

Zürcher Hochschule
für Angewandte Wissenschaften



Bachelorarbeit

Assessments zur Erfassung der Lebensqualität von Kindern und Jugendlichen

**Ein Literaturreview über die Reliabilität, Validität und
Praktikabilität des European Quality of Life, Pediatric Life Inventory
4.0 und KIDSCREEN-52**

Gürber Catherine S12-478-830

Departement:	Gesundheit
Institut:	Institut für Ergotherapie
Studienjahrgang:	2012
Eingereicht am:	31.05.2016
Betreuende Lehrperson:	Prof. Dr. phil. Gantschnig Brigitte

Diese Bachelorarbeit wurde genderneutral geschrieben.

Mit einem * markierte Wörter sind im Abkürzungsverzeichnis oder Glossar erläutert.

Es wird hauptsächlich von Kindern gesprochen, worunter Personen bis zum vollendeten 18. Lebensjahr verstanden werden. Werden Kinder und Jugendliche separat beschrieben, wird in der Samplebeschreibung die Gruppenzugehörigkeit definiert.

Ist die Rede von Lebensqualität, so wird darunter die gesundheitsbezogene Lebensqualität verstanden.

ABSTRACT

Die Lebensqualität gewinnt in der ergotherapeutischen Praxis und Forschung an Bedeutung. Einerseits wird die kindliche Lebensqualität mittels ergotherapeutischen Interventionen gefördert, andererseits wird die Lebensqualitätserfassung für gesundheitsökonomische Evaluationen empfohlen. Folglich ist die Verfügbarkeit von zuverlässigen und gültigen Assessments unabdingbar.

In dieser Arbeit werden die psychometrischen Eigenschaften des EuroQoL-5D-Youth, PedsQLTM4.0 und KIDSCREEN-52 überprüft und als Literaturreview von sechs Studien dargestellt. Abschliessend werden Handlungsempfehlungen und Implikation für die Ergotherapie ausgesprochen.

Die Literatur wurde mittels formulierten Schlüsselwörtern in den Datenbanken Cinahl, Medline, ERIC, Psych.Info, Cochrane Library, Web of Science und ProQuest gefunden. Dabei wurden aus 7'233 Treffern sechs Studien ausgewählt.

Im folgenden Literaturreview wurde für das EuroQoL-5D-Youth mehrheitlich moderate Test-Retest-Reliabilität, moderate Konvergenzvalidität bei Korrelationen zwischen zwei, psychologische Merkmale erfassende, Dimensionen und gute Praktikabilität festgestellt. Gute interne Konsistenz, Konstruktvalidität und Praktikabilität wurde dem PedsQLTM4.0 nachgewiesen. Das KIDSCREEN-52 wies gute interne Konsistenz und Test-Retest-Reliabilität, moderate bis hohe Konvergenzvalidität auf, zudem wurden die Eindimensionalität jeder Dimension und die multidimensionale Assessmentstruktur bestätigt.

Nach Analyse der Gütekriterien scheint das PedsQLTM4.0 für die Erfassung der kindlichen Lebensqualität am geeignetsten zu sein, weshalb dieses zur Verwendung im ergotherapeutischen Alltag empfohlen wird.

Schlüsselwörter: Lebensqualität, Kinder, EuroQoL-5D-Youth, KIDSCREEN-52, PedsQLTM4.0, Validität, Reliabilität, Praktikabilität

INHALTSVERZEICHNIS

1	<i>EINLEITUNG</i>	1
1.1	Thematische Einführung	1
1.2	Ergotherapeutische Relevanz.....	1
1.3	Relevanz des Themas und der Problemstellung	2
1.4	Fragestellung und Zielsetzung	2
1.5	Thematische Abgrenzung	3
2	<i>THEORETISCHER HINTERGRUND</i>	4
2.1	Lebensqualität	4
2.2	Assessments	4
2.3	Reliabilität, Validität und Praktikabilität	5
2.4	Gewählte Assessments	7
2.4.1	EuroQol-5D-Youth	7
2.4.2	PedsQL™4.0	8
2.4.3	KIDSCREEN-52	8
3	<i>METHODE</i>	10
3.1	Übergeordnete Methodik	10
3.2	Literatursuche und Assessmentauswahl.....	10
3.3	Literaturauswahl	12
3.4	Literaturbearbeitung	14
4	<i>RESULTATE</i>	15
4.1	Hauptstudie 1 – Ravens-Sieberer et al. (2010)	15
4.1.1	Ziel der Studie	15
4.1.2	Sample.....	15
4.1.3	Datenerfassung	15
4.1.4	Ergebnisse*	16
4.1.1	Limitationen	19
4.1.2	Zusammenfassung.....	19
4.2	Hauptstudie 2 – Canaway und Frew (2013)	20
4.2.1	Ziel der Studie	20
4.2.2	Sample.....	20
4.2.3	Datenerfassung	20
4.2.4	Ergebnisse	20

4.2.5	Limitationen	21
4.2.6	Zusammenfassung.....	22
4.3	Hauptstudie 3 – Varni et al. (2003).....	22
4.3.1	Ziel der Studie	22
4.3.2	Sample.....	22
4.3.3	Datenerfassung	23
4.3.4	Ergebnisse	24
4.3.5	Limitationen	26
4.3.6	Zusammenfassung.....	27
4.4	Hauptstudie 4 – Varni, Seid und Burwinkle (2006)	27
4.4.1	Ziel der Studie	27
4.4.2	Sample.....	27
4.4.3	Datenerfassung	28
4.4.4	Ergebnisse	29
4.4.5	Limitationen	31
4.4.6	Zusammenfassung.....	32
4.5	Hauptstudie 5 – Robitail et al. (2006)	32
4.5.1	Ziel der Studie	32
4.5.2	Sample.....	32
4.5.3	Datenerfassung	33
4.5.4	Ergebnisse	33
4.5.5	Limitationen	35
4.5.6	Zusammenfassung.....	35
4.6	Hauptstudie 6 – Ravens-Sieberer et al. (2008)	36
4.6.1	Ziel der Studie	36
4.6.2	Sample.....	36
4.6.3	Datenerfassung	37
4.6.4	Ergebnisse	37
4.6.5	Limitationen	39
4.6.6	Zusammenfassung.....	39
5	DISKUSSION	40
5.1	Diskussion der Hauptstudien	40
5.2	Diskussion der Assessments	41
5.2.1	EuroQoL-5D-Youth	43
5.2.2	PedsQL™4.0	45
5.2.3	KIDSCREEN-52	47

5.3	Limitationen der Arbeit	49
5.4	Weiterführende Forschung	50
5.5	Schlussfolgerung mit Implikation für die Ergotherapie.....	50
6	QUELLEN	53
6.1	Literaturverzeichnis.....	53
6.2	Abbildungsverzeichnis	59
6.3	Tabellenverzeichnis	59
A	WORTZAHL.....	61
B	ABKÜRZUNGSVERZEICHNIS.....	62
C	GLOSSAR.....	63
D	DANKSAGUNG.....	76
E	EIGENSTÄNDIGKEITSERKLÄRUNG.....	77
F	KRITISCHE WÜRDIGUNG DER HAUPTSTUDIEN.....	78
G	SEARCH-HISTORY.....	106

1 EINLEITUNG

Diese Bachelorarbeit gibt einen Überblick über die Reliabilität, Validität und Praktikabilität von drei Assessments zur Erfassung der Lebensqualität von Kindern. Die folgenden Kapitel dienen der thematischen Einordnung und Abgrenzung. Ebenfalls erläutert wird die Fragestellung und Zielsetzung.

1.1 Thematische Einführung

Die Lebensqualitätsforschung gewinnt in den letzten Jahren zunehmend an Bedeutung (Fayers & Machin, 2000; Ravens-Sieberer, Ellert & Erhart, 2007; Giani, Schanz & Hiemke, 2012). Dabei wurde der Begriff ‚Lebensqualitätsforschung‘ in den 1970er Jahren durch philosophische Arbeiten geprägt. Etwa zehn Jahre später wurden bereits Assessments entwickelt, welche wiederum circa zehn Jahre später in klinischen Studien an Klientinnen und Klienten getestet wurden (Giani et al., 2012). Seither wurden etablierte Verfahren adaptiert und zahlreiche neuartige Assessments entwickelt. Bereits 1998 waren schon über 800 Assessments zur Erfassung der Lebensqualität dokumentiert (Schmeck, 1998). Verglichen mit der erwachsenen Lebensqualitätsforschung, befinden sich die kindliche Lebensqualitätsforschung noch in ihren Kinderschuhen.

1.2 Ergotherapeutische Relevanz

Gemäss dem ErgotherapeutInnen-Verband Schweiz [EVS] (2012) wird die Lebensqualität von Klientinnen und Klienten durch ergotherapeutische Interventionen verbessert. Hierbei werden Alltagsfunktionen* trainiert, welche zur Selbstständigkeit beitragen (World Federation of Occupational Therapy [WFOT], 2012). Dabei darf die Selbstständigkeit als Teilaspekt der Lebensqualität betrachtet werden (Berting-Hüneke, n.d.).

Assessments sind Kernpunkte der Ergotherapie. Gemäss Fawcett (2007) werden mit den mittels Assessments erhobenen Daten ergotherapeutische Interventionen

geplant und deren Outcome überprüft. Dadurch wird ermöglicht, ergotherapeutische Interventionen auf deren Wirksamkeit zu überprüfen, was einen Teilaspekt der gesundheitsökonomischen Evaluationen* darstellt.

Gemäss Eiser und Morse (2001a) reicht es zudem nicht aus, wenn die Lebensqualität von Kindern ausschliesslich fremdeingeschätzt wird. Denn Eiser und Morse (2001b) zeigten Unterschiede zwischen den Antworten der Kinder und den Proxys (Konkordanz* der Assessmentmethoden*). Dadurch wird der immense Bedarf an Assessments, mit welchen Kinder ihre Lebensqualität selber beurteilen können, ist somit begründet.

1.3 Relevanz des Themas und der Problemstellung

Die besagten 800 Assessments zur Erfassung der Lebensqualität sind gemäss Schmeck (1998) nicht alle für die Anwendung bei gesunden Kindern geeignet, da Kinder im Vergleich zu Erwachsenen das Gesund- und Kranksein anders wahrnehmen (Seiffge-Krenke, 1990, zit. nach Ravens-Sieberer, 2000, S. 200). Als Erklärung dafür dient die emotionale und kognitive Entwicklung von Kindern (Ravens-Sieberer, 2000).

Gemäss Ravens-Sieberer (2000) wurde die kindliche Lebensqualitätsforschung in der Vergangenheit im Vergleich zur erwachsenen Lebensqualitätsforschung fälschlicherweise vernachlässigt. Die Erfassung der Lebensqualität wird jedoch zunehmend empfohlen, um einerseits ein übersichtlicheres Gesamtbild zu erhalten und andererseits, um die Therapieerfolge evaluieren zu können (Fayers & Machin, 2013). Somit stellt dieses Thema ein eher jüngeres Forschungsfeld dar (Ravens-Sieberer & Bullinger, 1998).

1.4 Fragestellung und Zielsetzung

Die Fragestellung dieser Bachelorarbeit lautet:

Sind die Assessments EuroQol-5D-Youth, PedsQLTM4.0 und KIDSCREEN-52 valide, reliabel und praktikabel?*

Hierzu werden die Gütekriterien der Assessments EuroQoL-5D-Youth (Wille et al., 2010), Pediatric Quality of Life Inventory™4.0 (Varni, 1998) und KIDSCREEN-52 (Ravens-Sieberer et al., 2005) analysiert und kritisch beurteilt. Die Darstellung der Reliabilität, Validität und Praktikabilität der drei Assessments erfolgt als Literaturreview*, welches sechs Hauptstudien* umfasst. Anschliessend werden Schlussfolgerungen und Implikationen für Forschung und Praxis abgeleitet.

1.5 Thematische Abgrenzung

Diese Arbeit thematisiert ausschliesslich Instrumente zur Erfassung der Lebensqualität von Kindern (das heisst von Personen, die ihr 18. Lebensjahr noch nicht beendet haben; Bundesamt für Gesundheit [BAG], n.d. a). Dabei wird auf die Erfassung von lebensqualitätsfördernden Interventionen* nicht eingegangen und es werden Assessments ausgeschlossen, die ausschliesslich auf Erwachsene zugeschnitten sind.

Es soll ein Überblick über drei Assessments gegeben werden, welche sowohl bei gesunden wie auch bei kranken Kindern anwendbar sind. Deshalb werden keine Assessments eingeschlossen, welche zur Erfassung der Lebensqualität von Kindern mit einer spezifischen Erkrankung (u.a. PedsQL™4.0 Cancer-Module; Varni, Burwinkle, Katz, Meeske & Dickinson, 2002) entwickelt wurden.

Zudem werden nur Assessments beurteilt, die für Kinder beider Geschlechter sowie mit unterschiedlichen kulturellen Hintergründen einsetzbar sind. Des Weiteren wurden Assessments ausgeschlossen, wovon zum Zeitpunkt der Auswahl keine deutsche Übersetzung existierte oder keine Selbstversion* entwickelt wurden, womit die Kinder ihre Lebensqualität selber beurteilen können.

2 THEORETISCHER HINTERGRUND

In den folgenden Kapiteln werden die relevanten Begriffe und Kernthemen der vorliegenden Arbeit eingeführt.

2.1 Lebensqualität

Da verschiedene Definitionen des Begriffes Lebensqualität existieren (in Anlehnung an BAG, n.d.b), wird im Rahmen dieser Arbeit folgende konsolidierte Definition verwendet:

Die gesundheitsbezogene Lebensqualität beschreibt das subjektive Wohlbefinden eines Menschen (Giani et al., 2012) und wird in mehrere Dimensionen unterteilt. Entsprechend beinhaltet die gesundheitsbezogene Lebensqualität physische, psychische, soziale, emotionale und verhaltensbezogene Komponenten (Schmeck, 1998). Die Begriffe gesundheitsbezogene Lebensqualität, subjektive Gesundheit und Lebensqualität werden in Anlehnung an Patrick und Erickson (1993, zit. nach Koller et al., 2009, S. 866) in dieser Arbeit synonym verwendet.

Gemäss Porzsolt (1997, zit. nach Giani et al., 2012, S. 35) ist die Lebensqualität vergleichbar mit der Differenz zwischen dem IST- und dem SOLL-Zustand* des psychischen Wohlbefindens. Hierbei beschreibt der SOLL-Zustand die individuellen Ansprüche der Person, der IST-Wert beschreibt die aktuelle Realität. Besteht eine grosse Abweichung zwischen SOLL- und IST-Wert, wobei der IST-Zustand* niedriger ist, deutet dies auf eine tiefe Lebensqualität hin.

2.2 Assessments

Mittels Assessments werden Informationen über die Performanz* einer Klientin oder eines Klienten beschrieben, analysiert und validiert (Centre for Advanced Palliative Care, 2005, zit. nach Fawcett, 2007, S. 4). Hierbei handelt es sich um einen Prozess, in welchem Ressourcen* und Schwierigkeiten* einer Person erfasst werden (Fawcett, 2007).

2.3 Reliabilität, Validität und Praktikabilität

Assessments können mittels Hauptkriterien (Reliabilität, Objektivität und Validität) und Nebenkriterien (u.a. Praktikabilität) beurteilt werden (Lienert & Raatz, 1998). In dieser Bachelorarbeit findet die Objektivität keine eigene Bedeutung, da laut Döring und Bortz (2016) die Objektivität als Voraussetzung für die Reliabilität gilt.

Gemäss Lienert und Raatz (1998) beschreibt die Reliabilität, wie zuverlässig die einzelnen Merkmale gemessen werden. Dabei ist es irrelevant, ob die gemessenen Merkmale bewusst oder unbewusst erfasst wurden. Nach Trochim (2006a) werden hauptsächlich vier Teilaspekte unterschieden: Interne Konsistenz, Test-Retest-Reliabilität, Parallel-Form-Reliabilität und Inter-Rater-Reliabilität (Tabelle 1).

Tabelle 1

Beschreibung der Reliabilität

	Beschreibung	Tests und Kennwerte	Quelle
Test-Retest-Reliabilität	Indem der Test nach einiger Zeit wiederholt wird, wird die Stabilität des Assessments gemessen. Die Ergebnisse werden miteinander korreliert und das Level der Übereinstimmung mittels Kappa-Koeffizient (Cohen, 1968) berechnet. Dabei muss die Testwiederholung unter möglichst identischen Bedingungen stattfinden, ansonsten werden die Messwerte verzerrt (Döring & Bortz, 2016).	Cohen's Kappa: <.20 schlecht, .21-.40 ausreichend, .41-.60 moderat, .61-.80 substantiell, ≥.81 fast perfekt Spearman-Rangkorrelation: $\rho = <.30$ tief, .30-.49 moderat, >.49 hoch	Landis & Koch (1977). Cohen (1968). Bartko (1966).
Inter-Rater-Reliabilität	Gepprüft wird, ob verschiedene Personen zu denselben Schlussfolgerungen gelangen. Somit wird gemessen, ob die Items gleich beantwortet werden, auch wenn unterschiedliche Rater den Test durchführen.	Spearman-Rangkorrelation: $\rho = <.30$ tief, .30-.49 moderat, >.49 hoch Pearson-Produkt-Moment-Korrelation: .10 klein, .30 medium, .50 gross	Bartko (1966). Cohen (1988).
Interne Konsistenz	Überprüft wird, ob alle Items, welche derselben Dimension angehören, dasselbe Merkmal messen und somit die Dimension gültig erfasst wird.	Cronbach's Alpha: .70 akzeptabel, >.89 sehr gut	Cronbach (1951) Nunnally & Bernstein (1994)

Anmerkung. Es werden ausschliesslich diejenigen Gütekriterien näher beschrieben, welche zum Verständnis der Hauptstudien relevant sind. Andernfalls wird auf das Glossar verwiesen.

Die Validität sagt aus, ob ein Test die Merkmale misst, die er zu messen vorgibt (Lienert & Raatz, 1998). Hierfür ist die Genauigkeit des Messerfolgs irrelevant. Nach Trochim (2006b) werden hauptsächlich drei Teilaspekte differenziert: Konstrukt*-, Kriteriums*- und Inhaltsvalidität (Tabelle 2).

Tabelle 2

Beschreibung der Konstruktvalidität

Konvergenzvalidität: Zwei Dimensionen, welche ähnliche Merkmale messen, werden miteinander korreliert. Dabei wird berechnet, inwiefern dieselben Merkmale gemessen werden. 1.0 (100%) würde bedeuten, dass dieselben Merkmale erfasst werden.	Spearman-Rangkorrelation: $\rho = <.30$ tief, $.30 - .49$ moderat, $>.49$ hoch	Bartko (1966).
Divergenzvalidität: Zwei Dimensionen, welche jeweils unterschiedliche Merkmale erfassen, werden miteinander korreliert. Dabei wird berechnet, inwiefern unterschiedliche Merkmale erfasst werden. Das Vorhandensein von divergenter Validität wird mit möglichst tiefen Korrelationen unterstützt.	Pearson-Korrelationen: $r = .1 - .3$ tief, $.31 - .50$ moderat, $>.50$ hoch	Cohen (1988).
Bekannt-Gruppen-Validität: Zwei sich in einem Merkmal unterscheidende Gruppen (z.B. Kinder mit versus Kinder ohne chronisch Erkrankung) werden auf Regelmässigkeiten (u.a. ob gesündere Kinder bessere Scores erzielen) überprüft.	χ^2 -Test: Effektgrössen (d)= $.20 - .50$ klein, $.51 - .80$ mittel, $>.80$ gross;	Cohen (1988)
	<i>t</i> -Test	

Anmerkung. Es werden ausschliesslich diejenigen Gütekriterien näher beschrieben, welche zum Verständnis der Hauptstudien relevant sind. Andernfalls wird auf das Glossar verwiesen.

In dieser Arbeit werden unter Praktikabilität die ökonomischen Eigenschaften der Tests verstanden. Gemäss Lienert und Raatz (1998) sind das folgende:

- Kurze Zeitdauer
- Effektive Verwendung von Ressourcen (wenig Material notwendig, Wiederverwendung des Materials möglich, Gruppenerfassungen möglich)
- Tiefe Kosten
- Einfache Handhabung (d.h. gute Verständlichkeit, wenig Fehler)
- Einfache und schnelle Auswertung der erfassten Informationen

2.4 Gewählte Assessments

In dieser Arbeit werden folgende Assessments untersucht: EuroQoL-5D-Youth (Wille et al., 2010), PedsQL™4.0 (Varni, 1998) und KIDSCREEN-52 (Ravens-Sieberer et al., 2005). Sie wurden aus den gemäss Schmeck (1998) etwa 800 existierenden Assessments zur Erfassung der Lebensqualität ausgewählt. Das Auswahlverfahren wird in Kapitel 3.2 detailliert aufgezeigt.

2.4.1 EuroQol-5D-Youth

Das EuroQoL-5D-Youth (Wille et al., 2010) ist eine für Kinder und Jugendliche zwischen acht und 18 Jahren modifizierte Version des EuroQoL-5D (EuroQoL Research Foundation, 1990). Dabei wird die Lebensqualität von Kindern anhand von fünf Dimensionen (Tabelle 3) mit je einem Item* erfasst, wobei jedes Item anhand einer Drei-Punkte-Likert-Skala* beantwortet wird. Anschliessend beurteilen die Teilnehmenden ihren aktuellen Gesundheitszustand mittels visueller Analogskala (VAS)*. Für dieses Assessment existiert eine Selbstbeurteilungsversion, welche gemäss EuroQoL Research Foundation (2016a) in 38 Sprachen erhältlich ist, und eine Proxyversion* zur Fremdbeurteilung der Lebensqualität der Kinder. Diese ist nur in englischer, dänischer und holländischer Sprache verfügbar (EuroQoL Research Foundation, 2016b).

Tabelle 3

Beschreibung des EuroQoL-5D-Youth

Beschreibung des Items	Mögliche Antworten ^a
Bewegung (herumlaufen)	Keine / Einige / Grosse Schwierigkeiten
Für sich selbst sorgen (sich waschen, anziehen)	Keine / Einige / Grosse Schwierigkeiten
Was ich normalerweise tue (z.B. Hobbies, Sport, Schule)	Keine / Einige / Grosse Schwierigkeiten
Schmerzen oder körperliche Beschwerden (unangenehme Gefühle im Körper)	Keine / Einige / Grosse Schmerzen
Sich unglücklich, traurig oder besorgt fühlen	Nicht / Etwas / Sehr unglücklich, traurig oder besorgt

Anmerkung. Diese Tabelle stellt nicht den ganzen Test dar, es wurden lediglich einzelne Beispiele zu Veranschaulichung aufgeführt. ^aDie Antworten werden in „Ich-Form“ gegeben (z.B. Ich bin nicht unglücklich...), wobei jeweils einer der drei Antwortsätze angekreuzt werden soll. Ist nicht erkennbar, welche Antwort ausgewählt wurde, gilt dies als Fehler.

Quelle: S. Sirini (Persönliche Kommunikation mit S. Sirini am 25.5.2016).

2.4.2 **PedsQL™4.0**

Mit dem PedsQL™4.0 (Varni, 1998) kann die Lebensqualität von Zwei- bis 18jährigen beurteilt werden. Es existieren drei verschiedene Selbstversionen für Fünf- bis 18jährige. Gleichzeitig wurden Proxyversionen entwickelt, mit welchen die Lebensqualität von Zwei- bis 18jährigen fremdbeurteilt werden kann. Die Erfassung dauert etwa vier Minuten (Varni, 2016a).

Die Lebensqualität wird anhand 23 Items erfasst, welche vier Funktionsskalen zugeteilt sind (Tabelle 4). Dabei werden die Teilnehmenden schriftlich gefragt, wie oft ihnen die gefragte Tätigkeit innert des letzten Monats Probleme bereiteten. Hierfür können die Items mittels Fünf-Punkte-Likert-Skala quantifiziert werden. Anhand der Angaben wird ein Gesamtscore* berechnet, welcher reversibel in eine lineare Funktion transformiert wird (vgl. Varni, 2016b). Folglich entsprechen tiefere Gesamtscores einer höheren Lebensqualität.

Derzeit existieren über 100 Übersetzungen, welche allerdings noch nicht für sämtliche Alterskategorien übersetzt und validiert wurden (Varni, 2016c).

Tabelle 4

Beschreibung des PedsQL™4.0

Dimension	Beschreibung der Funktionsskala	Anzahl Items
Physische Funktionsskala	körperliche und sportliche Performanz, Schmerzen, Level der körperlichen Energie	8
Emotionale Funktionsskala	Gefühle (u.a. Sorgen, Ängste, emotionale Schwierigkeiten)	5
Soziale Funktionsskala	Soziale Partizipation mit anderen Kindern,	5
Schulische Funktionsskala	Erinnerungsvermögen, Handlungsplanung, Steuerung der Aufmerksamkeit, Gründe für mögliches Fernbleiben des Unterrichts	5

Anmerkung. Der Begriff Funktionsskala wird im Glossar erläutert.

Quelle: Varni (1998)

2.4.3 **KIDSCREEN-52**

Anhand dieses Assessments wird gemäss Ravens-Sieberer et al. (2005) die Lebensqualität von Acht- bis 18jährigen mittels 52 Items erfasst. Es existieren dafür sowohl

eine Selbsterfassungsverfögen als auch eine Proxyverfögen. Laut KIDSCREEN-Group (2011a) kann dieses Assessment innerf 20 Minuten durchgeföhrt werden und ist in über 30 europöischen Sprachen verfögenbar (KIDSCREEN-Group, 2011b).

Die 52 Items (Tabelle 5) werden anhand einer Fünf-Punkte-Likert-Skala* beantwortet. Ein hoher Gesamtscore steht dabei für eine hohe Lebensqualität (Ravens-Sieberer et al., 2005). Des Weiteren existieren zwei kürzere Versionen mit 27*, respektive zehn* Items (Ravens-Sieberer et al., 2014).

Tabelle 5

Beschreibung der zehn Dimensionen des KIDSCREEN-52

Dimension	Beschreibung	Anzahl Items
1. Gesundheit und Bewegung	Körperliche Aktivitäten, Fitness, Energie.	5
2. Psychisches Wohlbefinden	Mentales Wohlbefinden, positive Emotionen, Lebenszufriedenheit	6
3. Stimmung	Negative Emotionen, Stressgeföhle, depressive Verstimmung.	7
4. Wie findest du dich?	Eigene positive oder negative Körperwahrnehmung, persönliches Aussehen, Umgang mit Kleidung und Accessoires	5
5. Freizeit	Selbstständiges Ermöglichen des Lebens (Selbstständigkeit).	5
6. Familie und Zuhause	Privatleben und Beziehungen zu den Eltern	6
7. Freunde	Beziehungen zu Gleichaltrigen, Freunden, Kameraden	6
8. Schule und Lernen	Schulischen Leistungen (Konzentration, Lernen, kognitive Kapazitäten) und eigene Geföhle bezüglich der Schule.	6
9. Du und die anderen	Soziale Integration in der Schule, Mobbing	3
10. Dein Geld	Subjektive Wahrnehmung der finanziellen Ressourcen, Umgang mit Geld, Dinge kaufen.	3

Anmerkung. Die Dimensionen wurden durch die Verfasserin der Einheitlichkeit halber übersetzt.

Quelle: Ravens-Sieberer et al. (2014).

3 METHODE

Im folgenden Kapitel wird das methodische Vorgehen dieser Arbeit erläutert.

3.1 Übergeordnete Methodik

Diese Arbeit wurde als Einzelarbeit zwischen September 2015 und Mai 2016 geschrieben. Sie ist in Form eines Literaturreviews basierend auf sechs Hauptstudien aufgebaut, wobei die Hauptstudien zwischen Oktober und November 2015 gesucht wurden.

3.2 Literatursuche und Assessmentauswahl

Um geeignete Assessments zur Erfassung der Lebensqualität von Kindern zu finden, wurde initial folgende Fragestellung formuliert: *Welche Assessments zur Erfassung der Lebensqualität sind für gesunde Kinder reliabel und valide?* Zu dieser Fragestellung wurden Schlüsselwörter (Tabelle 6) definiert, welche anschliessend für die Datenbankrecherche in Cinahl, Medline, ERIC, Psych.Info, Cochrane Library, Web of Science und ProQuest mittels Trunkierungszeichen* ergänzt und mittels Bool'schen Operatoren* kombiniert wurden.

Tabelle 6

Keywordtabelle zur Suche der Hauptstudien

Keyword dt.	Keyword engl.	Synonym
Lebensqualität	Quality of Life	QOL, HRQOL, „health related Quality of life“.
Kind	children	child*, kid*, young*
Jugendliche/r	adolescent	young adult, teen*, young*
Erfassungsinstrument	Assessment	Measure, evaluation, outcome measure,
Validität	Validity	Valuation, validation, concurrent validity
Reliabilität	Reliability	stability
Praktikabilität	Practicability	Utility, Feasibility
Ergotherapie	Occupational Therapy	ergo therapy
Euro-QoL	Euro-QoL	ED-5L-Y
PedsQL	PedsQL	Pediatric Quality of Life Inventory, PedsQL
KIDSCREEN-52	KIDSCREEN-52	KIDSCREEN

Anschliessend wurden unpassende Assessments anhand folgender Kriterien exkludiert: Es musste eine deutsche Übersetzung existieren und zum Assessment musste genügend quantitative Primärliteratur* verfügbar sein, welche die Erfassung der Gütekriterien erläutert. Anhand dieser Kriterien und Gesprächen mit Experten wurden drei Assessments ausgewählt (Abbildung 1).

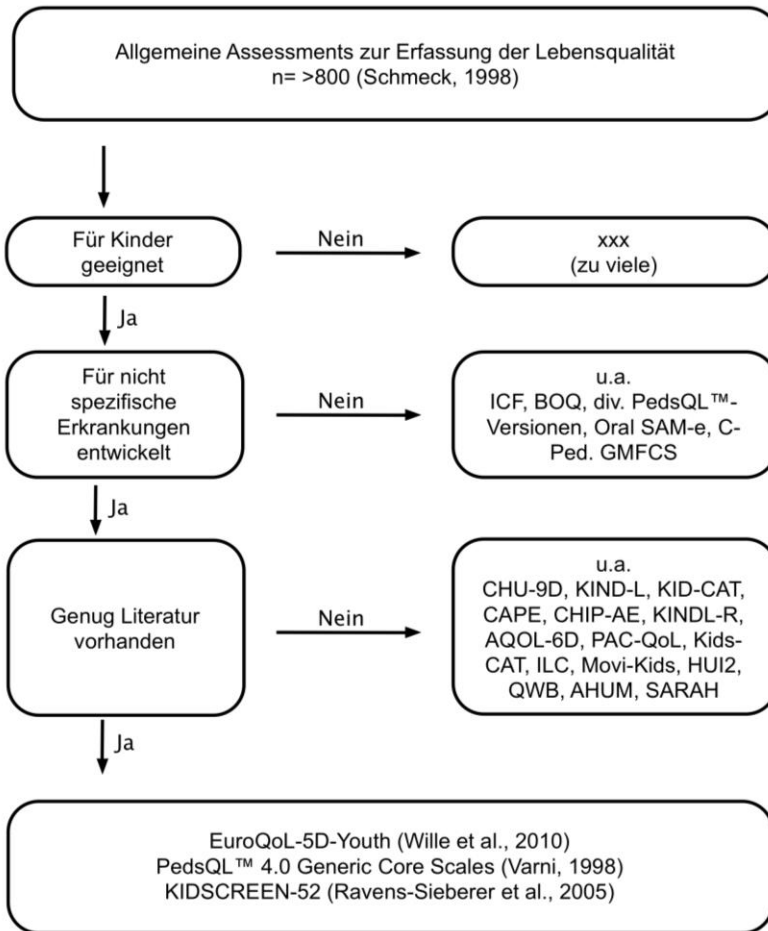


Abbildung 1. Darstellung des Auswahlprozesses der Assessments

Nach der Auswahl der Assessments wurde die Fragestellung spezifiziert. Die überarbeitete Fragestellung lautet:

Sind die Assessments EuroQoL-5D-Youth, Pediatric Quality of Life Inventory™ 4.0 und KIDSCREEN-52 reliabel, valide und praktikabel?

Gemäss dieser veränderten Fragestellung wurden die Schlüsselwörter mit den Namen der drei Assessments ergänzt. Es wurden erneut Datenbankrecherchen durchgeführt, um geeignete Primärliteratur zu finden. Die Search-History befindet sich in Anhang G.

3.3 Literatúrauswahl

Zur Literatúrauswahl wurden thematische (Tabelle 7) und methodische (Tabelle 8) Ein- und Ausschlusskriterien definiert. Es wurden nur Studien verwendet, welche die Gütekriterien eines dieser drei Assessments erläuterten. Zudem mussten pro Assessment mindestens zwei quantitative Primärquellen verfügbar sein, welche die Reliabilität, Validität und/oder Praktikabilität erfassen. Zusätzlich mussten die Hauptstudien innert den letzten zehn Jahren und bis November 2015 publiziert worden sein. Eine Ausnahme wurde für die Studie von Varni, Burwinkle, Seid und Skarr (2003) gemacht, da die Studie von Varni, Burwinkle und Seid (2006) darauf aufbaut.

Tabelle 7

Thematische Ein- und Ausschlusskriterien der Hauptstudien

Einschlusskriterien	<ul style="list-style-type: none">• Kinder und Jugendliche• Zwei- bis 18jährig• Untersuchung der Validität, Reliabilität und Praktikabilität des Assessments zur Erfassung der Lebensqualität.• Assessment mit deutscher Übersetzung
Ausschlusskriterien	<ul style="list-style-type: none">• Erwachsene• Assessments, die nicht die Lebensqualität beurteilen• Validierungen von Assessmentübersetzungen• Assessments zur Erfassung der Lebensqualität von Kindern mit spezifischen Erkrankungen• Überprüfung der Gütekriterien an einem Sample mit spezifischen Erkrankungen (u.a. ADHS, Krebs)
Irrelevant	<ul style="list-style-type: none">• Geschlecht der Teilnehmenden• Kultur der Teilnehmenden

Tabelle 8

Methodische Ein- und Ausschlusskriterien der Hauptstudien

Einschlusskriterien	<ul style="list-style-type: none"> • Primärliteratur • Quantitative Studien • Korrekter Aufbau der Studie • Verständlich erklärte Methode • Sinnvolle Verwendung der Studien • Publikationsjahr der Studien zwischen 2006 und 2016
Ausschlusskriterien	<ul style="list-style-type: none"> • Studien, welche nicht in Deutsch oder Englisch verfasst wurden • Keine Primärliteratur • Qualitative Primärliteratur

Anschliessend wurden die Titel und Abstracts der Studien gelesen und unpassende Artikel ausgeschlossen. Der Auswahlprozess wird auf Abbildung 2 dargestellt.

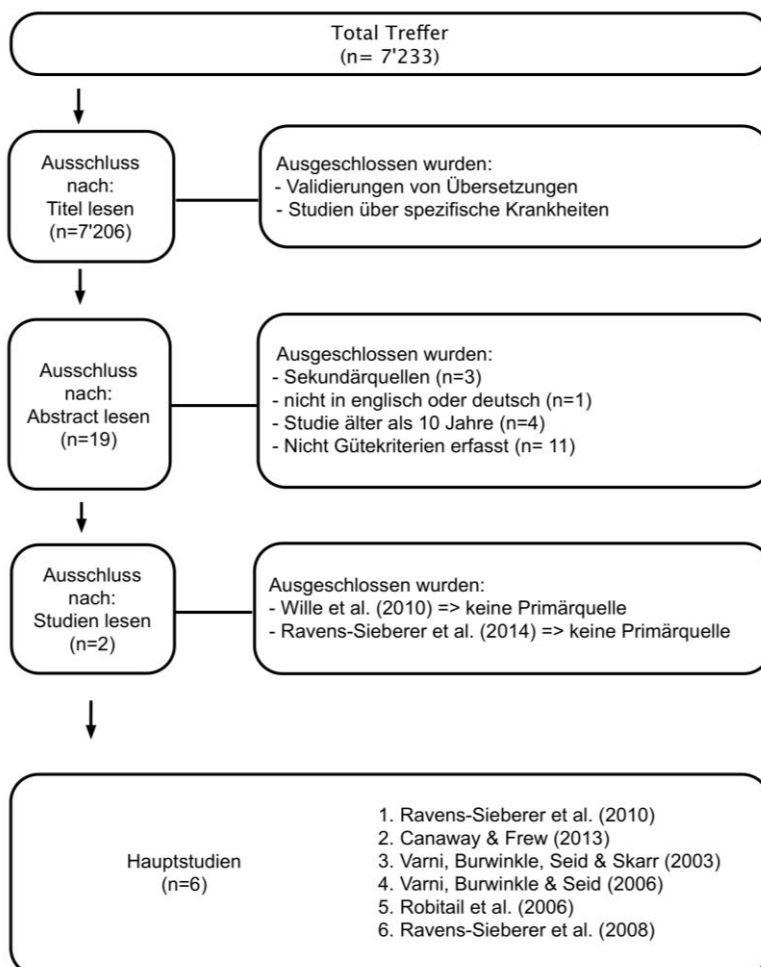


Abbildung 2. Darstellung des Auswahlprozesses der Hauptstudien

3.4 Literaturbearbeitung

Die ausgewählten sechs Hauptstudien wurden mittels der Vorlage zur kritischen Besprechung quantitativer Studien (Law et al., 1998) zusammengefasst und kritisch gewürdigt. Im Kapitel 4 werden relevante Aspekte der Hauptstudien beschrieben, die vollständigen kritischen Würdigungen der Studien befinden sich allerdings in Anhang F.

4 RESULTATE

Die isolierte Auswertung der Studien erfolgt angelehnt an die Vorlage zur kritischen Besprechung quantitativer Studien von Law et al. (1998). Dabei werden die Samples, die durchgeführten Methoden, Resultate und Messverzerrungen untersucht. Die ausführlichen Auswertungen pro Studie befinden sich in Anhang F, während in folgendem Kapitel die zentralen Resultate beleuchtet werden. Dabei werden alle Studien isoliert betrachtet. Die konsolidierten Auswertungen der Assessments sind in Kapitel 5 zu finden.

4.1 Hauptstudie 1 – Ravens-Sieberer et al. (2010)

Feasibility, reliability, and validity of the EuroQoL-5D-Youth: results from a multinational study
(Ravens-Sieberer et al., 2010)

4.1.1 *Ziel der Studie*

In dieser Studie wurden die Reliabilität, Validität und Praktikabilität des EuroQoL-5D-Youth (Wille et al., 2010) an einem kulturübergreifenden Sample analysiert.

4.1.2 *Sample*

Aus Regelschulen in fünf Ländern wurden 2'809 Teilnehmende (acht bis 19 Jahre) rekrutiert (Tabelle 9). Die Testwiederholung wurde mit 33% der spanischen und italienischen Teilnehmenden nach sieben bis zehn Tagen durchgeführt.

4.1.3 *Datenerfassung*

Das methodische Vorgehen war in den einzelnen Ländern sehr unterschiedlich, weshalb es in Tabelle 9 separat beschrieben wird.

Tabelle 9

Sample und Datenerfassung

	Deutschland	Italien	Spanien	Südafrika	Schweden
Teilnehmende	n=756 w: 371	n=415 w: 216	n=973 w:479	n=258 w: 128	n=407 w: 199
Alter	10 – 18 M=13.8 ± 1,9	8 – 15 M=11.8 ± 2,2	8 – 18 M=13.0 ± 2.7	13 – 19 M=15.5 ± 1,3	8 – 16 M=13.2 ± 2.7
Methodik	FB ohne pers. Unterstützung, in Schule, mit schriftl. Anleitung ausgefüllt.	FB mit pers. Unterstützung bei Fragen während des Unterrichts, mit schriftl. Anleitung ausgefüllt.	FB in Schule, ohne pers. Unterstützung, schriftl. Anleitung ausgefüllt.	FB ohne pers. Unterstützung, in Schule, mit schriftl. Anleitung ausgefüllt.	FB von der Gesundheitsbehörde an Familien versendet (inkl. Begleitbrief mit schriftl. Anleitung), ohne pers. Unterstützung ausgefüllt.
Sample	Zufällig	Angepasst	angepasst	angepasst	angepasst
Data	Konstruktvalidität, VAS	Konstruktvalidität, VAS, Test-Retest-Reliabilität	Konstruktvalidität, VAS, Test-Retest-Reliabilität	Konstruktvalidität, VAS	Konstruktvalidität, VAS

Anmerkung. w weiblich, M Mittelwert/Durchschnittsalter, FB Fragebogen, schriftl. schriftlich, pers. personelle, VAS visuelle Analogskala.

Quelle: Ravens-Sieberer et al., 2010, S. 889.

4.1.4 **Ergebnisse**

Test-Retest-Reliabilität: Es wurden mehrheitlich ausreichende bis moderate Cohen's Kappa-Koeffizienten* (Cohen, 1968; Landis & Koch, 1977) berechnet. Im italienischen Retest-Sample* waren alle Ergebnisse* für die Dimension ‚Selbstfürsorge‘ identisch, weshalb kein Kappa-Koeffizient berechnet werden konnte. Zusätzlich waren die Messungen der Dimension ‚Mobilität‘ nicht signifikant (Tabelle 10).

Tabelle 10

Test-Retest-Reliabilität des EuroQoL-5D-Youth (Wille et al., 2010)

Dimension	Ravens-Sieberer et al. (2010)	Canaway und Frew (2013)
,Mobilität'	IT: $K=.22$; $p=.3$ (91.5%)	$K=.39$; $p<.05$ (87.5%)
	ES: $K=-.00$; $p\leq.01$ (99.4%)	$\rho=.24$; $p>.1$
,Selbstfürsorge'	IT: $K=0.0^a$ (93.8%)	$K=.17$; $p>.1$ (70.8%)
	ES: $K=.67$; $p\leq.01$ (99.7%)	$\rho=.47$; $p<.05$
,Alltagsaktivitäten'	IT: $K=.55$; $p\leq.01$ (82.9%)	$K=.48$; $p<.05$ (89.6%)
	ES: $K=.56$; $p\leq.01$ (97.5%)	$\rho=-.61$; $p<.05$
,Schmerzen'	IT: $K=.35$; $p\leq.01$ (69.8%)	$K=.25$; $p<.1$ (77.1%)
	ES: $K=.44$; $p\leq.01$ (86.2%)	$\rho=.35$; $p>.1$
,Angst und Depression'	IT: $K=.35$; $p\leq.01$ (78.3%)	$K=.30$; $p<.05$ (72.9%)
	ES: $K=.47$; $p\leq.01$ (87.4%)	$\rho=.35$; $p>.1$
Intra-Klassen-Korrelation (Bartko, 1966) mit VAS	IT: $r=.82$ ES: $r=.83$	xxx

Anmerkung. Die Ergebnisse von Ravens-Sieberer et al. (2010) wurden denjenigen von Canaway & Frew (2013) gegenübergestellt, wodurch einerseits die Ergebnisse dargestellt werden und andererseits ein direktes Vergleichen für Kapitel 5 ermöglicht wird.

^aKappa konnte aufgrund identischer Ergebnisse nicht berechnet werden.

IT: Italienisches Retest-Sample, ES: Spanisches Retest-Sample, %: Prozent Übereinstimmung.

VAS visuelle Analogskala CHU-9D (Stevens, 2010), PedsQL™4.0 (Varni, 1998).

Quelle: Ravens-Sieberer et al., 2010; Canaway & Frew, 2013.

Konvergenzvalidität: Moderate bis hohe Spearman-Rang-Korrelationen (Cohen, 1988) wurden berechnet, wenn zwei Dimensionen miteinander korreliert wurden, die psychologische Merkmale erfassen. Zudem lassen die moderaten bis hohen Korrelationen zwischen VAS und dem KIDSCREEN-52 (Ravens-Sieberer et al., 2005) auf konvergente Validität schliessen (Tabelle 11).

Bekanntegruppenvalidität: Mittels χ^2 -Tests* wurden Unterschiede zwischen chronisch erkrankten und gesünderen Teilnehmenden aufgezeigt. Diesbezüglich wurde festgestellt, dass gesündere oder sich gesünder fühlende Kinder bessere Werte (= höhere Lebensqualität) erzielten (Tabelle 11).

Tabelle 11

Validität des EuroQoL-5D-Youth (Wille et al., 2010)

	Ravens-Sieberer et al. (2010)	Canaway & Frew (2013)																														
Konvergenzvalidität	Moderate bis hohe Korrelationen zwischen Dimensionen, welche psychisches Wohlbefinden thematisieren: $\rho = -.41$ bis $-.52$.	Bestätigung konvergenter Validität bei 8 von 9 Hypothesen.																														
	Moderate Korrelationen zwischen der ‚emotionalen FS‘ (PedsQL™4.0) und ‚Angst und Unwohlsein‘ : $\rho = -.47$.	<table border="1"> <thead> <tr> <th>EuroQoL-5D-Youth</th> <th>Korrelationsassessment</th> <th>ρ</th> </tr> </thead> <tbody> <tr> <td>Mobilität</td> <td>Activities^b</td> <td>.22^a</td> </tr> <tr> <td>Selbstfürsorge</td> <td>Routine^b</td> <td>.34^a</td> </tr> <tr> <td>Alltagsaktivitäten</td> <td>Activities^b</td> <td>.38^a</td> </tr> <tr> <td>Schmerzen</td> <td>Pain^b</td> <td>.54^a</td> </tr> <tr> <td>Sorgen</td> <td>Sad^b</td> <td>.38^a</td> </tr> <tr> <td>Sorgen</td> <td>Worried^b</td> <td>.34^a</td> </tr> <tr> <td>Schmerzen</td> <td>Phys. FS^c</td> <td>.35^a</td> </tr> <tr> <td>Mobilität</td> <td>Phys. FS^c</td> <td>.15</td> </tr> <tr> <td>Sorge</td> <td>Psych. FS^c</td> <td>-.29^a</td> </tr> </tbody> </table>	EuroQoL-5D-Youth	Korrelationsassessment	ρ	Mobilität	Activities ^b	.22 ^a	Selbstfürsorge	Routine ^b	.34 ^a	Alltagsaktivitäten	Activities ^b	.38 ^a	Schmerzen	Pain ^b	.54 ^a	Sorgen	Sad ^b	.38 ^a	Sorgen	Worried ^b	.34 ^a	Schmerzen	Phys. FS ^c	.35 ^a	Mobilität	Phys. FS ^c	.15	Sorge	Psych. FS ^c	-.29 ^a
EuroQoL-5D-Youth	Korrelationsassessment	ρ																														
Mobilität	Activities ^b	.22 ^a																														
Selbstfürsorge	Routine ^b	.34 ^a																														
Alltagsaktivitäten	Activities ^b	.38 ^a																														
Schmerzen	Pain ^b	.54 ^a																														
Sorgen	Sad ^b	.38 ^a																														
Sorgen	Worried ^b	.34 ^a																														
Schmerzen	Phys. FS ^c	.35 ^a																														
Mobilität	Phys. FS ^c	.15																														
Sorge	Psych. FS ^c	-.29 ^a																														
	Moderate bis hohe Korrelationen zwischen VAS und KIDSCREEN-27 / KIDSCREEN-10: $\rho = .33$ – $.56$.																															
	Tiefe Korrelation zwischen psychischer / physischer FS und Mobilität: $\rho = -.1$.																															
	Tiefe Korrelationen zwischen ‚körperliches Wohlbefinden‘ (KIDSCREEN-27) und ‚Mobilität‘ oder ‚Schmerzen und Unbehagen‘ ($\rho < .3$).																															
	Dimensionen ‚Selbstfürsorge‘ und ‚Alltagsaktivitäten‘ korrelierten mit keinem Assessment.																															
Be-kannte-Gruppen-Validität	<p>Weniger Schwierigkeiten bereiteten die gefragten Tätigkeiten den TN ...</p> <ol style="list-style-type: none"> mit höherem (versus TN mit tieferem) selbstbeurteilten Gesundheitszustand. ohne (versus TN mit) chronischer Beeinträchtigung. ohne (versus TN mit) mentalen Beeinträchtigungen. <p>(χ^2-Test-Ergebnisse wurden nicht angegeben)</p>	Gesündere TN erreichten höhere Werte (hatten weniger Schwierigkeiten bei gefragten Alltagstätigkeiten)																														

Anmerkung. Die Ergebnisse von Ravens-Sieberer et al. (2010) wurden denjenigen von Canaway & Frew (2013) gegenübergestellt, wodurch einerseits die Ergebnisse dargestellt werden und andererseits ein direktes Vergleichen für Kapitel 5 ermöglicht wird.

CHU-9D (Stevens, 2010), PedsQL™4.0 (Varni, 1998). ^a signifikant bei $p < .01$. ^b Dimension des CHU-9D. ^c Funktionsskala (FS) des PedsQL™4.0.

Quelle: Ravens-Sieberer et al., 2010; Canaway & Frew, 2013.

Praktikabilität: Es wurden 0-2% Fehler (= undeutlich oder nicht beantwortete Items) beim Beantworten der Items und 0-9% Fehler bei der Beurteilung des eigenen Gesundheitszustandes festgestellt, was einer guten Praktikabilität entspricht.

In allen Dimensionen wurden hohe Deckeneffekte* (>51%) beobachtet.

4.1.1 Limitationen

In Tabelle 12 werden die Limitationen und Bias* der Studie erläutert.

Tabelle 12

Limitationen und Bias in Ravens-Sieberer et al. (2010)

Limitationen	Bias
<ul style="list-style-type: none">• Nicht geprüft wurde, ob das Assessment alle Aspekte der gesundheitsbezogenen Lebensqualität beinhaltet (Inhaltsvalidität), welche für Kinder relevant sind.	<ul style="list-style-type: none">• Angaben bezüglich Signifikanzlevel fehlen.• Die Sample der Bekannte-Gruppen-Validität waren nicht gleich gross, somit können die Werte nicht direkt miteinander verglichen werden.• Die fehlende Angabe des Standard Error of Measurement könnte zu Messfehlern führen.• Test-Retest-Reliabilität nur an 2 Untersamples geprüft.• Sensitivität nicht geprüft.• Unterschiedliche Durchführungsmethoden zur Datenerfassung verwendet

Anmerkung. Die Tabellenstruktur wurde in Anlehnung an die Struktur von Fässler & Marchner (2010) geschrieben. Unter Limitationen werden Fehler verstanden, welche durch die Autorenschaft vermerkt wurden. Bias sind mögliche Fehler, welche nicht in den Studien dokumentiert wurden.

Test-Sensitivität wird im Glossar erläutert.

Quelle: Ravens-Sieberer et al. (2010).

4.1.2 Zusammenfassung

In der Studie wurden ausreichende bis moderate Kappa-Koeffizienten (Cohen, 1968) aufgezeigt. Ebenfalls wurden hohe Korrelationskoeffizienten (Cohen, 1988) zwischen zwei Dimensionen beobachtet, die psychologische Aspekte messen. Folglich korrelierte die ‚Mobilität‘ mit der ‚physischen Funktionsskala*‘ des PedsQL™4.0 (Varni, 1998) tief. Zuletzt enthielten wenig der retournierten Umfragen Fehler, was einer guten Praktikabilität entspricht.

4.2 Hauptstudie 2 – Canaway und Frew (2013)

Feasibility, reliability, and validity of the EuroQoL-5D-Youth: results from a multinational study
(Ravens-Sieberer et al., 2010)

4.2.1 *Ziel der Studie*

Es wurden die Gütekriterien des EuroQoL-5D-Youth (Wille et al., 2010) und des CHU-9D (Stevens, 2010) erfasst und miteinander verglichen. In der vorliegenden Arbeit werden ausschliesslich die Ergebnisse zum EuroQoL-5D-Youth dargestellt.

4.2.2 *Sample*

An der Studie nahmen 160 Sechs- und Siebenjährige aus den West Midlands (Birmingham) teil. 43% der Teilnehmenden waren Mädchen, 63% waren sechsjährig. Der Retest wurde an 24 Teilnehmenden am Nachmittag desselben Testtags durchgeführt.

4.2.3 *Datenerfassung*

Jedes Kind beantwortete die Selbstversionen zu drei Assessments (EuroQoL-5D-Youth, Wille et al., 2010; CHU-9D, Stevens, 2010; PedsQL™4.0, Varni, 1998) als Interview, welches durch Forschungsmitarbeitende geleitet wurde. Dabei erhielten alle Kinder eine anonyme Studien-Identität, welche mit den zu lösenden Assessments verlinkt wurde.

4.2.4 *Ergebnisse*

Test-Retest-Reliabilität: Ausser für die Dimension ‚Selbstfürsorge‘ ($\kappa=.17$; $p>.1$) wurden ausreichende Kappa-Koeffizienten (Cohen, 1968; Landis & Koch, 1977) berechnet (Tabelle 10). Zwei dieser Messungen waren statistisch nicht signifikant (‚Selbstfürsorge‘ und ‚Schmerzen und Unwohlsein‘; $p>.1$; Tabelle 10).

Konvergenzvalidität: Mittels mehrheitlich moderaten Spearman-Rang-Korrelationen (Cohen, 1988) wurde in acht von neun Korrelationshypothesen (Tabelle 11) konvergente Konstruktvalidität nachgewiesen. Auffällig war die tiefe Korrelation zwischen

der ‚physischen Funktionsskala‘ (PedsQL™4.0; Varni, 1998) und der Dimension ‚Mobilität‘ (EuroQoL-5D-Youth; Wille et al., 2010).

Praktikabilität: Die Durchführung des EuroQoL-5D-Youth (Wille et al., 2010) in Interviewform dauert drei Minuten, die Fragebogen wurden fehlerfrei retourniert. Zuletzt wurde das Item-Verständnis der Teilnehmenden durch die Interviewer beurteilt, wobei über 80% der Kinder die Aufgaben gut verstanden. Lediglich 7.1% der Teilnehmenden zeigten ein zu tiefes Sprachverständnis.

In allen Dimensionen wurden Deckeneffekte (>60%) festgestellt. Ebenfalls existiert ein Index für das Auftreten eines geringfügigen Bodeneffektes*.

4.2.5 *Limitationen*

In Tabelle 13 werden die Limitationen und Bias der Studie erläutert.

Tabelle13

Limitationen und Bias in Canaway und Frew (2013)

Limitationen	Bias
<ul style="list-style-type: none"> • Assessmentversion verwendet das Schema des EuroQoL-5D (EuroQoL Research Foundation, 1990) für Erwachsene, welches für Kinder nicht ideal ist. • Manchmal helfen Eltern, obwohl alles alleine geht (u.a. Selbstfürsorgedimension). Dadurch kann das Kind nicht sagen, wie gut es geht. 	<ul style="list-style-type: none"> • In der Diskussion wurde auf Kriteriumsvalidität Bezug genommen, aber keine Messdaten unter Ergebnisse. • Kurzer Zeitintervall zwischen Test und Retest => Erinnerungseffekt^a möglich. • Grosse Deckeneffekte schwächen die Reliabilität • Keine Inter-Rater-Reliabilität erfasst, obwohl Frageverständnis durch 5 verschiedene Rater beurteilt wurde. • Keine detaillierten Kriterien beschrieben, wonach das Sprachverständnis beurteilt wird => deutet auf Messfehler hin.

Anmerkung. ^aEin Erinnerungseffekt existiert, wenn sich die Teilnehmenden an die Testantworten des ersten Durchgangs erinnern. Dieses Phänomen soll nach Döring und Bortz (2016) vermieden werden.

Die Tabellenstruktur wurde in Anlehnung an die Struktur von Fässler & Marchner (2010) geschrieben. Unter Limitationen werden Fehler verstanden, welche durch die Autorenschaft vermerkt wurden. Bias sind mögliche Fehler, welche nicht in den Studien dokumentiert wurden. Erinnerungseffekt: Die Teilnehmenden erinnern sich an die gegebenen Antworten im Test (1. Durchgang).

Quelle: Canaway & Frew (2013)

4.2.6 Zusammenfassung

Mittels mehrheitlich ausreichenden Kappa-Koeffizienten (Cohen, 1968) wurde die Stabilität des Assessments bestätigt. Dabei waren die Dimensionen ‚Selbstfürsorge‘ und ‚Schmerz‘ nicht signifikant.

Anhand mehrheitlich moderaten Korrelationskoeffizienten (Cohen, 1988) wurde konvergente Validität nachgewiesen. Dabei korrelierte die Dimension ‚Mobilität‘ mit den korrelierten Dimensionen (Tabelle 11) und war nicht signifikant.

Gemäss der Autorenschaft ist der Test innert drei Minuten durchführbar und für Kinder in Interviewform gut verständlich, was einer guten Praktikabilität gemäss Essink-Bot, Krabbe und Bonsel (1997), Badia, Monserat, Roset und Herdman (1999) und Brazier und Devrill (1999) entspricht.

4.3 Hauptstudie 3 – Varni et al. (2003)

PedsQL™4.0: Reliability and Validity of the Pediatric Quality of Life Inventory™
Version 4.0 Generic Core Scales in Healthy and Patient Populations
(Varni et al., 2003).

4.3.1 Ziel der Studie

Es wurden die Reliabilität, Validität und Praktikabilität des PedsQL™4.0 (Varni, 1998) an Familien mit zwei- bis 16jährigen Kindern im Bundestaat Kalifornien überprüft.

4.3.2 Sample

Zur Analyse der Gütekriterien wurden 5'905 Selbstbeurteilungen von Fünf- bis 16jährigen und 7'021 Proxyversionen für Zwei- bis 16jährige verwendet. Das Durchschnittsalter betrug für die Selbstbeurteilenden 9.8 (5.0–16.4) Jahre, für die fremdbeurteilten Kinder 7.9 (2.0–16.4) Jahre.

Die teilnehmenden Familien waren unterschiedlicher Herkunft (Abbildung 3) und sprachen verschiedene Sprachen, weshalb der Fragebogen fünfsprachig administriert wurde (Abbildung 4). Durch das Erfragen von Informationen zu möglichen chronischen Erkrankungen wurde erfahren, dass 7.7% aller teilnehmenden Kinder chronisch erkrankt sind (Abbildung 5).

4.3.3 Datenerfassung

Mittels Willkommensanruf wurden Eltern über die folgende Umfrage und den vertrauenswürdigen Umgang mit persönlichen Daten informiert. Anschliessend wurden die Fragebogen mit Begleitbrief per Post versendet. Der Begleitbrief beinhaltete die Studienbeschreibung und eine Durchführungsanleitung des Tests. Fünf- bis Siebenjährige durften während des Tests durch ihre Eltern unterstützt werden, wenn diese die Proxyversion bereits beantwortet hatten. Alle anderen Kinder sollten die Umfrage alleine beantworten. Anschliessend wurden beide Assessmentmethoden ans Datacenter zurückgesendet.

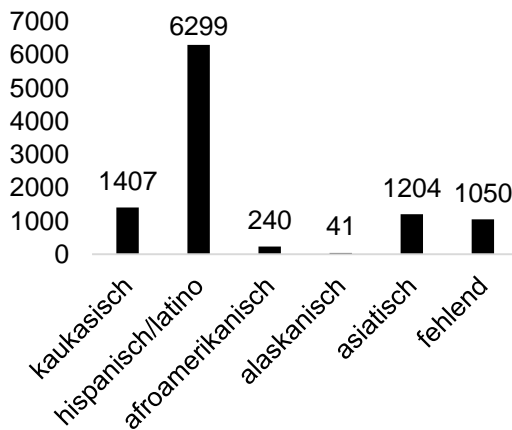


Abbildung 3. Herkunft der Teilnehmenden
Anmerkung. Dargestellt wird die Verteilung der Nationalitäten/Herkunft im Sample. Die meisten Teilnehmenden waren latin-amerikanischer/hispanischer Herkunft. Die Daten stammen von Varni et al. (2003).

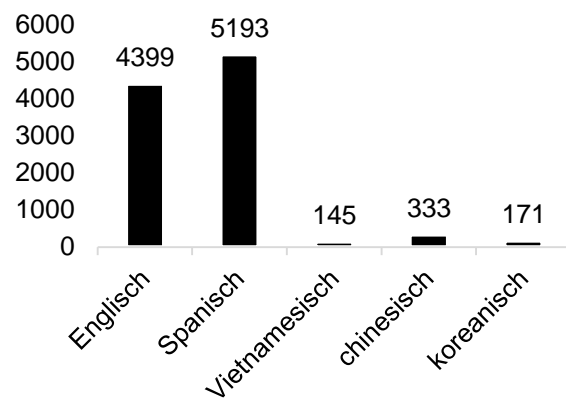


Abbildung 4. Die Verteilung der fünf Sprachversionen des Tests
Anmerkung. Dargestellt wird die Verteilung der ausgefüllten Sprachversionen im Sample. Es wurden hauptsächlich englische und spanische Versionen des Fragebogens zur Analyse der Gütekriterien verwendet. Die Daten stammen von Varni et al. (2003).

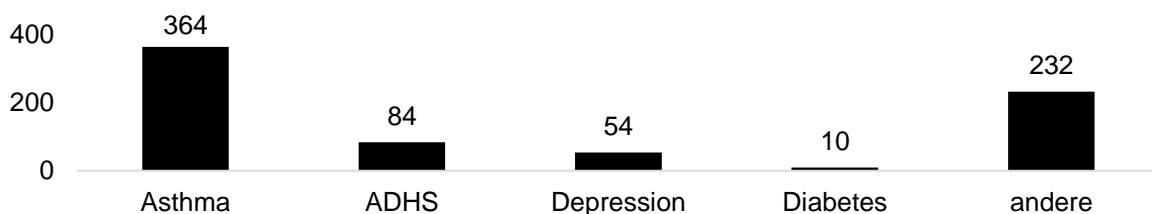


Abbildung 5. Verteilung der chronischen Krankheiten der Teilnehmenden
Anmerkung. Gesamthaft waren von 9'683 Teilnehmenden 744 chronisch in ihrer Gesundheit eingeschränkt. Von 103 Teilnehmenden fehlte die Angabe. Die Daten stammen von Varni et al. (2003).

4.3.4 *Ergebnisse*

Interne Stabilität: Für die berechneten Cronbach's Alpha-Koeffizienten* (Cronbach, 1951) wurden im Gesamtsample der Mindest-Wert* .70 (Nunnally & Bernstein, 1994) erzielt.

- Selbstversion: $\alpha=.71-.89$; $p<.05$
- Proxyversion: $\alpha=.74-.92$; $p<.05$

Drei einzelne Messwerte waren ungenügend ($p<.05$):

- ‚Schulische Funktionsskala‘ (Proxyversion, 2-4 Jahre): $\alpha=.58$
- ‚Körperliche Funktionsskala‘ (Selbstversion, 5-7 Jahre): $\alpha=.68$
- ‚Schulische Funktionsskala‘ (Selbstversion, 5-7 Jahre): $\alpha=.61$

Konkordanz Selbst- und Proxyversionen: Mittels Pearson-Korrelationen wurden moderate bis hohe Korrelationen (Cohen, 1988) berechnet ($p<.01$):

- Bei Fünf- bis Siebenjährigen: $r=.34-.70$
- Bei Acht- bis Zwölfjährigen: $r=.49-.77$
- Bei 13- bis 16jährigen: $r=.51-.79$

Die Werte waren gemäss der Autorenschaft konsistent mit früheren Studien (u.a. Varni, Seid & Kurtin, 2001).

Bekannt-Gruppen-Validität: Mittels *t*-Tests wurde festgestellt, dass weniger gesunde Kinder signifikant schlechtere Scores (= tiefere Lebensqualität) erzielten. Die Effektgrössen wurden gemäss Cohen (1988) beurteilt. Schlechtere Scores erzielten Kinder, welche...

- ... die Schule aufgrund Erkrankung häufiger versäumten: $t=18.89$
- ... an vier aufeinanderfolgenden Tagen Pflege benötigten: $t=20.46$
- ... zu krank für Spiele waren: $t=15.68$

Es wurden keine signifikanten Geschlechterunterschiede beobachtet, aber dafür Unterschiede zwischen den Altersgruppen:

- Fünf- bis Siebenjährige beurteilten ihre Lebensqualität mit tieferen Scores als Acht- bis 16jährige.
- Proxys von Zwei- bis Vierjährigen beurteilten die Lebensqualität der Kinder höher ein, als für Fünf- bis Siebenjährige.

Ebenfalls wurden Korrelationen zwischen der Sprache und erzielten Scores entdeckt.

- Englisch- und spanischsprachige Teilnehmende erzielten ähnliche Werte.
- Koreanische und chinesische Kinder schätzen ihre Lebensqualität höher ein als andere.

Auch Unterschiede bezüglich der Herkunft der Teilnehmenden wurden festgestellt:

- Asiatische Teilnehmende schätzten ihre Lebensqualität als höher ein als hispanische Teilnehmende.
- Kaukasische Eltern beurteilten die Lebensqualität ihrer Kinder als besser hispanische Eltern.

Von den versendeten Fragebogen wurden von der spanischen Sprachversionen die meisten zur Analyse der Gütekriterien weiterverwendet (58.5%; Tabelle 14). Die Teilnahmebereitschaft war bei hispanischen / latein-amerikanischen Familien am höchsten (53%; Tabelle 15).

Tabelle 14

Teilnahmebereitschaft in Bezug auf Sprachversion des Fragebogens

%	Sprachversionen
44.4	In englischer Version
58.5	Spanische Sprachversion
ca. 50*	Vietnamesische Sprachversion
ca. 50*	Koreanische Sprachversion
ca. 50*	Kantonesische Sprachversion

Anmerkung. Keine detaillierten Angaben.

In der Tabelle wird (in Prozent) gezeigt, wie viele Fragebogen ausgefüllt retourniert wurden.

Quelle: Varni et al. (2003).

Tabelle 15

Teilnahmebereitschaft in Bezug auf Nationalität der Teilnehmenden

%	Herkunft der Teilnehmenden
45.7	Nativ-amerikanische Familien
36.5	Afroamerikanische Familien
53	Hispanische / Latino-Familien
ca. 50*	Asiatische Familien
ca. 50*	Indisch-pazifische Familien
ca. 50*	Alaskische Familien

Anmerkung. Keine detaillierten Angaben.

In der Tabelle wird (in Prozent) gezeigt, wie viele der angefragten Familien teilnahmen in Bezug auf deren Nationalität.

Quelle: Varni et al. (2003).

Es wurde untersucht, ob die Lebensqualität durch das Erhalten von gesundheitsfördernden Dienstleistungen (u.a. Pflege) beeinflusst wird. Dazu wurden je zwei Gruppen mittels χ^2 -Test* verglichen:

- Familien mit chronisch erkrankten Kindern hatten mehr Mühe, Pflege zu erhalten als Familien mit gesunden Kindern ($\chi^2=161.2$; $p<.001$).
- Familien mit chronisch erkrankten Kindern, welche Pflegedienstleistungen nicht beziehen konnten, beurteilten die Lebensqualität signifikant tiefer als Familien, welche gesundheitliche Dienstleistungen erhielten ($\chi^2=117.4$; $p<.001$).

Mittels Multipler Regression* wurde aufgezeigt, dass die Lebensqualität durch das Erhalten von Pflegedienstleistungen beeinflusst wird (Höhere Werte bei Proxyversion: 2%, bei Selbstversion: 3%).

Praktikabilität: Die zur Analyse zugelassenen Fragebogen enthielten 1.8% Fehler (Selbstversion) und 3.4% Fehler (Proxyversion). Die meisten Fehler wurden für die ‚schulische Funktionsskala‘ bei Zwei- bis Vierjährige beobachtet.

4.3.5 Limitationen

In Tabelle 16 werden die Limitationen und Bias der Studie erläutert.

Tabelle 16

Limitationen und Bias in Varni et al. (2003)

Limitationen	Bias
<ul style="list-style-type: none"> • Zu wenig Kontakt zu den TN. Mit mehr Kontakt würden sich Fehler vermindern und die Teilnahmebereitschaft würde steigen. • Es wurden nicht validierte Sprachversionen verwendet, was zu Messverzerrungen führen könnte. 	<ul style="list-style-type: none"> • Fehlende Angaben betreffend Signifikanzlevel ($p<.05$?) • Eltern könnten eventuell trotzdem die Kinder unterstützt haben, obwohl die Aufforderung war, nur fünf- bis siebenjährige Kinder bei Bedarf zu unterstützen.

Anmerkung. Die Tabellenstruktur wurde in Anlehnung an die Struktur von Fässler & Marchner (2010) geschrieben. Unter Limitationen werden Fehler verstanden, welche durch die Autorenschaft vermerkt wurden. Bias sind mögliche Fehler, welche nicht in den Studien dokumentiert wurden. Quelle: Varni et al. (2003).

4.3.6 Zusammenfassung

Für beide Assessmentmethoden wurden Cronbach's Alpha-Koeffizienten (Cronbach, 1951) von $>.70$ berechnet, wobei für die ‚schulische Funktionsskala‘ die jeweils tiefsten Werte aufgezeigt wurden.

Zudem wurde eine moderate bis hohe Konkordanz (Cohen, 1988) zwischen den beiden Assessmentmethoden berechnet, wobei die Übereinstimmung mit zunehmendem Kindesalter stieg. Allerdings muss von Datenkontaminierung* ausgegangen werden.

Weiter erzielten gesündere Kinder Scores, die einer besseren Lebensqualität entsprechen. Die Lebensqualität ist kultur- und altersabhängig, jedoch geschlechterunabhängig.

Zuletzt wurde aufgrund der wenigen fehlerhaften Items gute Praktikabilität gemäss Essink-Bot et al. (1997) aufgezeigt.

4.4 Hauptstudie 4 – Varni, Seid und Burwinkle (2006)

The PedsQLTM4.0 as a school population health measure: Feasibility, reliability, and validity.

(Varni et al., 2006)

4.4.1 Ziel der Studie

Die vorliegende Studie überprüft die Reliabilität, Validität und Praktikabilität des PedsQLTM4.0 (Varni, 1998) als ein schulbasiertes Evaluationsmessinstrument.

4.4.2 Sample

Die Teilnehmenden wurden aus 304 Regelschulklassen in San Diego rekrutiert. Dabei wurden 2'436 Selbstversionen von Acht- bis 18jährigen und 4'227 Proxyversionen für Fünf- bis 18jährige ausgefüllt und insgesamt 2'415 Proxy-Kinder-Paare* gebildet.

Die Herkunft der Teilnehmenden (Abbildung 6), sowie die Sprachversion des Fragebogens (Abbildung 7) wurden erfasst. Zusätzlich wurden die Schulen in drei sozioökonomische Status (SES) eingeteilt: Schulen mit Kinder von tiefem, mit mitteltiefem oder mittlerem SES.

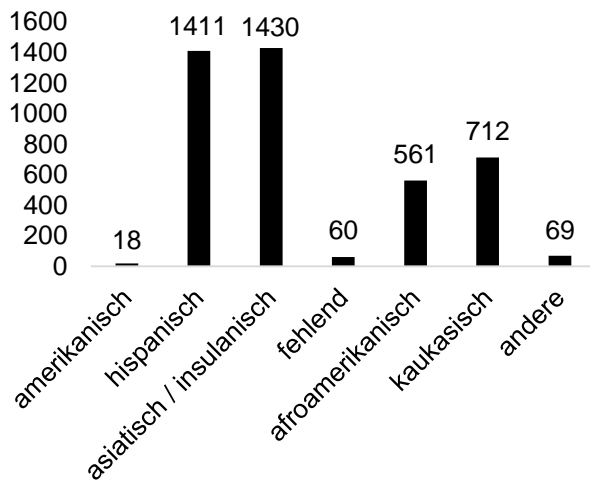


Abbildung 6. Die Herkunft/Nationalität der Teilnehmenden.

Anmerkung. Daten stammen von Varni et al. (2006).

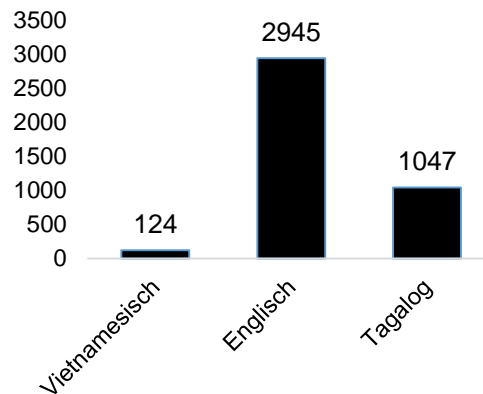


Abbildung 7. Der Test wurde in diesen Sprachen durchgeführt.

Anmerkung. Daten stammen von Varni et al. (2006).

4.4.3 Datenerfassung

Den Schülerinnen und Schülern wurde die Proxyversion, mit einer Einwilligungserklärung und einer schriftlichen Anleitung, verteilt, welche sie ihren Proxys überbrachten. Diese beantworteten die Fragen und retournierten die Umfrage beantwortet oder leer (leer = Teilnahmeverweigerung) an die Schule. Nachdem alle Proxyversionen retourniert wurden, wurden die Selbstversionen vor Ort von Forschungspersonal administriert.

50 Proxys mit niedrigem Bildungsniveau wurden in Versammlungen bei der Beantwortung der Fragen persönlich unterstützt. Zudem wurden Telefoninterviews mit Eltern geführt, welche sehr viele Items nicht beantwortet hatten.

4.4.4 *Ergebnisse*

Interne Konsistenz: Mittels Cronbach's Alpha-Koeffizienten (Cronbach, 1951) wurde ausreichende bis sehr gute (gemäss Nunnally & Bernstein, 1994) interne Konsistenz berechnet:

- Selbstversion: $\alpha=.70-.89$.
- Proxyversion: $\alpha=.76-.92$.

Ausschliesslich für die ‚schulische Funktionsskala‘ der Selbstversion von Acht- bis Zwölfjährigen wurde der Standardwert .70 nicht erreicht ($\alpha=.69$).

Konkordanz Proxy- und Selbstversion: Inwiefern die Antworten der beiden Assessmentmethoden übereinstimmen, wurde mittels Pearson-Produkt-Moment-Korrelation (Effektgrössen nach Cohen, 1988) und einer Intra-Klassen-Korrelation (Bartko, 1966) berechnet. Dabei wurden gemäss der Autorenschaft tiefere Werte als in Varni et al. (2003) erzielt. Hierbei stieg ebenfalls die Korrelation mit zunehmendem Alter der Kinder:

- Für Acht- bis Zwölfjährige: $r=.16-.30$; $p<.001$
- Für 13- bis 18jährige: $r=.29-.50$; $p<.001$

Konvergenz- und Divergenzvalidität: Das Stanford 9 (Harcourt Educational Measurement, 1996) und die schulische Funktionsskala korrelieren tief ($r=.26$; $p<.001$) miteinander (konvergente Validität), jedoch höher als die Korrelationen zwischen Stanford 9 und der emotionalen ($r = .07$; $p<.01$), der physischen ($r = .11$; $p<.001$) und der sozialen Funktionsskala ($r = .13$; $p<.001$) (divergente Validität).

Bekante-Gruppen-Validität: Teilnehmende ohne chronische Gesundheitseinschränkung und/oder mit mitteltiefem oder mittlerem SES erzielten bessere Scores (= höhere Lebensqualität; berechnet mittels *t*-Test):

- Selbstversion: $F=44.29$; $p<.001$
- Proxyversion: $F=62.92$; $p<.001$

Zwischen den Kategorien „mitteltiefer SES“ und „mittlerer SES“ wurden keine Unterschiede festgestellt.

Unterschiede betreffend Geschlecht, Alter, Herkunft/Nationalität: Es wurden keine signifikanten Unterschiede betreffend Geschlecht der Teilnehmenden beobachtet. Jedoch gaben Acht- bis Zwölfjährige eine leicht tiefere Lebensqualität an als 13- bis 18jährige. Zusätzlich wurden Unterschiede betreffend Sprache, Herkunft/Nationalität und SES erfasst, welche in Tabelle 17 beschrieben sind.

Tabelle 17

Ergebnisse der Bekannte-Gruppen-Validität

Kategorie	Selbsteinschätzung	Proxyversion
Herkunft, Nationalität	hispanische TN < kaukasische und afroamerikanische TN ^b	asiatische TN < kaukasische TN ^a hispanische TN < kaukasische und afroamerikanische TN ^a afroamerikanische TN < kaukasische TN ^c
Sprache	spanische TN < englische TN ^a	spanische und vietnamesische TN < englische und tagalogsprachige TN ^a vietnamesische TN < spanische TN ^a
SES	tiefer SES < mitteltiefer und mittlerer SES ^a	tiefer SES < mitteltiefer und mittlerer SES ^a

Anmerkung. Dargestellt werden Unterschiede zwischen den unterschiedlichen Nationalitäten, den verschiedensprachigen Fragebogenversionen, den SES und den erzielten Scores (Werte). Das < symbolisiert die tieferen, resp. höheren Werte. ^a $p < .001$, ^c $p < .05$. (Schlussfolgernd: ^b $p < .01$).
Quelle: Varni et al. (2006).

Mittels Multipler Regression mit drei Kontrollvariablen (mütterliche Herkunft/Nationalität, Sprache und Ausbildung (Abbildung 8) wurde die Beziehung zwischen dem Stanford 9 (Harcourt Educational Measurement, 1996) und der schulischen Funktionskala gemessen. Die Werte für beide Versionen waren signifikant ($p < .001$). Beobachtet wurde eine Erhöhung der Selbsteinschätzung, welche den Stanford 9 Wert um 1.93 Punkte erhöht. Die Erhöhung der Proxyversion um einen Punkt erhöhte den Stanford 9 Wert um 1.23 Punkte.

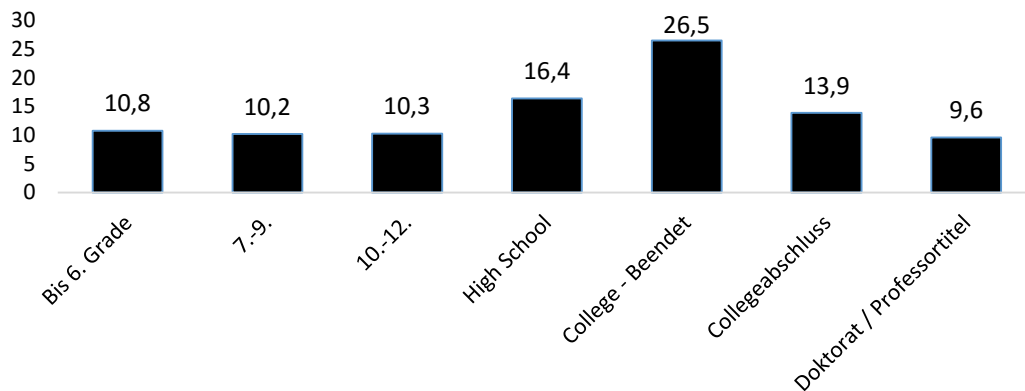


Abbildung 8. Die mütterliche Ausbildung als Kontrollvariable
 Anmerkung. Daten stammen von Varni et al. (2006) und wurden in Prozent angegeben.

Praktikabilität: Die Praktikabilität wurde anhand fehlerhaften oder nicht-beantworteten Items bestimmt und in Prozenten angegeben (nach Essink-Bot et al., 1997). Es wurden 0.001% Fehler bei der Selbstversion und 0.30% Fehler bei der Proxyversion festgestellt.

Es wurden keine Bodeneffekte und nur 2.5% Deckeneffekte gefunden, was einem minimalen bis moderaten Level entspricht.

4.4.5 Limitationen

In Tabelle 18 werden die Limitationen und Bias der Studie erläutert.

Tabelle 18

Limitationen und Bias in Varni et al. (2006)

Limitationen	Bias
<ul style="list-style-type: none"> • Anzahl verweigerter Schulen ist unbekannt. • Die Eltern der chronisch beeinträchtigten Kinder beurteilten deren chronische Erkrankung. Es wurden keine Zusatzinformationen durch Fachpersonen o.ä. eingeholt. • Aufteilung (SES) war unpräzise. • Nicht alle Übersetzungen des Assessments wurden bereits validiert. 	<ul style="list-style-type: none"> • In der Diskussion wurde auf die tiefen Konkordanz-Werte keinen Bezug genommen. Die Leserschaft weiss somit nicht, wie diese tiefen Werte begründet werden könnten. • Es nahmen mehr Kinder der Umfrage teil als Jugendliche.

Anmerkung. Die Tabellenstruktur wurde in Anlehnung an die Struktur von Fässler & Marchner (2010) geschrieben. Unter Limitationen werden Fehler verstanden, welche durch die Autorenschaft vermerkt wurden. Bias sind mögliche Fehler, welche nicht in den Studien dokumentiert wurden. Quelle: Varni et al. (2006).

4.4.6 Zusammenfassung

Für beide Assessmentmethoden wurden mehrheitlich gute Alpha-Koeffizienten (Cronbach, 1951) berechnet. Zudem korrelierten die beiden Assessmentmethoden auf tiefem bis moderatem Level (in Anlehnung an Cohen, 1988) miteinander, wobei die Konkordanz mit zunehmendem Kindesalter stieg.

Des Weiteren wird gemäss der Autorenschaft gesünderen Kindern aufgrund der erzielten Scores eine höhere Lebensqualität zugesprochen, welche geschlechterunabhängig ist, dafür aber durch Kultur und Alter beeinflusst wird. Zudem wurden tiefe Korrelationskoeffizienten zwischen der 'schulischen Funktionsskala' und dem Stanford 9 (Harcourt Educational Measurement, 1997) berechnet.

Zuletzt wurde in Anlehnung an Essink-Bot et al. (1997) gute Praktikabilität aufgezeigt, da die retournierten Fragebogen wenig Fehler enthielten.

4.5 Hauptstudie 5 – Robitail et al. (2006)

Validation of the European Proxy KIDSCREEN-52 Pilot Test Health-Related
Quality of Life Questionnaire: First Results.
(Robitail et al., 2006)

4.5.1 Ziel der Studie

Diese Studie erfasst die Validität und Reliabilität der Proxyversion des KIDSCREEN-52 (Ravens-Sieberer et al., 2005).

4.5.2 Sample

Die Teilnehmende (n=2'505 Proxy-Kinder-Paare) wurden in sieben europäischen Ländern (Österreich, Frankreich, Spanien, Deutschland, Schweiz, Niederlande, England) rekrutiert. Die Kinder der waren acht- bis elfjährig und zwölf- bis 18jährig (Jugendliche), wobei der Altersdurschnitt 12.4 ± 2.5 Jahre betrug. Insgesamt wurden von und für Knaben weniger Fragebogen beantwortet (48%) und 40% der Fragebogen wurden von und für Kinder ausgefüllt.

4.5.3 *Datenerfassung*

Die Daten wurden 2002 erhoben, wobei die Art der Datenerfassung je nach Land variierte. Dabei wurde die Datenerfassung nicht im Detail erläutert. Es wurde nur vage aufgezeigt, wie die Teilnehmenden für die Studie ausgesucht wurden:

- Deutschland: Alle Teilnehmende, welche die Einwilligung erteilten.
- Spanien: Teilnehmende, welche die Verweigerung nicht mitteilten.
- England: Der Teilnahmeentscheid oblag der Klassenlehrperson. Die Kinder übergaben die Proxyversion ihren Eltern, welche die Proxyversion wiederum via die Kinder retournierten.
- Niederlande: Den Teilnehmenden wurde die Umfrage brieflich zugestellt.

4.5.4 *Ergebnisse*

Interne Konsistenz: In allen Dimensionen wurden gute bis sehr gute Cronbach's Alpha-Koeffizienten ($\alpha=.76-.90$; Cronbach, 1951; Nunnally & Bernstein, 1994) berechnet, was die interne Konsistenz der Proxyversion unterstützt.

Merkmalsvalidität: Mittels Multitrait-Methode* wurde gute Merkmalsvalidität* berechnet, indem Korrelationen zwischen den zehn Assessmentdimensionen und den zwei Assessmentmethoden analysiert wurden. Es wurden Pearson-Korrelationskoeffizienten (Cohen, 1988) von $.12-.66$ festgestellt.

Konvergenz- und Divergenzvalidität: Es wurden Pearson-Korrelationen (Cohen, 1988) zwischen den zehn Assessmentdimensionen und den zehn Domänen der gesundheitsbezogenen Lebensqualität gemäss WHO berechnet. Die Pearson-Korrelationskoeffizienten reichten von $.12$ bis $.66$, was konvergente und divergente Validität aufgezeigt. Mittels MTMM-Matrix* wurden folgende Werte aufgezeigt:

- Konvergente Validität (Monotrait-Heteromethod-Korrelationen*): $r=.35-.55$
- Divergente Validität (Heterotrait-Heteromethod-Korrelationen*): $r=.02-.36$
- Divergente Validität (Heterotrait-Monomethod-Korrelationen*): $r=.07-.63$ (Selbstversion) und $r=.12-.66$ (Proxyversion)

Konkordanz: Es wurden mittels Intra-Klassen-Korrelation (Bartko, 1966) gute Konkordanz festgestellt. Die Übereinstimmung war bei Jugendlichen-Proxy-Paaren höher als bei Kinder-Proxy-Paaren und ebenfalls, bei Mädchen-Proxy-Paaren.

Psychometrische Eigenschaften: Es wurden kaum Bodeneffekte (maximal 2%) und mehrheitlich rare Deckeneffekte festgestellt (4.2 – 15.5%; ausser die Dimension ‚Soziale Akzeptanz‘: 33.6% Deckeffekte; Abbildung 9).

Table 2
Results for the 10 dimensions of the KIDSCREEN-52 proxy version

Dimension (item no)	MV	M ± SD	Range	Floor	Ceiling	Alpha	IIC	IDV	INFIT RRSR	NFIT PCM
PHY (5)	.5%	71.2 ± 17.5	6.3–100	.0%	6.0%	.80	.45–.65	100.0%	.81–1.15	.85–1.27
PWB (6)	1.2%	77.3 ± 14.3	5–100	.0%	7.4%	.88	.63–.73	100.0%	.86–1.23	.85–1.14
MOOD (7)	.8%	82.1 ± 13.7	14.3–100	.0%	9.1%	.83	.53–.66	100.0%	.80–1.19	.76–1.14
SELF (5)	.8%	78.1 ± 16.3	0–100	.1%	11.3%	.76	.47–.58	91.1%	.93–1.11	.83–1.12
AUTO (5)	.4%	78.1 ± 16.5	0–100	.0%	13.9%	.86	.64–.74	100.0%	.74–1.12	.82–1.13
PAR (6)	1.0%	79.0 ± 14.8	20.8–100	.0%	9.5%	.86	.56–.71	100.0%	.77–1.42	.81–1.23
SOC (6)	2.3%	67.8 ± 17.8	0–100	.2%	4.2%	.90	.67–.78	100.0%	.75–1.18	.76–1.16
SCH (6)	1.7%	70.5 ± 17.4	0–100	.1%	4.5%	.87	.60–.73	100.0%	.75–1.19	.84–1.20
BULL (3)	2.7%	84.7 ± 15.8	0–100	.0%	33.6%	.80	.57–.68	100.0%	.88–1.22	.79–1.23
PFIN (3)	3.5%	70.6 ± 22.4	0–100	.7%	15.5%	.83	.67–.69	100.0%	.96–1.06	.96–1.03

MV = Percentage of missing values; M ± SD = mean scores at baseline (standard deviation); range = range of scores; floor = floor effect (%); ceiling = ceiling effect (%); internal consistency alpha = (Cronbach alpha); IIC = item-internal consistency; IDV = item-discriminant validity; INFIT = Rasch fit statistic with PCM (Partial Credit Model) and RRSR (Rasch Rating Scale Model); PHY = physical well-being; PWB = psychological well-being; life satisfaction and positive emotions; MOOD = moods and emotions; SOC = social support and peers relation; PAR = parents relation and home life; SELF = self-perception: body image and self-esteem; AUTO = autonomy; SCH = cognitive and school functioning; BULL = bullying and social rejection; PFIN = perceived financial opportunities.

Scores ranging from 0 = lower HRQL to 100 = higher HRQL.

Abbildung 9. Relevante Ergebnisse der Konstruktvalidität

Anmerkung. Daten stammen von Robitail et al. (2006), S. 596.e4.

Konstruktvalidität: Mit einer Faktorenanalyse* wurde die Mehrdimensionalität des Assessments geprüft, indem die Angemessenheit der Proxy- und Selbstversionen miteinander analysiert wurden. Die Güte wurde einerseits mit der Maximum-Likelihood-Methode* ($\chi^2=14'167$; $df=1'229$; $RMSEA^*=.078$; $NNFI^*=.953$; $CFI^*=.957$) und der Weighted-Least-Square-Methode* ($\chi^2=11'944.30$; $df=1229$; $RMSEA=.71$; $NNFI=.959$; $CFI=.962$) geprüft.

Mittels Rasch-Analyse* wurde das Konstrukt jeder Dimension erfasst, wobei für alle Dimensionen (ausser Dimension ‚Privatleben und Beziehungen‘, INFIT* RRSR* .77 - 1.42) zufriedenstellende INFIT PCM*- und INFIT RRSR-Werte erzielt wurden (Abbildung 14).

Somit wurden fast alle Items der korrekten Dimension zugeteilt, was die Ergebnisse der Item-Diskriminanzvalidität* und die item-interne Konsistenz* bestätigten (Abbildung 9).

4.5.5 **Limitationen**

In Tabelle 19 werden Limitationen und Bias der Studie erläutert.

Tabelle 19

Limitationen und Bias in Robitail et al. (2006)

Limitationen	Bias
<ul style="list-style-type: none">• In der Studie werden keine Limitationen vermerkt	<ul style="list-style-type: none">• Möglicherweise liegt Datenkontaminierung vor, da die Daten zu Hause erfasst wurden und nicht ausgeschlossen werden kann, dass die Eltern und die Kinder die Fragebogen zusammen beantworteten.• Die Datenerfassung wurde nicht im Detail vermerkt!

Anmerkung. Die Tabellenstruktur wurde in Anlehnung an die Struktur von Fässler & Marchner (2010) geschrieben. Limitationen sind die Messfehler, welche durch die Autorenschaft dokumentiert wurden. Bias sind hier mögliche Limitationen, welche nicht in den Studien vermerkt wurden.

Quelle: Robitail et al. (2006).

4.5.6 **Zusammenfassung**

Für die Proxyversion wurde gute Interne Konsistenz aufgezeigt (in Anlehnung an Nunnally & Bernstein, 1994), was guter Reliabilität entspricht.

Zusätzlich korrelieren die zehn Assessmentdimensionen auf gutem Level mit den zehn Domänen der gesundheitsbezogenen Lebensqualität. Ebenfalls wurde die Eindimensionalität jeder Dimension, sowie die multidimensionale Assessmentstruktur bestätigt. Zuletzt ist die Konkordanz bei Jugendlichen-Proxy-Paaren höher als bei Paaren mit Kindern, und ebenfalls wenn die Selbstversion von Mädchen oder Kindern mit Beeinträchtigungen beantwortet wurden.

4.6 Hauptstudie 6 – Ravens-Sieberer et al. (2008)

The KIDSCREEN-52 Quality of Life Measure for Children and Adolescents:
Psychometric Results from a Cross-Cultural Survey in 13 European Countries.
(Ravens-Sieberer et al., 2008)

4.6.1 Ziel der Studie

Mittels dieser Studie wurde die Selbstversion des KIDSCREEN-52 (Ravens-Sieberer et al., 2005) auf dessen Reliabilität und Validität überprüft.

4.6.2 Sample

22'827 Personen aus 13 europäischen Ländern (Tabelle 20) nahmen an der Studie teil, wobei sie in Kinder (acht- bis elfjährig; $M^*=9.7$ Jahre) und in Jugendliche (zwölf- bis 18jährig; $M=14.4$ Jahre) eingeteilt wurden. Es nahmen ca. 2% mehr Mädchen teil als Knaben.

Die Test-Retest-Reliabilität wurde an 10% des initialen Samples in zehn Ländern nach sieben bis zehn Tagen überprüft.

Tabelle 20

Gruppeneinteilung und länderspezifische Datenerfassung der Studie

Gruppe	Zugehörigkeit	Studien-TN	Durchführung
Gruppe 1	Österreich, Deutschland, Schweiz, Spanien, Frankreich, Holland	Schülerinnen und Schüler; Eltern	Telefonische Erstkontaktierung, anschliessend Papierfragebogen ausfüllen und diese retournieren.
Gruppe 2	Griechenland, Ungarn, Irland, Polen, Schweden	Schülerinnen und Schüler von staatlichen, privaten, städtischen oder ländlichen Schulen; Eltern	Fragebögen wurden während der Schulzeit ausgefüllt; Elternversion per Schülerinnen und Schüler überbracht (in Polen per Post versendet). Elternversion wurde retourniert.
Gruppe 3	England Tschechische Republik	Schülerinnen und Schüler Zufälliges Sample der Bevölkerung.	 Kombination beider Verfahren

Anmerkung. Die Gruppeneinteilung der Länder erfolgte zwecks Übersicht durch die Verfasserin.
Quelle: Ravens-Sieberer et al. (2008).

4.6.3 **Datenerfassung**

Die Sample-Rekrutierung und die Datenerfassung waren in den einzelnen Ländern unterschiedlich, weshalb in Tabelle 20 darauf eingegangen wird.

4.6.4 **Ergebnisse**

Internale Konsistenz: Anhand der Cronbach's Alpha-Koeffizienten wurde überprüft, ob alle Items dasselbe Konstrukt messen. Dabei wurden für das Gesamtsample adäquate bis gute Alpha-Koeffizienten berechnet ($\alpha=.77-.89$).

Test-Retest-Reliabilität: Neben drei ungenügenden wurden mehrheitlich adäquate Intra-Klassen-Korrelationskoeffizienten* (Bartko, 1966) berechnet ($r=.56-.77$). Ungenügende Werte wurden für die Dimensionen ‚Stimmung und Emotionen‘ ($r=.58$), ‚Autonomie‘ ($r=.56$) und ‚soziale Akzeptanz‘ ($r=.57$) aufgezeigt.

Konvergente und divergente Konstruktvalidität: Bei allen Messungen wurde gute bis sehr gute Konvergenz- und Divergenzvalidität (gemäss Cohen, 1988) beobachtet. Alle Korrelationen waren signifikant ($p<.01$).

- Korrelationen mit den vier Funktionsskalen des PedsQL™4.0 (Varni, 1998): $r=.14-.50$.
- Korrelationen mit der Zufriedenheitsskala des CHIP-AE* (Starfield et al., 1994): $r=.24-.60$.
- Korrelationen mit der Wahrnehmungsskala des YQOL-S* (Patrick & Edwards, 2002): $r=.24-.61$.

Bekannte-Gruppen-Validität: In allen Dimensionen wurde unterstützende Bekannte-Gruppen-Validität anhand Effektgrössen d (Cohen, 1988) berechnet. Gesündere Teilnehmende erzielten höhere Scores, wobei alle Berechnungen signifikant ($p<.01$) waren:

- a. Teilnehmende mit und ohne mentaler Gesundheitseinschränkung wurden miteinander verglichen, wobei die Teilnehmenden mittels SDQ* (Goodman, 1997) und dem CSHCN* (Bethell et al., 2002) zu einer der beiden Gruppen eingeteilt wurden:
- Einteilung mit CSHCN: tiefe bis moderate Effektgrößen ($d=.17-.42$).
 - Einteilung mit SDQ: moderate bis hohe Effektgrößen ($d=.32-.72$).
- b. Teilnehmende mit tieferem versus höherem sozioökonomischem Status:
- Tiefe bis hohe Effektgrößen: $d=.10-.37$ (ausschliesslich für finanzielle Ressourcen $d=.75$).

Modell: Die Daten passten gut auf das Modell (RMSEA=.062; CFI=.976). Bei allen Items wurden Loading-Patterns* festgestellt, welche grösser als .62 waren. Modellparameter, d.h. Korrelationen zwischen den einzelnen Assessmentdimensionen, reichten von .16–.76. Weiter zeigten beide Selbstversionen ähnliche Goodness of Fit (RMSEAKinder & Jugendliche=.062; CFI_{Kinder}=.974, CFI_{Jugendliche}=.973).

Es wurden signifikante Unterschiede zwischen den beiden Versionen bezüglich Pattern-Koeffizienten, latente Faktorenkorrelation und Fehler-Terms gefunden ($\chi^2=3303,75$; $p<.001$). Diese Unterschiede zwischen den zugeordneten Pattern-Koeffizienten waren klein, was auf marginale bis moderate Effekte (.002–.278) nach der Fischer-Z-Transformation* schliessen lässt. Zudem war die Fischer-Z-Differenz klein (.11; Bortz & Döring, 2002).

Es wurde 97.8 – 100% Scaling-Success* für die verschiedenen Skalen erreicht. Dabei waren alle korrigierten Item-Korrelationen höher als $r=.42$. Die Resultate der Rasch-Analyse indizieren nachvollziehbare Güte (.80–1.27). Ausschliesslich vier von 52 Items (in den Dimensionen ‚körperliches Wohlbefinden‘, ‚Selbstwahrnehmung‘ und ‚soziale Unterstützung und Freunde‘) waren leicht über der Differential Item Functioning-Grenze* (DIF) von .035 (maximal .041).

Für Kinder und Jugendliche wurden die Assessmentanalysen wiederholt. Es zeigte sich, dass Jugendliche leicht höhere Alpha-Werte erzielten.

In einer früheren KIDSCREEN-52-Studie wurden gemäss der Autorenschaft keine signifikanten Bodeneffekte festgestellt und in ausschliesslich drei Dimensionen Deckeneffekte (Dimensionen ‚Privatleben und Zuhause‘ 15.5%, ‚Finanzielle Ressourcen‘ 24.5% und ‚soziale Akzeptanz‘ 49.1%) beobachtet.

4.6.5 Limitationen

In Tabelle 21 werden die Limitationen und Bias der Studie erläutert.

Tabelle 21

Limitationen und Bias in Ravens-Sieberer et al. (2008)

Limitationen	Bias
<ul style="list-style-type: none">Die Teilnehmende wurden mit Assessments in zwei Gruppen eingeteilt (kränkere und gesündere Teilnehmende). Es wurden jedoch keine weiteren Krankeninformatoren eingeholt. Dies könnte die Validität und Reliabilität des Assessments schmälern.	<ul style="list-style-type: none">Eventuell waren die erhobenen Daten kontaminiert: Sie wurden teilweise zu Hause und nicht unter Aufsicht aufgenommen.

Anmerkung. Die Tabellenstruktur wurde in Anlehnung an die Struktur von Fässler & Marchner (2010) geschrieben. Unter Limitationen werden Fehler verstanden, welche durch die Autorenschaft vermerkt wurden. Bias sind mögliche Fehler, welche nicht in den Studien dokumentiert wurden. Quelle: Ravens-Sieberer et al. (2008).

4.6.6 Zusammenfassung

Es wurden mehrheitlich adäquate Korrelationskoeffizienten (in Anlehnung an Bartko, 1966) beobachtet, was wiederum die Stabilität des Assessments aufzeigt. Zusätzlich wurden mehrheitlich gute Alpha-Koeffizienten (Cronbach, 1951; Nunnally & Bernstein, 1994) aufgezeigt, wodurch die Interne Konsistenz des KIDSCREEN-52 (Ravens-Sieberer et al., 2005) nachgewiesen wurde.

Die mehrheitlich moderaten bis hohen Korrelationskoeffizienten bestätigen indes die konvergente und divergente Konstruktvalidität. Ebenfalls wird die Konstruktvalidität mittels den Ergebnissen der Bekannte-Gruppen-Validität unterstützt. Dabei erzielten gesündere Kinder Scores, die einer höheren Lebensqualität entsprechen.

5 DISKUSSION

In der vorliegenden Arbeit wurden das EuroQol-5D-Youth (Wille et al., 2010), das PedsQL™4.0 (Varni, 1998) und das KIDSCREEN-52 (Ravens-Sieberer et al., 2005) auf deren Reliabilität, Validität und Praktikabilität überprüft. Zur Beantwortung der Fragestellung werden im Folgenden die in Kapitel 4 beschriebenen Hauptstudien kritisch betrachtet (Kapitel [5.1](#)), die wichtigsten Ergebnisse von Kapitel 4 diskutiert und assessmentspezifische Implikationen für die Praxis und Forschung abgeleitet (Kapitel [5.2](#)). Nach der Diskussion der Limitierungen (Kapitel [5.3](#)) und dem Aufzeigen weiterführender Forschungsschwerpunkte (Kapitel [5.4](#)), folgen in Kapitel [5.5](#) die Schlussfolgerungen und Handlungsempfehlungen.

5.1 Diskussion der Hauptstudien

Kinder müssen in Forschung und Praxis ihre Lebensqualität selber beurteilen, damit aussagekräftige Erkenntnisse erhalten werden. Zu diesem Schluss kamen bereits Eiser und Morse (2001b) und er wurde von Varni et al. (2006) bestätigt. In letztgenannter Studie wurde die Datenerfassung vorbildlich getrennt, um Kontaminierung vollständig ausschliessen zu können. Deswegen wurden wohl mehrheitlich tiefe bis moderate Korrelationen aufgezeigt. Folglich kann abgeleitet werden, dass die Datenerfassung ebenfalls unter Kontrolle und nicht unterstützt durch die Proxys stattfinden soll. Nach Ansicht der Verfasserin kam es in den anderen Studien (Varni et al., 2003; Robitail et al., 2006; Ravens-Sieberer et al., 2008) zu Messverzerrungen, da die Beurteilung mittels Proxy- und Selbstversion unsauber getrennt wurden und somit Proxys die Antworten der Kinder bewusst oder unbewusst verfälschten.

Des Weiteren wurden in Ravens-Sieberer et al. (2010) Teilnehmende im Alter zwischen 13 und 19 Jahren inkludiert und mit Kindern ab acht Jahren verglichen. Nach Ansicht der Verfasserin und in Anbetracht der Tatsache, dass jüngere Kinder eine vermeintlich tiefere Lebensqualität aufzeigen (vgl. Varni et al., 2003; Varni et al., 2006), führte dieser direkte Vergleich zu Fehlinterpretationen der Messergebnisse. Somit können diese beiden Alterskategorien nicht direkt verglichen werden.

Die Nachvollziehbarkeit und Verständlichkeit wird aufgrund verschiedener Aspekte erschwert. In allen Studien wurden nach Ansicht der Verfasserin relevante Daten nicht dargestellt. Dies könnte zu Missinterpretationen der dargestellten Ergebnisse führen, da teilweise die beschriebenen Daten im Fliesstext nicht mit denjenigen in den Tabellen übereinstimmen. Des Weiteren erschwert das nicht konsistente Verwenden von Fachbegriffen das Leseverständnis der Studie (vgl. Robitail et al., 2006).

Nach Ansicht der Verfasserin hätte das PedsQL™4.0 (Varni, Seid & Burwinkle, 2001) auf Wiederholbarkeit überprüft werden sollen, um zu erfahren, ob die Merkmale stabil erfasst werden. Mittels der überprüften internen Konsistenz konnte jedoch die zuverlässige Erfassung der Konstrukte durch die Items gewährleistet werden. Des Weiteren wird nicht klar, ob die Ergebnisse in den europäischen Kontext übertragen werden können. Dies besitzt indes Forschungsrelevanz für die Zukunft.

5.2 Diskussion der Assessments

Im folgenden Kapitel werden die untersuchten Assessments im Hinblick auf die Fragestellung der Arbeit untersucht (Tabelle 22).

Tabelle 22

Zusammenfassung der Ergebnisse

EuroQoL-5D-Youth	PedsQL™4.0	KIDSCREEN-52
Reliabilität		
1 κ^a = mehrheitlich ausreichend–moderat (ital. ‚Self ^b : kein κ berechenbar, da identische Antworten; span: κ = substantiell; ‚Mob ^c nicht signifikant).	3 gute–sehr gute α -Koeffizienten ^h im Gesamtsample. Signifikant höhere Werte für 2–4jährige (Proxyvers. ^d), ‚schul. FS‘ tiefste Werte allgemein; 5–7j. ungenügend in ‚schul. FS ^e und ‚phys. FS ^f (Selbstvers. ^g). Konkordanz: moderat-hoch	5 Gut bis sehr gute α -Koeffizienten ^h Akzeptable Konkordanz (ICC ⁱ)
2 κ = mehrheitlich ausreichend (‚Self‘ nicht signifikant. ‚Mobilität‘ signifikant)	4 gute-sehr gute α -Koeffizienten ^h für Gesamtsample (beide Assessmentmethoden); tiefste α für ‚schul. FS‘; $\alpha < .70$ für 8-12jährige bei ‚schul. FS‘. Konkordanz: tief-moderat.	6 Gute bis sehr gute α -Koeffizienten ^h . mehrheitlich akzeptable Stabilität ^l

Fortsetzung. Tabelle 22

Validität		
1 Konvergenzvalidität: hohe Korrelationen zwischen zwei, psychologische Aspekte erfassende Dimensionen; tiefe Korrelation zwischen ‚Mob‘ und ‚phys. FS‘ ^k ; keine Korrelationen für ‚Alltagsaktivitäten‘ und ‚Self‘ aufgezeigt.	3 BGV ^l : Gemäss den erzielten Scores haben gesündere TN ^m , ältere Kinder im Vergleich zu jüngeren und kaukasische TN die höhere LQ ⁿ .	5 Jede Dimension misst ein anderes Konstrukt; alle Dimensionen messen die gesundheitsbezogene LQ.
2 Moderate Korrelationen zwischen EuroQoL-5D-Youth und PedsQL TM 4.0 (Varni, 1998) und CHU-9D (Stevens, 2010).	4 BGV: Gesündere TN und TN mit höherem SES ^o besitzen höhere LQ. Tiefe Konvergenzvalidität zwischen Stanford 9 ^p und ‚schul. FS‘.	6 Konvergenzvalidität: Moderat-hoch BGV: gesündere TN mit höherer LQ
Praktikabilität		
1 Wenig Fehler (Items: 0-2%; VAS: 0-9%).	3 Selbstvers.: 1.8% Fehler; Proxyvers.: 3.4% Fehler	5 nicht geprüft
2 Dauert 3 Min; 0% Fehler; gute Item-Verständlichkeit.	4 Selbstvers.: <0.001% Fehler; Proxyvers.: 0.3% Fehler.	6 nicht geprüft
Anmerkung		
Der Test soll nicht mit Evaluationswerten des EuroQoL-5D-3L evaluiert werden (Van Reenen et al., 2014) ^q . In beiden Hauptstudien wurden in allen Dimensionen hohe Deckeneffekte beobachtet. Der Test wird durch die EuroQoL Research Foundation nach der Registrierung zugesendet und ist abhängig des Zwecks mit Kosten verbunden (EuroQoL Research Foundation, 2016).	Für nicht akademische Forschung ist die Verwendung kostenlos. Andernfalls sind die Kosten abhängig vom Umfang der Studie(n) (Varni, 2016d).	Praktikabilität wurde nicht gemessen. Die Verwendung des Tests bei gesundheitsökonomischen Evaluationen ist möglich (Ravens-Sieberer et al., 2008). Die Anwendung zu nicht-kommerziellen Zwecken kostet einmalig 50€ (KIDSCREEN-Group, 2011c).

Anmerkung: 1-6: Hauptstudien (siehe Kapitel 4). ^aKappa-Koeffizient (Cohen, 1968), ^b‚Selbstfürsorge‘ (Assessmentdimension), ^c‚Mobilität‘ (Assessmentdimension), ^d‚Proxyversion, ^e‚schulische Funktionsskala‘ (Assessmentdimension), ^f‚physische Funktionsskala‘ (Assessmentdimension), ^g‚Selbstversion, ^h(Cronbach, 1951), ⁱIntra-Klassen-Korrelation, ^j Test-Retest-Reliabilität, ^k‚physische Funktionsskala‘ (Assesmentdimension), ^lBekannte-Gruppen-Validität, ^mTeilnehmende, ⁿLebensqualität, ^osozioökonomischer Status, ^pStanford 9 (Harcourt Educational Measurement, 1997), ^qVan Reenen, Janssen, Oppe, Kreimeier & Greiner (2014). Alle Quellen befinden sich im Quellenverzeichnis im Kapitel 6.1.

5.2.1 EuroQoL-5D-Youth

Reliabilität

In Ravens-Sieberer et al. (2010) und in Canaway und Frew (2013) wurde mittels ausreichenden bis moderaten Kappa-Koeffizienten (Cohen, 1968; Landis & Koch, 1977) die Wiederholbarkeit des Assessments erfolgreich belegt (Tabelle [10](#)).

Zwischen den beiden Studien von Ravens-Sieberer et al. (2010) und Canaway und Frew (2013) wurden Unterschiede festgestellt. Allgemein betrachtet wurden in Ravens-Sieberer et al. (2010) bessere Kappa-Koeffizienten (Cohen, 1968; Landis & Koch, 1977) berechnet als in Canaway und Frew (2013). Folglich werden die Merkmale der kindlichen Lebensqualität bei Kindern ab acht Jahren stabiler gemessen, wodurch sich die Anwendung bei letztgenannter Altersgruppe besser eignet. Diese Schlussfolgerung muss jedoch mit weiterführender Forschung untersucht werden, da die besseren Werte ebenfalls auf die unterschiedlichen Samplegrössen zurückzuführen sind.

Zudem wurden ausschliesslich im spanischen Retest-Sample in Ravens-Sieberer et al. (2010) für die Dimension ‚Selbstfürsorge‘ signifikante Messungen erzielt, wodurch diese Dimension unzuverlässig gemessen zu werden scheint. Die Dimension ‚Mobilität‘ wurde ausschliesslich in Canaway und Frew (2013) signifikant getestet. Dies deutet auf Messverzerrungen in Ravens-Sieberer et al. (2010), welche aufgrund hoher Deckeneffekte entstanden sind, hin.

Validität

Konvergenzvalidität wurde von Ravens-Sieberer et al. (2010) und von Canaway und Frew (2013) festgestellt (Tabelle [11](#)). Gute Korrelationen wurden für zwei psychologische Aspekte erfassende Dimensionen berechnet. Zudem korreliert die ‚Mobilität‘ mit keiner, körperliche Aspekte erfassende, Assessmentdimension auf akzeptablem Level (Robitail et al., 2010; Canaway & Frew, 2013). Beide Tatsachen lassen die Vermutung zu, dass mit dem EuroQoL-5D-Youth (Wille et al., 2010) eher psychologische Aspekte erfasst werden.

Allerdings wurde in Ravens-Sieberer et al. (2010) keine Konvergenzvalidität bezüglich der Dimensionen ‚Selbstfürsorge‘ und ‚Alltagshandlungen‘ festgestellt. Hingegen

korrelierten in Canaway und Frew (2013) beide genannten Dimensionen mit moderatem Level gemäss Bartko (1966) mit ähnlichen Dimensionen des CHU-9D (Stevens, 2010). Folglich existiert für alle Dimensionen konvergente Validität.

Des Weiteren wurde in beiden Studien festgestellt, dass gesündere Kinder Scores erzielen, die einer höheren Lebensqualität entsprechen.

Praktikabilität

Das EuroQoL-5D-Youth (Wille et al., 2010) ist praktikabel, da in Ravens-Sieberer et al. (2010) wenig und in Canaway und Frew (2013) keine Fehler festgestellt wurden (Kapitel [4.1.4](#); [4.2.4](#)). Da in beiden Studien Daten unterschiedlich erfasst wurden, ist die Schlussfolgerung legitim, dass mit direkter, personeller Unterstützung einer Fachperson Fehler stark verringert werden können. Zudem sind die Items gemäss Canaway und Frew (2013) gut verständlich formuliert und der Test innert drei Minuten durchführbar, wodurch die Anwendung des EuroQoL-5D-Youth (Wille et al., 2010) in einer Ergotherapiesequenz möglich ist und sich für grossflächige Erfassungen der kindlichen Lebensqualität eignen würde.

Schlussfolgerung

Das EuroQoL-5D-Youth ist gemäss den Ergebnissen von Ravens-Sieberer et al. (2010) und von Canaway und Frew (2013) reliabel, valide und praktikabel. Dabei ist Notwendigkeit weiterer Forschung unumstritten, um die erhobenen Ergebnisse zu festigen und noch bestehende Lücken zu schliessen.

Allerdings wird von Canaway und Frew (2013), Van Reenen, Janssen, Oppe, Kreimeier und Greiner (2014) und der Verfasserin von der Verwendung dieses Assessments abgeraten, da keine geeigneten Richtwerte zu dessen Evaluation existieren. Die für die Erwachsenenversion entwickelten Richtwerte sollten nicht verwendet werden, da sich das EuroQoL-5D vom EuroQoL-5D-Youth (Wille et al. 2010) inhaltlich unterscheidet.

Sollte dennoch der dringende Anwendungswunsch bestehen, ist das EuroQoL-5D-Youth (Wille et al., 2010) eher für Kinder ab acht Jahren geeignet, da die Items für jüngere Kinder (Tabelle [10](#)) instabiler gemessen werden.

5.2.2 PedsQL™4.0

Reliabilität

In Varni et al. (2003) und Varni et al. (2006) wurde anhand mehrheitlich guten Cronbach's Alpha-Koeffizienten (Cronbach, 1951) die zuverlässige Erfassung jedes Konstrukts durch die Items bestätigt. Für die schulische Funktionsskala wurden dabei die jeweils tiefsten Alpha-Koeffizienten berechnet.

In Varni et al. (2003) wurden für drei Messungen allerdings für die Alterskategorien Zwei- bis Vierjährige und Fünf- bis Siebenjährige ungenügende Werte erzielt (Kapitel [4.3.4](#)). Unter anderem ist die ‚schulische Funktionsskala‘ zweimal davon betroffen.

Ebenfalls eine ungenügende Messung in der ‚schulischen Funktionsskala‘ wurde von Varni et al. (2006) – allerdings für Acht- bis Zwölfjährige – bestätigt.

Daraus kann gefolgert werden, dass schulbezogene Lebensqualitätsaspekte durch die ‚schulische Funktionsskala‘ erst bei 13jährigen zuverlässig gemessen werden. Allerdings widersprechen Varni et al. (2003) dieser Schlussfolgerung, da sie bereits für Acht- bis Zwölfjährige genügende Alpha-Koeffizienten (Cronbach, 1951; Nunnally & Bernstein, 1994) berechnen konnten.

Zudem wurde durch tiefe bis moderate Konkordanz (Kapitel [4.4.4](#); Varni et al., 2006) zwischen den beiden Assessmentmethoden die Dringlichkeit, dass Kinder ihre Lebensqualität selber beurteilen (Morse & Eiser, 2001b), erneut unterstützt. Ungeklärt bleibt demnach ebenfalls, wie aussagekräftig die Lebensqualitätsbeurteilung von Zwei- bis Vierjährigen ist, da hierfür ausschliesslich eine Proxyversion existiert.

Validität

Mittels Beurteilung der Bekannte-Gruppen-Validität in Varni et al. (2003) und Varni et al. (2006), sowie Korrelationsüberprüfungen zwischen Stanford 9 (Harcourt Educational Measurement, 1997) und den vier Funktionsskalen des PedsQL™.0 (Varni, 1998) wurde die Konstruktvalidität aufgezeigt. Dabei wurde in Varni et al. (2003) und in Varni et al. (2006) ein kultureller Einfluss auf die kindliche Lebensqualität festgestellt (Kapitel [4.3.4](#); [4.4.4](#)). Dass die Lebensqualität von und für kaukasische Teilnehmende jeweils am höchsten eingeschätzt wurde, lässt auf historische und geografische Gründe schliessen. Ebenso wenig verwunderlich ist, dass Kinder aus niedrigeren sozialen Schichten eine tiefere Lebensqualität besitzen (Kapitel [4.4.4](#))

In Varni et al. (2003) und Varni et al. (2006) wurden keine signifikanten Geschlechtsunterschiede festgestellt. Allerdings wurde Kindern zwischen fünf und sieben Jahren anhand deren erzielten Scores eine signifikant tiefere Lebensqualität zugesprochen als älteren Kindern (Varni et al., 2003; Varni et al., 2006).

Die tiefe Korrelation zwischen dem Stanford 9 (Harcourt Educational Measurement, 1997) und der ‚schulischen Funktionsskala‘ lässt darauf schliessen, dass die ‚schulische Funktionsskala‘ nicht die schulbezogenen Aspekte der kindlichen Lebensqualität erfasst. Dadurch werden die Schlussfolgerung der internen Konsistenz (siehe oben) sowie die Empfehlung einer Zweitbeurteilung (u.a. mittels KIDSCREEN-52; Ravens-Sieberer et al., 2005) zur Erfassung von schulbezogenen Aspekten der Lebensqualität begründet.

Praktikabilität

Die Fragebogen wurden sowohl in Varni et al. (2003), als auch in Varni et al. (2006) mit wenig Fehlern retourniert (Kapitel [4.3.4](#); [4.4.4](#)). Dies lässt auf eine einfache Handhabung und verständlich-formulierte Items schliessen. Allerdings wurden in Varni et al. (2006) deutlich weniger Fehler vermerkt, was daran liegen könnte, dass die Teilnehmenden in Varni et al. (2006) deutlich mehr unterstützt wurden als in Varni et al. (2003). Denn Varni et al. (2003) vermuteten, die Fehlerzahlen könnten mittels mehr persönlichem Kontakt reduziert werden, was wiederum in Varni et al. (2006) umgesetzt wurde. Folglich eignet sich die Beantwortung der Fragebogen mit personeller Unterstützung beispielsweise durch Fachpersonen der Ergotherapie.

Schlussfolgerung

Anlässlich der mehrheitlich guten Alpha-Koeffizienten (Cronbach, 1951) wird das PedsQL™4.0 (Varni, 1998) als reliabel betrachtet.

Sowohl in Varni et al. (2003), als auch in Varni et al. (2006) wurde der kulturelle Einfluss auf die Lebensqualität festgestellt. Des Weiteren wird in Varni et al. (2006) konvergente Konstruktvalidität aufgezeigt. Aufgrund dieser Punkte kann das PedsQL™4.0 (Varni, 1998) als valide bezeichnet werden.

Abschliessend ist das PedsQL™4.0 (Varni, 1998) aufgrund seiner tiefen Fehlerquote, was auf eine einfache Handhabung hindeutet, praktikabel. Die Beantwortung der Fragen sollte jedoch mit Unterstützung einer Drittperson durchgeführt werden.

5.2.3 KIDSCREEN-52

Reliabilität

Anhand der internen Konsistenz wurde die Reliabilität des KIDSCREEN-52 gemessen. Dabei wurden in Robitail et al. (2006) und Ravens-Sieberer et al. (2008) gute Alpha-Koeffizienten (Cronbach, 1951) berechnet. Allerdings wurde in Robitail et al. (2006) die Proxyversion überprüft, in Ravens-Sieberer et al. (2008) jedoch die Selbstversion. Dadurch können diese Ergebnisse nur ansatzweise miteinander verglichen werden. Dazu ist weitere Forschung notwendig.

In Ravens-Sieberer et al. (2008) wurde zusätzlich die Test-Retest-Reliabilität mittels Intra-Klassen-Korrelationen beurteilt, wobei lediglich in drei Dimensionen ungenügende Werte (Kapitel [4.6.4](#)) erzielt wurden. Gemäss Ravens-Sieberer et al. (2008) besteht weiterer Forschungsbedarf zur Widerlegung dieser ungenügenden Dimensionenwerte.

Des Weiteren wurden in Ravens-Sieberer et al. (2008) Unterschiede betreffend Konkordanz der beiden Assessmentmethoden festgestellt. Da die Unterschiede nur bei vier von 52 Items auftreten und zusätzlich klein sind ($DIF \leq .041$), muss diesem Phänomen keine Aufmerksamkeit geschenkt werden (persönliches Gespräch mit B. Gantschnig am 12.5.2016). Trotzdem wird erneut aufgezeigt, dass die Kinder ihre Lebensqualität dringend selber beurteilen sollen.

Validität

Gemäss Robitail et al. (2006) wird die Lebensqualität mittels zehn Dimensionen gültig gemessen, wobei jede Assessmentdimension einen anderen Aspekt der Lebensqualität erfasst. Dadurch eignet sich das KIDSCREEN-52 (Ravens-Sieberer et al., 2005) ebenfalls zur Anwendung bei gesundheitsökonomischen Forschungsprojekten. Lediglich ein Item korreliert hoch mit der „falschen“ Dimension (Abbildung [9](#)). Da die entsprechende Messung in Ravens-Sieberer et al. (2008) nicht erhoben wurde, kann diese Beobachtung nicht abschliessend geklärt werden.

In Ravens-Sieberer et al. (2008) wurden Korrelationen zwischen KIDSCREEN-52 (Ravens-Sieberer et al., 2005) und PedsQL™4.0 (Varni, 1998), der Wahrnehmungsskala des YQOL-S (Patrick & Edwards, 2002) und der Zufriedenheitsskala des CHIP-

AE (Starfield et al., 1994) festgestellt (Kapitel [4.6.4](#)). Aufgrund der höheren Korrelationen zwischen KIDSCREEN-52 und YQOL's oder CHIP-AE kann abgeleitet werden, dass diese drei Assessments ähnliche, aber nicht dieselben Merkmale erfassen.

Praktikabilität

Weder in Robitail et al. (2006) noch in Ravens-Sieberer et al. (2008) wurde die Praktikabilität überprüft. Anlässlich der Kriterien nach Lienert und Raatz (1998; Kapitel [2.3](#)) wird das KIDSCREEN-52 (Ravens-Sieberer et al., 2005) als nicht praktikabel betrachtet. Kritikpunkte sind die lange Zeitdauer (20 Minuten; KIDSCREEN-Group, 2011a) und die umständliche Auswertung der Resultate.

Schlussfolgerung

Das KIDSCREEN-52 (Ravens-Sieberer et al., 2005) wird anhand der aufgezeigten internen Konsistenz in Robitail et al. (2006) und Ravens-Sieberer et al. (2008), sowie der zusätzlich in Ravens-Sieberer et al. (2008) aufgezeigten, akzeptablen Test-Retest-Reliabilität als reliabel betrachtet. Trotzdem besteht weiterer Forschungsbedarf zur Überprüfung der berechneten Werte.

Zudem wird das KIDSCREEN-52 (Ravens-Sieberer et al., 2005) als valide betrachtet, da in Ravens-Sieberer et al. (2008) gute konvergente und divergente Validität zu drei anderen Assessments (Kapitel [4.6.4](#)) aufgezeigt wurde. Zusätzlich misst gemäss Robitail et al. (2006) jede Dimension des KIDSCREEN-52 ein anderes Subkonstrukt der Lebensqualität.

Abschliessend wird das Assessment gemäss der Praktikabilitätsdefinition nach Lienert und Raatz 1998) als nicht praktikabel eingestuft, da viel Zeit zur Durchführung benötigt wird und die Evaluation zeitintensiv ist. Somit muss dieses Assessment im Rahmen der vorliegenden Fragestellung als nicht durchführbar eingestuft werden. Aufgrund der genannten Aspekte eignet sich das Assessment zur Verwendung in der ergotherapeutischen Praxis gemäss Verfasserin nicht. Es könnte allerdings für Forschungszwecke verwendet werden, da gemäss Robitail et al. (2006) die Konstrukte der gesundheitsbezogenen Lebensqualität erfasst werden.

5.3 Limitationen der Arbeit

Als Limitationen der vorliegenden Arbeit werden folgende Punkte angesehen: Gemäss den formulierten Kriterien (Kapitel [3.3](#)) wurden für diese Bachelorarbeit ausschliesslich englische und deutsche Artikel verwendet. Somit kann nicht ausgeschlossen werden, dass anderssprachige, relevantere Studien existieren.

In Ravens-Sieberer et al. (2010) wurden Teilnehmende bis 19 Jahre inkludiert. Dies widerspricht jedoch den Ein- und Ausschlusskriterien* der vorliegenden Bachelorarbeit gemäss Kapitel [1.4](#). Da über 90% der Studienteilnehmer (2'551 von gesamthaft 2'809 Kindern) die Einschlusskriterien diese Arbeit erfüllen, wurde die besagte Studie dennoch berücksichtigt.

Ebenfalls wurden ausschliesslich Validierungsstudien verwendet, welche die Gütekriterien der drei Assessments überprüften. Damit wurden Literaturreviews konsequent ausgeschlossen. Es ist somit möglich, dass relevante Sekundärliteratur* existiert, aber nicht berücksichtigt wurde.

Des Weiteren mussten die Hauptstudien bis November 2015 publiziert worden sein und in den durchsuchten Datenbanken (siehe Search-History in Anhang G) angezeigt worden sein. Einerseits kann die Existenz von neueren Artikeln nicht ausgeschlossen werden. Andererseits wäre ein Publication Bias* möglich, da nicht alle Artikel in den Datenbanken angezeigt werden.

Für diese Arbeit wurde das Beurteilungsraster von Law et al. (1998) gewählt. Dieses stellte sich als nicht optimal heraus, da es für Interventionsstudien konzipiert wurde und nicht für Validierungsstudien. Teilweise mussten die zu beurteilenden Aspekte sinngemäss angepasst werden.

Zuletzt können Übersetzungs- oder Interpretationsfehler nicht ausgeschlossen werden. Einige Quellen waren sprachlich für die Verfasserin schwer verständlich und es wurden statistische Verfahren verwendet (vgl. Robitail et al., 2006; Ravens-Sieberer et al., 2008), welche im Rahmen des Bachelorstudiums nicht thematisiert wurden.

5.4 Weiterführende Forschung

Weiterer Forschungsbedarf ist unumstritten, da dieses Forschungsfeld bisher, im Vergleich zur Lebensqualitätsforschung bei Erwachsenen, ignoriert wurde. Insbesondere müssen vermehrt bereits etablierte Assessments auf Kinder angepasst und validiert werden. Um der elterlichen Überschätzung ihrer Kinder Herr zu werden, sollte das Augenmerk von der zukünftigen Forschung auf Assessments gelegt werden, welche die kindliche Lebensqualität mittels Selbstversion erfassen.

Ebenfalls sieht die Verfasserin erheblichen Bedarf zur Validierung von Übersetzungsversionen zahlreicher Assessments. Aus Praktikabilitätsgründen wurden in der Vergangenheit teilweise Assessments automatisch mittels eines Computers übersetzt. Gemäss Varni et al. (2003) und Varni et al. (2006) ist dieses Vorgehen legitim. Weitere Forschung zur Überprüfung, ob diese nicht-validierten Sprachversionen die kulturellen Aspekte der Lebensqualität erfassen, ist unabdingbar.

Im Rahmen der vorliegenden Arbeit wurde teilweise beobachtet, dass vereinzelte Gruppenvaliditäten aufgrund limitierter Samplegrösse nicht statistisch signifikant berechnet werden konnten. Dies sollte mit weiterführender Forschung nachgeholt werden. Dabei werden optimalerweise vergleichbare Daten erfasst, was wiederum die Effizienz der zukünftigen Projekte positiv beeinflusst.

Eine Vielzahl von vorliegenden Studien leidet, aufgrund der Sampleauswahl (mehrheitlich gesunde TN), unter erheblichen Deckeneffekten. Somit wird die Aussagekraft der Assessments bei beeinträchtigten Kindern nicht ausreichend untersucht. Für die ergotherapeutische Praxis sind jedoch Assessments nötig, welche den Einfluss von therapeutischen Massnahmen auf die kindliche Lebensqualität krankheits-unspezifisch, zuverlässig und abgestuft erfassen können.

5.5 Schlussfolgerung mit Implikation für die Ergotherapie

In der vorliegenden Arbeit wurden in Form eines Literaturreviews gute psychometrische Eigenschaften des EuroQoL-5D-Youth (Wille et al., 2010), des PedsQLTM4.0

(Varni, 1998) und des KIDSCREEN-52 (Ravens-Sieberer et al., 2005) aufgezeigt. Dabei wurden für das EuroQoL-5D-Youth, das Pediatric Quality of Life Inventory™ 4.0 und das KIDSCREEN-52 gute Reliabilität und Validität aufgezeigt, das PedsQL™ 4.0, sowie das EuroQoL-5D-Youth wiesen ebenfalls gute Praktikabilität auf.

Der Bezug zwischen Ergotherapie und Lebensqualitätsförderung wurde in Kapitel [1.2](#) bereits erläutert. Daher werden nun ausschliesslich markante Aspekte aufgegriffen: Lebensqualität, Assessments und Ergotherapie stehen in engem Zusammenhang. Gemäss WFOT (2012) soll jede ergotherapeutische Intervention auf die Förderung der Lebensqualität der Klientinnen und Klienten abzielen. Um aber die Effektivität dieser Interventionen zu überprüfen, werden Assessments benötigt (Fawcett, 2007). Anhand dieser Arbeit erhalten Praktizierende der pädiatrischen Ergotherapie eine Auswahl von drei reliablen und validen Assessments zur Erfassung der kindlichen Lebensqualität. Ebenso sind die in Kapitel [4](#) aufgezeigten psychometrischen Eigenschaften hilfreich zur Begründung der Assessmentanwendung (in Anlehnung an Gi-ani et al., 2012).

Bereits Eiser und Morse (2001a) erkannten den grossen Bedarf an der Lebensqualitätsforschung von Kindern und das Bedürfnis an geeigneten, lebensqualitätserfassenden Assessments. Dabei erläuterten sie die besondere Notwendigkeit an Erfassungsinstrumenten, mit welchen die Kinder ihre Lebensqualität selber beurteilen können. Diese Selbstversionen sind wichtig, da sich die Fremd- und Selbstbeurteilungen unterscheiden (vgl. Varni et al., 2006; Ravens-Sieberer et al., 2008). Folglich ist die Befragung der Kinder nach deren Lebensqualität von immenser Bedeutung. Im ergotherapeutischen Alltag sollte somit wenn immer möglich auf die Selbstbeurteilung zurückgegriffen werden. In begründeten Ausnahmefällen wird ebenfalls die Erfassung mittels Proxyversion empfohlen. Diese ist hilfreich, falls das Kind die eigene Lebensqualität nicht beurteilen kann oder möchte (Varni et al., 2003; Robitail et al., 2006). Für die Praxis eignet sich ebenfalls ein Assessment, welches sowohl als Proxy- als auch als Selbstversion verfügbar ist. Dabei gilt zu beachten, dass die beiden Datenerhebungen unabhängig voneinander durchgeführt werden. So kann einer einseitigen Über- oder Unterschätzung effektiv vorgebeugt werden.

Beide Versionen sollten nach Ansicht der Verfasserin mit Unterstützung einer Ergotherapeutin oder eines Ergotherapeuten beantwortet werden. Dies wird durch die Interpretation der Ergebnisse in Varni et al. (2003) und Varni et al. (2006) bekräftigt. Während die Daten in Varni et al. (2003) zu Hause erhoben wurden, korrelierten die beiden Methoden höher als in Varni et al. (2006). Allerdings wurde nur in Varni et al. (2006) darauf geachtet, dass die Daten separat erfasst wurden. Dies stützt die Vermutung, dass Eltern die Antworten ihrer Kinder bewusst oder unbewusst beeinflussen.

Für die Erfassung der Lebensqualität in der ergotherapeutischen Praxis und Forschung wird aufgrund seiner Kompaktheit (Umfang, Dauer) das PedsQLTM4.0 (Varni, 1998) empfohlen. Allerdings eignet sich das Assessment zur Evaluation der schulbezogenen Lebensqualitätsaspekte aufgrund ungenügender Werte und tiefer Konvergenzvalidität nicht. Ebenfalls wurden bisher kaum europäische Sprachversionen validiert. Anlässlich dieser Kritikpunkte empfiehlt sich – trotz des fehlenden Nachweises der Praktikabilität – die zusätzliche Anwendung des KIDSCREEN-52 (Ravens-Sieberer et al., 2005). Dieses misst zwar die kindliche Lebensqualität in deutlich höherem Item- und Zeitumfang (KIDSCREEN-Group, 2011a). Doch werden die Aspekte der gesundheitsbezogenen Lebensqualität gemäss Robitail et al. (2006) und Ravens-Sieberer et al. (2008) erfasst.

6 QUELLEN

6.1 Literaturverzeichnis

- Badia, X., Monserrat, S., Roset, M., & Herdman, M. (1999). Feasibility, validity and test-retest reliability of scaling methods for health states: The visual analogue scale and the time trade-off. *Quality of Life Research*, 8, 4, 303–310.
- Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19, 1, 3–11.
- Berting-Hüneke, C. (n.d.). Theoretische Grundlagen - Bedeutung von Selbstständigkeit für die Lebensqualität. In C. Scheepers, U. Steding-Albrecht & P. Jehn. (Hrsg.). *Ergotherapie – Vom Behandeln zum Handeln* (3. Aufl., S. 578–579) (2007). Stuttgart: Georg-Thieme Verlag.
- Bethell, C.D., Read, D., Stein, R.E., Blumenberg, S.J., Wells, N. & Newacheck, P.W. (2002). Identifying children with special health care needs: development and evaluation of a short screening instrument. *Ambulatory Pediatrics*, 2, 1, 38–48.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation* (3. Edition). Berlin: Springer-Verlag.
- Brazier, J., & Deverill, M. (1999). A checklist for judging preference-based measures of health related quality of life: Learning from psychometric measures. *Health Economics*, 8, 41–51.
- Bundesamt für Gesundheit [BAG]. (n.d.a). Definitionen. Heruntergeladen von <http://www.bag.admin.ch/jugendprogramme/10038/10039/index.html?lang=de> am 01.03.2016
- Bundesamt für Gesundheit [BAG]. (n.d.b). Lebensqualität. Heruntergeladen von <http://www.bag.admin.ch/transplantation/00692/02582/12400/index.html?lang=de> am 29.05.2016
- Canaway, A.G. & Frew, E.J. (2013). Measuring preference-based Quality of Life in children aged 6 – 7 years: a comparison of the performance of the CHU-9D and EuroQoL-5D-Youth – the WAVES Pilot Study. *Quality of Life Research*, 22, 173–183. doi:10.1007/s11136-012-0119-5

- Cohen, J. (1968). Weighted kappa: Nominal Scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213–220.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hilldale: Erlbaum.
- Cronbach, L.J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika*, 16, 297–334.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Ausgabe). Berlin-Heidelberg: Springer-Verlag.
- Eiser, C. & Morse, R. (2001a). A review of measures of quality of life for children with chronic illness. *Archives of Disease in Childhood*, 84, 205–211.
- Eiser, C. & Morse, R. (2001b). Can parents rate their child's health-related quality of life? Results of a systematic review. *Quality of Life Research*, 10, 347–357.
- Essink-Bot, M., Krabbe, P., & Bonsel, G. A. N. (1997). An empirical comparison of four generic health status measures: The Nottingham health profile, the medical outcomes study 36-item short-form health survey, the COOP/WONCA charts, and the EuroQol Instrument. *Medical Care*, 35, 522–537.
- ErgotherapeutInnen-Verband Schweiz [EVS]. (2012). Definitionen der Ergotherapie. Heruntergeladen von http://www.ergotherapie.ch/resources/uploads/Ergoinfo/Définitions%20de%20l'ergothérapie_D.pdf am 21.02.2016
- EuroQoL Research Foundation. (1990). EuroQoL – a new facility for measurement health-related quality of life. *Health Policy*, 16,3, 199–208.
- EuroQoL Research Foundation. (2016a). EuroQoL-5D-Youth. Heruntergeladen von <http://www.EuroQol.org/eq-5d-products/EuroQoL-5D-Youth-youth.html> am 28.3.2016
- EuroQoL Research Foundation. (2016b). Proxy Paper. Heruntergeladen von <http://www.EuroQol.org/eq-5d-products/EuroQoL-5D-Youth-youth/proxy-paper.html> am 28.3.2016
- Fawcett, A.J.L. (2007). *Principles of Assessment and Outcome Measurement for Occupational Therapists and Physiotherapists—Theory, Skills and Application*. North Yorkshire: John Wiley & Sons Ltd.

- Fayers, P.M. & Machin, D. (2000). *Quality of Life. Assessment, Analysis and Interpretation*. West Sussex: John Wiley & Sons Ltd.
- Fayers, P.M. & Machin, D. (2013). *Quality of Life: The Assessment, Analysis, and Interpretation of Patient-reported Outcomes* (2. Aufl.). [KINDLE-Edition].
Heruntergeladen von https://www.amazon.com/Quality-Life-Assessment-Interpretation-Patient-reported-ebook/dp/B00D0NG68E?ie=UTF8&me=&ref_=mt_kindle
- Fässler, S. & Marchner, M. (2013). *Validität, Reliabilität und Praktikabilität der Assessments FIM, PRPP, AMPS und BI – Erfassung der Selbstständigkeit in Aktivitäten des täglichen Lebens bei Klienten mit erworbener Hirnschädigung* (nichtpublizierte Bachelorarbeit). Zürcher Hochschule für Angewandte Wissenschaften: Winterthur.
- Giani, A.M., Schanz, B. & Hiemke, C. (2012). Lebensqualität am Beispiel depressiver Patienten nach stationärer Behandlung. *Psych. Pflege*, 18, 34 – 41. doi: 10.1055/s-0031-1301032
- Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38, 581–596.
- Harcourt Educational Measurement. (1996). *Stanford Achievement Test Series, Ninth Ed.* San Antonio, TX: Harcourt Educational Measurement.
- KIDSCREEN-Group. (2011a). Durchführung. Heruntergeladen von <http://www.kidscreen.org/deutsch/fragebögen/kidscreen-52-langform/> am 28.03.2016
- KIDSCREEN-Group. (2011b). Existierende Sprachversionen. Heruntergeladen von <http://www.kidscreen.org/deutsch/sprachen/existierende-sprachversionen/> am 28.03.2016
- Koller, M., Neugebauer, E.A.M., Augustin, M., Büssing, A., Farin, E., Klinkhammer-Schalke, M., ... & Wieseler, B. (2009). Die Erfassung von Lebensqualität in der Versorgungsforschung – konzeptuelle, methodische und strukturelle Voraussetzungen. *Gesundheitswesen*, 71, 864–872.
- Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 1, 159–174.

- Law, M., Stewart, D., Pollock, N., Letts, L., Bosch, J. & Westermorland, M. (1998). Formular zur kritischen Besprechung quantitativer Studien. Heruntergeladen von <http://srs-mcmaster.ca/wp-content/uploads/2015/04/Critical-Review-Form-Quantitative-Studies-German.pdf> am 29.5.2016
- Lienert, G. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6.Ausg.). Weinheim: Psychologie Verlags Union.
- Nunnally, J.C. & Bernstein, I.R. (1994). *Psychometric Theory* (3. Aufl.). New York: McGraw-Hill.
- Patrick, D.L. & Edwards, T.C. (2002). Youth Quality of Life Instrument–Short Form (YQOL-SF). Heruntergeladen von <http://www.midss.org/content/youth-quality-life-instrument-short-form-yqol-sf> am 20.5.2016
- Ravens-Sieberer, U. (2000). Verfahren zur Erfassung der gesundheitsbezogenen Lebensqualität bei Kindern und Jugendlichen. *Bundesgesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz*, 43, 198–209.
- Ravens-Sieberer, U. & Bullinger, M. (1998). Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. *Quality of Life Research*, 7, 399–407.
- Ravens-Sieberer, U., Ellert, U. & Erhart, M. (2007). Gesundheitsbezogene Lebensqualität von Kindern und Jugendlichen in Deutschland – Eine Normstichprobe für Deutschland aus dem Kinder- und Jugendgesundheitsurvey (KIGGS). *Gesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz*, 50, 810–818.
doi:10.1007/s00103-007-0244-4
- Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Duer, W., Auquier, P., ... & the European KIDSCREEN Group. (2005). KIDSCREEN-52 quality of life measure for children and adolescents. *Expert Review of Pharmacoeconomics & Outcomes Research*, 5, 3, 353–364.
- Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Power, M., Duer, W., ..., the KIDSCREEN–Group. (2008). The KIDSCREEN-52 Quality of Life Measure for Children and Adolescents: Psychometric Results from a Cross-Cultural Survey in 13 European Countries. *Value in Health*, 11, 4 ,645–658.

- Ravens-Sieberer, U., Herdmann, M., Devine, J., Otto, C. Bullinger, M., Rose, M. & Klasen, F. (2014). The European KIDSCREEN Approach to measure quality of life and well-being in children: development, current application, and future advances. *Quality of Life Research*, 23, 791–803.
- Ravens-Sieberer, U., Wille, N., Badia, X., Bonsel, G., Burström, K., Cavrini, G., Devlin, N., ... & Greiner, W. (2010). Feasibility, reliability, and validity of the EuroQoL-5D-Youth: results from a multinational study. *Quality of Life Research*, 19, 887–897. doi:10.1007/s11136-010-9649-x
- Robitail, S., Simeoni, M-C., Erhart, M., Raven-Sieberer, U., Bruil, J., Auquier, P. & the European Kidscreen Group. (2006). Validation of the European Proxy KIDSCREEN-52 Pilot Test Health-Related Quality of Life Questionnaire: First Results. *Journal of Adolescent Health*, 39, 596.e1–596.e10.
- Schmeck, K. (1998). Lebensqualität in der Kinder- und Jugendpsychiatrie. In K. Schmeck, F. Pouska & H. Katschnig. (Hrsg.) (1989). *Qualitätssicherung und Lebensqualität in der Kinder- und Jugendpsychiatrie*. Wien: Springer-Verlag.
- Starfield, B., Bergner, M., Ensminger, M., Riley, A.W., Green, B.F. & Ryan, S. (1994). *Child Health and Illness Profile-Adolescent Edition (CHIP-AE)*. Baltimore: The John Hopkins University.
- Stevens, K. (2010). Assessing the performance of a new generic measure of health related quality of life for children and refining it for use in health state valuation. *Applied Health Economics and Health Policy*, 8, 3, 157–169.
- Trochim, W.M.K. (2006a). Types of Reliability. Heruntergeladen von <http://www.socialresearchmethods.net/kb/reotypes.php> am 21.01.2016
- Trochim, W.M.K. (2006b). Measurement Validity Types. Heruntergeladen von <http://www.socialresearchmethods.net/kb/measval.php> am 28.01.2016
- Van Reenen, M., Janssen, B., Oppe, M., Kreimeier, S. & Greiner, W. (2014). EQ-5D-Y User Guide – Basic Information on how to use the EQ-5D-Y instrument. Heruntergeladen von http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-Y_User_Guide_v1.0_2014.pdf am 24.04.2016
- Varni, J.W. (1998). PedsQL™ – Pediatric Quality of Life Inventory™ 4.0. Heruntergeladen von <http://www.pedsqol.org/pedsqol13.html> am 23.05.2016

- Varni, J.W. (2016a). About the Model. Heruntergeladen von http://www.pedsq.org/about_pedsq.html am 28.03.2016
- Varni, J.W. (2016b). Scoring Instructions. Heruntergeladen von <http://www.pedsq.org/score.html> am 29.5.2016
- Varni, J.W. (2016c). Existing Translations. Heruntergeladen von <http://www.pedsq.org/PedsQL-Translation-Tables.pdf> am 14.5.16
- Varni, J.W. (2016d). PedsQL™ Conditions of Use. Heruntergeladen von <http://pedsq.org/conditions.html> am 29.5.2016
- Varni, J.W., Burwinkle, T.M. & Seid, M. (2006). The PedsQL™4.0 as a school population health measure: Feasibility, reliability, and validity. *Quality of Life Research*, 15, 203–215. doi:10.1007/s11136-005-1388-z
- Varni, J.W., Burwinkle, T.M., Seid, M. & Skarr, D. (2003). The PedsQL™4.0 as a Pediatric Population Health Measure: Feasibility, Reliability, and Validity. *Ambulatory pediatrics*, 3, 329–341.
- Varni, J.W., Burwinkle, T.M., Katz, E.R., Meeske, K. & Dickinson, P. (2002). The PedsQL™ in Pediatric Cancer – Reliability and Validity of the Pediatric Quality of Life Inventory™ Generic Core Scales, Multidimensional Fatigue Scale, and Cancer Module. *Cancer*, 94, 2090–2106.
- Varni, J.W., Seid, M. & Kurtin, P. (2001). The PedsQL™4.0: Reliability and validity of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales in healthy and patient populations. *Medical Care*, 39, 800–812.
- Wille, N., Badia, X., Bonsel, G., Burström, K., Cavrini, G., Devlin, N., Egmar, A-C., ... & Ravens-Sieberer, U. (2010). Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Quality of Life Research*, 19, 875–886. doi: 10.1007/s11136-010-9648-y
- World Federation of Occupational Therapy [WFOT]. (2012). Definition of Occupational Therapy. Heruntergeladen von <http://www.wfot.org/AboutUs/AboutOccupationalTherapy/DefinitionofOccupationalTherapy.aspx> am 05.10.2015

6.2 Abbildungsverzeichnis

Abbildung 1. Darstellung des Auswahlprozesses der Assessments.....	11
Abbildung 2. Darstellung des Auswahlprozesses der Hauptstudien.....	13
Abbildung 3. Herkunft der Teilnehmenden	23
Abbildung 4. Die Verteilung der fünf Sprachversionen des Tests.....	23
Abbildung 5. Verteilung der chronischen Krankheiten der Teilnehmenden	23
Abbildung 6. Die Herkunft/Nationalität der Teilnehmenden.	28
Abbildung 7. Der Test wurde in diesen Sprachen durchgeführt.....	28
Abbildung 8. Die mütterliche Ausbildung als Kontrollvariable	31
Abbildung 9. Relevante Ergebnisse der Konstruktvalidität	34

6.3 Tabellenverzeichnis

Tabelle 1 Beschreibung der Reliabilität	5
Tabelle 2 Beschreibung der Konstruktvalidität.....	6
Tabelle 3 Beschreibung des EuroQoL-5D-Youth.....	7
Tabelle 4 Beschreibung des PedsQL™4.0.....	8
Tabelle 5 Beschreibung der zehn Dimensionen des KIDSCREEN-52	9
Tabelle 6 Keywordtabelle zur Suche der Hauptstudien	10
Tabelle 7 Thematische Ein- und Ausschlusskriterien der Hauptstudien	12
Tabelle 8 Methodische Ein- und Ausschlusskriterien der Hauptstudien	13
Tabelle 9 Sample und Datenerfassung	16
Tabelle 10 Test-Retest-Reliabilität des EuroQoL-5D-Youth (Wille et al., 2010)	17
Tabelle 11 Validität des EuroQoL-5D-Youth (Wille et al., 2010)	18
Tabelle 12 Limitationen und Bias in Ravens-Sieberer et al. (2010)	19
Tabelle 13 Limitationen und Bias in Canaway und Frew (2013)	21
Tabelle 14 Teilnahmebereitschaft in Bezug auf Sprachversion des Fragebogens ...	25
Tabelle 15 Teilnahmebereitschaft in Bezug auf Nationalität der Teilnehmenden	25
Tabelle 16 Limitationen und Bias in Varni et al. (2003)	26
Tabelle 17 Ergebnisse der Bekannte-Gruppen-Validität.....	30
Tabelle 18 Limitationen und Bias in Varni et al. (2006)	31
Tabelle 19 Limitationen und Bias in Robitail et al. (2006)	35

Tabelle 20 Gruppeneinteilung und länderspezifische Datenerfassung der Studie....	36
Tabelle 21 Limitationen und Bias in Ravens-Sieberer et al. (2008)	39
Tabelle 22 Zusammenfassung der Ergebnisse	41

Anhang

A WORTZAHL

Abstract: 190

Bachelorarbeit: 7'920

B ABKÜRZUNGSVERZEICHNIS

Aufl.	Auflage, Edition.
BGV	Bekannte-Gruppen-Validität
CFA	Confirmatory Factor Analysis
CSHCN	Child Special Health Care Needs
d.h.	das heisst
etc.	et cetera, und so weiter
ICC	Intra-Klassen-Korrelation
IIC	Item-interne Konsistenz
IDV	Item-diskriminante Validität
INFIT	Information-Mean-Square-Fit
LQ	Lebensqualität
n.d.	nicht datiert
NNFI	Non-normed-Fit-Index
PCM	Partial Credit Measurement
PedsQL™4.0	Pediatric Quality of Life Inventory™ 4.0
RMSEA	Root-Mean-Squar-Estimation
RRSM	Rasch-Rating-Scale-Models
SDQ	Strenght and Difficulties Questionnaire
SES	sozioökonomischer Status
TN	Teilnehmende der Studie
u.a.	Unter anderem
VAS	visuelle Analogskala
vgl.	vergleichen mit der folgenden Quelle
WLS	Weghted-Least-Square-Estimation

C GLOSSAR

Aktivität	Aktivitäten sind Tätigkeiten, die für die Klientinnen und Klienten keine spezielle Bedeutung haben oder auch als gleichgültig betrachtet werden. Beispiele sind Putzen oder Aufräumen, wobei es die Individualität der Klientinnen und Klienten zu beachten gilt.
Alltagsfunktionen	Unter Alltagsfunktionen werden unter anderem Aktivitäten des täglichen Lebens verstanden. Alltagsfunktionen können sowohl Betätigungen (siehe Betätigung) oder Aktivitäten (siehe Aktivität) sein. Diese Alltagsfunktionen werden benötigt, um Tätigkeiten für sich (d.h. für die Klienten) zufriedenstellend auszuführen.
Assessmentmethode	Hierbei werden die beiden Möglichkeiten verstanden, wie die Lebensqualität von Kindern beurteilt werden kann. Unter der Assessmentmethode werden demnach die Selbst- und die Proxyversion verstanden.
Betätigung	Betätigungen sind Aktivitäten, welche für Klientinnen und Klienten bedeutsam sind. Betätigungen sind individuell, weshalb sie nicht mit spezifischen Tätigkeiten verallgemeinert in Verbindung gebracht werden können.
Bias	Zu Bias (Verzerrungen) kommt es, wenn systematische Fehler in der Datenerfassung existieren, welche die Ergebnisse negativ oder positiv beeinflussen. Dadurch wird ein Ergebnis fehlinterpretiert.
Bodeneffekt	Ein Bodeneffekt tritt dann auf, wenn zu viele Teilnehmende eines Tests die tiefste Antwortmöglichkeit (meistens "sehr schwierig" oder "trifft immer zu" auswählen. Dies lässt auf sehr schwierig oder unverständlich formulierte Testitems schliessen.

Bool'sche Operatoren	Boolesche Operatoren sind die Worte AND (und), OR (oder) und NOT (nicht). Mittels diesen Operatoren werden Schlüsselwörter / Suchwörter bei der Datenbankrecherche miteinander kombiniert, um möglichst spezifische Treffer zu erhalten.
CFI	Der "Comparative Fit-Index" ist ein statistisches Gütekriterium der Strukturgleichungsmodelle. Mittels CFI kann angegeben werden, "wie gut" ein statistisches Modell eine Menge Beobachtungen erklären kann. Neben dem CFI hat sich u.a. ebenfalls der "nicht-normierte Fit-Index" (siehe NNFI) etabliert. Ein CFI-Wert von >0.90 unterstützt die Anpassungsgüte des Modells (= Goodness of fit).
CHIP	Das „Children Health and Illness Profile“ (Starfield et al., 1994) ist ein Erfassungsinstrument der Lebensqualität für Kinder im Alter von sechs bis 17 Jahren. Zur Beurteilung existiert sowohl eine Selbst-, als auch eine Proxyversion. Verfügbar sind verschiedene Sprachen, jedoch existiert keine deutsche Version. Mehr Informationen sind unter www.childhealthprofile.org beschrieben.
Cohen's Kappa-Koeffizient	Mittels Kappa-Koeffizient (Cohen, 1968) wird die Stabilität (Test-Retest-Reliabilität) des Assessments bestimmt. Anhand diesem Koeffizienten wird ebenfalls die Berechnung der Inter-Rater-Reliabilität möglich. Landis & Koch (1977) schlagen folgende Werte zur Beurteilung der Kappa-Koeffizienten vor: 0.0-.20 leichte (slight) Übereinstimmung (= Agreement), .21-.40 ausreichende (fair) Übereinstimmung; .41-.60 moderate Übereinstimmung; .61-.80 beachtliche (substantial) Übereinstimmung; >.80 fast perfekte (almost perfect) Übereinstimmung.
Cronbach's Alpha-	Der Cronbach's Alpha-Koeffizient (Cronbach, 1951) wird hauptsächlich in der Sozialwissenschaft und Psychologie angewendet.

Koeffizient	Anhand diesem Koeffizienten wird die Interne Konsistenz berechnet. Dabei gilt der Standardwert von .70 (Nunnally & Bernstein, 1994).
CSHCN	Das „Child Special Health Care Need Screener“ (Bethell, 2002) Assessment dient zur Identifikation von körperlichen und geistigen Beeinträchtigungen bei Kindern. Dieses Assessment enthält fünf Dimensionen à insgesamt, wobei jedes nach einer chronischen Beeinträchtigung der Gesundheit fragt.
Datenkontaminierung	Kontaminierung kann auftreten, wenn beispielsweise zwei Gruppen miteinander verglichen werden. Dabei erhält Gruppe 1 die spezielle Therapie, Gruppe 2 nicht. Tauschen sich die beiden Gruppen aus, könnten die Teilnehmenden anderes Verhalten zeigen, wodurch die Ergebnisse verzerrt werden. Im Falle dieser Arbeit kann Kontaminierung aufgetreten sein, wenn die Proxys und die Kinder die beiden Assessmentmethoden gemeinsam beantworteten und sich währenddessen austauschten.
Deckeneffekt	Ein Deckeneffekt tritt dann auf, wenn zu viele Teilnehmende einer Umfrage die beste Antwort (meistens "immer" oder "nie") anwählen. Dabei wird eine vorgängige Definition der zulässigen "Top-of-Antworten" überschritten.
Differential Item Functioning	Mit dem Different Item Functioning (DIF) wird angegeben, ob Unterschiede zwischen den Items der Proxyversion und denselben der Selbstversionen vorliegen. Gemäss B. Gantschnig (persönliches Gespräch am 12.5.16) darf der Unterschied zwischen den beiden Methoden (siehe Assessmentmethode) .041 betragen. Zusätzlich darf der DIF als unwichtig betrachtet werden, wenn weniger als 10% der Items einen höheren DIF-Wert aufweisen.

Dimension-Score	Ein Dimensions-Score wird aus den summierten Item-Scores (siehe Item-Score) berechnet. Mittels diesen Dimension-Scores werden die Assessments anschliessend ausgewertet.
Ein- und Ausschlusskriterien	Diese Kriterien werden zur Auswahl der Studien formuliert. Anhand diesen Kriterien können unpassende Studien aussortiert werden.
Ergebnisse	In dieser vorliegenden Arbeit werden Ergebnisse von Resultaten getrennt betrachtet. Von Ergebnissen wird gesprochen, wenn die Ergebnisse der Hauptstudien aufgezeigt werden. Mittels Resultaten hingegen wird die Fragestellung anschliessend beantwortet. Daher sind Ergebnisse die Resultate der Studien und die Resultate die Erkenntnisse dieser vorliegenden Arbeit.
Faktorenanalyse	Bei einer Beobachtung werden viele verschiedene Merkmale erfasst. Mittels Faktorenanalyse (CFA = Confirmatory Factor Analysis) werden die, den vielen erfassten Merkmalen zugrunde liegenden, latenten Variablen analysiert.
Fisher-z-Transformation	Da empirische Korrelationskoeffizienten nicht normalverteilt sind, muss vor der Berechnung der Konfidenzintervalle eine Verteilungskorrektur mittels Fisher-z-Transformation vorgenommen werden. Durch die Fisher-Z-Transformation werden die nicht-normalverteilten Korrelationskoeffizienten zur Variable z transformiert, welche normalverteilt ist.
Funktionsskala	Hierbei handelt es sich um eine Dimension. Beim PedsQLTM4.0 wird eine Dimension Funktionsskala genannt.
Gesamtscore	Darunter wird die Summe aller Dimensionsscores (siehe Dimensionsscore) verstanden.

Gesundheits- ökonomische Evaluationen	Unter diesem Begriff werden Evaluationen des Gesundheitswesens verstanden. Z.B. die Überprüfung von ergotherapeutischen Interventionen auf deren Wirksamkeit oder die Entwicklung von präventiven Massnahmen, zur Verminderung der zunehmenden Kosten im Gesundheitswesen.
Hauptstudien	Anhand diesen Studien wird die Fragestellung beantwortet. Im Falle der vorliegenden Arbeit handelt es sich um quantitative Primärliteratur, welche die drei gewählten Assessments (EuroQoL-5D-Youth, PedsQLTM4.0 und KIDSCREEN-52) auf deren Validität, Reliabilität und Praktikabilität überprüfen.
Heterotrait-He- teromethod- Korrelationen	Diese Erfassungsmethode ist Teil der Multitrait-Methode (siehe Multitrait-Methode). Mittels diesen Angaben wird divergente Validität aufgezeigt.
Heterotrait-Mo- nomethod-Kor- relationen	Diese Erfassungsmethode ist Teil der Multitrait-Methode (siehe Multitrait-Methode). Mittels diesen Angaben wird divergente Validität aufgezeigt.
INFIT	Der "Information Mean Square Fit" wird zur Evaluation der Anpassungsgüte verwendet. Dabei befindet sich ein akzeptabler Wert zwischen .70 und 1.30 (Wright & Stone, 1979).
Intra-Klassen- Korrelation	Mit diesem statistischen Verfahren wird die Übereinstimmung von verschiedenen Ratern beurteilt. Hierbei werden intervallskalierte Daten vorausgesetzt. Die Korrelationskoeffizienten können dabei Werte von -1 bis +1 annehmen. Es gilt, je näher +-1, desto besser.
IST-Zustand	Der IST-Zustand (Gegenteil von SOLL-Zustand; siehe SOLL-Zustand) beschreibt das aktuelle Wohlbefinden. Wenn der IST-Zustand dabei vom SOLL-Zustand (siehe SOLL-Zustand) abweicht, existiert eine Einschränkung. Z.B. Ein Kind möchte und sollte die

Schuhe selber anziehen können, kann dies jedoch nicht, weicht der IST-Zustand (das Kind kann die Schuhe nicht anziehen) vom SOLL-Zustand (das Kind möchte die Schuhe selber anziehen können) ab und beeinträchtigen dadurch die Lebenszufriedenheit.

Item	Ein psychologischer Test enthält verschiedene Items, woran ein Konstrukt erfasst wird. Dabei werden Items mit Testaufgaben oder Testfragen gleichgesetzt.
Item-diskriminante Validität (IDV)	Anhand der IDV wird überprüft, ob die Items eher mit einer anderen Dimensionsscore korrelieren. Dazu wird der Item-Score (siehe Item-Score) mit allen Dimensions-Scores korreliert. IDV wird bestätigt, wenn die Items mit der "eigenen" Dimension am höchsten korrelieren.
item-interne Konsistenz (IIC)	Mittels der IIC wird überprüft, ob alle Items der korrekten Dimension angehören. Dabei werden die Item-Scores (siehe Item-Score) mit den dazugehörenden Dimensionsscore (siehe Dimensionsscore) korreliert.
Item-Score	Mittels Item-Scores werden die Items beurteilt, damit sie ausgewertet werden können. Im Falle der gewählten drei Assessments dieser Arbeit werden Punkte zwischen 0 und 5 verteilt. Dabei ist für jedes Assessment separat definiert, wie die Items quantifiziert werden.
KIDSCREEN-10	Das KIDSCREEN-10 (auch genannt Gesundheits-Index) ist eine abgeänderte stark verkürzte Version des KIDSCREEN-52. Dieses Erfassungsinstrument erfasst die Lebensqualität von Kindern mittels 10 Items, welche einer Dimension angehören. Mehr Informationen sind unter www.kidscreen.org verfügbar.

KIDSCREEN-27	Das KIDSCREEN-27 ist eine abgeänderte Version des KIDSCREEN-52, jedoch mit ausschliesslich 27 Items. Diese 27 Items werden zu fünf Dimensionen zugeteilt. Weitere Informationen sind der Internetseite des KIDSCREEN-27 (www.kidscreen.org) zu entnehmen.
Konkordanz	Unter Konkordanz wird die Übereinstimmung der Antworten der beiden Assessmentmethoden (siehe Assessmentmethode). Die Konkordanz wird berechnet, indem die Selbst- mit der Proxymethode korreliert werden.
Konstrukt	Ein Konstrukt ist ein Cluster gleicher oder ähnlicher Merkmale. Dabei werden Konstrukte wiederum in Sub-Konstrukte unterteilt. Z.B. Lebensqualität: Das Konstrukt ist die Lebensqualität. In diesem Konstrukt sind verschiedene Aspekte enthalten, beispielsweise die körperlichen Aspekte, die psychischen Aspekte, die schulischen Aspekte. Jeder dieser Aspekt gilt als Sub-Konstrukt.
Kriteriumsvalidität	Von Kriteriumsvalidität wird gesprochen, wenn möglichst hohe Korrelationen zwischen dem gemessenen Merkmal und den berechneten Messergebnissen. Kurz, es wird überprüft ob das gewünschte Merkmal gemessen wird (Haupt, 2007).
Lebensqualitätsfördernde Massnahmen	Unter Lebensqualitätsfördernden Interventionen werden Massnahmen verstanden, welche die Lebensqualität verbessern. Diese sind abhängig von Alter und Vorlieben einer Person. Kinder beispielsweise sehen Spiele als lebensqualitätsfördernde Interventionen an.
Likert-Skala	Eine Likert-Skala wird in Evaluationsumfragen oft verwendet. Hierbei werden je nach Typ unterschiedlich viele Antwortmöglichkeiten angeboten. Z.B. nie - selten - manchmal - oft - immer (Beispiel für

eine Fünf-Punkte-Likert-Skala, es werden fünf Antwortmöglichkeiten gegeben.

Literaturreview Ein systematisches Literaturreview fasst die Erkenntnisse verschiedener Studien zusammen. In diesem Fall besteht das Literaturreview aus insgesamt sechs Studien, den sogenannten Hauptstudien (siehe Hauptstudien).

Loading-Patterns Loading-Patterns beschreiben das Muster der Faktorladungen (bei einer Faktorenanalyse) auf dem Konstrukt, welches erfasst wird.

M Mittelwert. In diesem Fall ist der Altersdurchschnitt der Kinder gemeint.

Maximum-Likelihood Methode Diese Methode ist eine von vielen statistischen Schätzverfahren und dient zur Identifikation der Anpassungsgüte (= Goodness of Fit), welche u.a. mittels CFI (siehe CFI) und NNFI (siehe NNFI) angegeben wird. Da die Befragung der vollständigen Population aus Kosten- und Zeitgründen unmöglich ist, werden Stichproben gezogen und diese berechnet. Anhand der berechneten Werte wird nun geschätzt, ob diese berechneten Werte auf die gesamte Population zutreffen.

Merkmalsvalidität Es wird überprüft, ob der Untersuchte Merkmalsbereich gültig wiedergegeben wird (siehe auch Kriteriumsvalidität).

Monotrait-Heteromethod-Korrelationen Diese Erfassungsmethode ist Teil der Multitrait-Methode (siehe Multitrait-Methode). Mittels diesen Angaben wird konvergente Validität aufgezeigt.

Multiple Regression	Mittels Multipler Regression wird überprüft ob weitere Aspekte das gemessene Merkmal beeinflussen. Im Falle dieser Arbeit wurde überprüft, ob der mütterliche Bildungsgrad die Beurteilung der Kinder beeinflusst. Wie wird diese berechnet?
Multitrait-Methode (MTMM)	Die Multitrait-Methode beschreibt verschiedene statistischer Tests zur Aufzeigung von Konstruktvalidität (konvergente und divergente Validität). Bei dieser Methode werden verschiedene Messmethoden und Persönlichkeitsmerkmale miteinbezogen. Die Anordnung der Konstruktvalidität erfolgt anhand der Multitrait-Multimatrix. Beispiel: Mit dieser Methode kann das Verhalten eines Kindes hinsichtlich Konzentrationsfähigkeit und Sozialverhalten aus Sicht der Eltern, der Lehrperson oder des Kindes selber fremd- und selbstbeurteilt werden.
NNFI	Der "Non-Normed-Fit-Index" ist, wie das CFI (siehe CFI), ein statistisches Gütekriterium zur Identifikation der statistischen Güte eines Modells. Ein Wert von $>.95$ unterstützt die Anpassungsgüte eines Modells (= Goodness of fit).
PCM	Das „Partial Credit Model“ ist ein Teilaspekt der Rasch-Analyse (siehe Rasch-Analyse). Mittels PCM wird das eigentliche Verhalten der Testpersonen während der Testdurchführung berechnet.
Performanz	Die Performanz beschreibt, wie Klientinnen und Klienten eine Tätigkeit ausführen können. In der Ergotherapie werden Tätigkeiten (siehe Aktivität, siehe Betätigung) geübt, damit die Ausführung (wieder) für die Klientin / den Klienten zufriedenstellend ist.

Primärliteratur	Primärquellen (auch Primärstudien) sind Literaturquellen, die empirische Daten erfassen. Es wird somit eine komplette Studie durchgeführt, im Gegensatz zu Sekundärquellen (siehe Sekundärliteratur).
Proxy-Kinder-Paare	Proxy-Kinder-Paare bedeuten im Fall dieser Studie, dass beide Assessmentmethoden (siehe Assessmentmethode) von Eltern und deren Kinder beantwortet wurden.
Proxyversion	Mittels Proxyversion wird die Lebensqualität durch Proxys (siehe Proxy) fremdbeurteilt. Diese Methode beinhalten dieselben Items (siehe Item) wie die Selbstmethode, welche aber leicht anders formuliert sind. Z.B. Selbstversion: <i>Für mich ist es schwierig, um mehr als einen Block zu laufen.</i> Proxyversion: <i>Wie oft hatte Ihr Kind Probleme mit um mehr als einen Block zu laufen?</i>
Publication Bias	Ein „Publication Bias“ liegt meistens vor, wenn die Publikationen ausgesucht werden. Denn es ist nicht möglich, dass alle Publikationen angezeigt werden können und somit potentiell besser geeignete Quellen vorhanden sind.
Rasch-Analyse	Das Rasch-Modell ist ein statistisches Testverfahren der Parameterschätzung. Mit diesem Testverfahren werden die Vergleiche der Assessments unabhängig von den teilnehmenden Testpersonen möglich.
Ressourcen	Persönliche Ressourcen sind ergotherapeutisch gesehen die Stärken einer Person.
Resultate	Ist von Resultaten die Rede, sind die Erkenntnisse dieser Studie (u.a. die Beantwortung der Fragestellung) gemeint. Erkenntnisse der Studien werden als Ergebnisse (siehe Ergebnisse) definiert.

Retest-Sample	Die Test-Retest-Reliabilität wird an einer kleineren Gruppe des initialen Samples überprüft. Dabei wird in dieser vorliegenden Arbeit unter Retest-Sample die Gruppe, an welcher der Test wiederholt wurde, verstanden.
RMSEA	Beim "Root Mean Square Error of Approximation" handelt es sich um ein statistisches Testverfahren, welches zur Berechnung des Goodnes of Fit verwendet wird. Ein RMSEA-Wert $<.08$ bedeutet akzeptable Anpassungsgüte des Modells.
RRSM	Das RRSM ist ein Aspekt der Rasch-Analyse (siehe Rasch-Analyse).
Scaling-Success	Von Scaling-Success, also Skalierungserfolg, wird gesprochen, wenn überprüft wird, wie treffend die empirischen Daten auf einer Skala eingeordnet werden können.
Schwierigkeiten	Ergotherapeutisch gesehen sind Schwierigkeiten Probleme bei der Performanz (siehe Performanz), welche sich in der Ausführung der Aktivitäten zeigen.
SDQ	Mittels "Strength and Difficulties Questionnaire" von Goodman (1997) kann die Anwesenheit einer chronischen Beeinträchtigung diagnostiziert werden. Gefragt wird mit 25 Items, welche zu insgesamt fünf verschiedenen Dimensionen zugeteilt wurden.
Sekundärliteratur	Unter Sekundärliteratur wird Literatur verstanden, welche Primärliteratur inkludiert. Beispielsweise sind dies Literaturreviews oder einzelne Fachbücher.

- Selbstversion** Mittels dieser Assessmentmethode beurteilen die Kinder ihre eigene Lebensqualität selber. Dabei kann der Begriff 'Selbstversion' synonym mit 'Selbstbeurteilung' oder 'Selbsteinschätzung' verwendet werden. Der Begriff wurde im Rahmen dieser Arbeit durch die Verfasserin definiert, um der Leserschaft das Verständnis der Arbeit zu erleichtern.
- SOLL-Zustand** Der SOLL-Zustand ist das Gegenstück zum IST-Zustand (siehe IST-Zustand). Darunter wird ein Zustand verstanden, der eigentlich vorhanden sein sollte. Ist er es nicht wird von IST-Zustand gesprochen.
- Test-Sensitivität** Die Test-Sensitivität misst, wie korrekt der Test misst. D.h. Es wird die Anzahl korrekter und die Anzahl falscher Testergebnisse berechnet.
Beispiel: Der Test soll messen, ob eine chronische Krankheit vorliegt. Stimmen viele Testergebnisse mit der physischen Konstitution der Testpersonen überein, zeigt der Test gute Sensitivität.
- Trunkierungszeichen** Trunkierungszeichen werden in der Literaturrecherche verwendet, um spezifische Treffer zu erhalten. Gemeint sind Anführungszeichen (") und ein Sternchen (*). Mittels Anführungszeichen werden die zwischen Anführungs- und Schlusszeichen definierten Wörter als Wortkombination gesucht. Z.B. "gesundheitsbezogene Lebensqualität". Würde ohne Anführungs- und Schlusszeichen gesucht, würde nach gesundheitsbezogene und Lebensqualität separat gesucht werden. Mit der Ergänzung des Sterns können Wörter mit unterschiedlichen Wortendungen gesucht werden. Z.B. wird mit effec* sowohl nach effect, als auch nach efficacy gesucht.

Visuelle Analogskala	Mittels visueller Analogskala (VAS) wird im Spital o.ä. das subjektive Wohlbefinden der Patientinnen und Patienten ermittelt. Meist wird gefragt "Auf einer Skala von 0 bis 10, wie fühlen Sie sich heute". Diese Frage kann ebenfalls anhand einer Smiley-Skala mit unterschiedlichen Gesichtsausdrücken oder einer Thermometer-Darstellung" (wie im Falle des EuroQoL-5D-Youth; Wille et al., 2010) visuell / schriftlich gefragt werden.
Wert	Unter Werte werden berechnete Ergebnisse verstanden. Hingegen wird von Scores gesprochen, wenn es sich um die erzielten Punkte handelt, die in den Assessments erreicht wurden.
Weighted-Least-Square-Methode	Die „Methode der kleinsten Quadrate“ ist ein mathematisches Standardverfahren zur Ausgleichsrechnung.
YQOL-S	Mit dem „Youth Quality of Life-Surveillance Instrument“ kann die Lebensqualität von Kindern zwischen elf und 18 Jahren erfasst werden. Es existieren dafür eine Proxy- und eine Selbstversion. Weitere Infos sind der Internetseite (http://depts.washington.edu/seaqol/) zu entnehmen.
χ^2 -Quadrat-Test	Ein Chi-Quadrat-Test wird untersucht, ob die erhobenen Daten regelmässig verteilt wurden. Im Falle der vorliegenden Arbeit wurde überprüft, ob die Scores der kränkeren Kinder Regelmässigkeiten aufwiesen. Dabei wurde festgestellt, dass gesündere Kinder bessere Scores erzielten.

D DANKSAGUNG

Zuerst möchte ich Frau Gantschnig für die engmaschige Betreuung danken. Sie unterstützte mich während der gesamten Zeit bei der Bearbeitung der Fragestellung, beantwortete jederzeit meine Fragen oder gab Anregungen zur Schreibqualität. Zusätzlich gab sie das Thema als Partnervorschlag vor und steckte mich mit dem Interesse an der Lebensqualitätsforschung an.

Gerne möchte ich mich auch bei Frau M. Huber bedanken, welche mich in statistischen Fragen gut betreut hat und mich mit Rat und Antwort unterstützte.

Zudem gilt ein grosses Dankeschön den Lektoren – speziell gemeint sind B. Schulthess, S. Scherer und R. Zehnder. Dank ihren Inputs – und hauptsächlich dank S. Scherer – ist der Aufbau meiner Bachelorarbeit strukturiert, geordnet und logisch. Des Weiteren bedanke ich mich herzlich bei meinen Freunden, welche mich mit fachlichen Inputs im Schreibprozess unterstützt haben.

Zuletzt möchte ich mich herzlich bei meiner Familien und meinen Freunden für die soziale Unterstützung bedanken. Sie ermutigten mich in jeder Situation, mit der Arbeit fortzufahren.

E EIGENSTÄNDIGKEITSERKLÄRUNG

„Ich Catherine Gürber erkläre hiermit, dass ich die vorliegende Arbeit selbstständig, ohne Mithilfe Dritter und unter Benutzung der angegebenen Quellen verfasst habe.“

.....

Winterthur, 31.Mai 2016

F KRITISCHE WÜRDIGUNG DER HAUPTSTUDIEN

Anhand des Vorlage nach Law et al. (1998) wurden alle sechs Hauptstudien kritisch geprüft und gewürdigt. Da diese Vorlage für Interventionsstudien konzipiert wurde, mussten Anpassungen vorgenommen werden: Bei ‚Massnahmen‘ wird die Datenerfassung beschrieben. Bei ‚Outcome-Messung‘ wird beschrieben, mit welchen statistischen Tests die Gütekriterien erfasst wurden. Diesbezüglich konnte kein Bereich definiert werden.

Ravens-Sieberer et al. (2010)

Es folgt in die kritische Würdigung der Studie von Ravens-Sieberer et al. (2010).

Formular zur kritischen Besprechung	Dieses Dokument wurde anhand der Vorlage nach Law et al. (1998) durch die Verfasserin generiert.
Titel	Feasibility, reliability, and validity of the EuroQoL-5D-Youth: results from a multinational study
Publikation	Ravens-Sieberer et al., 2010, <i>Quality of Life Research</i> , 19, 887-897.
Zweck der Studie	
Wurde der Zweck klar angegeben?	Ja
Skizzieren Sie den Zweck der Studie. Inwiefern bezieht sich die Studie und/oder auf Ihre Forschungsfrage?	Die Studie überprüft die Reliabilität, Validität und Praktikabilität des EuroQoL-5D-Youth (Wille et al., 2010).
Literatur	
Wurde relevante Hintergrundliteratur gesichtet?	Ja
Geben Sie an, wie die Notwendigkeit der Studie gerechtfertigt wurde.	Es besteht Bedarf an Assessments zur Erfassung der Lebensqualität von Kindern, da mehrheitlich Assessments zur Erfassung der erwachsenen Lebensqualität existieren. Diese sind für Kinder jedoch ungeeignet. Die vorliegende Studie stellt nach Ansicht der Autorenschaft und der Verfasserin die erste grossflächige Überprüfung des EuroQoL-5D-Youth (Wille et al., 2010) bei Kindern dar.
Design	
Welches Design wurde verwendet?	Prospektives, kulturübergreifendes Design und deskriptive Statistik

Beschreiben Sie das Studiendesign. Entsprach das Design der Studienfrage (z.B. im Hinblick auf den Wissensstand zur betreffenden Frage, auf Ergebnisse, auf ethische Aspekte)?

Die Autorenschaft erhob empirische Daten zur statistischen Überprüfung der Reliabilität, Validität und Praktikabilität in 5 Ländern (Deutschland, Spanien, Italien, Schweden und Südafrika).

Spezifizieren Sie alle systematischen Fehler (Verzerrungen, Bias), die vielleicht aufgetreten sein könnten, und in welche Richtung sie die Ergebnisse beeinflussen.

Die TN, an welchen die Test-Retest-Reliabilität überprüft wurde, hätten sich an die gegebenen Antworten des ersten Testdurchlaufs erinnern können. Folglich hätten sie dieselben Antworten erneut auswählen können, was zu einer Verzerrung der Messwerte führen würde. Zusätzlich besteht die bekannte Gefahr, dass sich die zu erfassenden Merkmale zwischen Test und Retest (7-10 Tage bei der vorliegenden Studie) signifikant verändern, was natürlich die entsprechenden Ergebnisse erheblich verzerren würde. Die Retest-Reliabilität wurde lediglich in zwei Studienländern erhoben. Die Fragebogenadministration wurde in den verschiedenen Ländern unterschiedlich erteilt, was sich negativ auf das Leseverständnis der Studie und somit auf die Fehlerwahrscheinlichkeit (Drop-Outs) ausgewiesen haben könnte.

Stichprobe

N=2'809

Wurde die Stichprobe detailliert beschrieben?

Ja. 2'809 TN, zwischen 8 und 19 Jahren, mehr Knaben als Mädchen, von Deutschland, Schweden, Spanien, Italien und Südafrika stammend. Nur Regelschulen inkludiert. Test-Retest-Reliabilität in 33% des Samples in Italien und Spanien überprüft.

Wurde die Stichprobengröße begründet?

Ja. Es wurden mindestens 200 TN pro Land rekrutiert, um die Signifikanz der Ergebnisse zu garantieren.

Ergebnisse (Outcomes)

Geben Sie an, wie oft Outcome Messungen durchgeführt wurden (also vorher, nachher, bei Nachbeobachtungen...).	Outcome Bereiche	Listen Sie die verwendeten Messungen auf.
	Es wurden mehrheitlich einmal Daten erhoben. Dabei wurde die Test-Retest-Reliabilität an 33% des italienischen und spanischen Samples überprüft, wobei in diesen Ländern erneut Daten erhoben wurden.	<p>Test-Retest-Reliabilität: Cohen's Kappa-Koeffizient (Cohen, 1988) und Intra-Klassen-Korrelation (Bartko, 1966) zur Korrelationsberechnung der Item-Antworten und des selbstbeurteilten Gesundheitszustands (VAS).</p> <p>Konvergenzvalidität: Spearman-Rang-Korrelationen zwischen EuroQoL-5D-Youth (Wille et al., 2010) und KIDSCREEN-27 (AUTOR) oder KIDSCREEN-10 (AUTOR).</p> <p>Bekannte-Gruppen-Validität: χ^2-Test</p>

Waren die Messungen gültig und zuverlässig?	gültig: Ja, wurde als Ziel der Studie nachgewiesen	zuverlässig: Ja, wurde als Ziel der Studie nachgewiesen
---------------------------------------------	-----------------------------------------------------------	----------------------------------------------------------------

Massnahmen

Wurden die Massnahmen detailliert beschrieben? Ja.	<p>Die Datenerfassung in den verschiedenen Ländern war unterschiedlich. Jedoch wurde sie nachvollziehbar beschrieben.</p> <p>Italien: Nur gebürtige Italienerinnen und Italiener wurden zur Administration während des Unterrichts mit schriftlicher Anleitung und personeller Unterstützung (Forschungsassistent) zugelassen. An 33% des italienischen Samples fand nach 7-10 Tage die Testwiederholung statt.</p> <p>Spanien: Mittels schriftlicher Anleitung wurden die Umfragen während des Unterrichts ohne Investigator administriert, dabei wurde die Umfrage von den TN selbstständig gelöst.</p> <p>Deutschland: Die Administration erfolgte ohne Investigator während des Unterrichts mittels schriftlicher Anleitung.</p> <p>Schweden: Die Umfrage wurde per Post versendet, wobei die TN die Umfrage zu Hause alleine beantworten sollten. Die 2 mitgesendeten Informationsbriefe und die schriftliche Anleitung informierten dabei ebenfalls die Eltern, welche ihre Einwilligung zur Teilnahme erteilen mussten. Nach zwei Wochen wurde ein Reminder versendet, welcher noch einmal zur Teilnahme aufforderte.</p> <p>Südafrika: Die Umfrage wurde während des Unterrichts mittels schriftlicher Anleitung und ohne Investigator administriert.</p>	
-----------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

Wurde Kontaminierung vermieden?	Entfällt. Kontaminierung war kein relevanter Punkt. Die miteinander verglichenen Gruppen konnten sich durch interne Absprachen nicht „verbessern“.	
---------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------	--

Wurden Ko-Interventionen vermieden?	Entfällt. Alle TN beantworteten dieselben Fragebogen, wodurch keine Gruppe zu anderen Behandlungen kam.	
-------------------------------------	----------------------------------------------------------------------------------------------------------------	--

Ergebnisse

Wurde die statistische Signifikanz der Ergebnisse angegeben?	Ja. Es wurden aber nicht alle Signifikanzlevels angegeben, wobei nicht alle Auswertungen eine Berechnung des Signifikanzlevels erfordern. Mehrheitlich wurden die Signifikanzlevel $p \leq .01$, selten auch $p \leq .05$, gewählt. Dabei wurde beides als signifikant erachtet.	
--------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

Waren die Analysemethoden geeignet?	Ja. Nach Ansicht der Verfasserin waren die Analysemethoden geeignet und zielführend. Die Autorenschaft kommentierte diesen Punkt jedoch nicht explizit.	
-------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------	--

Welches waren die Ergebnisse? Waren sie statistisch signifikant?	<p>Test-Retest-Reliabilität: Mehrheitlich ausreichende bis moderate Test-Retest-Reliabilität wurde festgestellt. Dabei war im italienischen Sample die ‚Mobilität‘ nicht signifikant und für die ‚Selbstfürsorge‘ konnte kein Kappa-Koeffizient (Cohen, 1988) aufgrund identischer Test- und Retestantworten berechnet werden. Im spanischen Sample wurden mehrheitlich moderate bis substantielle Kappa-Koeffizienten berechnet (ausser ‚Mobilität‘: $-.003$; $p \leq .01$, 99.4% Übereinstimmung).</p>	
---------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

<p>War die Studie gross genug, um einen eventuell auftretenden wichtigen Unterschied anzuzeigen? Falls es um viele Ergebnisse ging: Wurde dies bei der statistischen Analyse berücksichtigt?</p>	<p>Konvergenzvalidität: Moderate bis hohe Korrelationen wurden zwischen psychologisch-geprägten Dimensionen festgestellt ($\rho = -.41$ bis $-.52$; signifikant). Tiefe Korrelationen wurden in der Überprüfung der körperlichen und der psychischen Funktionsskala (PedsQL™4.0) und der Dimension Mobilität festgestellt ($-.10$; $p < .05$). Die VAS korreliert mit ausreichenden bis hohen Werten mit den zu korrelierenden Kontrollassessments. Für die Dimension Mobilität wurde keine Konvergenzvalidität festgestellt.</p> <p>Bekannte-Gruppen-Validität: Gesundere TN hatten im Vergleich zu TN mit einer chron. Beeinträchtigung weniger Schwierigkeiten bei den gefragten Tasks. V.a. hatten die TN mit einer chron. Beeinträchtigung mehr Schmerzen oder waren traurig / besorgt.</p> <p>Praktikabilität: Die Items wurden bis maximal 2% falsch ausgefüllt, die VAS wurde bis 9% falsch beantwortet. (Falsch= nicht oder undeutlich beantwortet) In allen Dimensionen wurden Deckeneffekte ($\geq 69.5\%$) festgestellt, jedoch keine Bodeneffekte.</p>
<p>Wurde die klinische Bedeutung angegeben?</p>	<p>Ja. Die Ergebnisse werden korrekt interpretiert und Schlussfolgerungen abgeleitet. Die hohen Deckeneffekte können erwartet werden, wenn das EuroQoL-5D-Youth an einem gesunden Sample überprüft wird. Dies verhält sich analog der Überprüfung des EuroQoL-5D (EuroQoL Research Foundation, 1990) an einem erwachsenen Sample.</p>
<p>Welches war die klinische Bedeutung der Ergebnisse? Waren die Unterschiede zwischen Gruppen klinisch von Bedeutung?</p>	<p>Es wurden Unterschiede zwischen kränkeren und gesünderen TN berechnet. Hierbei besteht die klinische Relevanz, dass kränkere TN schlechtere Werte erzielten, wodurch ihnen eine tiefere Lebensqualität zugesprochen wird.</p>
<p>Wurden Fälle von Ausscheiden aus der Studie angegeben?</p>	<p>Nein. Es wurden keine Dropouts dokumentiert. Drop-Outs sind nach Ansicht der Verfasserin jedoch wahrscheinlich, da die Umfragen mit teilweise bis 9% Fehlern retourniert wurden und falschausgefüllte Items nicht zur Analyse der Gütekriterien verwendet werden können.</p>
<p>Schlussfolgerungen und klinische Implikation</p>	
<p>Waren die Schlussfolgerungen angemessen im Hinblick auf Methoden und Ergebnisse der Studie?</p>	<p>Ja. Es wurden alle Ergebnisse interpretiert und bei tiefen Werten mögliche Gründe dafür angegeben. Die Begründungen erscheinen der Verfasserin als angemessen und nachvollziehbar. Es wurden zusätzlich die unterschiedlichen Vorgehensweisen diskutiert, wobei sich der anwesende Investigator als Vorteil herausgestellt hat.</p>

<p>Zu welchem Schluss kam die Studie? Welche Implikationen haben die Ergebnisse für die ergotherapeutische Praxis? Welches waren die hauptsächlichen Begrenzungen oder systematischen Fehler der Studie?</p>	<p>Das EuroQoL-5D-Youth (Wille et al., 2010) ist für Kinder zwischen 8 und 19 Jahren reliabel, valide und praktikabel. Jedoch dient es nicht zur Identifikation von Beeinträchtigungen der Lebensqualität (vgl. hohe Deckeneffekte) und sollte daher nur bei Kindern mit verschiedenen chronischen Erkrankungen verwendet werden.</p> <p>Limitationen: Die Testdurchführung in den einzelnen Ländern wurde nicht vermerkt, was zu Verzerrungen führen könnte. Ebenfalls konnten keine persönlichen Daten zur psychischen und physischen Gesundheit der TN eingeholt werden. Zusätzlich kann aufgrund der nicht erfassten Inhaltsvalidität nicht beurteilt werden, ob alle Aspekte der gesundheitsbezogenen Lebensqualität erfasst werden.</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Schlussfolgerung der Verfasserin: In dieser Studie wurde die Reliabilität, Validität und Praktikabilität des EuroQoL-5D-Youth (Wille et al., 2010) an einem Sample mit über 2'800 Personen überprüft. Dabei wurde gut begründet, wie es zu einer solchen Samplegrösse kam. Die Autorenschaft gaben bis auf eine Ausnahme alle Daten an, was einen berechtigten, aber nebensächlichen Kritikpunkt darstellt.

Canaway und Frew (2013)

Es folgt die kritische Würdigung der Studie von Canaway und Frew (2013).

Formular zur kritischen Besprechung	Dieses Dokument wurde anhand der Vorlage nach Law et al. (1998) durch die Verfasserin generiert.
Titel	Measuring Preference-based quality of life in children aged 6-7 years: a comparison of the performance of the CHU-9D and EQ-5D-y – The WAVES Pilot study
Publikation	Canaway & Frew (2013). <i>Quality of Life Research</i> , 22, 175–183.
Zweck der Studie	
Wurde der Zweck klar angegeben?	Ja
Skizzieren Sie den Zweck der Studie. Inwiefern bezieht sich die Studie und/oder auf Ihre Forschungsfrage?	Anhand dieser Studie wird die Praktikabilität, Reliabilität und Validität des EuroQoL-5D-Youth an Sechs- und Siebenjährigen in England (West Midlands) überprüft und mit weiteren Assessments (CHU-9D, PedsQL™4.0) verglichen. Das übergeordnete Ziel der Studie war die Auswahl eines geeigneten Assessments vor dessen anschliessender Verwendung innerhalb eines grossen Forschungsprojekts.
Literatur	
Wurde relevante Hintergrundliteratur gesichtet?	Ja

Geben Sie an, wie die Notwendigkeit der Studie gerechtfertigt wurde.

Es gibt keine relevanten Vorstudien, welche die Anwendung des Assessments an Kindern zwischen 6 und 7 Jahren unterstützen. Die Fokussierung auf Selbstversion der Assessments löst das bekannte Problem der elterlichen Überschätzung des Kindes und ermöglicht somit neue Erkenntnisse direkt aus Kindessicht. In dieser Pilotstudie wird das EuroQoL-5D-Youth (Wille et al., 2010) validiert und mit dem CHU-9D (Stevens, 2010) verglichen, bevor es für ein Forschungsprojekt verwendet werden sollte.

Design

Welches Design wurde verwendet?

Prospektives Design mit Anteil einer Randomisierung.

Beschreiben Sie das Studiendesign. Entsprechend das Design der Studienfrage (z.B. im Hinblick auf den Wissensstand zur betreffenden Frage, auf Ergebnisse, auf ethische Aspekte)?

Es wurden empirische Daten erhoben, um das EuroQoL-5D-Youth (Wille et al., 2010) zu validieren. Obwohl die Autorenschaft das gewählte Design nicht kommentierte, stimmt dieses nach Ansicht der Verfasserin da es vergleichbar zu ähnlichen Studien ist.

Spezifizieren Sie alle systematischen Fehler (Verzerrungen, Bias), die vielleicht aufgetreten sein könnten, und in welche Richtung sie die Ergebnisse beeinflussen.

Die Tatsache, dass den TN anonyme Studienidentität zugeteilt wurde, welche mit den zu beantwortenden Assessments verlinkt wurde, lässt auf einen Anteil einer Randomisierung schliessen.
Es wurden von allen TN dieselben Fragebogen beantwortet. Dabei willigten die Eltern vorgehend in die Studienteilnahme ein.

Am Nachmittag desselben Tages, an welchem am Morgen die Datenerfassung durchgeführt wurde, fand die Testwiederholung statt. Dadurch wäre es möglich, dass die Kinder über den Mittag Unterstützung erhalten haben und die Werte dadurch verfälscht worden sind. Zudem wurden bemerkenswert hohe Deckeneffekte in vier Dimensionen entdeckt wodurch die Test-Retest-Reliabilität negativ beeinflusst werden könnte.
Verschiedene Interviewer befragten die Kinder und schätzten deren Itemverständnis ein. Dies könnte zu unterschiedlichen Ansichten geführt haben, wobei einzelne TN schlechter oder besser eingeschätzt wurden. Diesen Umstand gilt es kritisch zu beachten, da die Autorenschaft keine messbaren Kriterien beschreibt, wonach das Frageverständnis der TN beurteilt wurde.

Stichprobe

n=160

Kinder (sechs- und siebenjährig), 43% Mädchen, 63% sechsjährig

Wurde die Stichprobe detailliert beschrieben?

Ja. Die Autorenschaft beschreibt folgendes:

- Alter der TN und deren Verteilung (siehe oben)
- Geschlechterverteilung (siehe oben)
- Freiwillige Studienteilnehmende, die Einverständniserklärung wurde vorgängig bei den Eltern eingeholt.
- Die Daten der Studie wurden im der Englischen Region „West Midlands“ (Gebiet um Birmingham) erfasst.

Wurde die Stichprobengrösse begründet?

Nein. Die Begründung der Samplegrösse fehlt gänzlich. Nach Ansicht der Verfasserin ist diese Grösse für eine Pilotstudie jedoch ausreichend, da statistisch signifikante Ergebnisse erzielt wurden.

Ergebnisse (Outcomes)		
Geben Sie an, wie oft outcome Messungen durchgeführt wurden (also vorher, nachher, bei Nachbeobachtungen...).	<p>Outcome Bereiche</p> <p>Es wurde das EuroQoL-5D-Youth (Wille et al., 2010) auf dessen Reliabilität, Validität und Praktikabilität in der Anwendung an sechs- und siebenjährigen, gesunden Kindern überprüft.</p> <p>136 Kinder wurden einmal befragt, 24 Kinder wurden am selben Nachmittag erneut befragt (Test-Retest-Reliabilität).</p>	<p>Listen Sie die verwendeten Messungen auf.</p> <p>Test-Retest-Reliabilität: Prozentuale Übereinstimmung, Spearman-Rang-Korrelation (Bartko, 1966) und Cohen's Kappa (Landis & Koch, 1977; Cohen, 1988).</p> <p>Konvergenzvalidität: Spearman-Rang-Korrelationen zwischen EuroQoL-5D-Youth (Wille et al., 2010) und CHU-9D (Stevens, 2010) oder PedsQL™4.0 (Varni, Seid & Burwinkle, 2001).</p> <p>Bekannt-Gruppen-Validität: Verteilung der Antworten, Utility Score.</p> <p>Praktikabilität: Benötigte Zeit, Itemverständnis der TN und Anzahl Fehler (ungenau oder nicht beantwortete Items) pro Item in Prozent.</p>
Waren die Messungen gültig und zuverlässig?	gültig: Ja, die Validität wurde nachgewiesen.	zuverlässig: Ja, die Reliabilität wurde nachgewiesen..
Massnahmen		
Wurden die Massnahmen detailliert beschrieben?	Ja. Der gesamte Vorgang wurde detailliert beschrieben. 5 Forscher wurden ausgebildet, um die Umfrage in Interviewform während der Schulzeit durchzuführen. Einzig die Beurteilungskriterien zur Itemverständlichkeit sind nicht gänzlich dokumentiert.	
Wurde Kontaminierung vermieden?	Entfällt. Es war keine Interventionsstudie, deshalb sind Kontaminierungspräventionen nicht notwendig.	
Wurden gleichzeitig Ko-Interventionen vermieden?	Entfällt. Es war keine Interventionsstudie. Dieser Punkt ist irrelevant, da alle TN alle drei Fragebogen beantworteten.	
Ergebnisse		
Wurde die statistische Signifikanz der Ergebnisse angegeben?	Ja. Die p -Werte wurden mehrheitlich angegeben. Dabei betrachtet die Autorenschaft Werte mit $p < .1$ als signifikant. Nach der Verfasserin ist dies für die vorliegende Pilotstudie zulässig, obwohl ansonsten eine Grenze von $p < .05$ gilt.	
Waren die Analysemethoden geeignet?	Ja. Die Analysemethoden waren nach Ansicht der Verfasserin geeignet, da die Messungen zielführend und effizient waren. Diesen Punkt wurde durch die Autorenschaft nicht explizit kommentiert wird aber durch deren eigene Würdigung klar.	

Welches waren die Ergebnisse? Waren sie statistisch signifikant? Ja. War die Studie gross genug, um einen eventuell auftretenden wichtigen Unterschied anzuzeigen? Ja.	Test-Retest-Reliabilität: In allen Dimensionen wurde >70.5% Übereinstimmung berechnet. Dabei wurden moderate Kappa-Koeffizienten kalkuliert, welche mehrheitlich statistisch signifikant waren (exkl. ‚Schmerzen‘ $p < .1$ und ‚Selbstfürsorge‘ $p > .1$). Zusätzlich wurden mehrheitlich gute Spearman-Rang-Korrelationen berechnet, wobei drei Dimensionen nicht signifikant waren (‚Mobilität‘, ‚Schmerzen‘ und ‚Angst‘, je $p > .1$). Konvergenzvalidität: Die ‚Mobilität‘ korreliert tief mit der ‚physischen Funktionsskala‘ ($-.15$; $p = .07$) des PedsQL™4.0 (Varni, 1998), 8 moderate Korrelationen ($-.35$ und $.34 - .45$; $p < .01$). Bekannte-Gruppen-Validität: Unterschiede festgestellt. Keine eindeutige Angabe, welche Gruppe bessere Werte erzielten. Anhand der Abbildung und mittels Vorwissen kann abgeleitet werden, dass Kinder ohne chronische Beeinträchtigung „weniger Schwierigkeiten“ rapportieren. Praktikabilität: Die Testdauer beträgt weniger als 3 Minuten. 0% fehlerhafte Items wurden festgestellt. Es wurden aber keine Korrelationen zwischen der VAS und den Items überprüft (vgl. Ravens-Sieberer et al., 2010). (jedoch keine Korrelation mit VAS). >81% der TN verstanden die Items gut bis sehr gut. Nur 7.1% der TN verstanden die Items schlecht bis sehr schlecht. Die Inter-Rater-Reliabilität wurde nicht gemessen.
Falls es um viele Ergebnisse ging: Wurde dies bei der statistischen Analyse berücksichtigt? Nein.	
Wurde die klinische Bedeutung angegeben?	Ja. Die Autorenschaft erläutert die Ergebnisse und zieht korrekte Schlussfolgerungen.
Welches war die klinische Bedeutung der Ergebnisse? Waren die Unterschiede zwischen Gruppen klinisch von Bedeutung?	Die Ergebnisse zeigen, dass dieses Assessment für Kinder (6 und 7 Jahre) reliabel, valide und praktikabel ist, wenn es als Interview durchgeführt wird. Es gab signifikante Unterschiede zwischen kränkeren und gesünderen TN (klinische Bedeutung: Kränkere TN besitzen schlechtere Lebensqualität).
Wurden Fälle von Ausscheiden aus der Studie angegeben? Schieden Teilnehmende aus der Studie aus? Warum? Wurden Gründe angegeben und Fälle von Ausscheiden angemessen gehandhabt?	Ja. Es wurden keine Drop-Outs vermerkt. Es wird jedoch von keinen Drop-outs ausgegangen, da von den Kindern keine fehlerhaften Fragebogen retourniert wurden.

Schlussfolgerungen und klinische Implikation

Waren die Schlussfolgerungen angemessen im Hinblick auf Methoden und Ergebnisse der Studie?	Ja. Das EuroQoL-5D-Youth (Wille et al., 2010) ist für Kinder (6 und 7 Jahre) in Interviewform reliabel, praktikabel und valide.
Zu welchem Schluss kam die Studie? Welche Implikationen haben die Ergebnisse für die ergotherapeutische Praxis?	Die Anwendung wird durch die Autorenschaft nicht empfohlen, da keine geeigneten Evaluationswerte entwickelt wurden und diejenigen des EuroQoL-5D-3L (für die Erwachsenenform) nicht verwendet werden sollten (Van Reenen, 2014). Zudem erzielte das in der Studie ebenfalls verglichene CHU-9D (Stevens, 2010) bessere Wert, wodurch das EuroQoL-5D-Youth nicht für die Folgestudie verwendet wurde.
Welches waren die hauptsächlichen Begrenzungen oder systematischen Fehler der Studie?	Es wurden hohe Deckeneffekte entdeckt, welche jedoch erwartet wurden. Diese treten ebenfalls auf, wenn das EuroQoL-5D bei einem gesunden Erwachsenenprobe überprüft wird. Dadurch eignet sich dieses Assessment zur Evaluation von Beeinträchtigungen der Lebensqualität nicht. Weitere Forschung wird benötigt, um a) die Deckeneffekte zu reduzieren und b) die Verwendung bei 6- und 7jährigen detaillierter zu erforschen.

Fazit der Verfasserin: Dies war die erste Studie über EuroQoL-5D-Youth auf sechs- und siebenjährige Kinder angewendet. Einerseits ist die Darstellung der Reliabilität gut gelungen, denn es wurden mehrheitlich gute Kappa-Koeffizienten berechnet. Andererseits reflektieren die beiden Autoren diese Studie korrekt. Schlechtere Ergebnisse kommentieren sie und geben Begründungen an, weshalb diese Werte so ausfallen. Beispielsweise benennen sie, dass die Kinder teilweise von ihren Eltern bei Aktivitäten, welche die Selbstfürsorgedimension betreffen, aus Zeitgründen unterstützt wurden, obwohl sie die Aktivität (sich ankleiden) eigentlich selber hätten machen können. Ebenfalls werden alle Daten angegeben.

Diese Studie besitzt jedoch auch negative Aspekte. Einerseits wird nicht klar definiert, ob die Kinder mit einer chronischen Krankheit eine schlechtere oder bessere Lebensqualität besitzen. Die Antwort kann jedoch etwas umständlich aus den gezeigten Daten abgeleitet werden.

Varni et al. (2003)

Es folgt die kritische Würdigung der Studie von Varni et al. (2003).

Formular zur kritischen Besprechung	Dieses Dokument wurde anhand der Vorlage nach Law et al. (1998) durch die Verfasserin generiert.
Titel	The PedsQL™4.0 as a Pediatric Population Health Measure: Feasibility, Reliability, and Validity
Publikation	Varni, Burwinkle, Seid & Skarr. (2003). <i>Ambulatory Pediatrics</i> , 3, 6, 329–341.
Zweck der Studie	
Wurde der Zweck klar angegeben?	Ja
Skizzieren Sie den Zweck der Studie. Inwiefern bezieht sich die Studie und/oder auf Ihre Forschungsfrage?	In der vorliegenden Studie wird die Reliabilität (interne Konsistenz), die Validität (Bekannte-Gruppen-Validität) und die Praktikabilität des PedsQL™4.0 (Varni, 1998) an Kindern zwischen 8 und 18 Jahren untersucht. Die Datenerfassung fand im ganzen Bundesstaat California (US) statt.
Literatur	
Wurde relevante Hintergrundliteratur gesichtet?	Ja
Geben Sie an, wie die Notwendigkeit der Studie gerechtfertigt wurde.	Um den Gesundheitsstatus der Kinder im Bundesstaat California zu erfassen wurde ein geeignetes Assessment benötigt. Die Autoren untersuchen in dieser Studie, ob das Assessment PedsQL™4.0 (Varni, 1998) eine geeignete Wahl ist und somit für spätere großflächige Untersuchungen genutzt werden kann.
Design	
Welches Design wurde verwendet?	Prospektives Design mit deskriptiver Statistik
Beschreiben Sie das Studiendesign. Entsprechend dem Design der Studienfrage (z.B. im Hinblick auf den Wissensstand zur betreffenden Frage, auf Ergebnisse, auf ethische Aspekte)?	Es wurden Fragebogen versendet, welche ausgefüllt retourniert wurden. Dabei wurde die Lebensqualität von Kindern mittels der Selbstversion eingeschätzt oder mittels Proxyversion durch die Eltern fremdbeurteilt. Dabei sollten die Kinder (8-18 Jahre) den Fragebogen zu Hause ohne Unterstützung der Eltern beantworten. Die Kinder zwischen 5 und 7 Jahren durften durch ihre Eltern unterstützt werden, nachdem die Eltern die Proxyversion bereits beendet hatten. Nach Ansicht der Verfasserin wurde dieses Design korrekt gewählt, da die Art der Datenerfassung gerechtfertigt und zielführend erscheint. Mittels statistischen Tests wurde indes bestätigt, dass das PedsQL™4.0 (Varni, 1998) die Lebensqualität der Kinder zuverlässig und gültig misst. Ebenfalls wird so zielführend untersucht ob das Assessment praktikabel ist.

Spezifizieren Sie alle systematischen Fehler (Verzerrungen, Bias), die vielleicht aufgetreten sein könnten, und in welche Richtung sie die Ergebnisse beeinflussen.

Es wäre möglich, dass Eltern und Kinder den Fragebogen gemeinsam beantworteten und dadurch abgesprochene Antworten rapportierten. Somit korrelierten die zwei Methoden (Selbst- und Proxyversion) möglicherweise höher miteinander. Diese Verunreinigung kann in diesem Fall nicht ausgeschlossen werden, weil die Datenerfassung nicht überwacht wurde.

Stichprobe

N=10'241

10'241 Familien nahmen an der Umfrage teil, davon wurden 7'021 Proxyversionen für Kinder (5-16 Jahre) und 5'905 Selbstversionen von Kindern zwischen 5 und 16 Jahren zur Analyse der Gütekriterien weiterverwendet.

Wurde die Stichprobe detailliert beschrieben?

Ja. Erfasst wurde die Nationalität/Herkunft der TN, die gesprochene Sprache und ob die TN (Kinder) chronisch erkrankt sind – falls ja, welche Erkrankung das Kind hat. Ebenfalls wurde das Geschlecht und Alter der Kinder erfasst. Die Autorenschaft beschrieb das Sample sehr detailliert.

Allerdings wäre die Darstellung mittels Tabelle übersichtlicher und sinnvoller gewesen. Hingegen wurden mittels Tabellen die Gütekriterien dargestellt, wobei wiederum die Anzahl Stichproben aufgeführt wurde. Diesbezüglich wurden nicht immer alle Studien-TN zur Analyse zugelassen oder die Angaben fehlten. Beispielsweise wurden 5'905 Fragebogen zur Analyse zugelassen, werden nun aber die Selbstversionen in Tabelle 1 (S. 334) summiert, ergibt dies 4'972 Selbstversionen (siehe unten, Ausscheidungen aus der Studie). Also wurden in Tabelle 1 fehlende Angaben nicht berücksichtigt. Indes wird es für die Leserschaft schwierig, das Vorgehen nachzuvollziehen. Die Verfasserin geht hingegen davon aus, dass diese an der Anzahl Studienteilnehmenden gemessenen Abweichungen keine Auswirkungen auf die statistische Evaluation haben.

Wurde die Stichprobengrösse begründet?

Nein. Es wurde lediglich kommentiert, welche Familien zur Umfrage zugelassen wurden (a. Kinder zwischen 2 und 16 Jahren, b. Beantragung einer Krankenversicherung im Februar und März 2001, c. Gesprochene Sprache war englisch, spanisch, vietnamesisch, koreanisch oder kantonesisch).

Die Vielzahl an untersuchten Gruppen-Validitäten rechtfertigt in den Augen der Verfasserin die bemerkenswerte hohe Anzahl Studienteilnehmende.

Ergebnisse (Outcomes)		
Geben Sie an, wie oft outcome Messungen durchgeführt wurden (also vorher, nachher, bei Nachbeobachtungen...).	<p>Outcome Bereiche</p> <p>Es wurde die Fragebogenmethoden validiert, womit die Lebensqualität von Kindern erfasst werden kann. Diesbezüglich wurden einmalige Messungen durchgeführt.</p>	<p>Listen Sie die verwendeten Messungen auf.</p> <p>Interne Konsistenz: Cronbach's Alpha-Koeffizienten.</p> <p>Bekannte-Gruppen-Validität: Effektgrößen (Cohen, 1988), t-Test, χ^2-Test; Multiple Regression</p> <p>Praktikabilität: Prozentsatz von Fehlern pro Item.</p>
Waren die Messungen gültig und zuverlässig?	gültig: Ja, die Validität wurde als Studienziel überprüft	zuverlässig: Ja, die Reliabilität wurde als Studienziel überprüft
Massnahmen		
Wurden die Massnahmen detailliert beschrieben?	<p>Ja. Mittels Willkommensanruf wurden die Familien über die per Post folgende Umfrage informiert. Anschliessend wurden beide Assessmentmethoden inklusive zweier Begleitbriefe, welche die schriftlichen Anleitungen beinhaltete, versendet. Die Proxys und Kinder mussten anschliessend den jeweiligen Fragebogen selbstständig beantworten. Nur Kinder zwischen 5 und 7 Jahren durften durch ihre Eltern unterstützt werden, wenn diese die Proxyversion bereits beendet hatten. Beide Methoden wurden ans Datacenter zurückgesendet.</p>	
Wurde Kontaminierung vermieden?	<p>Nein. Es wurde nicht angegeben, ob die Familien den Fragebogen tatsächlich separat beantworteten. Ein gemeinsames Beantworten der Items hätte die bewusste oder unbewusste Beeinflussung bedeutet und die Werte vermutlich verbessert. Allerdings war die Kontrolle dieses Aspekts unmöglich, da die Datenerhebung nicht durch Forschungspersonal überwacht wurde.</p>	
Wurden Ko-Interventionen vermieden?	<p>Entfällt. Es handelt sich um eine Validierungsstudie, weshalb dieser Punkt nach Ansicht der Verfasserin irrelevant ist.</p>	
Ergebnisse		
Wurde die statistische Signifikanz der Ergebnisse angegeben?	<p>Ja. Es wurden p-Werte dokumentiert, wobei einige Angaben jedoch fehlten. Die Berechnung der p-Wertewird allerdings nicht bei allen Messungen benötigt.</p>	
Waren die Analysemethoden geeignet?	<p>Ja. Nach Ansicht der Verfasserin waren die Analysemethoden geeignet und zielführend. Sie wurden allerdings durch die Autoren nicht gesondert kommentiert. Leider wurde die Test-Retest-Reliabilität von den Autoren nicht untersucht.</p>	

Welches waren die Ergebnisse? Waren sie statistisch signifikant? War die Studie gross genug, um einen eventuell auftretenden wichtigen Unterschied anzuzeigen? Falls es um viele Ergebnisse ging: Wurde dies bei der statistischen Analyse berücksichtigt?

Interne Konsistenz: Für das Gesamtsample wurden gute bis sehr Cronbach's Alpha-Koeffizienten (Cronbach, 1951) berechnet, wobei für die ‚schulische Funktionsskala‘ die jeweils tiefsten Alpha-Koeffizienten aufgezeigt wurden. Nur 3 Dimensionen (Proxyversion für 2-4j., schulische Funktionsskala, $\alpha=.58$; Selbstversion, 5-7j., schulische Funktionsskala, $\alpha=.61$; Selbstversion, 5-7j., physische Funktionsskala: $\alpha=.68$) erzielten ungenügende Werte.

Konkordanz: Es wurden moderate bis hohe Pearson-Korrelationen ($r_{5-7 \text{ Jahre}}=.34-.70$; $r_{8-12 \text{ Jahre}}=.49-.77$; $r_{12-16 \text{ Jahre}}=.51-.79$) berechnet. Dabei wurde beobachtet, dass die Übereinstimmung grösser war, wenn die Kinder älter waren.

Bekannte-Gruppen-Validität: Gesündere Kinder im Vergleich zu weniger gesunden Kindern erzielten bessere Werte. Eltern mit kaukasischem Hintergrund beurteilten die Lebensqualität ihrer Kinder als höher als andere Eltern. Asiatische Kinder beurteilten ihre Lebensqualität als höher als Kinder mit hispanischem Hintergrund. Kinder mit asiatischen Sprachen (koreanisch, chinesisches) erzielten bessere Werte als englisch- und spanischsprachige Kinder. Proxys von 2-4 jährigen Kindern beurteilten deren Lebensqualität signifikant höher als Proxys von älteren Kindern. Jüngere Kinder beurteilten ihre Lebensqualität als tiefer als ältere Kinder. Das Nichterhalten von benötigten Pflegedienstleistungen beeinflusst die Lebensqualität negativ.

Praktikabilität: Es wurden 1.8% Fehler in den Selbstversionen und 3.4% Fehler in den Proxyversionen aufgezeigt.

Zusammengefasst werden die gute Reliabilität, die gute Praktikabilität und die akzeptable Bekannte-Gruppen-Validität dargestellt. In Betrachtung der Bekannte-Gruppen-Validität wurden kulturbedingte Unterschiede entdeckt. Es muss jedoch davon ausgegangen werden, dass in Europa eine andere Bevölkerungszusammensetzung besteht, wodurch die Übertragbarkeit der Ergebnisse in den europäischen Kontext nicht vollständig gewährleistet wird.

Wurde die klinische Bedeutung angegeben? Welches war die klinische Bedeutung der Ergebnisse? Waren die Unterschiede zwischen Gruppen klinisch von Bedeutung?

Ja. Die Autorenschaft beschreibt die Werte und zieht angemessene Schlussfolgerungen, wobei sie den Bezug zur Praxis im Detail darstellt. Gut beschrieben wurde die klinische Bedeutung bezüglich der Konkordanz zwischen Proxy- und Selbstversionen. Allerdings fehlt die Dokumentation zur internen Konsistenz.

Wurden Fälle von Ausscheiden aus der Studie angegeben? Schieden Teilnehmende aus der Studie aus? Warum? Wurden Gründe angegeben und Fälle von

Ja. Einerseits wurden nicht alle Fragebogen zurückgesendet, nur 51% der angefragten Familien retournierten die Umfrage. Andererseits wurden Fragebogen mit Fehlern (1.8% Fehler in den Selbstversionen und 3.4% Fehler in den Proxyversionen) zurückgesendet. Nach Ansicht der Verfasserin wurden diese TN ebenfalls ausgeschlossen, da

Ausscheiden angemessen gehandhabt? diese Items nicht zur Analyse weiterverwendet werden konnten. Diesbezüglich wurden jedoch keine Angaben gemacht.

Schlussfolgerungen und klinische Implikation

Waren die Schlussfolgerungen angemessen im Hinblick auf Methoden und Ergebnisse der Studie? **Ja.** Die wichtigsten Ergebnisse wurden erneut aufgegriffen und mit Literatur ergänzt diskutiert. Die Ergebnisse wurden indes gemäss den genannten Leitlinien korrekt interpretiert.

Welches waren die hauptsächlichsten Begrenzungen oder systematischen Fehler der Studie? Es nahmen nur 51% der Familien teil. Gemäss Varni et al. (2003) könnte die Teilnahmequote durch mehr Klientenkontakt vergrössert werden.
Zur Datenerhebung wurden nicht validierte Assessmentversionen (u.a. vietnamesische Sprachversion) verwendet, welche jedoch durch eine Übersetzungssoftware übersetzt wurden. Die fehlende Kontrolle zum getrennten Ausfüllen der Selbst- und Proxyversion muss ebenfalls als möglicher systematischer Fehler betrachtet werden.

Zu welchem Schluss kam die Studie? Welche Implikationen haben die Ergebnisse für die ergotherapeutische Praxis? Das PedsQL™4.0 (Varni, 1998) ist für Kinder valide, reliabel und praktikabel. Des Weiteren ist die Sichtweise der Kinder zur Erfassung der Lebensqualität von Kindern unbedingt zu berücksichtigen, da zwischen Proxy- und Selbstversionen Unterschiede existieren (vgl. Konkordanz). Ergänzend fügen Varni et al. (2003) an, dass die Proxyversion sich gut zur Erfassung der kindlichen Lebensqualität eignet, wenn ein Kind die Selbstversion nicht ausfüllen kann oder will. Zuletzt eignet sich das PedsQL™4.0 für grossflächige gesundheitsökonomische Evaluationen.

Schlussfolgerung der Verfasserin: Die Autorenschaft zeigte die psychometrischen Eigenschaften des PedsQL™4.0 (Varni, 1998) gut auf. Mittels guten Cronbach's Alpha-Koeffizienten wurde die Erfassung des Konstrukts durch die Items bestätigt. Es hätte allerdings noch die Test-Retest-Reliabilität geprüft werden müssen, womit die Wiederholbarkeit des PedsQL™4.0 überprüft worden wäre.

Des Weiteren kann eine Kontaminierung der Daten nicht ausgeschlossen werden. Es ist vorstellbar, dass Eltern ihren Kindern halfen auch wenn das für Kinder über sieben Jahren nicht erlaubt gewesen wäre. Des Weiteren könnten sich die Eltern von Fünf- bis Siebenjährigen und deren Kinder abgesprochen haben und somit gleiche oder ähnliche Antworten ausgewählt haben. Darum ist das Beantworten einer Umfrage dieser Wichtigkeit zu Hause aus Sicht der Verfasserin nicht optimal.

Varni et al. (2006)

Es folgt die kritische Würdigung der Studie von Varni et al. (2006).

Formular zur kritischen Besprechung	Dieses Dokument wurde anhand der Vorlage nach Law et al. (1998) durch die Verfasserin generiert.
Titel	The PedsQL™4.0 as a school population health measure: Feasibility, reliability, and validity
Publikation	Varni et al. (2006). <i>Quality of Life Research</i> , 15, 203–215.
Zweck der Studie	
Wurde der Zweck klar angegeben?	Ja
Skizzieren Sie den Zweck der Studie. Inwiefern bezieht sich die Studie und/oder auf Ihre Forschungsfrage?	Die vorliegende Studie erfasst die Reliabilität, Validität und Praktikabilität des PedsQL™4.0 (Varni, 1998) als ein schulbasiertes Evaluationsmessinstrument zur Erfassung der Lebensqualität bei acht- bis 18-jährigen Kindern. Weiter gibt die Autorenschaft an, dass für die grossflächige Identifikation von gesundheitspolitischen und präventiven Massnahmen die Verfügbarkeit eines geeigneten Assessments von zentraler Bedeutung ist.
Literatur	
Wurde relevante Hintergrundliteratur gesichtet?	Ja
Geben Sie an, wie die Notwendigkeit der Studie gerechtfertigt wurde.	Es existieren verschiedene Assessments zur Erfassung der kindlichen Lebensqualität. Die Autorenschaft nennt dafür Quellen, wobei sie unter anderem die Studie von Varni et al. (2003) referenziert. Zusätzlich wurde bisher keine Studie publiziert, welche die Eignung des PedsQL™4.0 als schulbasiertes Erfassungsinstrument zur Evaluation der kindlichen Lebensqualität aufzeigt.
Design	
Welches Design wurde verwendet?	Deskriptive Statistik und prospektives Design
Beschreiben Sie das Studiendesign. Entsprech das Design der Studienfrage (z.B. im Hinblick auf den Wissensstand zur betreffenden Frage, auf Ergebnisse, auf ethische Aspekte)?	Mittels Fragebogen wurden die Assessments gelöst, welche anschliessend auf deren Reliabilität, Validität und Praktikabilität überprüft wurden. Somit wurde Data erfasst, was zu einem prospektiven Studiendesign passt. Die anschliessende statistische Auswertung legt das Studiendesign ‚deskriptive Statistik‘ nahe.

Spezifizieren Sie alle systematischen Fehler (Verzerrungen, Bias), die vielleicht aufgetreten sein könnten, und in welche Richtung sie die Ergebnisse beeinflussen.

An der Umfrage nahmen mehr Schulklassen der Grundstufe (bis Grad 6) teil. Ein direkter Vergleich der Untergruppen wird dadurch erschwert.

Die Verfasserin möchte betonen, dass eine vorbildliche Trennung zwischen Selbst- und Proxyversion stattgefunden hat (siehe unten, Massnahmen).

Stichprobe

N=6'663 Es wurden 2'415 Proxy-Kinder-Paare gebildet.

Wurde die Stichprobe detailliert beschrieben? **Ja.** Es wurden 2'436 Selbstversionen von Kindern (8-18 Jahre) und 4'227 Proxyversionen für Kinder (5-18 Jahre) beantwortet. Alle Teilnehmenden gingen im Schuldistrikt San Diego (California, US) zur Schule. Erfasst wurde Geschlecht, Alter, Nationalität/Herkunft, die gesprochene Sprache und ob eine chronische Erkrankung der Kinder vorliegt. Es nahmen am meisten Familien mit hispanischem Hintergrund teil. Zusätzlich wurden die teilnehmenden Schulen und somit die Kinder, in drei sozioökonomische Status eingeteilt: tiefer, mitteltiefer und mittlerer sozioökonomischer Status. Des Weiteren wurde der Bildungsstand der Mutter erfasst, um den Einfluss des mütterlichen Bildungsstands auf die kindliche Lebensqualität zu überprüfen.

Wurde die Stichprobengrösse begründet? **Nein**, aber nach Ansicht der Verfasserin bewegt sich die Stichprobengrösse im Rahmen der anderen, für diese Arbeit analysierten Studien, und kann somit als gerechtfertigt betrachtet werden.

Ergebnisse (Outcomes)

Outcome Bereiche	Listen Sie die verwendeten Messungen auf.
Geben Sie an, wie oft outcome Messungen durchgeführt wurden (also vorher, nachher, bei Nachbeobachtungen...).	<p>Es wurden keine Outcome-Messungen von Interventionen erhoben. Die Datenerfassung erfolgte einmalig.</p> <p>Interne Konsistenz: Cronbach's Alpha-Koeffizient (Cronbach, 1951).</p> <p>Bekannte-Gruppen-Validität: t-Test und Effektgrößen (Cohen, 1988), Gruppenvergleiche mittels ANOVA.</p> <p>Konvergenzvalidität: Überprüfung auf mögliche Korrelationen zwischen den vier Funktionsskalen des PedsQL™4.0 (Varni, 1998) und des Stanford 9 (Harcourt Educational Measurement, 1997) mittels Intra-Klassen-Korrelationen und Pearson-Produkt-Moment-Korrelationen.</p> <p>Praktikabilität: Prozentuale Fehler pro Item (Fehler= unbeantwortetes Item)</p>
Waren die Messungen gültig und zuverlässig?	gültig: Ja , die Validität wurde als Studienziel überprüft zuverlässig: Ja , die Reliabilität wurde als Studienziel überprüft

Massnahmen

Wurden die Massnahmen detailliert beschrieben?	<p>Ja. Forschungsassistenten verteilten den Schülerinnen und Schülern die Proxyversion für ihre Eltern. Die Eltern beantworteten die Items und retournierten die Umfrage an die Schule. Nachdem alle Proxyversionen retourniert wurden, wurde die Selbstversion während des Unterrichts ebenfalls mit anwesendem Forschungspersonal administriert.</p> <p>In jedem Fall sollte die Proxyversion retourniert werden, wurde sie leer zurückgegeben, verweigerten die Eltern die Teilnahme an der Studie.</p> <p>Mit Eltern, welche auffallend viele Items unbeantwortet retournierten, wurden Telefoninterviews geführt. Ebenfalls wurden 50 Elternpaare mit verminderter Bildung an Versammlungen bei der Beantwortung des Fragebogens unterstützt.</p>
Wurde Kontaminierung vermieden?	Ja. Die Proxy- und Selbstversionen wurden zeitversetzt und an unterschiedlichen Orten administriert.
Wurden Ko-Interventionen vermieden?	Entfällt. Es wurden keine Interventionen geprüft, wodurch keine Untergruppe andere Interventionen erhielt.

Ergebnisse

Wurde die statistische Signifikanz der Ergebnisse angegeben?	Ja. Mehrheitlich wurden die p -Werte dokumentiert, wobei nicht alle Messungen eine Berechnung des p -Werts verlangen.
Waren die Analysemethoden geeignet?	Ja. Die Autorenschaft kommentierte die gewählte Analysemethoden nicht, welche aber nach Einschätzung der Verfasserin geeignet waren, da die gewählten Methoden mit anderen ähnlichen Studien deckungsgleich sind.
Welches waren die Ergebnisse? Waren sie statistisch signifikant? War die Studie gross genug, um einen eventuell auftretenden wichtigen Unterschied anzuzeigen? Falls es um viele Ergebnisse ging: Wurde dies bei der statistischen Analyse berücksichtigt?	<p>Interne Konsistenz: Es wurden gute bis sehr gute Cronbach's-Alpha-Koeffizienten (Cronbach, 1951) berechnet. Für die schulische Funktionsskala des PedsQL™4.0 (Varni, 1998) wurden die jeweils tiefsten Werte berechnet.</p> <p>Konkordanz: Es wurden tiefe Korrelationen zwischen den Proxy- und den Selbstversionen berechnet. Dabei korrelierten die Proxy- und Selbstversionen für und von 8-12jährigen tiefer als für 13-18jährige. Stimmt die Versionen von und für 8- bis 12jährige weniger gut überein als die Methoden von und für 13-18jährige. Mittels Intra-Klassenkorrelationen wurden die tiefen Werte bestätigt.</p> <p>Bekannte-Gruppen-Validität: Gesundere Kinder erzielten signifikant bessere Werte als weniger gesundere Kinder (ANCOVA; $F=44.29$ Selbstversion; $F=62.92$ Proxyversion; $p<.001$). Schülerinnen und Schüler von einer Schule mit tiefem SES erzielten schlechtere Werte als Schülerinnen und Schüler einer Schule mit mitteltiefem oder mittlerem SES. / Keine Geschlechterunterschiede / jüngere Kinder erzielten schlechtere Werte als ältere TN.</p>

	<p>Konvergenzvalidität: Tiefe konvergente Validität zwischen der ‚schulischen Funktionsskala‘ des PedsQL™4.0 (Varni, 1998) und dem ‚Stanford 9‘ (Harcourt Educational Measurement, 1997) / gute divergente Validität zwischen dem Stanford 9 und den anderen 3 Dimensionen des PedsQL™4.0.</p> <p>Praktikabilität: Für die Selbstversion wurden <0.001%, für die Proxyversion wurden 0.3% Fehler festgestellt.</p>
Wurde die klinische Bedeutung angegeben?	Ja. Die klinische Bedeutung wurde für notwendige Werte dokumentiert.
Welches war die klinische Bedeutung der Ergebnisse? Waren die Unterschiede zwischen Gruppen klinisch von Bedeutung?	Analog den aufgezeigten Ergebnissen werden die Konstrukte anhand der Items zuverlässig gemessen. Des Weiteren müssen Kinder ihre Lebensqualität – wenn möglich – selber einschätzen, da tiefe Übereinstimmungen berechnet wurden. Ebenfalls erzielen gesellschaftlich schwächere und/oder weniger gesunde TN schlechtere Werte (= tiefere Lebensqualität). Zuletzt wurden bestätigt, dass mehr Klientenkontakt die Fehlerreduktion begünstigt (vgl. Varni et al., 2003).
Wurden Fälle von Ausscheiden aus der Studie angegeben? Schieden Teilnehmende aus der Studie aus? Warum? Wurden Gründe angegeben und Fälle von Ausscheiden angemessen gehandhabt?	Ja. Die Umfrage wurde von 59% der TN retourniert, somit schieden 41% aus der Studie aus. Jedoch enthielten die retournierten Fragebogen wenige Fehler. Da unbeantwortete Items nicht zur Analyse verwendet werden können, werden weitere Ausschlüsse von der Verfasserin vermutet.
Schlussfolgerungen und klinische Implikation	
Waren die Schlussfolgerungen angemessen im Hinblick auf Methoden und Ergebnisse der Studie?	Ja. Die Ergebnisse wurden analog den vorgegebenen Leitlinien korrekt interpretiert. Es wurden korrekte Schlussfolgerungen gezogen. Als Negativpunkt gilt zu betrachten, dass die Praktikabilität zusätzlich anhand der Verteilung der Item-Antworten gemessen werden sollte. Die Autorenschaft kommentiert diese Ergebnisse aber nicht.
Welches waren die hauptsächlichsten Begrenzungen oder systematischen Fehler der Studie? Zu welchem Schluss kam die Studie? Welche Implikationen haben die Ergebnisse für die ergotherapeutische Praxis?	Das PedsQL™4.0 (Varni, 1998) ist ein geeignetes Messinstrument zur Erfassung der (schulischen) Lebensqualität von Kindern (8-18 Jahre) in der Schule. Limitationen: Es ist unbekannt, wie viele Schulen die Teilnahme verweigerten. Die Eltern beurteilten den Gesundheitszustand ihrer Kinder und gaben keine Auskunft über die Art oder Schwere der Krankheit. Die Test-Retest-Reliabilität wurde nicht verifiziert.

Die Lebensqualität von Kindern kann mittels PedsQL™4.0 (Varni, 1998) zuverlässig, gültig und praktikabel gemessen werden. Und das PedsQL™4.0 (Varni, 1998) eignet sich zur Identifikation von Beeinträchtigungen der Lebensqualität.

Schlussfolgerung der Verfasserin: Die Autorenschaft zeigte die gute Reliabilität des Assessments auf. Dabei werden die Konstrukte anhand der Items gut erfasst. Die ‚schulische Funktionsskala‘ des PedsQL™4.0 (Varni, 1998) wurde indes speziell überprüft, indem sie mit dem Stanford 9 (Harcourt Educational Measurement, 1997) korreliert wurde. Dabei wurde festgestellt, dass diese beiden Instrumente sehr unterschiedliche Aspekte erfassen. Fraglich bleibt, ob die ‚schulische Funktionsskala‘ dadurch ungenau misst. Dies würde die tieferen Reliabilitätswerte erklären. Bemerkenswert ist die vorbildliche Vermeidung von ungewollten Korrelationen der Selbst- und Proxyversion durch die zeitliche Trennung der Durchführung. Die so gemessenen tiefen Korrelationswerte legen nahe, dass die dokumentierten Korrelationswerte anderer und methodisch schwächeren Studien fragwürdig hoch sind. Des Weiteren hätte die Test-Retest-Reliabilität überprüft werden sollen, um die Wiederholbarkeit des PedsQL™4.0 (Varni, 1998) zu überprüfen.

Robitail et al. (2006)

Es folgt die kritische Würdigung der Studie von Robitail et al. (2006).

Formular zur kritischen Besprechung	Dieses Dokument wurde anhand der Vorlage nach Law et al. (1998) durch die Verfasserin generiert.
Titel	Validation of the European Proxy KIDSCREEN-52 Pilot Test Health-Related Quality of Life Questionnaire: First Results.
Publikation	Robitail et al. (2006). <i>Journal of Adolescent Health</i> , 39, 596.e1 – 596.e10.
Zweck der Studie	
Wurde der Zweck klar angegeben?	Ja
Skizzieren Sie den Zweck der Studie. Inwiefern bezieht sich die Studie und/oder auf Ihre Forschungsfrage?	Die Autoren überprüften die Validität und Reliabilität der Proxyversion des KIDSCREEN-52 (Ravens-Sieberer et al., 2005) und legen somit die Grundlage für weitere Untersuchungen zwischen Jugend- und Proxyberichten.
Literatur	
Wurde relevante Hintergrundliteratur gesichtet?	Ja

Geben Sie an, wie die Notwendigkeit der Studie gerechtfertigt wurde.

Die Autorenschaft vermerkt, dass ein deutliches Bedürfnis an Erfassungsinstrumenten zur Erfassung der Lebensqualität von Kindern besteht. Zudem bestehe ein wachsendes Interesse am Aufzeigen der gesellschaftlichen Gesundheit. Ebenso stellt die vorliegende Studie die erste detaillierte Untersuchung von KIDSCREEN-52 (Ravens-Sieberer et al., 2005) dar.

Design

Welches Design wurde verwendet?

deskriptive Statistik und prospektives Design.

Beschreiben Sie das Studiendesign. Entsprechend dem Design der Studienfrage (z.B. im Hinblick auf den Wissensstand zur betreffenden Frage, auf Ergebnisse, auf ethische Aspekte)?

Die Studie wählte das korrekte Design. Es wurden empirische Daten erhoben, wodurch die Gütekriterien berechnet wurden.

Spezifizieren Sie alle systematischen Fehler (Verzerrungen, Bias), die vielleicht aufgetreten sein könnten, und in welche Richtung sie die Ergebnisse beeinflussen.

Aufgrund der unterschiedlichen Verfahren zur Datenerfassung kann die Objektivität nicht gewährleistet werden, wodurch Verzerrungen nicht ausgeschlossen werden können. Des Weiteren kann Kontaminierung der beiden Methoden nicht ausgeschlossen werden, da unklar ist, wie die Daten erhoben wurden.

Nach Ansicht der Verfasserin kam es zu keinen weiteren systematischen Fehlern, da „ausschliesslich“ die Gütekriterien der Erfassungsmethode analysiert wurden.

Alle TN beantworteten dieselbe Umfrage.

Stichprobe

N = 2'505

Es wurde die Daten von 2'505 Proxy-Kinder-Paare erhoben, dabei wurden nur Daten der Selbstversion gezeigte, welche relevant für die Daten der Proxyversion waren (u.a. Konkordanz).

Wurde die Stichprobe detailliert beschrieben?

Ja. 2'505 Proxy-Kinder/Jugendliche-Paare. Das Durchschnittsalter der Kinder/Jugendliche* betrug 12.4 ± 2.5 Jahre. Es nahmen 48.1% Knaben teil und 40.8% waren im Kindesalter (8 – 11 Jahre). Die Proxys waren leibliche Eltern (Mütter + Väter, 84.1% + 7.1%), Stiefeltern (Stiefmütter + Stiefväter, 5.3% + 0.5%) oder jemand anderes (2.9%). 680 Familien gehörten dem tiefen SES an, 1'221 Familien dem mittleren und 479 Familien dem höheren. 278 Kinder waren von einer chronischen Krankheit betroffen.

Jedoch variiert die Angabe in Tabelle mit der Angabe im Fliesstext (Differenz 24 Paare).

Wurde die Stichprobengrösse begründet?

Nein, aber die Stichprobengrösse liegt der Verfasserin nach im Rahmen vergleichbarer Studien.

Ergebnisse (Outcomes)

Geben Sie an, wie oft outcome Messungen durchgeführt wurden (also vorher, nachher, bei Nachbeobachtungen...).	Outcome Bereiche Es wurden die Assessmentdimensionen auf Validität und Reliabilität überprüft.	Listen Sie die verwendeten Messungen auf. Konstruktvalidität: Analysiert wurde die Item-interne Konsistenz und die item-diskriminante Validität zur Erfassung der Eindimensionalität jeder Dimension. Mittels Faktorenanalyse wurde die mehrdimensionale Assessmentstruktur aufgezeigt. Zudem wurde die Anpassungsgüte mit verschiedenen Kennwerten berechnet (χ^2 -Test, RMSEA, NNFI, CFI, WLS). Konstrukt jeder Dimension wurde mit einer Rasch-Analyse (PCM und RSM) überprüft. Interne Konsistenz: Cronbach's Alpha Koeffizient (Cronbach, 1951). Externe Validität: Merkmalsvalidität wurde mit der Multitrait-Methode analysiert. Dabei wurden Korrelationen zwischen den zehn Assessmentdimensionen den beiden Erfassungsmethoden (Selbst- und Proxyversion) berechnet. Konvergente und divergente Validität wurde überprüft, indem die HRQoL-Domänen (HRQoL = health related quality of life) und die zwei Erfassungsmethoden korreliert wurden. Die Konkordanz wurde mittels Intra-Klassen-Korrelationen berechnet.
---------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Waren die Messungen gültig und zuverlässig?	gültig: Ja (die Validität wurde überprüft)	zuverlässig: Ja (die Reliabilität wurde überprüft)
---------------------------------------------	---------------------------------------------------	-----------------------------------------------------------

Massnahmen

Wurden die Massnahmen detailliert beschrieben?	Nein. Die Erfassung der psychometrischen Eigenschaften wird detailliert und begründet beschrieben. Als Einschränkung bleibt zu bemerken, dass die Durchführung in den einzelnen Ländern unklar beschrieben wurde. Dadurch wird die Studie für die Leserschaft nicht reproduzierbar.
Wurde Kontaminierung vermieden?	Nicht angegeben. Die Datenerfassung wurde ungenau beschrieben, wodurch Kontaminierung nicht ausgeschlossen werden kann. Beispielsweise beantworteten die Proxys im schwedischen Sample die Fragebogen gemeinsam mit den Kindern, wodurch die Daten verunreinigt wurde.
Wurden Ko-Interventionen vermieden?	Entfällt. Alle TN beantworteten denselben Fragebogen.

Ergebnisse

Wurde die statistische Signifikanz der Ergebnisse angegeben?	Ja. Die statistische Signifikanz wurde für die Berechnungen der Validität angegeben. Sie fehlen jedoch für die Angaben der Reliabilität. Die Berechnung der <i>p</i> -Werte ist allerdings nicht für jede Messung erforderlich.
--------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Waren die Analysemethoden geeignet?	Ja. Die verwendeten statistischen Analysemethoden waren effektiv und deshalb nach Ansicht der Verfasserin geeignet. Die Autoren dokumentierten jedoch die Eignung dieser Methoden nicht explizit. Ebenfalls wird die Praktikabilität nicht gesondert untersucht (siehe unten, Schlussfolgerungen).
Welches waren die Ergebnisse? Waren sie statistisch signifikant? War die Studie gross genug, um einen eventuell auftretenden wichtigen Unterschied anzuzeigen? Falls es m viele Ergebnisse ging: Wurde dies bei der statistischen Analyse berücksichtigt?	Interne Konsistenz: Die Proxyversion zeigt adäquate bis sehr gute Alpha-Koeffizienten ($\alpha = .76 - .90$). Es werden aber keine Daten gezeigt, welche die Selbstversion betreffen. Konstruktvalidität: Es wurden rare Boden- und Deckeneffekte beobachtet (ausser 33.6% bei der Dimension soziale Akzeptanz). Item-Interne Konsistenz: mindestens .45 (körperliche Dimension). Externe Validität: Die Merkmalsvalidität wurde mit dem MTMM berechnet, welche die Korrelationen zwischen den 10 Dimensionen und den zwei Methoden (SV und PV) messen. Dabei wurde konvergente und divergente Validität ($r = .12 - .66$; $p < .001$) bestätigt. Konkordanz zwischen Proxy- und Selbstversionen: Es wurden zufriedenstellende Intra-Klassen-Korrelationskoeffizienten berechnet ($> .40$). Diesbezüglich wird der Leserschaft detaillierte Data vorenthalten. Gemäss Autorenschaft sind die Übereinstimmungen für Kinder-Proxy-Paare tiefer als für Jugendliche-Proxy-Paare.
Wurde die klinische Bedeutung angegeben? Welches war die klinische Bedeutung der Ergebnisse?	Ja. Für die statistischen Messungen wurde die Relevanz mehrheitlich nicht aufgezeigt. Als Ausnahme betrachtet werden soll die Überprüfung der Einzigartigkeit jeder Dimension, sowie die multidimensionale Assessmentstruktur. Des Weiteren wurden die Werte der Reliabilität und Validität erläutert. Es wurde angegeben, dass anhand der Proxyversion die Lebensqualität von Kindern gut fremdbeurteilt werden kann. Des Weiteren kann dieses Assessment in drei Ausprägungen verwendet werden: Die Proxys können die Lebensqualität ihrer Kinder beurteilen falls...: a) ...wenn die Kinder ihre Lebensqualität selber nicht einschätzen können. b) ... um weitere Informationen zu sammeln, welche zur ergotherapeutischen Interventionsplanung verwendet werden kann. c) ... um relevante Informationen für gesundheitsökonomische Evaluationen zu erfassen.
Waren die Unterschiede zwischen Gruppen klinisch von Bedeutung?	Ja. Es wurde die Proxyversion des KIDSCREEN-52 validiert. Dabei wurde ausschliesslich überprüft, ob die Übereinstimmung der Antworten altersabhängig war. Festgestellt wurde, dass die Übereinstimmungen zwischen Proxy-Jugendliche höher waren (s.o. „Konkordanz zwischen Proxy- und Selbstversionen“).

Wurden Fälle von Ausscheiden aus der Studie angegeben? Schieden Teilnehmende aus der Studie aus? Warum? Wurden Gründe angegeben und Fälle von Ausscheiden angemessen gehandhabt?	Ja. Einerseits wurden nicht alle verteilten Fragebogen retourniert, andererseits mussten 21 Proxy-Kinder-Paare aufgrund nichtbeantworteten Items ausgeschlossen werden.
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Schlussfolgerungen und klinische Implikation

Waren die Schlussfolgerungen angemessen im Hinblick auf Methoden und Ergebnisse der Studie?	Ja. Es wurden die wichtigen Ergebnisse erneut aufgegriffen und diskutiert. Des Weiteren wurde die Bedeutung der Ergebnisse für die Praxis erläutert. Es wurden keine Limitationen dieser Arbeit erwähnt, wodurch die Güte dieser Publikation geschmälert wird. Die Praktikabilität des KIDSCREEN-52 (Ravens-Sieberer et al., 2005) wurde nicht explizit gemessen. Da jedoch dieses Assessment mit verschiedenen Verfahren (Delphi-Methode, Fokusgruppen-Interviews, Expertengespräche) entwickelt wurde, wird davon ausgegangen, dass die Praktikabilität bereits untersucht wurde.
---------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Zu welchem Schluss kam die Studie? Welche Implikationen haben die Ergebnisse für die ergotherapeutische Praxis? Welches waren die hauptsächlichsten Begrenzungen oder systematischen Fehler der Studie?	Die Proxyversion des KIDSCREEN-52 ist ein valides Assessment, um die Lebensqualität von Kindern zu erfassen. Dabei kann diese Version in drei Hauptbereichen verwendet werden: 1. Anhand dieser Version können mittels Selbstversion erhobene Informationen ergänzt werden, was zur Interventionsplanung dient. 2. Für Kinder, welchen die Durchführung der Selbstversion nicht möglich ist, kann die Lebensqualität anhand dieser Proxyversion fremdeingeschätzt werden. 3. Diese Proxyversion kann für Planungen und Begründungen der Interventionen gegenüber dem öffentlichen Gesundheitswesen verwendet werden.
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Schlussfolgerung der Verfasserin: Die Autorschaft zeigte gute Werte der Proxyversion des KIDSCREEN-52 (Ravens-Sieberer et al., 2005) auf. Sie geben auch mögliche Begründungen für die Ergebnisse an. Des Weiteren leiten sie einen nachvollziehbaren Zusammenhang zur Praxis her. Dabei erklärt die Autorschaft, was anhand dieser Assessmentmethode ermöglicht wird (siehe Schlussfolgerung). Gelungen sind ebenfalls die Tabellen, welche sparsam aber adäquat verwendet wurden

Der nicht konsistente Gebrauch von Fachbegriffen macht die Studie jedoch sehr schwer lesbar. Ebenfalls werden Data vorenthalten (vgl. Resultate der internen Konsistenz), wodurch die Studie nicht reproduzierbar wird.

Ravens-Sieberer et al. (2008)

Es folgt die kritische Würdigung der Studie von Ravens-Sieberer et al. (2008).

Formular zur kritischen Besprechung	Dieses Dokument wurde anhand der Vorlage nach Law et al. (1998) durch die Verfasserin generiert.
Titel	The KIDSCREEN-52 Quality of Life Measure for Children and Adolescents: Psychometric Results from a Cross-Cultural Survey in 13 European Countries
Publikation	Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Power, M., Duer, W., ... & the KIDSCREEN Group. (2008). <i>Value in Health</i> , 11, 4, 645–658.
Zweck der Studie	
Wurde der Zweck klar angegeben?	Ja
Skizzieren Sie den Zweck der Studie. Inwiefern bezieht sich die Studie und/oder auf Ihre Forschungsfrage?	Bei der vorliegenden Studie handelt es sich um die erste multinationale Studie des KIDSCREEN-52 (Ravens-Sieberer et al., 2005). Anhand dieser Studie wurde die Anwendung des KIDSCREEN-52 europaweit überprüft. Zudem werden die Reliabilität und Validität aufgezeigt.
Literatur	
Wurde relevante Hintergrundliteratur gesichtet?	Ja.
Geben Sie an, wie die Notwendigkeit der Studie gerechtfertigt wurde.	Es besteht Bedarf an der Entwicklung von Assessments, womit Kinder ihre Lebensqualität selber einschätzen können. Während die Lebensqualitätsforschung für Erwachsene stetig wuchs, wurde die Lebensqualitätsforschung für Kinder fälschlicherweise ignoriert. Mittels dieser Studie wird die Anwendung des KIDSCREEN-52 (Ravens-Sieberer et al., 2005) erstmals in 13 europäischen Ländern überprüft.
Design	
Welches Design wurde verwendet?	Prospektives Design und deskriptive Statistik
Beschreiben Sie das Studiendesign. Entsprechend das Design der Studienfrage (z.B. im Hinblick auf den Wissensstand zur betreffenden Frage, auf Ergebnisse, auf ethische Aspekte)?	Die gesammelten Daten werden genau auf die Forschungszielsetzung zugeschnitten. Es wird erfasst, ob das Assessment reliabel und valide ist. Somit werden Daten erfasst, die diese Zielsetzung bestätigen oder widerlegen. Das Design wurde richtig gewählt. Es wurde jedoch nicht explizit erwähnt, dass es sich um eine „prospektive Studie“ handelt, was aber der fachkundigen Leserschaft klar sein sollte.

Spezifizieren Sie alle systematischen Fehler (Verzerrungen, Bias), die vielleicht aufgetreten sein könnten, und in welche Richtung sie die Ergebnisse beeinflussen.

Die Datenerfassung in den Ländern Österreich, Schweiz, Spanien, Frankreich und Holland wurde zu Hause durchgeführt. Dabei könnten die Eltern die Umfrage mit den Kindern gemeinsam ausgefüllt haben, wodurch die Eltern (bewusst oder unbewusst) die Antworten der Kinder beeinflusst haben könnten und somit die Kinder ihre Lebensqualität verfälscht einschätzten.

Stichprobe

N=22'827

Wurde die Stichprobe detailliert beschrieben? **Ja.** Die TN wurden aus 13 europäischen Ländern (Österreich, Schweiz, Spanien, Frankreich, Deutschland und Holland, Griechenland, Ungarn, Irland, Polen und Schweden, Tschechien und England) rekrutiert und waren zwischen 8 und 18 Jahre alt. Die Autorenschaft beschrieb die Stichprobe mittels Tabelle (Tabelle 1, S. 650) detailliert und übersichtlich. Des Weiteren wurde der sozioökonomische Status erfasst.

Wurde die Stichprobengrösse begründet? **Nein.** Die Begründung der Samplegrösse fehlt. Ein grosses Sample wird jedoch benötigt, um die Vielzahl untersuchter Gruppen statistisch signifikant überprüfen zu können.

Ergebnisse (Outcomes)

Geben Sie an, wie oft outcome Messungen durchgeführt wurden (also vorher, nachher, bei Nachbeobachtungen...).

Outcome Bereiche (z.B. Selbstversorgung, Produktivität, Freizeit)

Listen Sie die verwendeten Messungen auf.

Es handelt sich um die Validierungsstudie des KIDSCREEN-52 (Ravensieberer et al., 2005), weshalb keine Outcome-Messungen durchgeführt wurden.

Test-Retest-Reliabilität: Intra-Klassen-Korrelationen (Bartko, 1966).

An 90% aller TN wurden einmalig Daten erhoben, an 10% der TN wurde die Test-Retest-Reliabilität überprüft, wodurch die Datenerfassung wiederholt wurde.

Interne Konsistenz: Cronbach's Alpha Koeffizient (Cronbach, 1951).

Konvergenzvalidität: Pearson-Korrelationskoeffizient (Cohen, 1988) zwischen PedsQL™4.0 und YQOL-S (Wahrnehmungsskala; Edwards, 2002) oder CHIP-AE (Zufriedenheitsdomäne; Starfield et al., 1994).

Bekannte-Gruppen-Validität (Gruppeneinteilung mit CSHCN (Bethell et al., 2002, und SDQ; Goodman, 1997): Kinder mit mentaler Beeinträchtigung wurden mit Kindern ohne mentale Beeinträchtigung verglichen. Familien mit tiefem SES erzielten schlechtere Werte (tiefere Lebensqualität).

Waren die Messungen gültig und zuverlässig? gültig: **Ja** (die Validität wurde überprüft) zuverlässig: **Ja** (die Reliabilität wurde überprüft)

Massnahmen

Wurden die Massnahmen detailliert beschrieben?	<p>Ja. Die durchgeführten statistischen Messungen und die Datenerfassung wurden detailliert beschrieben. Als Kritikpunkt gilt die Datenerhebung in den Ländern Tschechien und England zu erachten, welche gemäss der Verfasserin zu ungenau beschrieben wurden.</p> <p>In Österreich, Schweiz, Spanien, Frankreich, Deutschland und Holland wurden die Umfragebogen nach telefonischer Erstkontaktierung per Post versendet und zu Hause selbstinitiativ administriert, bevor sie wieder zurückgesendet wurden.</p> <p>In Griechenland, Ungarn, Irland, Polen und Schweden wurden die Samples in der Schule erfasst und die Proxyversion den Kindern verteilt, damit diese die Proxyversion ihren Eltern übergeben konnten. Diese retournierten die Version ans Datacenter.</p> <p>In England wurden die beiden Verfahren kombiniert und in Tschechien wurden die Daten in einem randomisierten -Haushalts-Sample erhoben.</p>
Wurde Kontaminierung vermieden?	<p>Nicht angegeben. Die Autorenschaft vermerkte keine Kontaminierung. Nach Ansicht der Verfasserin hätte jedoch unter anderem im tschechischen Sample die Datenerhebung verunreinigt werden können. Hier wurde die Umfrage zu Hause beantwortet, wobei Proxys und Kinder die Fragebogen in gemeinsamer Absprache gelöst haben könnten.</p>
Wurden Ko-Interventionen vermieden?	<p>Entfällt. Alle TN beantworteten dieselbe Umfrage, weshalb dieser Punkt nach Ansicht der Verfasserin irrelevant ist.</p>

Ergebnisse

Wurde die statistische Signifikanz der Ergebnisse angegeben?	<p>Ja. Die Autorenschaft dokumentiert die statistische Signifikanz der Validitätsmessungen. Jedoch wird die statistische Signifikanz der Reliabilitätsmessungen nicht vermerkt. Dabei wird die Berechnung der p-Werte nicht für alle Messungen benötigt.</p>
Waren die Analysemethoden geeignet?	<p>Ja. Die gewählten Analysemethoden waren zielführend und effektiv. Die Autorenschaft kommentierten die Eignung der Analysemethoden jedoch nicht explizit. Hingegen wurde die Praktikabilität nicht gesondert untersucht. Die grosse Stichprobenzahl in Kombination mit der tief anzunehmenden Fehlerquote lässt darauf schliessen, dass sich das Assessment in der Praxis als praktikabel erwiesen hat.</p>
Welches waren die Ergebnisse? Waren sie statistisch signifikant? War die Studie gross genug, um einen eventuell auftretenden wichtigen Unterschied anzuzeigen? Falls es um viele Ergebnisse ging: Wurde dies bei	<p>Test-Retest-Reliabilität: Mehrheitlich akzeptable Intra-Klassen-Korrelationskoeffizienten (Bartko, 1966) wurden berechnet, wobei für ausschliesslich 3 Dimensionen (‚soziale Akzeptanz‘ .57, ‚Autonomie‘ .56 und ‚Stimmung und Emotionen‘ .58) leicht ungenügende Intra-Klassen-Korrelationskoeffizienten berechnet wurden.</p> <p>Interne Konsistenz: Im Gesamtsample wurden gute Alpha-Koeffizienten (Cronbach, 1951) aufgezeigt (.77–.89).</p> <p>Konvergenzvalidität: Es wurden moderate Korrelationen zwischen dem KIDSCREEN-52 (Ravens-Sieberer et al., 2005) und dem PedsQL™4.0 (Varni, 1998) berechnet. Das KIDSCREEN-52 korreliert mit dem YQOL-S</p>

der statistischen Analyse berücksichtigt? (Edwards, 2002) oder CHIP-AE (Starfield et al., 1994) auf moderatem bis hohem Niveau. In allen Korrelationen wurde zudem unterstützende divergente Validität beobachtet.

Bekannte-Gruppen-Validität: Wurden die TN mittels CSHCN (Bethell et al., 2002) in zwei Gruppen eingeteilt, wurden tiefe bis moderate Effektgrößen (Cohen, 1988) berechnet. Fand die Einteilung mittels SDQ (Goodman, 1997) statt, wurden moderate bis hohe Effektgrößen (Cohen, 1988) festgestellt.

Es wurden für vier von 52 Items kleine DIF-Unterschiede festgestellt. Gemäss Gantschnig (persönliches Gespräch mit Gantschnig, B., am 12.5.2016) könnten diese Unterschiede als nichtig betrachtet werden. Es bestätigt jedoch erneut, dass Kinder – wenn möglich – zur Beurteilung der eigenen Lebensqualität selber gefragt werden müssen (vgl. Eiser & Morse, 2001; Varni et al., 2006).

Wurde die klinische Bedeutung angegeben?

Welches war die klinische Bedeutung der Ergebnisse? Waren die Unterschiede zwischen Gruppen klinisch von Bedeutung?

Ja. Die Autorenschaft interpretierte die Bedeutung der Werte. Dabei wurden die Ergebnisse analog der Beurteilungskriterien eingeordnet. Die Leserschaft erhält dadurch ein Verständnis, wie die Werte zu deuten sind. Ebenfalls wird erwähnt, dass sich das KIDSCREEN-52 (Ravens-Sieberer et al., 2005) aufgrund der erfassten Werte zur Erfassung von Beeinträchtigungen der kindlichen Lebensqualität eignet.

Wurden Fälle von Ausscheiden aus der Studie angegeben? Schieden Teilnehmende aus der Studie aus? Warum? Wurden Gründe angegeben und Fälle von Ausscheiden angemessen gehandhabt?

Ja. Es wurden nicht alle Fragebogen retourniert. Allerdings wurden während der Datenanalyse keine Fragebogen exkludiert, wodurch während der offiziellen Studie keine Ausschlüsse zu vermerken waren.

Schlussfolgerungen und klinische Implikation

Waren die Schlussfolgerungen angemessen im Hinblick auf Methoden und Ergebnisse der Studie?

Ja. Die Ergebnisse wurden korrekt nach den Leitlinien interpretiert, welche die Autorenschaft im Methodenteil beschrieben hatte. Dabei wurden korrekte und für die Leserschaft nachvollziehbare Schlussfolgerungen abgeleitet.

Zu welchem Schluss kam die Studie? Welche Implikationen haben die Ergebnisse für die ergotherapeutische Praxis? Welches waren die hauptsächlichsten Begrenzungen oder systematischen Fehler der Studie?	<p>Es ist weitere Forschung nötig, um die Ergebnisse zu bestätigen und die Sensitivität zu erfassen. Mittels weiterer Forschung könnten die ungenügenden Intra-Klassen-Korrelationskoeffizienten (Bartko, 1966) widerlegt und die Wiederholbarkeit des KIDSCREEN-52 (Ravens-Sieberer et al., 2005) unterstützt werden.</p> <p>Das KIDSCREEN-52, sowie die beiden kürzeren Ableitungen eignen sich zur Erfassung der Lebensqualität von Kindern in Europa. Mittels diesem Assessment können ebenfalls ergotherapeutische Interventionen auf deren Wirksamkeit überprüft werden.</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Schlussfolgerung der Verfasserin: Der theoretische Hintergrund ist sehr ausführlich und strukturiert aufgebaut. Die Autorenschaft beschrieb detailliert, wofür und wie die statistische Erfassung durchgeführt wurde. Des Weiteren wurde die Datenerfassung mehrheitlich detailliert beschrieben. Als Ausnahme gilt die Beschreibung der Datenerhebung in England und Tschechien. Nach Ansicht der Verfasserin wurde im Allgemeinen undeutlich beschrieben, wie die Daten erhoben wurden.

G SEARCH-HISTORY

Es folgt die Search-History.

DB	Dat.	Keywords	Tref.	Rel.
Cinahl Com- plete	21.10.15	Assessment2 AND children1 AND "quality of life"1 OR "health related quality of life"1 NOT caregiver2 NOT mother2 NOT father2	4983	xxx
		qol3 AND children3 AND assessment3 AND validity3 AND reliability3	8	3
		"quality of life" AND children3 AND assessment3 AND va- lidity3 AND reliability3	1238	xxx
	02.11.15	"quality of life" AND children3 AND "pediatric quality of life inventory"3 AND validity3	2	0
		pedsq1 4.01 AND reliability2	9	3
		pedsq1 4.01 AND reliability AND validity AND feasibility	2	0
		Kidscreen1 AND kidscreen-526	2	0
		kidscreen2 AND validity AND reliability	11	1
		eq-5d-y AND validity AND reliability	1	0
		KINDL6 AND validity AND reliability AND children	5	0
KINDL6 AND feasibility6	3	0		
ERIC via ProQuest	26.10.15	"quality of life"4 AND "PedsQL 4.0"	5	5
ERIC via Ovid	02.11.15	KINDL6	3	1
		kidscreen-522	2	0
		Euoqol	1	1
		"pediatric quality of life inventory" AND validity	3	1
		"pediatric quality of life inventory" AND reliability	4	1
"pediatric quality of life inventory" AND feasibility OR prac- ticability	1	0		
Psych. Info	26.10.15	"quality of life" AND "children	8	8
		"quality of life" 2 AND child* 2 AND assessment	307	xxx
	28.10.15	quality of life5 AND child* 5 AND assessment 5	64	8
		measure 5 AND quality of life 5 children	11	0
	02.11.15	"Pediatric Quality of life inventory" AND validity AND rela- bility	50	1
		"Kidscreen-52" AND validity AND reliability	30	1
	"EuroQoL-5D-Youth" OR "EuroQol Youth" AND reliability AND validity	11	0	
	KINDL6 AND validity AND reliability	13	0	
Medline via Pro- Quest	29.10.15	"child quality of life" 6 AND "pedsq1 4.0" 6 OR "pediatric quality of life inventory 4.0" 6 AND children6 OR adole- scents6 AND validation6 OR validity6 AND reliability6 And practicability6 OR feasibility6)	5	4
		"quality of life" AND "pedsq1 4.0" OR "pediatric quality of life inventory 4.0" AND children OR adolescents AND vali- dation OR validity AND reliability And practicability OR feasibility	21	11
		"PedsQL" AND "German Version" AND validity	0	0

Fortsetzung. SEARCH-HISTORY

DB	Datum	Suchwortkombinationen	Tref.	Rel.
	30.10.15	"inventory for the assessment of Quality of life in Children and Adolescents" ² OR "IQL" OR "ILK" AND "quality of life" OR "QOL"	0	0
		KINDL-R6 AND validity	13	0
		"Validity of KINDL-R"	0	0
		KINDL-R5	26	0
		Kidscreen-52	70	21
Medline via Pro-Quest	31.10.15	"ED-5D" AND "children" OR "adolescents" AND "validity" OR "reliability" NOT "caregivers" ⁴	109	10
		"ED-5D-Y" AND "children" OR "adolescents" AND "validity" OR "reliability" NOT "caregivers" ⁴	12	8
		"KINDL" AND "children" OR "adolescents" AND "validity" OR "reliability" OR "feasibility" NOT "caregivers" ⁴	42	0
		"KIDSCREEN-52" AND "children" OR "adolescents" AND "validity" OR "reliability" NOT "caregivers" ⁴	21	10
Medline via Web of Science	06.11.15	kidscreen-52 AND valid* AND europ*	10	1
		kidscreen AND feasibility	22	0
		kidscreen-52 AND validity AND europ*	6	0
Pubmed	02.11.15	kidscreen-52 AND reliability AND europ*	6	0
		kidscreen-52 AND feasibility AND europ*	1	1
		eq-5d-y AND feasibility AND	1	0
		eq-5d-y AND reliability AND	1	0
		eq-5d-y AND validity AND	12	0
		pedsq 4 0 AND validity AND preschool AND reliability AND feasibility	20	3
	03.11.15	kindl AND validity AND reliability	42	0
		kindl AND feasibility	5	0
		"pedsq 4.0" AND validity AND reliability	3	21
Cochrane library	01.11.15	"pedsq 4.0" AND feasibility	2	11
		"Kidscreen-52" AND feasibility OR practicability OR validity OR reliability	0	xxx
		"EUROQOL" AND children AND valid*	3	xxx
		"EUROQOL" AND children AND practicability/reliability/feasibility	0	xxx
		KINDL AND validity AND reliability	1	0
		KINDL AND feasibil*	2	0

Anmerkung. ¹Suche in Titel, ²MW Word in Subject, ³select a field optinal; ⁴Schlagwort, ⁵Suche in Keywords, ⁶Suche in Abstract. Tref. Anzahl Treffer, Rel. Anzahl relevanter Treffer. Xxx zu viele Treffer um relevante Treffer zu berücksichtigen. Relevante Treffer wurden schon von Duplikaten reduziert und anhand Titel ein- oder ausgeschlossen.