that a clinician should consider the individual score to be true score ± 4.49 and that individual has to change at least 6.34 points on the total, to be judged as having really changed. On the same note, whereas the minimal important difference as the smallest difference in the score perceived as beneficial is 8.95, an individual has to change nearly 9 points on the Q-LES-Q – SF to be considered as having clinically meaningful change. Finally, considering the three responsiveness indexes, reliable change proportion (RCP) or sensitivity to change, MID proportion, and specificity to change, the QLEQ – SF is highly sensitive and specific. Based on the data, the questionnaire could detect in 80% of the cases those who have really changed in their HRQOL according to the external criteria, and could in 100% exclude those who have not really changed.

Nevertheless, we should be aware that the above parameters, although robust and trustworthy, do not necessary generalize to all situations and they were observed in a small group of patients (126, 146). Specially, other external criteria that will serve as anchors of change might be used for MID or clinicians might set a priori the MID of the Q-LES-Q – SF (146). Additionally, if someone intends to use

the measure for group comparisons, he or she must take into account the group effect, SDC group (57).

Finally, the measurement characteristics of item No 15 and No 16, medications and overall life satisfaction, are worth mentioning. The distribution data were satisfactory and they were highly correlated with the total Q-LES-Q – SF, but both possess inappropriate reliability for individual comparisons. Therefore, the responsiveness was not evaluated and these items shall be used with precautions. Comparing these findings with other validation studies of the Q-LES-Q, the following was observed. The internal consistency reliability coefficients of this version or the general activities scale, an equivalent to the short form, are similar between the original (0.90) (92), Czech (0.90) (147), Hebrew (0.95) (148), Italian (0.92) (149), and the Serbian version (0.9). However, the test-retest reliability coefficient is highest for the Serbian Q-LES-Q – SF (0.93). Finally, the authors that validated the Q-LES-Q – SF for attention deficit – hyperactivity disorder reported that the minimally important difference anchored by clinical ratings is 3 points on the raw score, what equals the SEM observed in this study (150). In two studies with bipolar disorder and generalized anxiety disorder, the minimum clinically important Q-LES-Q (SF) score change was identified to be 11.89 and 6.80 points, respectfully (146, 151). These findings indicates that for measures such as the Q-LES-Q (SF) the magnitude of the MID may vary depending on the specific population of interest and as well as of the purpose of assessment.

There are several limitations of the study. First, the heterogeneity of the sample was (gender, age, and diagnoses) could explain high reliability coefficients. Second, a small number of participants did not allowed to study changes in mental health in those who deteriorated during the study period. Third, other aspects of validity, like construct and predictive, were not evaluated and it were not including other HRQOL measures, whereas such are unavailable in Serbian. Finally, the sample in overall was small and this limit the generalizability of the study to other settings.

Article 8

The overall HRQOL among adolescents with favorable seizure control was reported on the QOLIE-AD 48 as generally good to satisfactory, where all the participants placed positive values and perceptions on their current health and adjustments to epilepsy, although with greater variances between the subjects.

The highest scores were observed on the School behavior and Physical functioning, higher on the Health perceptions, Epilepsy impact, Stigma, and Memory/concentrations, and the lowest on the Social support and Attitudes towards the epilepsy domain. At the 50th percentile level, these scores were presented considerably high, showing a good overall HRQOL. Nevertheless, some domains showed significant variances in the HRQOL evaluation, the Stigma, Epilepsy impact, Social support, and Attitudes, indicating significant differences in health status adjustments between the subjects. Besides the Epilepsy impact, where significant variances in the scores were observed among the girls, the boys generally had higher and less variable scores. The girls perceived stigma better and reported better on social support and conducts, but no significant differences in the HRQOL domains were observed between them. In the previous reporting (152-154), adolescent females with well-controlled seizures had a slightly poorer quality of life and showed less favorable adjustments to epilepsy, but both genders showed good adjustments and coping with stigma and attitudes towards epilepsy (155-157). Compared to the results from the other studies using the QOLIE-AD 48, this sample had a significantly better overall HRQOL score than the one in the study of Cramer et al. (94) and a slightly better than the one in the study of Benavente-Aguilar et al. (158). Apart from cultural differences and socioeconomic development, this could be explained by a longer seizure-free period, its impact on the HRQOL and good adjustment to epilepsy.

The following was observed based on the results of the regression analysis. AEDs, epilepsy concern and gender were predictors to develop a poorer or better HRQOL, and generally, the more AEDs taken and the higher concerns of having seizures again, the poorer HRQOL perceived. In the overall score, near 60% of the

variance could be explained by the number of AEDs and epilepsy concern. Only taking AEDs could explain the variance in the Health perception, Social support, and Memory/concentration domain, but along with the epilepsy concern, they were responsible for the Physical functioning, Stigma, School behavior, and Attitudes towards epilepsy score variances. Finally, the model that included gender (female), AEDs, and epilepsy concerns explained the variances of the Epilepsy impact. The analysis reported that age and school achievement was not predictors of the HRQOL in adolescents with a favorable seizure control. Contrary, the predictors of the HRQOL among adolescents with different epilepsy were age, gender, seizure severity, and neurotoxicity (154, 158-160); where girls and older youth with epilepsy experienced poorer quality of life, self-esteem and competence and less adjustment to epilepsy (154, 161).

All these findings imply that the subjects might have a compromised health-related quality of life in social aspects and conducts and that they developed generally poorer attitudes towards their epilepsy compared to the other domains of importance. Additionally, considering the correlations between the domains, the Social support showed insignificant correlations with the others, except for the overall score, what might infer a secondary role in the concept of the HRQOL or in the development of some coping strategies apart from the environment where they belonged. Low scores for the Attitudes towards epilepsy picture insufficient knowledge about seizures and epilepsy, what significantly affected the HRQOL and the functioning. An important finding was also that, in spite of being seizure-free, concerns about experiencing seizures remained and strongly influenced adjustment to epilepsy and the HRQOL perception. On the other hand, physical health and condition were perceived as the most satisfactory in the HRQOL, along with good school behavior and adaptation. This implies that school might be an important aspect in health and with a good physical health these might be the best indicators of favorable seizure control. Finally, good general health, improved functioning in some cognition aspects (memory and concentration), and well-perceived stigma and impacts of epilepsy were principal constituents for coping well with epilepsy.

In the theoretical model of the HRQOL in youth with epilepsy suggested recently (161), the domains studied here are represented on the level of impairment (e.g. AEDs) and as intermediate variables. However, it is advisable to consider stigma and attitudes towards epilepsy as separated intermediate variables, because these domains have important influences on the HRQOL evaluation and epilepsy adjustments in adolescents with a favorable seizure control.

The study has some limitations. It is a cross-sectional study and no group comparisons were made, the number of subjects was small, the sample was made as a highly homogenous one because of the study criteria, and only one measure was used, which all might be limiting in the future use of the data. One particular limitation was that a great majority of studies in the past, evaluated the quality of life of adolescents with well-controlled epilepsy rather than health-related quality of life, what is compromising to compare the findings properly.

Conclusions and recommendations

Through a systematic approach to translation, cultural adaptation, pre-testing, and psychometric evaluation of several most frequently used pediatric, HRQOL questionnaires, it was operationalized a model of the cross-cultural adaptation of pediatric HRQOL questionnaires. The model includes the following steps: pre-translation, translation and cultural adaptation, pre-testing, psychometric, and finalization phase. Within these steps, it were suggested various methods for achieving satisfactory levels of equivalence between the original and translating versions, with minimal levels of measurement bias. Through this systematic approach, it is possible to determine to what levels a translated questionnaire could be appropriately used, in terms of reliable and valid HRQOL measurements, and to determine further steps in the cross-cultural adaptation of those questionnaires.

The main conclusions, with some directions for further work, considering the cross-culturally adapted questionnaires in this thesis are the following:

- The Serbian KINDL is a feasible, short and easily scored questionnaire for HRQOL assessments in children and adolescents. The basic measurements properties of the items are satisfactory and none of the items significantly deviates from the others. However, the reliability is not sufficient for several sub-scales and the questionnaire shall not be used for any evaluations, but only for basic research purposes, like the HRQOL concept development or sorting subjects. Additionally, the concepts of the sub-scales might be overlapped, besides that the concepts of the items across the hypothesized sub-scales are generally unique. This relative weakness of the subscales' internal consistency and validity allows using only the total KINDL-S score for interpreting HRQOL. The Serbian KINDL possesses appropriate reproducibility for group compressions, but priorities should be given to the total score. The subscales should be used with precautions, considering that some of them are not stable in producing reliable results in repeated assessments. A CFA failed to confirm the original model of the KINDL and its six subscales, so its construct validity

remained unsupported for valid HRQOL assessments in healthy children and adolescents. Generally, it could be inferred the Serbian KINDL could produce relatively reliable, but insufficiently valid HRQOL assessments in healthy children and adolescents. Consider some negative findings, it is advised to replicate the study to ensure whether the current KINDL measurement model is appropriate or not for HRQOL assessments in healthy children and adolescents in Serbia. In the meanwhile, the psychometric properties of the translation for HRQOL assessments in different population with chronic diseases will be reported that would add clearer insights into its measurement properties and direct eventual revisions. Additionally, it is underway a cross-cultural study evaluating the measurement equivalence of the questionnaire using Serbian and Iranian samples (190). The aim of this study is to evaluate which items showed different item functioning and to determine whether the questionnaire in its revised form could be used for cross-cultural comparisons.

- The general psychometric properties of the Serbian family of the KIDSCREEN questionnaire, KIDSCREEN-52, KIDSCREEN-27, and KIDSCREEN-10 Index, are acceptable and closely resemble those found for other language versions, make it available for screening, monitoring and evaluation purposes in Serbia. The questionnaires had appropriate internal consistency, sound criterion validity, and good convergent/discriminant validity. Additionally, suitable levels of agreement between children and parents' ratings in the main measurement model represented by the KIDSCREEN-52 were found. The Self-Perception domain of the KIDSCREEN-52 only had questionable psychometric properties. Finally, it is demonstrated that the KIDSCREEN holds promises to be used in cross-cultural comparisons, due to its sound measurement invariance (192). However, future works are warranted that include more participants recruited from various populations and implementation of other statistical procedures to further evaluate psychometric properties of the versions.

- The Serbian PedsQL is a feasible, short, and easily scored questionnaire for HRQOL assessments in children and adolescents. The scales have appropriate internal consistency reliability, sufficient for group evaluations, and good

convergent validity against psychological constructs. However, the current structure is confirmed only for the Psychosocial Health Score, and not for the entire measure, so more work is needed regarding its true construct validity. At the present, an international project is being conducted under the auspices of the International Child mental health Study Group (ICMH-SG) aiming to test the measurement invariance of the PedsQL structure and its appropriateness for cross-cultural HRQOL compressions (191; for details see https://www.facebook.com/pages/International-Child-Mental-Health-Study-Group/423569974422042). The ICMH-SG is a non-profit research-oriented organization of child and adolescent psychiatrists, psychologists, and other mental health practitioners from undeveloped and developing countries. The present study shows that the measurement model is cross-culturally non-invariant and it is not suitable for cross-cultural comparisons.

- The CHEQOL-25 Serbian version is an acceptable measure, the originally hypothesized subscales generally possess sufficient reliability, with only one subscale with insufficient internal consistency, and a parent form could be used as a proxy measure. In the next study, we will assess the construct validity, factor structure, reproducibility, and responsiveness in order to demonstrate that the Serbian CHEQOL-25 is appropriate for HRQOL assessments in children with epilepsy.

- The QOLIE-AD- 48 Serbian version is a comprehensive and feasible questionnaire and fully represents the domains it claims to measure, which can be assessed in 15 minutes; it can be used in clinical practice, as well as in HRQOL research. However, with awareness of the strengths of the present study, further investigations will be directed so that the cultural adaptation and further psychometric evaluation are wider, and cross-cultural comparisons and competition with an equivalent model are included (17).

- The preliminary analysis of the Q-LES-Q – SF demonstrated that it is appropriate HRQOL questionnaire for routine, clinical assessments of individuals with psychiatric illnesses. It showed that the measure could

produce reliable, valid, and sensitive assessments of the individuals' QOL. Additionally, it was demonstrated that the questionnaires is suitable for HRQOL in pediatric populations (185). Future research will be directed to evaluate its measurement properties in samples that are more homogeneous, with different groups of patients. In this way, the Q-LES-Q - SF will evolve into a gold standard for HRQOL evaluations in routine psychiatric practice, as it is in research.

- Considering the data about risk factors for HRQOL in adolescent epilepsy, gender differences were noticed in the perception of epilepsy impact, while only AEDs and epilepsy concern were in strong associations with the HRQOL. However, there were great variances among the subjects in their evaluation and perception of the HRQOL and their adjustments to their epilepsy, what indicates the need for individual approach in measuring and evaluation, so that they can better understand their self-perceptions and expectations about their well-being (23).

The operational model adopted in this thesis for the cross-cultural adaptation of pediatric HRQOL questionnaires offers new opportunities and challenges for pediatric HRQOL research. However, it would be important to evaluate the importance and relevance of each step proposed in the phases in follow-up studies to claim that this model is "*good research practice*" to the cross-cultural adaptation process. Additionally, it would be important to develop specific, consensus-based checklists for the cross-cultural adaptation. Finally, it would be important to evaluate measurement invariance of other questionnaires presented here, besides the KINDL and PedsQL whose analyses are underway, as an important step in the cross-cultural adaptation process, because only invariant questionnaires can be used in cross-cultural compressions.

References

1. Acquadro C, Berzon R, Dubois D, Leidy N, Marquis P, Revicki D et al. Incorporating the Patient's Perspective into Drug Development and Communication: An Ad Hoc Task Force Report of the Patient-Reported Outcomes (PRO) Harmonization Group Meeting at the Food and Drug Administration, February 16, 2001. Value Health. 2003; 6: 522-531.

2. Food US. Drug Administration: Guidance for industry: Patient-reported outcome measures—Use in medical product development to support labeling claims. Federal Reg. 2009; 235: 65132-65133.

3. Ravens-Sieberer U, Erhart M, Wille N, Wetzel R, Nickel J, Bullinger M. Generic health-related quality-of-life assessment in children and adolescents: Methodological considerations. Pharmacoeconomic. 2006; 24: 1199-1220.

4. Verdugo MA, Schalock RL, Keith KD, Stancliffe RJ. Quality of life and its measurement: important principals and guidelines. J Intellect Disabil Res. 2005; 49: 707-717.

5. Chassany O, Sagnier P, Marquis P, Fullerton S, Aaronson N. Patient-reported outcomes: the example of health-related quality of life—a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. Drug Info J. 2002; 36(1): 209-238.

6. Committee for Medicinal Products for Human Use, 25-7-2005. Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products. http://www.emea.europa.eu/pdfs/human/ewp/13939104en.pdf. 25-7-2005.

7. Bullinger M. The challenge of cross-cultural quality of life assessment. Psychol Health. 1997; 12(6): 815-825.

8. Schmidt S, Bullinger M. Current issues in cross-cultural quality of life instrument development. Arch Phys Med Rehabil. 2003; 84: S29-S34.

9. Acquadro C, Conway K, Giroudet C, Mear I. Linguistic Validation Manual for Patient-Reported Outcomes (PRO) Instruments. Lyon: Mapi Research Institute; 2004.

10. Mear I. Difficulties of international clinical trials: cultural adaptation of quality of life questionnaires. In: Chassany O, Caulin C eds. Health-Related Quality of Life and Patient-Reported Outcomes: Scientific and Useful Criteria. Paris: Springer; 2002.

11. Robitail S, Ravens-Sieberer U, Simeoni MC, Rajmil L, Bruil J, Power M et al. Testing the structural and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire. Qual Life Res. 2007; 16(8): 1335-1345.

12. Acquadro C, Conway K, Hareendran A, Aaronson N; European Regulatory Issues and Quality of Life Assessment (ERIQA) Group. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. Value Health. 2008; 11(3): 509-521.

13. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of Health-Related Quality of Life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417–32.

14. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25:3186–91.

15. Stewart AL, Nápoles-Springer A. Health-related quality-of-life assessments in diverse population groups in the United States. Med Care. 2000;38:II102-24.

16. Dasen PR, Berry JW, Sartorius N, eds. Health and cross-cultural psychology: Toward applications. Newbury Park, Calif: Sage Publications Inc; 1988.

17. Hutchinson JF. Quality of life in ethnic groups. In: Spilker B, ed. Quality of life and pharmacoeconomics in clinical trials. Philadelphia, Pa: Lippincott-Raven; 1996:587–593.

18. Sartorius N, Kuyken W. Translation of health status instruments. In: Orley J, Kuyken W, eds, Quality of Life Assessment: International Perspectives. Berlin Heidelberg: Springer–Verlag, 1994.

19. WHOQOL Group, Study protocol For the World Health Organization project to develop a quality of life assessment instrument (WHOQOL). Qual Life Res. 1993, 2: 152-159.

20. Bullinger M. The Challenge of Cross-Cultural Quality of Life Assessment. Paper presented at the International Workshop on Researching Well-being in Developing Countries, Delmenhorst, Germany, 2-4th July 2004.

21. Gandek B, Ware JE Jr, IQOLA Group. Methods for validating and norming translations of health status questionnaires: the IQOLA project approach. J Clin Epidemiol. 1998; 51: 953–959.

22. Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol. 1998; 51: 945–952.

23. Bullinger M, Alonso J, Apolone G, Leplège A, Sullivan M, Wood-Dauphinee S, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol. 1998; 51(11):913-923.

24. Ravens-Sieberer, U., the European KIDSCREEN Group. (2006). The KIDSCREEN Questionnaires - Quality of life questionnaires for children and adolescents – Handbook. Lengerich: Papst Science Publisher.

25. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: The universalist approach. Qual Life Res. 1998; 7: 323–335.

26. Kuyken W, Orley J, Hudelson P, Sartorius N. Quality of life assessment across cultures. Int J Ment Health. 1994; 23: 5–27.

27. Berry JW. On cross-cultural comparability. Int J Psychol. 1969; 4: 119 –128.

28. Triandis HC, Marı´n G. Etic plus emic versus pseudoetic: A test of a basic assumption of contemporary cross-cultural psychology. J Cross-Cult Psychol. 1983; 14: 489 –500.

29. Johnson TP. Methods and frameworks for cross-cultural measurement. Med Care. 2006; 44: S17-20.

30. Van de Vijver FJR, Leung K. Equivalence and Bias: A Review of Concepts, Models, and Data Analytic Procedures. In Matsumoto D, Van de Vijver FJR (eds.), Cross-cultural research methods in psychology. New York, NY: Cambridge University Press, 2011: pp. 17-455

31. Van de Vijver FJR. Bias and equivalence: Cross-cultural perspectives. In Harkness JA, Van de Vijver FJR, Mohler PP (eds.), Cross-cultural survey methods. New York: Wiley; 2003: 143–155.

32. Johnson TP. Approaches to equivalence in cross-cultural and cross-national survey research. ZUMA-Nachrichten spezial. 1998; 3: 1-40.

33. Poortinga YH. Equivalence of cross-cultural data: An overview of basic issues. Int J Psychol. 1989; 24: 737–756.

34. Herdman M, Fox-Rushby J, Badia X. 'Equivalence' and the translation and adaptation of health-related quality of life questionnaires. Qual Life Res. 1997;6:237-247.

35. Bowden A, Fox-Rushby JA. A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. Soc Sci Med. 2003; 57: 1289-1306.

36. Marquise P, Keininger D, Acquadro C, de la Loge C. Translating and evaluating questionnaires: cultural issues for international research. In: Fayers PM, Hays

RD (eds), Assessing Quality of Life in Clinical Trials: Methods and Practice (2nd edn.). Oxford, Oxford University Press. 2005; 77-93.

37. Van der Vijer F, Hambleton RK. Translating tests: some practical guidelines. Eur Psychol. 1996; 1: 89-99.

38. Camilli G, Shepard LA. Methods for identifying biased test items. Thousand Oaks, CA: Sage, 1994.

39. Cella DF, Lloyd SR, Wright BD. Crosscultural instrument equating: Current research and future directions. In: Spilker B, ed. Quality of life and pharmacoeconomics in clinical trials. Philadelphia, Pa: Lippincott-Raven; 1996: 707–715.

40. Loh LC. Racial Differences Influence Health-Related Quality-of-Life Measurements: Racial Difference and Quality of Life. CHEST. 2012; 141(2): 570-571.

41. Hunt SM. Cross-cultural issues in the use of quality of life measures in randomized controlled clinical trials. In: Staquet MJ, Hays RD, Fayers PM, eds. Quality of life assessment in clinical trials. New York, NY: Oxford University Press; 1998: 51– 67.

42. Gonza´lez-Calvo J, Gonza´lez VM, Lorig K. Cultural diversity issues in the development of valid and reliable measures of health status. Arthritis Care Res. 1997; 19: 448–456.

43. Stewart AL, Hays RD, Ware JE. Methods of constructing health measures. In: Stewart AL, Ware JE, eds. Measuring functioning and well-being: The Medical Outcomes Study approach. Durham, NC: Duke University Press; 1992: 67– 85.

44. Byrne BM, Shavelson RJ, Muthen B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychol Bull. 1989; 105: 456–466.

45. Meredith W. Measurement invariance, factor analysis and factorial invariance. Psychometrika. 1993; 58: 525–543.

46. Dancer LS, Anderson AJ, Derlin RL. Use of log-linear models for assessing differential item functioning in a measure of psychological functioning. J Consult Clin Psychol. 1994; 62: 710 –717.

47. Swaine-Verdier A1, Doward LC, Hagell P, Thorsen H, McKenna S. Adapting quality of life instruments. Value Health. 2004; 7:S27–30.

48. McKenna SP. A new theoretical approach to the measurement of quality of life. Drug Inf J. 1994; 28: 13–18.

49. Hagell P, Hedin PJ, Meads DM, Nyberg L, McKenna SP. Effects of method of translation of patient-reported health outcome questionnaires: a randomized study of the translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden. Value Health. 2010; 13: 424-30.

50. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005; 8: 94–104.

51. Barofsky I. The role of cognitive equivalence in studies of health-related quality-of-life assessments. Med Care. 2000; 38:II-125–II-129.

52. Sudman S, Bradburn NM, Schwarz N. Thinking about answers: The application of cognitive processes to survey methodology. San Francisco, Calif: Jossey-Bass Inc; 1996.

53. Hughes KA. Comparing Pretesting Methods: Cognitive Interviews, Respondent Debriefing, and Behavior Coding. Paper presented at the Annual Meeting of the Federal Committee on Statistical Methodology Arlington, VA, 17-19 Nov, 2003.

54. Morgan DL. Focus groups as qualitative research: Volume 16, Qualitative research methods. Newbury Park, Calif: Sage Publications; 1988.

55. Wagner L. Focus groups: When and how to use them: A practical guide. 2nd ed. Stanford, Calif: Health Promotion Resource Center, Stanford Center for Research in Disease Prevention; 1992.

56. Leidy NK, Revicki DA, Genesté B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. Value Health. 1999; 2(2):113-27.

57. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60(1):34-42.

58. Fayers PM, Machin D. Quality of Life: The assessment, analysis and interpretation of patient-reported outcomes. 2nd edition. Chichester: John Wiley & Sons; 2007.

59. Streiner DL, Norman GR. Health Measurement Scales : A practical guide to their development and use. Oxford University Press, 2008.

60. Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, Rothman M. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res. 2000; 9(8):887-900.

61. Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002; 11: 193-205.

62. Nunnally J, Bernstein IJ. Psychometric Theory (3rd ed.). New York: Mc-Graw-Hill; 1994.

63. Ware JE, Gandek BL, Keller SD, for the IQOLA Project Group. Evaluating instruments used cross-nationally: Methods from the IQOLA project. In: Spilker B, ed. Quality of life and pharmacoeconomics in clinical trials. 2nd ed. Philadelphia, Pa: Lippincott-Raven; 1996: 681– 692.

64. Haynes SN, Richard DSC, Kubany ES. Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. Psychol Assess. 1995, 7: 238-247.

65. Gorsuch RL. Exploratory factor analysis: Its role in item analysis. J Personal Assess. 1997; 68(3): 532-560.

66. Johnson DE. Applied Multivariate Methods for Data Analysts. 1998.

67. Juniper EF, Guyatt GH, Streiner DL, King DR. Clinical impact versus factor analysis for quality of life questionnaire construction. J Clin Epidemiol. 1997; 50(3):233-8.

68. Costello AB, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval. 2005; 10:7.

69. Joreskog KG. A general method for estimating a linear structural equation system. In: Goldberger AS, Duncan OD, eds. Structural equation models in the social sciences. New York, NY: Seminar Press/Harcourt Brace; 1973:85–112.

70. Hambleton RK, Swaminathan H and Rogers HJ. Fundamentals of Item Response Theory. Thousand Oaks, CA, SAGE Publications; 1991.

71. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. Med Care. 2001; 38: II28-42.

72. Hays RD, Liu H, Spritzer K, Cella D. Item response theory analyses of physical functioning items in the medical outcomes study. Med Care. 2007; 45(5):S32-8.

73. Teresi JA, Ocepek-Welikson K, Kleinman M, Cook KF, Crane PK, Gibbons LE, Morales LS, Orlando-Edelen M, Cella D. Evaluating measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): applications (with illustrations) to measures of physical functioning ability and general distress. Qual Life Res. 2007; 16: 43-68.

74. Davis E, Waters E, Mackinnon A, Reddihough D, Graham HK, Mehmet-Radji O, Boyd R. Peadiatric quality of life instruments: a review of the impact of the conceptual framework on outcomes. Dev Med Child Neurol. 2006; 48: 311–318.

75. De Civita M, Regier D, Alamgir AH, Anis AH, Fitzgerald MJ, Marra CA. Evaluating health-related quality-of-life studies in paediatric populations: some conceptual, methodological, and developmental considerations and recent applications. Pharmacoeconomic. 2005; 23: 659-685.

76. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. Health Technol Assess. 2001; 5:1-157.

77. Frisen A. Measuring health-related quality of life in adolescence. Acta Pediatr. 2007; 96: 963-968.

78. Barnes PM, Jenney MEM. Measuring quality of life. Curr Paediatr. 2002; 12: 476-480.

79. Christakis DA, Johnston BD, Connell FA. Methodologic issues in pediatric outcomes research. Amb Pediatr 2001; 1: 59-62.

80. Eiser C, Morse R. Can parents rate their child's health-related quality of life? Results of a systematic review. Qua Life Res. 2001; 10: 347-357.

81. Erling A. Methodological considerations in the assessment of health-related quality of life in children. Acta Paediatr. 1999; 88: S106-S107.

82. Waters E, Davis E, Ronen GM, Rosenbaum P, Livingston M, Saigal S. Quality of life instruments for children and adolescents with neurodisabilities: how to choose the appropriate instrument. Dev Med Child Neurol. 2009; 51: 660-669.

83. Solans M, Pane S, Estrada MD, Serra-Sutton V, Berra S, Herdman M, et al. Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. Value Health. 2008; 11(4):742-64.

84. Rajmil L, Roizen M, Psy AU, Hidalgo-Rasmussen C, Fernández G, Dapueto JJ; Working Group on HRQOL in Children in Ibero-American Countries. Health-related quality of life measurement in children and adolescents in Ibero-American countries, 2000 to 2010. Value Health. 2012; 15(2):312-22.

85. Jankovic S, Vukicevic J, Djordjevic S, Jankovic J, Marinkovic J, Basra MK. The Cardiff Acne Disability Index (CADI): linguistic and cultural validation in Serbian. Qual Life Res. 2013; 22(1):161-6.

86. Stevanovic D. Impact of the pediatric epilepsy scale and its responsiveness. Epilepsy Behav. 008; 13:714

87. Stevanovic D. Quality of life in adolescents with epilepsy assessed with a measure for adults. Acta Neurol Scand. 2008;118: 343.

88. Stevanovic D. Choosing the appropriate quality-of-life measurement. Epilepsy Behav. 2008; 13: 416.

89. Stevanovic D. Is use of the QOLIE-31 in nonadult populations appropriate? Epilepsy Behav. 2007; 10: 511.

90. Ravens-Sieberer U, Bullinger M. Assessing health-related quality of life in chronically ill children with the German KINDL: First psychometric and content analytical results. Qual Life Res. 1998; 7: 399–407.

91. Varni JW, Seid, Kurtin PS. PedsQL 4.0TM: Reliability and validity of the pediatric quality of life inventory TM version 4.0 Generic Core Scales in healthy and patient populations. Med Care. 2001; 39: 800–812.

92. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull. 1993; 29: 321–326.

93. Ronen GM, Streiner DL, Rosenbaum P, the Canadian Pediatric Epilepsy Network. Health-related quality of life in children with epilepsy: development

and validation of self-report and parent proxy measures. Epilepsia. 2003; 44: 598–612.

94. Cramer JA, Westbrook L, Devinsky O, Perrine K, Glassman M, Camfield C. Development of the quality of life in epilepsy inventory for adolescents: the QOLIE-AD-48. Epilepsia. 1999; 40: 1114–21.

95. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. Psychol Assess. 1995; 7: 286-299.

96. Streiner DL. Figuring out factors: the use and misuse of factor analysis. Can J Psychiatr. 1994; 39: 135-140.

97. de Vet HCW, Ader HJ, Terwee CB, Pouwer F. Are factor analytical techniques appropriately used in the validation of health status questionnaires? A systematic review on the quality of factor analyses of the SF-36. Qual Life Res. 2005; 14: 1203-1218

98. Ravens-Sieberer U, Bullinger M. Manual KINDLr. Berlin, Hamburg, Germany; 2000.

99. Schreiber JB. ore reporting practices in structural equation modeling. Res Social Adm Pharm. 2008; 4: 83-97.

100. Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. J Roy Stat Soc. 2002; 165: 233-261.

101. Nevitt J, Hancock GR. Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. Struct Equ Model. 2001; 8: 353-377.

102. Boehmer S, Luszczynska A. Two kinds of items in quality of life instruments: 'Indicator and causal variables' in the EORTC QLQ-C30. Qual Life Res. 2006; 15: 131-141.

103. Meuleners LB, Lee AH, Binns CW, Lower A. Quality of life for adolescents: assessing measurement properties using structural equation modeling. Qual Life Res. 2003; 12: 283-290.

104. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Power, et al. The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Value Health. 2008; 11: 645-658.

105. Ravens-Sieberer U, Erhart M, Rajmil L, Herdman M, Auquier P, Bruil J, et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: A short measure for children and adolescents' well-being and health-related quality of life. Qual Life Res. 2010; 19: 1487–1500.

106. Ravens-Sieberer U, Auquier P, Erhart M, Gosch A, Rajmil L, Bruil J, et al. The KIDSCREEN-27 quality of life measure for children and adolescents: Psychometric results from a cross-cultural survey in 13 European countries. Qual Life Res. 2007; 16: 1347–1356.

107. Robitail S, Simeoni M C, Erhart M, Ravens-Sieberer U, Bruil J, Auquier P, et al. Validation of the European proxy KIDSCREEN-52 pilot test health-related quality of life questionnaire: First results. J Adol Health. 2006; 39: 596.e1.

108. Robitail S, Simeoni MC, Ravens-Sieberer U, Bruil J, Auquier P, the KIDSCREEN Group. Children proxies' quality-of-life agreement depended on the country using the European KIDSCREEN-52 questionnaire. J Clin Epidemiol. 2007; 60: 469–478.

109. Robitail S, Ravens-Sieberer U, Simeoni MC, Rajmi IL, Bruil J, Power M, et al. Testing the structural and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire. Qual Life Res. 2007; 16: 1335–1345.

110. Erhart M, Ottova V, Gaspar T, Jericek H, Schnohr C, Alikasifoglu M, et al. Measuring mental health and wellbeing of school-children in 15 European countries using the KIDSCREEN-10 Index. Int J Pub Health. 2011; 54: 160–166.

111. Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates; (1988).

112. Goodman R. The strengths and difficulties questionnaire: A research note. J Child Psychol Psychiatr. 1997; 38: 581–586.

113. Limbers CA, Newman DA, Varni JW. Factorial invariance of child self-report across healthy and chronic health condition groups: A confirmatory factor analysis utilizing the PedsQL™ 4.0 Generic Core Scales. J Pediatr Psychol. 2008; 33: 630–639.

114. Petersen S, Ha¨gglo¨f B, Stenlund H, Bergstro¨m E. Psychometric properties of the Swedish PedsQL, pediatric quality of life inventory 4.0 Generic Core Scales. Acta Paediatr. 2009; 98: 1504–1512.

115. Harter S. The self-perception profile for children: revision of the perceived competence scale for children. Denver: Univ. of Denver; 1985.

116. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979; 86: 420–428.

117. Velarde-Jurado E, Avila-Figueroa C. Consideraciones metodologicas para evaluar la calidad de vida. Salud Publica Mex. 2002; 44: 448–463.

118. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chron Dis. 1985; 38: 27–36.

119. Eriksson KJ, Koivikko MJ. Prevalence, classification, and severity of epilepsy and epileptic syndromes in children. Epilepsia. 1997; 38: 1275–1282.

120. Raty LK, Soderfeldt BA, Larsson G, Wilde-Larsson BM. The relationship between illness severity, sociodemographic factors, general self-concept, and illness-specific attitude in Swedish adolescents with epilepsy. Seizure. 2004; 13: 375–382.

121. Johnson AL. Some statistical issues in quality of life measurements. In: Trimble MR, Dodson WE, editors. Epilepsy and quality of life. New York: Raven Press; 1994; pp. 65–84.

122. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders – Diagnostic Criteria for Research. World Health Organization, Geneva; (1993).

123. Sprangers MA1, Moinpour CM, Moynihan TJ, Patrick DL, Revicki DA. Assessing meaningful change in quality of life over time: a users' guide for clinicians. Mayo Clin Proc. 2002; 77: 561–571.

124. Crosby RD, Kolotkin RL, Williams GH. An integrated method to determine meaningful changes in health-related quality of life. J Clinical Epidemiol. 2004; 57: 1153– 1160.

125. Williams JI, Naylor CD. How should health status measures be assessed? Cautionary notes on procrustean frameworks. J Clinl Epidemiol. 1992; 45: 1347–1351.

126. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol. 2004; 57: 1008–1018.

127. de Vet HCW, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. International Q-LES-Q – SF in clinical practice. J Tech Assess Health Care. 2001; 17: 479–487.

128. Cowan J, Baker GA. A review of subjective impact measures for use with children and adolescents with epilepsy. Qual Life Res. 2004; 13(8): 1435-1443.

129. Helseth S, Lund T. Assessing health-related quality of life in adolescents: some psychometric properties of the first Norwegian version of KINDL®. Scand J Caring Sci. 2005; 19(2): 102-109.

130. Wee HL, Ravens-Sieberer U, Erhart M, Li SC. Factor structure of the Singapore English version of the KINDL® children quality of life questionnaire. Health Qual Life Outcomes. 2007; 5(1): 4.

131. Rajmil L, Serra-Sutton V Fernandez-Lopez JA, Berra S, Aymerich M, et al. The Spanish version of the German health-related quality of life questionnaire for children and adolescents: the KINDL. Anal pediatr. 2003; 60: 514.

132. Lee P H, Chang LI, Ravens-Sieberer U. Psychometric evaluation of the Taiwanese version of the Kiddo-KINDL® generic children's health-related quality of life instrument. Qual Life Res. 2008; 17(4): 603-611.

133. Haraldstad K, Christophersen K A, Eide H, Nativg GK., Helseth S, KIDSCREEN Group Europe. Health related quality of life in children and adolescents: Reliability and validity of the Norwegian version of KIDSCREEN-52 questionnaire, a cross sectional study. Inter J Nur Stud. 2011; 48: 573–581.

134. Hong SD, Yang JW, Jang WS, Byun H, Lee MS, Kim HS, et al. The KIDSCREEN-52 quality of life measure for children and adolescents (KIDSCREEN-52-HRQOL): Reliability and validity of the Korean version. J Korean Med Sci. 2007; 22: 446–452.

135. Tzavara C, Tzonou A, Zervas I, Ravens-Sieberer U, Dimitrakaki C, Tountas Y. Reliability and validity of the KIDSCREEN-52 health-related quality of life questionnaire in a Greek adolescent population. Ann Gen Psychiatr. 2012; 13(11): 3.

136. Berra S, Bustingorry V, Henze C, Dı´az Mdel P, Rajmil L, Butinof M. Cross-cultural adaptation of the KIDSCREEN questionnaire to measure the health related quality of life in the 8 to 18 year-old Argentinean population. Arch Argent Pediatr. 2011; 107: 307–314.

137. Varni JW, Burwinkle TM, Seid M. The PedsQL 4.0 as a school population health measure: Feasibility, reliability, and validity. Qual Life Res. 2006; 15: 203–215.

138. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQLTM 4.0 as a pediatric population health measure: Feasibility, reliability, and validity. Amb Pediatr. 2003; 3: 329–341.

139. Gkoltsiou K, Dimitrakaki C, Tzavara C, Papaevangelou V, Varni JW, Tountas Y. Measuring health-related quality of life in Greek children: Psychometric properties of the Greek version of the pediatric quality of life inventory TM 4.0 Generic Core Scales. Qual Life Res. 2008; 7(2), 299–305.

140. Hao Y, Tian Q, Lu Y, Chai Y, Rao S. Psychometric properties of the Chinese version of the pediatric quality of life inventory 4.0 Generic Core Scales. Qual Life Res. 2010; 9(8):1229-1233.

141. Beckstead JW, Yang CY, Lengacher CA. Assessing cross-cultural validity of scales: A methodological review and illustrative example. Inter J Nursing Stud. 2008; 45(1): 110–119.

142. Bastiaansen D, Koot HM, Bongers IL, Varni JW, Verhulst FC. Measuring quality of life in children referred for psychiatric problems: Psychometric properties of the PedsQL 4.0 Generic Core Scales. Qual Life Res. 2004; 13: 489–495.

143. Papavasiliou A, Rizou I, Kotsalis C, Panou C, Demakakos P. Quality of life in children and adolescents with epilepsy in Greece (pilot study). Paediatr Croat. 2003; 47(2): 53 - 56.

144. Benavente IA, Morales C, Rubio E. Adaptacio´n transcultural del cuestionario de medida de calidad de vida (QOLIE-AD-48) en adolescentes epile´pticos. Psiquis 2002;23:226–37.

145. Mucic-Pucic B, Basnec A, Barisic N, Jurin M. Health-related quality of life in children with epilepsy in Croatia. In: Proceeding 8th Mediterranean Epilepsy Conference; 17–20 November 2004; Marrakech, Morocco. Morocco: MLAE, 2004.

146. Wyrwich K, Harnam N, Revicki DA, Locklear JC, Svedsäter H, Endicott J. Assessing health-related quality of life in generalized anxiety disorder using the Quality Of Life Enjoyment and Satisfaction Questionnaire. Inter Clin Psychopharmacol. 2009; 24: 289–295.

147. Müllerová H, Libigerová E, Prouzová M, et al. Cross-cultural transfer and validation of the Quality of Life Enjoyment and Satisfaction Questionnaire in the population of depressed patients. Psychiatrie. 2001; 5: 80–87.

148. Ritsner M, Kurs R, Kostizky H, Ponizovsky A, Modai I. Subjective quality of life in severely mentally ill patients: a comparison of two instruments. Qual Life Res. 2002; 11: 553–561.

149. Rucci P1, Rossi A, Mauri M, Maina G, Pieraccini F, Pallanti S, Camilleri V, et al. Validity and reliability of Quality of Life, Enjoyment and Satisfaction Questionnaire, Short Form. Epidemiol E Psichiatria. 2007; 16: 82–89.

150. Mick E, Faraone SV, Spencer T, Zhang HF, Biederman J. Assessing the validity of the quality of life enjoyment and satisfaction questionnaire short form in adults with ADHD. J Att Disorder. 2008; 11: 504–509.

151. Endicott J, Rajagopalan K, Minkwitz M, BOLDER Study Group. A randomized, double-blind, placebo-controlled study of quetiapine in the treatment of bipolar I and II depression: improvements in quality of life. Inter Clin Psychopharmacol. 2007; 22: 29–37.

152. Ronen GM, Streiner DL, Rosenbaum P. Health-related quality of life in childhood epilepsy: moving beyond seizure control with minimal adverse effects. Health Qual Life Outcomes. 2003;1:36.

153. Austin JK, Huster GA, Dunn DW, Risinger MW. Adolescents with active or inactive epilepsy or asthma: a comparison on quality of life. Epilepsia. 1996; 37: 1228–1238.

154. Raty LKA, Wilde Larsson BM, Soderfeldt BA. Health-related quality of life in youth: a comparison between adolescents with uncomplicated epilepsy and randomly selected healthy controls. J Adolesc Health. 2003; 33: 252–258.

155. Norrby U, Carlsson J, Beckung E, Nordholm L. Self-assessment of well-being in a group of children with epilepsy. Seizure. 1999; 8: 228–234.

156. MacLeod JS, Austin JK. Stigma in the lives of adolescents with epilepsy: a review of the literature. Epilepsy Behav. 2003; 4: 112–117.

157. Kemp S, Morley S, Anderson E. Coping with epilepsy: do illness representations play a role? Br J Clin Psychol. 1999; 38: 45–58.

158. Benavente-Aguilar I, Morales-Blanquez C, Rubio EA, Rey JM. Quality of life of adolescents suffering from epilepsy living in the community. J Pediatr Psychol. 2004; 40: 110–113.

159. Miller V, Palermo TM, Grewe SD. Quality of life in pediatric epilepsy: demographic and disease-related predictors and comparison with healthy controls. Epilepsy Behav. 2003; 4: 36–42.

160. Cramer JA, Westbrook LE, Devinsky O, Perrine K, Glassman MB, Camfield C. Development of the quality of life in epilepsy inventory for adolescents: the QOLIE-AD-48. Epilepsia. 1999; 40: 1114–1121.

161. Lach LM, Ronen GM, Rosenbaum PL, Cunningham C, Boyle MH, Bowman S, Streiner DL. Health-related quality of life in youth with epilepsy: theoretical model for clinicians and researchers. Part I: The role of epilepsy and co-morbidity. Qual Life Res. 2006; 15: 1161–1171.

162. Greenfield PM, Keller H, Fuligni A, Maynard A. Cultural pathways through universal development. Ann Rev Psychol. 2003; 54: 461–490.

163. Nikapota A, Rutter M. Sociocultural/Ethnic Groups and Psychopathology. In Rutter's Child and Adolescent Psychiatry, 5th Edition, Edited by M. Rutter, D. V.

M. Bishop D. S. Pine, S. Scott, J. Stevenson, E. Taylor and A. Thapar. Blackwell Publishing Limited. 2008.

164. Cerović S, Zivković Z, Milenković B, Stojanović JJ, Bajec AO, Vukaŝinović Z, Veković V. The Serbian version of the pediatric asthma quality of life questionnaire in daily practice. J Asthma. 2009 ;46(9): 936-939.

165. Susic G, Ruperto N, Stojanovic R, Gacic D, Pilipovic N, Pasic S, et al. The Serbian version of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ). Clin Exp Rheumatol. 2001; 19: S168-172.

166. Moorthy LN, Weiss E, Peterson MG, Hassett AL, Lehman TJ; International SMILEY Collaborative Group. An update on cross-cultural adaptation of US English SMILEY. Lupus. 2012; 21(13): 1450-4.

167. Stevanovic D. Pediatric health-related quality of life instrument. Part I: general guidelines for selection. Vojnosanit Pregle. 2008: 65; 469-72.

168. Simeoni MC, Schmidt S, Muehlan H, Debensason D, Bullinger M; DISABKIDS Group. Field testing of a European quality of life instrument for children and adolescents with chronic conditions: the 37-item DISABKIDS Chronic Generic Module. Qual Life Res. 2007; 16(5): 881-93

169. Gurung RAR. Health Psychology: A Cultural Approach, 2nd Edition. University of Wisconsin, Green Bay, 2010.

170. Brislin RW. Comparative research methodology: crosscultural studies. Inter J Psychol. 1976; 11: 215–229.

171. Behling O, Law KS. Translating Questionnaires and other Research Instruments: Problems and Solutions. Sage Publications, Inc., Thousand Oaks, CA, USA; 2000.

172. Cha ES, Ki LH, Erlen JA. Translation of scales in cross-cultural research: issues and techniques. J Adv Nurs. 2007; 58(4): 386–395.

173. Brislin RW. Back-translation for cross- cultural research. J Cross-Cul Psychol. 1970: 1(3): 185–216.

174. Brislin RW. Translation and Content Analysis of Oral and Written Material. In: H.C. Triandis and J.W. Berry (eds.), Handbook of Cross-Cultural Psychology, Boston: Allyn & Bacon; 1980: 389-444.

175. Van de Vijver FJR, Leung K. Methods and Data Analysis of Comparative Research. In: W. Berry, Y.H. Poortinga and J. Pandey: (eds.), Handbook of CrossCultural Psychology Boston: Allyn & Bacon. 1997: 257-300.

176. Hambleton RK. Issues, Designs and Technical Guidelines for Adapting Tests Into Multiple Languages and Cultures. In R. K. Hambleton, P. F. Merenda and C. D. Spielberger (Eds.). Adapting Psychological and Educational Tests for Cross-Cultural Assessment. NJ: Lawrence Erlbaum; 2005.

177. Geisinger KF. Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. Psychol Assess. 1994; 6: 304–312.

178. Acquadro C, Jambon B, Ellis D, Marquis P. Language and Translation Issues. In: B. Spilker (ed.), Quality of Life and Pharmacoeconomics in Clinical Trials (2nd edition). Philadelphia: Lippincott-Raven; 1996.

179. House J. A Model for Translation Quality Assessment. Tübingen: Narr; 1977.

180. Geisinger KF. Testing and Assessment in Cross-Cultural Psychology. In: Weiner IB (ed), Handbook of psychology: Assessment psychology, Hoboken, New Jersey, John Wiley & Sons, Inc; 2003.

181. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. J Adv Nurs. 2004; 48(2): 175–186.

182. King KM, Khan N, LeBlanc P, Quan H. Examining and establishing translational and conceptual equivalence of survey questionnaires for a multi-ethnic, multi-language study. J Adv Nurs. 2011; 67(10): 2267-2274.

183. Van Widenfelt BM, Treffers PD, de Beurs E, Siebelink BM, Koudijs E. Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. Clin Child Fam Psychol Rew. 2005; 8(2): 135-147.

184. Endicott J, Nee J, Yang R, Wohlberg C. Pediatric quality of life enjoyment and satisfaction questionnaire (PQ-LES-Q): reliability and validity. JACAPAP 2006; 45(4): 401-407.

185. Stevanovic D, Lakic A. Pediatric Quality of Life Enjoyment and Satisfaction Questionnaire (PQ-LES-Q): a psychometric study form Serbia. 19th World Congress of IACAPAP; 2010; Bejing, China.

186. Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. Exp Aging Res. 1992; 18: 117-144.

187. Byrne BM, Watkins D. The issue of measurement invariance revisited. J Cross-Cul Psychol. 2003; 34(2): 155-175.

188. Baars RM, Atherton CI, Koopman HM, Bullinger M, Power M. The European DISABKIDS project: development of seven condition-specific modules to measure health related quality of life in children and adolescents. Health Qual Life Outcomes. 2005; 5: 70.

189. Fayed N, De Camargo O K, Kerr E, Rosenbaum P, Dubey A, Bostan, et al. Generic patient-reported outcomes in child health research: a review of conceptual content using World Health Organization definitions. Dev Med Child Neurol. 2012; 54(12): 1085-1095.

190. Jafari P, Stevanovic D. A cross-cultural study to assess measurement equivalence of the KINDL across Iranian and Serbian children: A differential item functioning analysis. 2014; (personal communication)

191. Stevanovic D, Maroco J, Atilola O, Vostanis P, Singh-Balhara Y, Franic T, et al. Can we compare quality of life among adolescents across the globe? Testing the measurement invariance of the PedsQL™ 4.0 across different countries. Qual Life Res. 2014; in review.

192. Stevanovic D, Jafari P. A cross-cultural study to assess measurement invariance of the KIDSCREEN-27 questionnaire across Serbian and Iranian children and adolescents. 2014; (personal communication).

193. SPSS for Windows. Version 10 & 18.0. Chicago, SPSS Inc.

Appendix list

Appendix I Methods for investigating equivalence and eliminating bias for

Methods for minimizing bias (as adapted from (30))

Construct bias • Simultaneously developing the same questionnaire in several cultures • Independent within-culture development of questionnaires and subsequent cross-cultural administration of all questionnaires • Consult informants with expertise in local culture and language • Use samples of bilingual participants • Conduct local pilot studies for content analyses of free-response questions • Nonstandard questionnaire administration (e.g., “thinking aloud”)
Method bias • Convergent/discriminant validity studies, and monotrait-multimethod studies • Connotation of key phrases • Extensive training of interviewers and administrators • Detailed manual/protocol for administration, scoring, and interpretation • Detailed instructions • Include background and contextual variables • Gather collateral information (e.g., test-taking behavior or test attitudes) • Assessment of response styles • Conduct test–retest, training, and/or intervention studies
Item bias • Judgmental methods (e.g., linguistic and psychological analysis) • Psychometric methods (e.g., differential item functioning analysis)

Methods for investigating equivalence (as adapted from (15, 25, 39-45))

Conceptual equivalence

- Review literature, especially ethnographic and anthropological, in target cultures for ways in which the HRQOL construct is operationalized
- Conduct interviews and focus groups of persons from target group to learn how they think about and define the construct
- Consult experts to rate items and constructs in terms of relative importance, equivalence, relevance, appropriateness, and acceptability, and identify missing items

Semantic equivalence

- Use structured qualitative methods with target population to identify meaning they ascribe to the construct (focus groups, expert panels)
- Resolve discrepancies using cognitive testing with probes to determine what subjects think items mean
- Apply semantic differential techniques across groups to define semantic space in which word is located
- Translation methods include forward and backward translation.

Operational equivalence

- Pretest and debrief; include probes about difficulty and appropriateness of survey.
- Use cognitive testing methods (e.g. in-depth interviewing or think-aloud interviews) to identify whether cognitive processes involved in interpreting and answering questions differ across groups
- Compare effects of different methods of administration on scores
- Assess cultural norms regarding ways to address people and ways of framing questions
- Have expert panel consider whether data-gathering approach is consistent with culture to which it is being applied

Item equivalence

- Differential item functioning analysis using item response theory methods or log linear models
- Re-examine item relevance in target culture via ratings by experts or laypeople, or use mathematical approach
- Scale items relative to the central tendencies of the culture studied. Rank items in both cultures using an external scale or referent to compare intervals between ranks. Compare ranking of measures by subgroups to determine comparability across cultures
- Thurstone's method of equal-appearing intervals or Stevens' magnitude estimation method

Psychometric/measurement equivalence

- Statistical methods to deal with variability, reliability, validity, measurement invariance, sensitivity to change, and responsiveness

Appendix II Short instructions for translation and cultural adaptation

Item, response options, and instructions to be translated according the following:

- During the translation/adaptation process, maintaining the content of an instruction/item/response is the paramount aim,
- Semantic-syntax rules should follow accordingly; the correctness of grammar and phrasing should be also ensured,
- Use decentering as a translation procedure that does not require direct translation if the original content and meaning can be kept in translated version. It implies that an item is "diluted" to discover the original concept and then to translate the concept appropriately.
- Short and simple sentences with active voice should be preferred. Avoid using as much as possible colloquialisms, the subjective mode, adverbs and prepositions indicating time or position, possessive forms, vague terms and sentences with more than one suggested variable action,
- Vocabulary should be sensitive to children and/or adolescents,
- Vocabulary should be sensitive to our language/culture,
- The concepts covered in the item at about the same level of abstraction in the two language versions,
- Consider semantic rearrangements, supplementing an item, or substituting a specific term with a synonym, maintaining the original construction/layout /format is preferable,
- The correctness of culturally specific aspects should be ensured, and
- Readability levels for children aged 8 years.

Biography

Dejan Stevanovic was born in 1979 in Serbia. He graduated Faculty of Medicine, University of Belgrade in 2004, where he also finished academic specialty studies in Neuropsychology (2007) and Child and Adolescent Psychiatry (2013). Dejan was a Fellow of the Serbian Ministry of Science in 2003-2006. Since his graduation, he attended several ECNP workshops and seminars, Salzburg CHOP seminars Berlin Summer School, VIREPA Education Program in Epilepsy, and received Donald Cohen Fellowship award. His main research interests are cross-cultural child psychiatry, cognitive neuropsychiatry, and pharmacoeconomics. Dejan serves editorial boards of several journals, is a member of different research and professional societies, and is a consultant for medical research. He has publications in various national and international journals. Currently affiliated as a clinician with Clinic for Neurology and Psychiatry for Children and Youth Belgrade, Serbia. Interested in social and cultural diversities, life coaching, traveling, and hedonism. He is married and has two sons.

Prilog 1.

Izjava o autorstvu

Potpisani-a: Dejan Stevanović

broj upisa _____

Izjavljujem

da je doktorska disertacija pod naslovom

„Applying the principles of good research practice for cross-cultural adaptation to pediatric health-related quality of life questionnaires" – „Примена принципа добре истраживачке праксе за културолошку адаптацију на педијатријске упитнике о квалитету живота"

- rezultat sopstvenog istraživačkog rada,
- da predložena disertacija u celini ni u delovima nije bila predložena za dobijanje bilo koje diplome prema studijskim programima drugih visokoškolskih ustanova,
- da su rezultati korektno navedeni i
- da nisam kršio/la autorska prava i koristio intelektualnu svojinu drugih lica.

Potpis doktoranda

U Beogradu, 16.5.2014. god

_____

Prilog 2.

Izjava o istovetnosti štampane i elektronske verzije doktorskog rada

Ime i prezime autora: Dejan Stevanović

Broj upisa:

Studijski program:

Naslov rada: „Applying the principles of good research practice for cross-cultural adaptation to pediatric health-related quality of life questionnaires" – „Примена принципа добре истраживачке праксе за културолошку адаптацију на педијатријске упитнике о квалитету живота"

Mentor: Prof. Dr Aneta Lakić

Potpisani: Dejan Stevanović

izjavljujem da je štampana verzija mog doktorskog rada istovetna elektronskoj verziji koju sam predao/la za objavljivanje na portalu **Digitalnog repozitorijuma Univerziteta u Beogradu.**

Dozvoljavam da se objave moji lični podaci vezani za dobijanje akademskog zvanja doktora nauka, kao što su ime i prezime, godina i mesto rođenja i datum odbrane rada.

Ovi lični podaci mogu se objaviti na mrežnim stranicama digitalne biblioteke, u elektronskom katalogu i u publikacijama Univerziteta u Beogradu.

Potpis doktoranda

U Beogradu, 16.5.2014. god

Prilog 3.

Izjava o korišćenju

Ovlašćujem Univerzitetsku biblioteku „Svetozar Marković“ da u Digitalni repozitorijum Univerziteta u Beogradu unese moju doktorsku disertaciju pod naslovom:

„Applying the principles of good research practice for cross-cultural adaptation to pediatric health-related quality of life questionnaires“ – „Примена принципа добре истраживачке праксе за културолошку адаптацију на педијатријске упитнике о квалитету живота“

koja je moje autorsko delo.

Disertaciju sa svim prilozima predao/la sam u elektronskom formatu pogodnom za trajno arhiviranje.

Moju doktorsku disertaciju pohranjenu u Digitalni repozitorijum Univerziteta u Beogradu mogu da koriste svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons) za koju sam se odlučio/la.

1. Autorstvo

2. Autorstvo - nekomercijalno

3. Autorstvo – nekomercijalno – bez prerade

4. Autorstvo – nekomercijalno – deliti pod istim uslovima

5. Autorstvo – bez prerade

6. Autorstvo – deliti pod istim uslovima

(Molimo da zaokružite samo jednu od šest ponuđenih licenci, kratak opis licenci dat je na poleđini lista).

Potpis doktoranda

U Beogradu, 16.5.2014.