

Aus dem Institut für Gesundheits- und Pflegewissenschaft der Medizinischen Fakultät der Martin-Luther-Universität Halle-Wittenberg
(Direktorin: Prof. Dr. phil. Gabriele Meyer)

Externe Einrichtung, in der die Arbeit mitbetreut wurde: Graduiertenkolleg Demenz am Netzwerk
Altersforschung (NAR) der Ruprecht-Karls-Universität Heidelberg/Neckar
(Direktor: Prof. Dr. rer. nat. Dr. h.c. Konrad Beyreuther)

Organisationskontext und Forschungsanwendung in deutschen Pflegeheimen messen: Übersetzung, Adaptation und psychometrische Testung dreier kanadischer Assessmentinstrumente

Dissertation

zur Erlangung des akademischen Grades

Doktor rerum medicarum (Dr. rer. medic.) für das Fachgebiet

Pflegewissenschaft

vorgelegt

der Medizinischen Fakultät

der Martin-Luther-Universität Halle-Wittenberg

von Matthias Hoben
geboren am 16.10.1976 in Mutlangen/Ostalbkreis

Betreuer: 1. Prof. Dr. phil. (habil.) Johann Behrens
2. Prof. Dr. phil. (habil.) Dr. h.c. Andreas Kruse

Gutachter: 1. Prof. Dr. phil. (habil.) Johann Behrens
2. Prof. Dr. phil. Hermann Brandenburg
3. Prof. Dr. phil. (habil.) Oliver Schilling

Eröffnung: 11.02.2014
Verteidigung: 30.04.2014

Referat

Zielsetzung: Auch in der Altenpflege ist die Umsetzung wissenschaftlicher Erkenntnisse ethisch geboten. Das *Alberta Context Tool (ACT)*, *Estabrooks' Kinds of Research Use (RU) items* und die *Conceptual Research Use (CRU) scale* sind drei international breit genutzte Forschungsinstrumente zur Untersuchung des Zusammenhangs zwischen Organisationskontext und der Nutzung wissenschaftlicher Erkenntnisse in der stationären Altenpflege. Auf Deutsch liegen keine Instrumente mit vergleichbarer Qualität und Eignung vor. Ziel dieser Arbeit war daher, die Instrumente auf Deutsch zu übersetzen, sie an den Kontext der deutschen stationären Altenpflege anzupassen und umfassend psychometrisch zu testen.

Methoden: Phase 1: Übersetzung und Adaptation der Instrumente auf Basis internationaler *Best-Practice*-Leitlinien. Phase 2: Untersuchung der Verständlichkeit. Kognitive Interviews mit 16 Pflegehilfskräften (PHK), 5 Pflegefachkräften (PFK), 7 Therapie-/Betreuungskräften (TBK), 5 Experten (Exp.) und 6 Führungskräften (FK) aus 5 Pflegeheimen. Phase 3: Statistische Validierung. Fragebögen von 273 PHK, 196 PFK, 152 TBK, 6 Exp., 129 FK, 65 Altenpflegeschülern (APS) aus 38 Pflegeeinrichtungen wurden eingeschlossen. Die Validierung umfasste a) konfirmatorische Faktorenanalysen, b) Reliabilitätsbestimmung auf Basis der Faktormodellparameter, c) Faktorinvarianzanalysen, d) Mehrebenenfaktorenmodelle sowie e) Strukturgleichungsmodelle.

Ergebnisse: Phase 1: Die involvierten Experten bescheinigten den Instrumenten akzeptable inhaltliche Validität. Phase 2: Nach Anpassung problematischer Items (PHK-Gruppe) waren die Instrumente für alle Probanden verständlich. Phase 3: Die zustimmungsskalierten ACT-Items weisen eine Sieben-, die Häufigkeitsskalierten ACT-Items eine Sechs-Faktoren-Struktur auf. Die Items der CRU-Skala reflektieren einen Faktor. Beide ACT-Modelle sind partiell stark, das CRU-Modell partiell strikt messinvariant. Die Reliabilität der ACT-Konzepte ist mehrheitlich akzeptabel ($>0,7$), die der CRU-Skala sehr gut (0,897-0,963). Die Aggregierbarkeit der Antworten auf Einrichtungs-/Wohnbereichsebene ist auf Basis der Mehrebenenfaktorenmodelle nicht gegeben. Die Strukturgleichungsmodelle ergaben einige signifikante Zusammenhänge zwischen individuellen Merkmalen sowie ACT-Konzepten und den verschiedenen RU-Konzepten. Allerdings waren auch einige der Konzepte nicht signifikant oder in unerwarteter Richtung mit RU assoziiert.

Schlussfolgerungen: Verschiedene Ergebnisse verweisen auf akzeptable Validität der deutschen Instrumente: Experten beurteilten die Inhalte als adäquat, die Zielpersonen verstanden die Items wie intendiert, akzeptable konfirmatorische Faktorenmodelle verweisen auf strukturelle Validität und die Instrumente erwiesen sich als partiell stark (ACT) bzw. partiell strikt (CRU-Skala) messinvariant. Die Aggregierbarkeit der Daten wird allerdings von den Befunden dieser Arbeit nicht gestützt und die Evidenz bzgl. der Zusammenhänge zwischen den Variablen ist ambivalent.

Hoben, Matthias: Organisationskontext und Forschungsanwendung in deutschen Pflegeheimen messen: Übersetzung, Adaptation und psychometrische Testung dreier kanadischer Assessmentinstrumente, Halle (Saale), Univ., Med. Fak., Diss., 110 Seiten, 2014

Inhaltsverzeichnis

1	Einleitung.....	1
2	Forschungsstand.....	6
2.1	Bedeutung des Organisationskontext für DI-Prozesse in der stationären Altenpflege.....	6
2.2	Instrumente zur Erfassung des Organisationskontexts und der Anwendung von Forschungswissen in der stationären Altenpflege.....	10
2.2.1	Kontextassessmentinstrumente.....	10
2.2.2	<i>Research-Use-Instrumente</i>	12
2.3	Eigenschaften und Qualität der ausgewählten Instrumente.....	13
2.3.1	Das <i>Alberta Context Tool</i>	13
2.3.2	<i>Estabrooks' Kinds of Research Use items</i>	21
2.3.3	Die <i>Conceptual Research Use scale</i>	25
3	Zielstellung.....	27
4	Material und Methodik.....	27
4.1	Bezugsrahmen der psychometrischen Testung: <i>The Standards for Educational and Psychological Testing</i>	27
4.2	Genereller Aufbau der Studie.....	29
4.3	Phase 1: Übersetzung und Adaptation der Instrumente.....	29
4.4	Phase 2: <i>Cognitive Debriefing</i> der übersetzten Instrumente.....	35
4.4.1	<i>Cognitive-Debriefing-Sample</i>	35
4.4.2	Auswahl der zu testenden Fragebogenitems.....	36
4.4.3	Datensammlung.....	37
4.4.4	Datenanalyse und Instrumentenmodifikation.....	38
4.5	Phase 3: Psychometrische Testung der übersetzten Instrumente.....	38
4.5.1	Sample psychometrische Testung.....	38
4.5.2	Fragebögen und erhobene Variablen.....	41
4.5.3	Datensammlung.....	43
4.5.4	Datenaufbereitung.....	43
4.5.5	Datenanalyse.....	46
4.5.5.1	Analyse der beobachteten Skalenscores.....	46
4.5.5.2	Überprüfung der Faktorenstruktur und der Faktorinvarianz.....	46
4.5.5.3	Bestimmung der Reliabilität.....	51
4.5.5.4	Aggregierbarkeit der ACT-Daten auf Einrichtungs- bzw. Wohnbereichsebene.....	51
4.5.5.5	Zusammenhänge zwischen den Variablen: Strukturgleichungsmodelle.....	53
5	Ergebnisse.....	56
5.1	Erkenntnisse aus dem Übersetzungsprozesses.....	56

5.2	Resultate der <i>Cognitive Debriefings</i>	59
5.2.1	Beschreibung des <i>Cognitive-Debriefing</i> -Samples	59
5.2.2	<i>Cognitive Debriefings</i> mit den Pflegehilfskräften.....	60
5.2.3	<i>Cognitive Debriefings</i> mit den anderen Berufsgruppen	63
5.3	Psychometrische Güte der übersetzten Instrumente.....	63
5.3.1	Beschreibung der Stichprobe und der beobachteten Skalenscores	63
5.3.2	Faktorenstruktur und Reliabilität	68
5.3.3	Faktorinvarianz	72
5.3.4	Aggregierbarkeit der ACT-Daten	75
5.3.5	Zusammenhänge zwischen den Variablen	78
6	Diskussion	80
6.1	Übersetzung und Adaptation der Instrumente.....	81
6.1.1	Einordnung der Erkenntnisse des Übersetzungsprozesses	81
6.1.2	Limitationen des Übersetzungsprozesses.....	82
6.2	<i>Cognitive Debriefing</i> der übersetzten Instrumente.....	83
6.2.1	Einordnung der <i>Cognitive-Debriefing</i> -Ergebnisse.....	83
6.2.2	Limitationen des <i>Cognitive-Debriefing</i> -Teils	85
6.3	Psychometrische Testung der übersetzten Instrumente.....	85
6.3.1	Faktorenstruktur und Reliabilität	85
6.3.2	Faktorinvarianz	87
6.3.3	Aggregierbarkeit der ACT-Daten	87
6.3.4	Zusammenhänge zwischen den Variablen	88
6.4	Limitationen der psychometrischen Testung	90
6.4.1	Stichprobe	90
6.4.2	Modellfitevaluation	91
6.4.3	Modellmodifikation.....	93
6.4.4	Partielle Faktorinvarianz	93
6.4.5	Zusammenhänge zwischen den Variablen	94
7	Zusammenfassung	95
	Literaturverzeichnis	98
	Thesen	107
	Anhang	A1
	Lebenslauf	A6
	Selbstständigkeitserklärung	A7
	Erklärung über frühere Promotionsversuche	A8
	Danksagung	A9

Abkürzungsverzeichnis

ACT	Alberta Context Tool
ANOVA	Analysis of variance
ARU	Attitudes Towards Research Use
ATRU	Attitudes Towards Research Use (latenter Faktor)
BMS	Between-group mean square
BSTR	Behavioural Stress (latenter Faktor)
BURN	Burnout (latenter Faktor)
BW	Baden-Württemberg
CAMP	Connections Among People (latenter Faktor)
CFI	Comparative Fit Index
CI	Konfidenzintervall
CORU	Conceptual Research Use (latenter Faktor)
CRU	Conceptual Research Use
CSTR	Cognitive Stress (latenter Faktor)
CULT	Culture (latenter Faktor)
df	Degrees of freedom
DIF	Disseminations- und Implementierungsforschung
diff.	Differenz
DI-Prozess(e)	Disseminations- und Implementierungsprozess(e)
EBP	Evidenzbasierte Praxis
EL	Einrichtungsleitung
Exp.	Experte
FEED	Feedback (latenter Faktor)
FG	Frei-gemeinnützig
FINT	Formal Interactions (latenter Faktor)
FK	Führungskraft
G	Groß
GEE	General Estimating Equation
GEHE	General Health (latenter Faktor)
HE	Hessen
HLM	Hierarchisch-lineares Modell
ICC	Interclass correlation
IINT	Informal Interactions (latenter Faktor)
INRU	Instrumental Research Use (latenter Faktor)
IRU	Instrumental Research Use
ISPOR	International Society of Pharmacoeconomics and Outcomes Research

JOSA	Job Satisfaction (latenter Faktor)
K	Klein
LEAD	Leadership (latenter Faktor)
M	Mittel
MANOVA	Multivariate analysis of variance
MCAR	Missing completely at random
MTMM	Multitrait-Multimethod-Modelle
MW	Mittelwert
NFI	Normed Fit Index
Ö	Öffentlich
ORU	Overall Research Use
OVRU	Overall Research Use (latenter Faktor)
P	Privat
PARIHS	Promoting Action on Research Implementation in Health Services
PDL	Pflegedienstleitung
PFK	Pflegefachkraft
PFVI	Partielle Faktorvarianz-Invarianz
PERMANOVA	Permutation-based multivariate analysis of variance
PERU	Persuasive Research Use (latenter Faktor)
PHK	Pflegehilfskraft
PRU	Persuasive Research Use
RMSEA	Root Mean Square Error of Approximation
RP	Rheinland-Pfalz
RU	Research use
SA	Standardabweichung
SPAC	Space (latenter Faktor)
SRMR	Standardized Root Mean Square Residual
SSB	Sum of squares between groups
SST	Sum of squares total
STAF	Staffing (latenter Faktor)
STER	Structural/Electronic Resources (latenter Faktor)
TBK	Therapie-/Betreuungskraft
The Standards	The Standards for Educational and Psychological Testing
TIME	Time (latenter Faktor)
TLI	Tucker Lewis Index
TP	Therapeut
TREC	Translating Research in Elder Care

WB	Wohnbereich
WBL	Wohnbereichsleitung
WMS	Within-group mean square

1 Einleitung

Bei der professionellen Unterstützung hilfe- und pflegebedürftiger Menschen (gleich ob in einem pflegerischen, therapeutischen oder medizinischen Kontext) ist evidenzbasierte Praxis (EBP) ein zentrales Element (Behrens und Langer, 2010; Straus et al., 2011; Howlett et al., 2013). EBP in einem umfassenden Sinne verstanden bedeutet, professionelle Entscheidungen gemeinsam mit den hilfe- und pflegebedürftigen Menschen sowie ggf. ihren Angehörigen zu treffen (*shared decision making*) und dabei Informationen aus verschiedenen Quellen zu integrieren: a) den derzeit besten verfügbaren wissenschaftlichen Erkenntnisstand, b) Erfahrungen, Sichtweisen und Präferenzen der hilfe- und pflegebedürftigen Menschen wie auch ihrer Angehörigen, c) die eigene professionelle Erfahrung und Expertise und d) die strukturellen und institutionellen Anreize und Rahmenbedingungen (Behrens und Langer, 2010; Straus et al., 2011; Howlett et al., 2013). Insbesondere wissenschaftliche Erkenntnisse jedoch, so wird seit Jahren länder-, setting- und berufsgruppenübergreifend beklagt, werden nicht ausreichend berücksichtigt (Grimshaw et al., 2012; Meyer und Köpke, 2012). Suboptimale, wirkungslose oder schädigende Interventionen erhöhen die öffentlichen Gesundheitskosten, verringern die Rentabilität öffentlicher Forschungsausgaben und beeinträchtigen v.a. das Wohl der Betroffenen, ja kosten ggf. sogar Leben (Eccles et al., 2009; Behrens und Langer, 2010; Grimshaw et al., 2012).

Evidenzbasierte Neuerungen in die Praxis einzuführen (z.B. in Form von Leitlinien, Standards, Konzepten, Verfahren, Instrumenten oder allgemeinen EBP-Schulungen) erweist sich regelmäßig als Herausforderung (z.B. Edwards et al., 2003; Colón-Emeric et al., 2006; Müller-Hergl, 2006; Capezuti et al., 2007; Berta et al., 2010; Gitlin et al., 2010; Hoben, 2010; Sterns et al., 2010). Vielfältige Barrieren, Akteure und heterogene, z.T. widersprüchliche Interessenlagen tragen dazu bei, dass Implementierungsprozesse hochkomplex und damit schwer zu planen und zu steuern sind (Greif et al., 2004; Greenhalgh et al., 2005; Kitson, 2009). International setzt sich ein immer größer werdender Fundus an Forschungsarbeiten mit der Frage auseinander, wie wissenschaftliche Evidenz möglichst effektiv in die klinische Praxis implementiert werden kann – subsumiert unter unterschiedlichsten Begriffen, wie z.B. *translational research*, *research utilization*, *knowledge translation*, *innovation diffusion*, *implementation research* uvm. (Graham et al., 2006; Green et al., 2009; McKibbin et al., 2010). Für diese Begriffe existieren unzählige Definitionen, die teils deutliche Unterschiede für ein und denselben Begriff, z.T. wiederum weitgehende Überschneidungen unterschiedlicher Begriffe implizieren (Graham et al., 2006; McKibbin et al., 2010). Bei aller Begriffsvielfalt ist jedoch – vor allem für Personen, die mit dem Feld vertraut sind – ein gemeinsamer Kern dieser Arbeiten erkennbar: die wissenschaftliche Analyse von Prozessen, in denen wissenschaftliche Erkenntnisse in die Praxis überführt werden sollen (Graham et al., 2006; Estabrooks et al., 2008a).

In Deutschland beginnt sich für diesen Gegenstandsbereich der Begriff „Disseminations- und Implementierungsforschung“ (DIF) zu etablieren (DGP, 2013; Roes et al., 2013), ähnlich wie im

US-amerikanischen Diskurs (Green et al., 2009; NIH, 2013). Diese Bezeichnung wird in dieser Arbeit durchgängig verwendet. Dabei ist es hilfreich, zwischen den Begriffen „Diffusion“, „Dissemination“ und „Implementierung“ zu unterscheiden (vgl. zum Folgenden Rabin und Brownson, 2012, sowie die Definitionen in Abb. 1). Diffusion bezeichnet die passive, nicht geplante oder gesteuerte Verbreitung einer Neuerung, z.B. indem Praktiker darüber lesen und diese weiterkommunizieren. Dissemination hingegen ist ein aktiver, gezielter und systematischer Ansatz, Informationen zu verbreiten, z.B. durch Schulungen, indem schriftliches oder elektronisches Informationsmaterial verteilt wird oder durch gezielten Einsatz von Massenmedien. Implementierung schließlich ist der aktivste der drei Ansätze und umfasst neben Disseminationsinterventionen eine große Palette weiterer Strategien, wie z.B. Praxistrainings, Monitoring der Ist-Situation und anschließende Rückmeldung der Ergebnisse (*audit and feedback*), Beratung, Coaching oder umfassende Unterstützung (*facilitation*) durch das Management bzw. durch Experten. Folglich bezeichnet Disseminationsforschung die Untersuchung von Disseminationsprozessen, der darin wirksamen Faktoren und effektiver Disseminationsmethoden. Die Implementationsforschung untersucht Implementierungsprozesse einschließlich Einflussfaktoren, effektiver Implementierungsstrategien, Adaptation und Umsetzungstreue (*implementation fidelity*) der Neuerung etc.

„**Diffusion** is the passive, untargeted, unplanned, and uncontrolled spread of new interventions. Diffusion is part of the diffusion-dissemination-implementation continuum, and it is the least focused and intense approach.” (Rabin and Brownson, 2012, 25)

„**Dissemination** is an active approach of spreading evidence-based interventions to the target audience via determined channels using planned strategies.” (Rabin and Brownson, 2012, 26)

„**Implementation** is the process of putting to use or integrating evidence-based interventions within a setting.” (Rabin and Brownson, 2012, 26) **Implementation** = „dissemination plus action to actively encourage the adoption of recommendations contained in a message.” (Mowatt et al., 1998, 696)

„**Dissemination research** is the systematic study of processes and factors that lead to widespread use of an evidence-based intervention by the target population. Its focus is to identify the best methods that enhance the uptake and utilization of the intervention.” (Rabin and Brownson, 2012, 31)

„**Implementation research** seeks to understand the processes and factors that are associated with successful integration of evidence-based interventions within a particular setting (e.g., a worksite or school). Implementation research assesses whether the core components of the original intervention were faithfully transported to the real-world setting (i.e., the degree of fidelity of the disseminated and implemented intervention with the original study) and is also concerned with the adaptation of the implemented intervention to local context. Another [...] essential component of implementation research involves the enhancement of readiness through the creation of effective climate and culture in an organization or community.” (Rabin and Brownson, 2012, 31)

Abb. 1: Definition zentraler Schlüsselbegriffe der DIW

Der systemspezifische Fokus dieser Arbeit liegt auf dem Sektor der stationären Altenpflege. Diesem kommt im Kontext der DIF eine besondere Bedeutung zu. Verschiedene Arbeiten zeigen auch in diesem Bereich Diskrepanzen auf zwischen wissenschaftlichen Erkenntnissen (*best prac-*

tice) und Kenntnisstand bzw. Handeln der in der Praxis Tätigen – hierzulande (Bredthauer et al., 2005; Kuske et al., 2009; Meyer et al., 2009; Majic et al., 2010; Treusch et al., 2010; Wilborn and Dassen, 2010; Jordan et al., 2012; Sommer et al., 2012; Reuther et al., 2013) wie auch international (Ishii et al., 2009; Masso and McCarthy, 2009; Boström et al., 2012). 2011 existierten in Deutschland 12.354 Pflegeheime, in denen knapp 30% aller pflegebedürftigen Menschen (743.120 Personen) versorgt wurden (Statistisches Bundesamt, 2013). Diese Personengruppe zeichnet sich aus durch einen hohen und komplexen Pflegebedarf. Wie z.B. Balzer et al. (2013) zeigen, leiden Pflegeheimbewohner durchschnittlich an vier bis fünf Erkrankungen zugleich, 50-70% sind von einer Demenz betroffen, 70-80% urin- oder stuhlinkontinent, 50-90% weisen Mund- und Zahn- bzw. Zahnprothesenprobleme auf (Zahnbeläge, Karies, parodontale Entzündungen, schlechter Zustand der Zahnprothesen), 25-50% leiden unter Depressivität, Schlafstörungen oder zeigen agitierte oder apathische Verhaltensweisen. Hinzu kommen funktionale Einschränkungen (z.B. Seh- oder Hörprobleme, eingeschränkte Sprechfähigkeit, reduzierte Mobilität) und sie sind hochgradig bedroht von Infektionen, Stürzen, Mangelernährung, Dekubitus etc. (Lahmann et al., 2010; Volkert et al., 2011; Balzer et al., 2013). Pflegeheimbewohner sind somit eine vulnerable Personengruppe, die von Qualitätsproblemen besonders betroffen ist. Gerade der stationäre Altenhilfebereich jedoch wurde im Kontext der DIF bislang noch wenig beachtet (Masso and McCarthy, 2009; Boström et al., 2012). International wurden daher große Forschungsprogramme ins Leben gerufen, um diese Lücken zu schließen (z.B. Estabrooks et al., 2009a; Cranley et al., 2011; Seers et al., 2012). In Deutschland findet DIF dagegen eher vereinzelt statt und kaum im stationären Altenpflegebereich (Bartholomeyczik, 2008; Roes et al., 2013).

Ein wesentlicher Bestandteil der oben zitierten Forschungsprogramme ist die Definition und standardisierte Messung geeigneter Outcomes. Dies gilt als eine der dringlichsten Aufgaben, um den noch jungen Bereich der DIF voranzubringen, und zugleich als eine der größten Herausforderungen dieses Gebiets (Graham et al., 2010; Proctor et al., 2011; Proctor and Brownson, 2012). Diesem DIF-Themenbereich widmet sich auch die vorliegende Arbeit. Die standardisierte Erfassung von Variablen in DI-Prozessen mittels reliabler und valider Messinstrumente ermöglicht es, theoretische Annahmen über Zusammenhänge in diesen Prozessen sowie die Effektivität von DI-Strategien mittels statistischer Methoden in großen Stichproben zu testen – und trägt damit zu einem besseren Verständnis dieser Prozesse und zur Theoriebildung bei (Proctor and Brownson, 2012). Auch im Rahmen der Interventionsforschung sind diese Messinstrumente bedeutsam (vgl. zum Folgenden Fixsen et al., 2005; Carroll et al., 2007; Graham et al., 2010; Proctor and Brownson, 2012): Die „Wirksamkeit“ einer evidenzbasierten Intervention (z.B. eine klinische Leitlinie oder die EBP im Allgemeinen) hängt auch davon ab, ob und wie gut sie umgesetzt wird – und Letzteres wird wiederum maßgeblich durch den DI-Prozess beeinflusst (Abb. 2).

Klassische Ergebnismaße zur Wirksamkeitsüberprüfung sind Patienten-/Bewohneroutcomes (z.B. bessere Lebensqualität, Schmerzreduktion, besserer Ernährungszustand, ...) oder System-

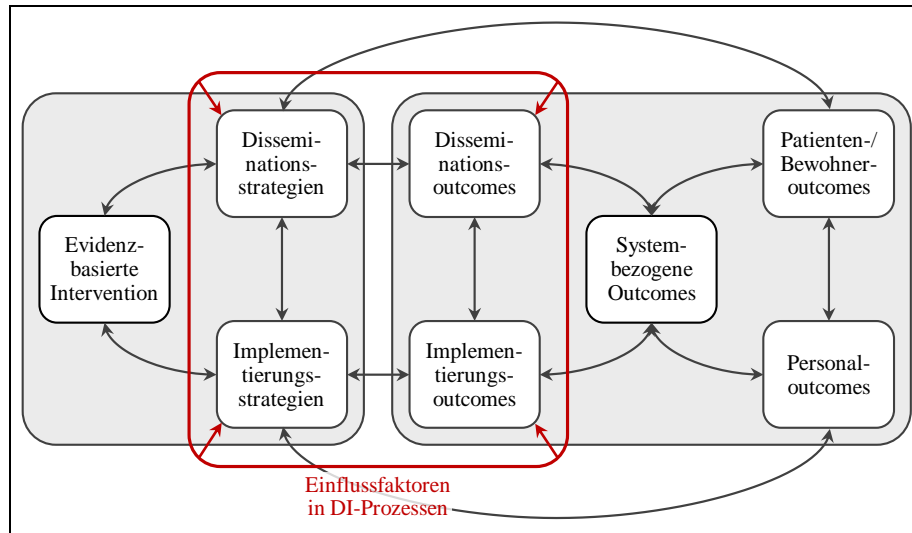


Abb. 2: Outcome-Ebenen in der DIF
(nach Proctor and Brownson, 2012, 262, eigene Adaptation)

outcomes (z.B. Kostenreduktion, verringerte Wartezeiten, reibungslosere Prozesse, ...) (Proctor and Brownson, 2012). Diese Outcomes lassen jedoch nur sehr begrenzt Aussagen darüber zu, ob die zu untersuchende Neuerung von allen Personen angewendet wurde, die sie anwenden sollten, bei allen Personen, die zur Zielgruppe gehören (und nur bei diesen), auf die vorgesehene Weise (Häufigkeit, Anlass/Zeitpunkt, Dauer, Vollständigkeit, Korrektheit, ...) etc. (Carroll et al., 2007). Um die Aussagekraft von Interventionsstudien zu erhöhen und deren Ergebnisse besser interpretieren zu können, wird daher gefordert, nicht nur die Effektivität der Intervention zu untersuchen, sondern auch den DI-Prozess detailliert darzustellen und zu evaluieren. DI-Outcomes sind z.B. Nutzung von Evidenz im Allgemeinen (*research use*), Akzeptanz der Neuerung durch die Zielpersonen, *Fidelity* (Grad, in dem die Neuerung wie vorgesehen umgesetzt wird) oder Nachhaltigkeit (dauerhafte Nutzung) (Proctor and Brownson, 2012). Auch auf Ebene des Personals können die Auswirkungen von DI-Prozessen erfasst werden. Arbeitszufriedenheit, Kompetenzzuwachs, Belastung, Stress, Burnout u.ä. sind exemplarische Outcomes (Graham et al., 2010). In wie weit schließlich DI-Strategien wirksam sind, hängt ab von unzähligen Einflussfaktoren, die auf verschiedenen Ebenen interagieren: Merkmale der Neuerung selbst, Eigenschaften der Pflegebedürftigen und ihrer Angehörigen, Eigenschaften der Professionellen, Merkmale der Organisation und externe Rahmenbedingungen (Chaudoir et al., 2013; Quasdorf et al., 2013).

Reliable und valide Instrumente zur Erfassung DI-spezifischer Outcomes sind somit bedeutsam für die Erforschung von DI-Prozessen im Allgemeinen, wie auch für die Evaluation dieser Prozesse im Rahmen klinischer Studien. Insbesondere der Zusammenhang zwischen Faktoren des Organisationskontexts (z.B. Führungsqualität, Miteinander im Team, Organisationskultur u.ä.) und der Nutzung wissenschaftlicher Evidenz kristallisiert sich hier als bedeutsames Forschungsthema heraus (Dopson und Fitzgerald, 2005; Greenhalgh et al., 2005; Meijers et al., 2006). Organisationskontext wird in dieser Arbeit definiert als: „the environment or setting in which people

receive health care services” (McCormack et al., 2002, 96), bzw. zugespitzt auf DI-Prozesse: „the environment or setting in which the proposed change is to be implemented” (Kitson et al., 1998, 150). Der Kontext einer Organisation beeinflusst die Situation ihrer Akteure (z.B. Arbeitszufriedenheit, Arbeitsbelastung, Stress) ebenso wie ihre Einstellung und ihr Handeln (Glisson, 2002; Greenhalgh et al., 2005; Meijers et al., 2006; Kaplan et al., 2010; Aarons et al., 2012) – und dies nicht nur bei Individuen, sondern bei ganzen Teams (Cummings et al., 2010; Estabrooks et al., 2011b; Estabrooks et al., 2011c; Aarons et al., 2012). Zugleich sind viele dieser Organisationskontextfaktoren besser modifizierbar als manch andere Faktoren, auf die sich die DIF lange konzentrierte (z.B. Berufserfahrung oder Ambiguitätstoleranz von Individuen, Komplexität oder subjektiver Nutzen der Innovation oder globale Rahmenbedingungen wie die politische und ökonomische Situation) (Estabrooks et al., 2007; Estabrooks et al., 2008b; Estabrooks et al., 2009c). International stehen mehr und mehr standardisierte Assessmentinstrumente zur Verfügung, um abhängige DI-Variablen (z.B. Nutzung von Forschungswissen in der Praxis, Squires et al., 2011c) wie auch Einflussfaktoren in DI-Prozessen (Chaudoir et al., 2013) zu erfassen. Robuste Instrumente für den stationären Altenpflegebereich sind jedoch rar – insbesondere was die Erfassung des Organisationskontexts angeht (Estabrooks et al., 2009c).

Das *Alberta Context Tool (ACT)* schließt diese Lücke (Squires et al., 2013c). Es wurde 2006 entwickelt, da die Urheberinnen kein anderes Instrument identifizieren konnten, das a) auf robuster DI-Theorie und -Evidenz basierte, b) potenziell modifizierbare Organisationskontextfaktoren in verschiedenen Versorgungssettings erfasste, c) kurz genug war, um von professionellen Akteuren mit knappen Zeitressourcen akzeptiert zu werden und d) eine akzeptable psychometrische Güte aufwies (Hutchinson et al., 2008; Estabrooks et al., 2009b; Squires et al., 2009). Es liegt unter anderem in einer Version für den stationären Altenpflegebereich vor, beinhaltet spezifische Fragebögen für sechs verschiedene Berufsgruppen (Pflegehilfskräfte, Pflegefachkräfte, Betreuungs-/Therapeuten, Experten wie Qualitätsbeauftragte oder Fort- und Weiterbildungsdozenten, Führungspersonen sowie Ärzte), wurde erfolgreich in verschiedenen Studien eingesetzt und die psychometrischen Eigenschaften wurden und werden weiterhin ausführlich untersucht (Squires et al., 2013c) (Details siehe Abschnitt 2.3.1). Um die abhängige Variable – *research use* – zu erfassen, wurde in den angesprochenen Studien zwei weitere Instrumente eingesetzt: *Estabrooks’ Kinds of Research Use (RU) items* (Estabrooks, 1999a, 1999b) und die *Conceptual Research Use (CRU) scale* (Squires et al., 2011b). Beide Instrumente liegen ebenfalls in Versionen für die genannten Berufsgruppen der stationären Altenpflege vor und auch sie konnten sich in forschungspraktischer wie psychometrischer Hinsicht bewähren (Squires et al., 2011b; Squires et al., 2011c). Zum Zeitpunkt der Konzeption dieser Arbeit (2010) waren keine Instrumente mit vergleichbaren Eigenschaften in deutscher Sprache verfügbar (vgl. Abschnitt 2.2). Für das ACT trifft dies bis heute zu. Deutschsprachige Instrumente zur Erfassung forschungsbasierter Praxis wurden inzwischen von Breimaier et al. (2011), Schnittger et al. (2012) und Köpke et al. (2013) entwickelt,

basierend auf der Übersetzung und Adaptation internationaler Skalen und selbst entwickelter Items. Alle drei Instrumente wurden jedoch für den Krankenhausbereich konzipiert und statistische Daten zur psychometrischen Güte publizierten bislang nur Schnittger et al. (2012) (interne Konsistenzreliabilität – Cronbach's α und dimensionsreduzierende Faktorenanalysen).

Hauptziele dieser Arbeit waren daher: a) die deutsche Übersetzung der drei kanadischen Instrumente – des ACT, der *Estabrooks' Kinds of RU items* und der CRU-Skala – jeweils in der Ausführung für den stationären Altenpflegebereich, b) die Adaptation der Instrumente an die kulturellen und strukturellen Gegebenheiten der deutschen stationären Altenpflege sowie c) die ausführliche psychometrische Testung der übersetzten Tools. Abhängig davon, ob die Ergebnisse der psychometrischen Überprüfung entsprechende Aussagen zulassen würden, sollte – als Unterziel – ein erster Einblick gewonnen werden, wie Faktoren des Organisationskontexts sowie die Anwendung wissenschaftlicher Erkenntnisse in deutschen Altenpflegeheimen ausgeprägt sind und ob ein Zusammenhang zwischen diesen Größen besteht.

2 Forschungsstand

2.1 Bedeutung des Organisationskontext für DI-Prozesse in der stationären Altenpflege

Einen der umfassendsten Einblicke in den Zusammenhang zwischen Organisationskontext und Einführung von Neuerungen im Gesundheitswesen geben Greenhalgh et al. (2005) in ihrem metanarrativen Review. Lange Zeit konzentrierte sich die DIF vorwiegend auf Eigenschaften individueller Akteure (z.B. Berufserfahrung, Qualifikation, Haltungen und Einstellungen u.ä.), die sich allerdings als wenig erklärungskräftig erwiesen – sind doch DI-Prozesse stets soziale Prozesse und damit eingebettet in einen sozialen Kontext (Greenhalgh et al., 2005). 20 der insgesamt 499 einbezogenen Arbeiten beleuchten Fragen des Organisationskontexts: drei Metaanalysen (insgesamt 64 Publikationen) sowie 17 Einzelarbeiten (darunter eine nichtrandomisiert-kontrollierte Studie, drei mit Pre-Post-Design, fünf Querschnittsstudien, zwei *Mixed-Methods*-Untersuchungen und sechs qualitative Arbeiten). Nur eine der Einzelstudien wurde im Pflegeheimbereich durchgeführt, 15 in anderen Gesundheitssettings (vorwiegend in Krankenhäusern) und eine entstammt einem Schulsetting. Auch die Metaanalysen bezogen nicht nur Studien aus Gesundheitssettings ein, exakte Details sind jedoch nicht berichtet.

Kontext, so Greenhalgh et al. (2005, 134-135), könne man sich vorstellen als „medium through which any organisational innovation must pass in order for it to be assimilated and routinised, which affects the rate and direction of adoption“. Harte, greifbare Strukturkomponenten (z.B. Organisationsgröße) zählen ebenso zum Organisationskontext wie weiche, weniger klare Konstrukte (z.B. Organisationsklima oder -kultur) (Greenhalgh et al., 2005). Folgende strukturelle Faktoren spielen eine Rolle: Innovationen erfolgreich zu assimilieren, gelingt Organisationen besser, die a) groß sind, b) seit langem bestehen, c) funktionell stark ausdifferenziert sind und d) ihr professionelles Wissen besonders spezialisiert haben. Allerdings erklären diese Faktoren

nur knapp 15% der Innovativitätsvarianz (Greenhalgh et al., 2005). Zwei weitere wichtige Einflussfaktoren auf die Innovativität sind nichtstruktureller Natur: die Aufnahmefähigkeit (*absorptive capacity*) für neues Wissen und ein für Veränderungen empfänglicher Kontext (*receptive context for change*) (Greenhalgh et al., 2005). Im Gegensatz zu den strukturellen Faktoren, die jeweils als Einzelkonstrukte konzipiert sind, handelt es sich bei diesen beiden Konstrukten um komplexe Konglomerate verschiedener Faktoren. Für die Aufnahmefähigkeit neuen Wissens sind insbesondere bereits vorhandene Wissensbestände (v.a. *tacit knowledge*) und Fertigkeiten, verfügbare Technologien, eine lernende Organisationskultur sowie eine proaktive, auf Wissensaustausch gerichtete Führung ausschlaggebend. Für einen veränderungsempfänglichen Kontext relevante Komponenten sind: starke Führung, klare strategische Vision, gute Beziehungen zwischen Führungspersonen und Untergebenen, visionäres Personal in Schlüsselpositionen, ein Klima, das Experimentieren und Risikobereitschaft fördert sowie ein effektives Datenerfassungs- und -überwachungssystem (Greenhalgh et al., 2005). Die Bereitschaft und Fähigkeit zur Veränderung (*readiness for change*) ist eine weitere Komponente. Sie ist eine komplexe Funktion aus Veränderungsdruck, Innovations-Organisations-Passung, Einschätzung der mit der Innovation verbundenen Implikationen, Unterstützung und Befürwortung der Innovation, verfügbaren Ressourcen (Zeit, Geld, Personal, Wissen, ...) und Fähigkeit, die Innovation zu evaluieren (Greenhalgh et al., 2005). Für die Verstetigung von Innovationen sind insbesondere die Fähigkeit, diese anzupassen, Kommunikation sowie effektive Feedbackmechanismen relevant (Greenhalgh et al., 2005).

Zu ganz ähnlichen Ergebnissen kommen Kaplan et al. (2010) in ihrer systematischen Übersichtsarbeit. Sie untersuchten den Zusammenhang zwischen Organisationskontextfaktoren und dem Erfolg von Qualitätsentwicklung. Die Autoren schlossen 47 Studien ein: drei nichtrandomisiert-kontrollierte Studien, sieben Kohortenstudien, eine Fall-Kontroll-Studie, zwei Pre-Post-Studien und 34 Querschnittstudien. Zehn der Studien bezogen sich auf stationäre Altenpflegesettings, die Ergebnisse sind jedoch Settingübergreifend dargestellt und lassen keine gesonderte Interpretation für einzelne Bereiche zu. Bedeutsame Faktoren waren, in Übereinstimmung mit gängigen Organisations- und Veränderungstheorien, Führung durch das Topmanagement, Organisationskultur, datenbezogene Infrastruktur und Informationssysteme sowie Erfahrung mit formaler Qualitätsentwicklung (Kaplan et al., 2010). Weniger eindeutige, aber dennoch potenziell relevante Faktoren waren: Involvierung von Ärzten in Qualitätsentwicklung, Motivation im Team, Ressourcen und Führung auf Teamebene. Die verfügbaren Studien, so Kaplan et al. (2010), waren vor allem dadurch limitiert, dass sie ihren Überlegungen kein theoretisches Modell zugrunde legten, kontextuelle Faktoren nicht klar definierten oder unzureichend validierte Erhebungsinstrumente einsetzten.

Meijers et al. (2006) führten ein systematisches Review durch zum Zusammenhang zwischen Organisationskontext und Forschungsanwendung im Pflegebereich. Sie schlossen zehn Studien ein: eine nichtrandomisiert-kontrollierte Studie, acht Querschnittstudien und eine *Mixed-Me-*

thods-Studie. Nur eine Arbeit bezog Pflegende aus dem stationären Altenpflegebereich ein. Die Ergebnisse sind jedoch für dieses Setting nicht isolierbar. Alle eingeschlossenen Studien wiesen methodische Limitationen bzgl. Design, Sampling, Outcomemessung oder statistischer Analyse auf. Nur drei der Studien legten ein theoretisches Modell zugrunde (Meijers et al., 2006). Insgesamt hingen sechs Kontextfaktoren wiesen statistisch signifikant mit der Anwendung von Forschungswissen zusammen: a) Rolle der Pflegenden (Verantwortungsbereich und Involvierung in Qualitätsentwicklungsaufgaben), b) Zugriff auf Forschungswissen, personelle und materielle Ressourcen, c) Organisationsklima, d) Support, e) Zeit und e) EBP-bezogene Bildungsangebote (Meijers et al., 2006). All diese Faktoren ließen sich subsumieren unter die Kategorien eines bekannten DIF-Modells (Abb. 3) – des PARIHS-Frameworks (PARIHS steht für *Promoting Action on Research Implementation in Health Services*) (Kitson et al., 1998; Rycroft-Malone et al., 2002). Dieses geht davon aus, dass der Erfolg von DI-Prozessen eine Funktion ist aus der a) Evidenz der einzuführenden Neuerung, b) der *Facilitation* im DI-Prozess und c) dem Kontext, in dem der Prozess stattfindet. Kontext wiederum wird operationalisiert in die Subdimensionen Kontext, Kultur, Führung und Evaluation. Die Subsumierbarkeit der gefundenen Faktoren wertete die Autoren als eine gewisse Unterstützung der Modellannahmen; allerdings fand sich kein Faktor zur Subdimension Evaluation (Meijers et al., 2006).

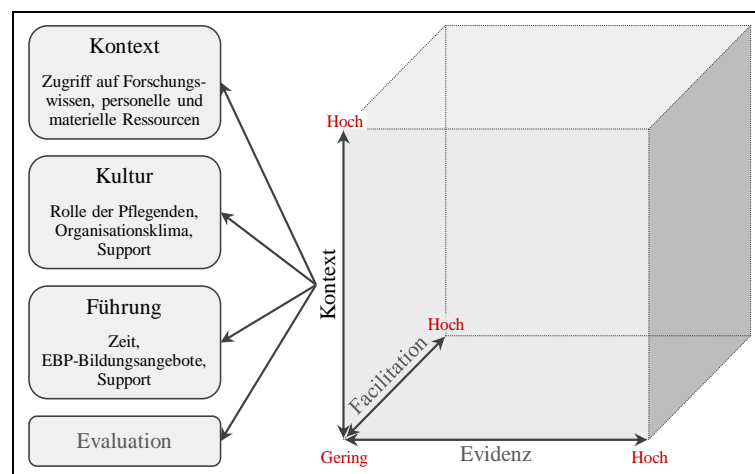


Abb. 3: Zuordnung der von Meijers et al. (2006) gefundenen Kontextfaktoren zu den Kontextdimensionen des PARIHS-Frameworks (nach Kitson et al., 1998 u. Meijers et al., 2006, eigene Adaptation)

Metaanalysen von Interventionsstudien kommen zu etwas anderen Ergebnissen als die zitierten Reviews. Dijkstra et al. (2006) untersuchten mittels Metaregressionsmodellen, welche DI-Strategien im Rahmen der Leitlinienimplementierung in Krankenhäusern effektiv sind. Die Methodik folgte den Empfehlungen der *Cochrane Effective Practice and Organization of Care (EPOC) review group* (Bero et al., 2001). Eine für diese Arbeit interessante Nebenfragestellung war, inwieweit Organisationskontextfaktoren die Effektivität dieser Strategien moderieren. Folgende Kontextfaktoren wurden einbezogen: a) Führung (Support durch das Management ja/nein),

b) lernende Umgebung (akademisches versus nicht akademisches Krankenhaus), c) funktionale Differenzierung (mehr als eine Spezialisierung ja/nein, mehr als eine Art Ärzte vorhanden ja/nein, multiprofessionelle Aufgabenbearbeitung ja/nein) sowie d) konsensuale Prozesse (Leitlinie im Krankenhaus versus extern entwickelt, DI-Strategien im Krankenhaus versus extern entwickelt) (Dijkstra et al., 2006). 53 Studien wurden in die Analyse einbezogen: 32 clusterrandomisierte Studien, sieben individualrandomisierte Studien, sieben nichtrandomisiert-kontrollierte Studien, sieben Pre-Post-Studien. 19 der Studien untersuchten stationäre, 28 ambulante Krankenhausbereiche und sechs bezogen gemischte Settings ein (Dijkstra et al., 2006). Nur zwei der sieben Kontextfaktoren beeinflussten die Effektivität der eingesetzten DI-Strategien signifikant (Tab. 1): Akademische Krankenhäuser schnitten, was den Erfolg der DI-Strategien angeht, signifikant besser ab als nichtakademische – allerdings nicht im Gesamtstudiensample, sondern nur bei getrennter Betrachtung stationärer und ambulanter Bereiche, und im stationären Bereich auch nur verglichen mit nichtakademischen Einrichtungen öffentlicher Trägerschaft. Extern entwickelte DI-Strategien erwiesen sich zudem als erfolgreicher als intern entwickelte – im Gesamtsample und insbesondere in ambulanten Krankenhausbereichen.

Tab. 1: Effekte moderierender Organisationskontextvariablen
(Auszug aus Dijkstra et al., 2006, eigene Übersetzung)

Moderatorvariable	Gesamt		Stationäre Bereiche		Ambulante Bereiche	
	n	OR (95%CI)	n	OR (95%CI)	n	OR (95%CI)
Lernende Umgebung						
Akademisches Krankenhaus	44	2,15 (1,62-2,86)	9	3,42 (2,25-5,19)	32	1,91 (1,41-2,59)
Nichtakademisches Krankenhaus	37	2,11 (1,49-2,98)	20	1,44 (0,95-2,19)*	9	3,90 (2,87-5,29)*
Konsensuale Prozesse						
DI-Strategie intern entwickelt	59	1,78 (1,36-2,32)	16	2,56 (1,56-4,21)	36	1,85 (1,32-2,60)
DI-Strategie extern entwickelt	22	4,62 (2,82-7,57)**	13	1,37 (0,60-3,11)	5	12,7 (4,99-32,20)**

Signifikante Unterschiede für vorhandenen/nicht vorhandenen Moderator: *p < 0,05 **p < 0,01
95%CI: 95%-Konfidenzintervall, n: Zahl der Studien; OR: Odds Ratio

Ein Cochrane-Review (Flodgren et al., 2012) untersuchte die Studienlage zur Effektivität organisationaler Infrastrukturinterventionen zur Beförderung von EBP. Organisationale Infrastruktur definierten die Autoren als „the underlying foundation or basic framework through which clinical care is delivered and supported“ (Flodgren et al., 2012, 3). DI-Strategien, die an einer Modifikation dieser Faktoren ansetzen, sind z.B. der Einsatz von EBP-Mentoren/-Champions, Kooperation zwischen klinischen und akademischen Institutionen, Verfügbarmachen von Ressourcen für EBP, Support, EBP-Besprechungen, Journal-Clubs, schriftliche Fixierung der EBP-Ziele in der Organisationsphilosophie, EBP-bezogene Fort- und Weiterbildung sowie Gratifikation erfolgreicher EBP-Implementierung. In den vorangegangenen Reviewversionen (Foxcroft und Cole, 2000, 2003) konnte keine einzige einschließbare Studie identifiziert werden. Sieben Studien wurden beschrieben, aus methodischen Gründen jedoch ausgeschlossen. In der aktuellen

Version (Flodgren et al., 2012) wurde eine Pre-Post-Studie mit sehr schlechter Qualität eingeschlossen, die aufgrund fehlender Informationen und hohen Biasrisiken keine Aussage zur Effektivität der DI-Strategien (Poster zur Weiterbildung des Personals, Eins-zu-eins-Trainings durch Peers, so genannte Auszugspräsentationen – also prägnante Erklärungen eines komplexen Sachverhalts in kurzer Zeit sowie Dokumentationsaudits) zuließ.

Die zitierten Übersichtsarbeiten, so das Zwischenfazit, zeigen eindeutig, dass insbesondere quantitative Studien zum Zusammenhang von Organisationskontext und Anwendung von Forschungswissen unter methodischen Limitationen leiden. Die Hauptursachen hierfür sind unzureichende theoretische Fundierung und Definition der Konzepte sowie Erhebungsinstrumente unzureichender Güte. Narrative Reviews und qualitative Studien betonen die Bedeutung des Organisationskontexts, methodisch suffiziente Überprüfungen dieser Hypothesen mittels statistischer Methoden sind bislang jedoch rar. Insgesamt fokussieren die Studien vor allem den Krankenhausbereich. Im stationären Altenpflegebereich ist die Studienlage sehr dünn, was auch Boström et al. (2012) in ihrem *Scoping Review* unterstreichen.

2.2 Instrumente zur Erfassung des Organisationskontexts und der Anwendung von Forschungswissen in der stationären Altenpflege

2.2.1 Kontextassessmentinstrumente

In englischer Sprache sind, wie Chaudoir et al. (2013) in ihrer systematischen Übersichtsarbeit aufzeigten, 62 Assessmentinstrumente zur Erfassung DI-relevanter Einflussfaktoren erhältlich. Diese Instrumente erfassen Einflussfaktoren auf einer bzw. mehreren der folgenden Ebenen: a) Eigenschaften der Innovation (z.B. Komplexität, Testbarkeit, Wahrnehmbarkeit der Auswirkungen, relativer Nutzen im Vergleich zu vorher, ...), b) Eigenschaften individueller Pflegebedürftiger (z.B. krankheits-/therapiebezogenes Wissen, Selbstbewusstsein, Einfordern bestimmter Leistungen/Vorgehensweisen), c) Eigenschaften individueller Professioneller (z.B. Einstellungen/Haltungen, Qualifikation, Berufserfahrung, ...), d) Faktoren des Organisationskontexts (z.B. Feedbackmechanismen, Miteinander im Team, Führung, ...) und e) strukturelle Bedingungen bzw. Umweltfaktoren (z.B. politische, juristische oder ökonomische Situation) (Chaudoir et al., 2013; Quasdorf et al., 2013).

Nach Wissen des Autors dieser Arbeit sind auf Deutsch nur zwei Instrumente verfügbar, die spezifisch dafür konstruiert wurden, DI-relevante Einflussfaktoren zu erfassen: die *Barriers to Research Utilization Scale (BARRIERS)* (Funk et al., 1991) und der *Questionnaire on Utilisation of Nursing Research* (kurz: *Parahoo-Questionnaire*) (Parahoo, 1998). Die *BARRIERS scale* wurde zunächst im Rahmen einer schweizerischen Studie (Saxer, 2002; Walker Schläefli, 2005) auf Deutsch übersetzt und untersucht auf a) Inhaltsvalidität bzw. Verständlichkeit (Pretest mit 10 Pflegeexpertinnen in Ausbildung), b) Interne-Konsistenz-Reliabilität (Cronbach's α , Werte zwischen 0,66 und 0,91) sowie c) strukturelle Validität (Hauptkomponentenanalysen, Varimaxrota-

tion mit Kaiser-Normalisierung, bestätigte die Faktoren-Struktur des Originalinstruments). Auch in einer bundesdeutschen Studie (Schubert und Wrobel, 2009) wurde die *BARRIERS scale* eingesetzt und mittels „Faktorenanalysen“ untersucht (bestätigte die Vier-Faktoren-Struktur), wobei keine weiteren methodischen Details berichtet werden (z.B. ob es sich um ein exploratorisches oder konfirmatorisches Modell handelt – wobei die Aussagen eher auf ein exploratorisches Vorgehen hindeuten, ob im ersten Fall eine Hauptkomponentenanalyse (regressionsorientiert) vorgenommen wurde oder eine exploratorische Faktorenanalyse im eigentlichen Sinne (Residualvarianzen werden geschätzt), welche Rotationsmethode verwendet wurde, welche Extraktionskriterien zugrunde gelegt wurden etc.). Der *Parahoo-Questionnaire* wurde im Rahmen einer österreichischen Studie übersetzt (Breimaier et al., 2011). Die übersetzte Version beinhaltet sieben offene und 14 geschlossene Fragen zu den Bereichen: a) Anwendung von Forschungswissen in der Praxis, b) Bedürfnisse, Wünsche, Forderungen der Pflegenden bzgl. Pflegewissenschaft, Pflegeforschung und Implementierung von Forschungsergebnissen in die Praxis, c) forschungsbezogene Inhalte in der Ausbildung der Pflegenden, d) Einstellungen der Pflegenden zur Pflegeforschung und deren Anwendung sowie e) Faktoren, die die Anwendung von Forschungswissen fördern oder behindern. Bei letzteren handelt es sich um offene Fragen aus der *BARRIERS scale* (Breimaier et al., 2011). Die Interne-Konsistenz-Reliabilität wurde ermittelt (Cronbach's α , Werte zwischen 0,69 und 0,86), Inhaltsvalidität wurde mittels Evaluation des übersetzten Fragebogens durch EBP-Experten sichergestellt und die Verständlichkeit des Bogens wurde in einem Pretest überprüft (54 Pflegende eines anderen Krankenhauses, keine Modifikationen erforderlich, Interviewdetails nicht berichtet) (Breimaier et al., 2011).

Um die Rolle des Organisationskontexts für die Anwendung von Forschungswissen in stationären Altenpflegeeinrichtungen besser zu verstehen – und um ggf. Ansatzpunkte zu finden, diesen Kontext gezielt zu beeinflussen, um EBP und Pflegequalität zu fördern –, sind u.a. methodisch hochwertige Studien erforderlich, die diese Fragen mit statistischen Methoden untersuchen. Dafür sind Erhebungsinstrumente nötig, die idealerweise a) auf Basis robuster DI-Theorie und -Evidenz sowie anerkannter Instrumentenentwicklungsmethoden konstruiert wurden, b) spezifisch potenziell modifizierbare Organisationskontextfaktoren des stationären Altenpflegebereichs erfassen, c) für verschiedene Berufsgruppen dieses Settings geeignet sind, d) sich in diesen Settings bewähren konnten und e) eine akzeptable psychometrische Güte aufweisen. Keines der beiden Instrumente erfüllt diese Kriterien in angemessener Weise. Beide deutschen Instrumentenversionen wurden bisher nur im Krankenhausbereich und ausschließlich mit examinieren Pflegenden angewendet, für keines der beiden deutschen Tools liegen Ergebnisse fortgeschrittener statistischer Validierungsuntersuchungen vor (z.B. Regressionsanalysen, konfirmatorische Faktorenanalysen, multivariate Modelle o.ä.) und keines der beiden Instrumente erfasst eine breite Palette DI-relevanter und potenziell modifizierbarer Organisationskontextfaktoren. Was die *BARRIERS scale* angeht, kommen Kajermo et al. (2010, 1) außerdem in ihrem systematischen Review zu dieser

Skala zu folgendem Schluss:

„The BARRIERS scale is a nonspecific tool for identifying general barriers to research utilization. The scale is reliable as reflected in assessments of internal consistency. The validity of the scale, however, is doubtful. There is no evidence that it is a useful tool for planning implementation interventions. We recommend that no further descriptive studies using the BARRIERS scale be undertaken. Barriers need to be measured specific to the particular context of implementation and the intended evidence to be implemented.“

Instrumente, die nicht-DI-spezifische Organisationskontextfaktoren erfassen, sind auf Deutsch zahlreicher. In ihrem Handbuch organisationspsychologischer Instrumente listen Sarges et al. (2010) 40 verschiedene Verfahren auf. Nur eines davon, das Tätigkeits- und Arbeitsanalyseverfahren (TAA) (Glaser et al., 2008), liegt in einer Version für die stationäre Altenpflege vor; alle anderen Verfahren sind nicht für Gesundheitssettings konzipiert. Auch das TAA Altenpflege ist jedoch eher unzureichend validiert: Interne-Konsistenz-Reliabilität (Cronbach's α , Werte zwischen 0,67 und 0,88) und bivariate Korrelationen der TAA-Skalen mit Skalen vergleichbarer Instrumente (nur fünf von 75 Korrelationen weisen Werte um 0,5 auf, alle anderen sind bedeutend kleiner; p-Werte sind nicht berichtet) (Glaser et al., 2008). Außerdem fokussiert das TAA – genau wie andere auf Deutsch erhältliche Kontextassessmentinstrumente für Gesundheitssettings (z.B. der revised Nursing Work Index, Aiken et al., 2001; Schubert et al., 2008) – eher Belastungsfaktoren als DI-spezifische Einflussfaktoren.

Daher wurden die verfügbaren englischsprachigen Instrumente näher analysiert. Es zeigte sich, dass das *Alberta Context Tool (ACT)* (Hutchinson et al., 2008; Estabrooks et al., 2009b; Squires et al., 2009; Squires et al., 2013c) die oben angesprochenen Merkmale nahezu optimal erfüllt (Details siehe Abschnitt 2.3.1). Da kein adäquates deutsches Instrument verfügbar war, zugleich mit dem ACT aber ein sehr vielversprechendes internationales Tool identifiziert werden konnte, fiel die Entscheidung, das ACT auf Deutsch zu übersetzen und umfassend psychometrisch zu untersuchen.

2.2.2 Research-Use-Instrumente

Auch Skalen zur Erfassung der Anwendung von Forschungswissen (im Folgenden kurz RU-Skalen) sind auf Deutsch rar. Zum Zeitpunkt der Konzeption dieser Arbeit (2010) konnte kein solches Instrument ausfindig gemacht werden. Inzwischen liegen – nach Wissen des Autors – drei RU-Skalen in deutscher Sprache vor: die RU-Skala des im vorangegangenen Abschnitt vorgestellten *Parahoo-Questionnaires* (Breimaier et al., 2011), eine von Schnittger et al. (2012) selbst entwickelter EBP-Fragebogen und ein Fragebogen von Köpke et al. (2013), basierend auf der Übersetzung und Adaptation internationaler Skalen sowie selbst entwickelter Items. Lediglich für den EBP-Fragebogen wurde – neben der Internen-Konsistenz-Reliabilität (Cronbach's α , Werte zwischen 0,72 und 0,90) – eine statistische Validitätsüberprüfung berichtet: „dimensionsreduzierende

Faktorenanalysen“, die auf eine Vier-Faktoren-Struktur hindeuteten (Schnittger et al., 2012). Auch hier sind die genauen methodischen Details unklar und Parameter der Faktoranalyse werden nicht berichtet. Köpke et al. (2013) evaluierten a) die Inhaltsvalidität des Fragebogens durch das Expertenurteil von vier Pflege- bzw. Gesundheitswissenschaftlerinnen, b) die *Response-Process-Validität* im Rahmen kognitiver Interviews mit vier Pflegenden sowie c) Praktikabilität und Verständlichkeit des Bogens in Form eines Pretests (21 Pflegenden zweier Stationen). Auch für diese Instrumente liegen somit unzureichende psychometrische Daten vor und Erfahrungen bestehen nur für die Anwendung mit Pflegenden im Krankenhausbereich.

Squires et al. (2011c) identifizierte in ihrem systematischen Review 60 RU-Skalen, die für eine Selbsteinschätzung durch berufliche Akteure des Gesundheitswesens konzipiert waren. 12 der Skalen wurden ausgeschlossen, da keine Validitätsdaten verfügbar waren. Nur eine der Skalen wurde a) in einem akzeptablen Rahmen validiert und b) erfolgreich im stationären Altenpflegebereich mit verschiedenen dort tätigen Akteuren angewendet (im Rahmen der *Translating Research in Elder Care (TREC)* Studie, Estabrooks et al., 2009c): *Estabrooks' Kinds of RU items* (Estabrooks, 1999a, 1999b). In der TREC-Studie kam außerdem eine weitere Skala zum Einsatz, die in diesem Rahmen entwickelt und validiert wurde: die *Conceptual Research Use (CRU) scale* (Squires et al., 2011b). Diese beiden Skalen wurden daher in die vorliegende Studie einbezogen.

2.3 Eigenschaften und Qualität der ausgewählten Instrumente

2.3.1 Das Alberta Context Tool

Das ACT wurde 2006 für den Krankenhausbereich zur Versorgung Erwachsener entwickelt (Estabrooks et al., 2008c; Estabrooks et al., 2009b) und später adaptiert für den pädiatrischen Krankenhausbereich (Hutchinson et al., 2008; Estabrooks et al., 2009b), die stationäre Altenpflege (Squires et al., 2009) und die ambulante Pflege (Squires et al., 2013c). Jede Version enthält spezifische Fragebögen für folgende Berufsgruppen: a) Pflegehilfskräfte, b) Pflegefachkräfte, c) Therapie- und Betreuungskräfte, d) klinische Experten (z.B. Qualitätsbeauftragte, Dozenten für Fort- und Weiterbildung oder *Advanced Practice Nurses*), e) Führungspersonen und f) Ärzte (Squires et al., 2013c). Die Fragebögen unterscheiden sich leicht bzgl. Anzahl der enthaltenen Items (in der Version für die stationäre Altenpflege 56-59), Struktur und Formulierung der Einleitungstexte und den verwendeten Beispielen (Estabrooks et al., 2011b). Die Items reflektieren zehn Konzepte (*Leadership, Culture, Evaluation, Social Capital, Informal Interactions, Formal Interactions, Structural/Electronic Resources* und *Organizational Slack* – letzteres repräsentiert durch die drei Subkonzepte *Staff, Space* und *Time*) (Estabrooks et al., 2011b). Die Originaldefinition der Konzepte und jeweils ein Beispielitem sind in Tab. 2 dargestellt.

Drei *Technical Reports* (Estabrooks et al., 2008c; Hutchinson et al., 2008; Squires et al., 2009) beschreiben ausführlich Entwicklung und Pilottests des ACT (Tab. 3). Zur Sicherung der Inhaltsvalidität bezogen die Entwicklerinnen Experten in verschiedenen Entwicklungsphasen ein und

Tab. 2: Konzepte des kanadischen ACT stationäre Altenpflege (Estabrooks et al., 2011b)

Konzept	Definition	Beispielitem
Leadership ¹	The actions of formal leaders in an organization (unit) to influence change and excellence in practice, items generally reflect emotionally intelligent leadership	The leader calmly handles stressful situations
Culture ¹	The way that “we do things” in our organizations and work units; items generally reflect a supportive work culture	My organization effectively balances best practice and productivity
Evaluation ¹	The process of using data to assess group/team performance and to achieve outcomes in organizations or units (i.e., evaluation)	Our team routinely monitors our performance with respect to the action plans
Social Capital ¹	The stock of active connections among people. These connections are of three types: bonding, bridging, and linking	People in the group share information with others in the group
Informal Interactions ²	Informal exchanges that occur between individuals working within an organization (unit) that can promote the transfer of knowledge	[How often do you interact with] people in the following roles or positions? - Someone who champions research and its use in practice
Formal Interactions ²	Formal exchanges that occur between individuals working within an organization (unit) through scheduled activities that can promote the transfer of knowledge	How often do these activities occur? - Team meetings
Structural/Electronic Resources ³	The structural and electronic elements of an organization (unit) that facilitate the ability to assess and use knowledge	How often do you use/attend the following? - Notice Boards
Organiz. Slack	The cushion of actual or potential resources which allows an organization (unit) to adapt successfully to internal pressures for adjustments or to external pressures for changes	
Staff ¹		Enough staff to deliver quality care
Space ¹		Use of designated space
Time ¹		Time to do something extra for residents

¹Skala 1: (1) strongly disagree, (2) disagree, (3) neither agree or disagree, (4) agree, (5) strongly agree

²Skala 2: (1) never, (2) rarely, (3) occasionally, (4) frequently, (5) almost always

³Skala 3: Wie Skala 1 + (6) not accessible

nutzen das *PARIHS-Framework* (Kitson et al., 1998; Rycroft-Malone et al., 2002) sowie zwei wichtige DIF-Reviews (Fleuren et al., 2004; Greenhalgh et al., 2004) als theoretische Basis (Squires et al., 2013c). Fokusgruppen (Krankenhaus Erwachsene) bzw. individuelles Feedback (stationäre Altenpflege) gaben Aufschluss über die Verständlichkeit der Items, Hauptkomponentenanalysen lieferten erste Einblicke in die Faktorenstruktur des ACT und mittels bivariater Korrelationen wurde der Zusammenhang der ACT-Skalen mit *Estabrooks' Kinds of RU items* (Estabrooks et al., 2008c; Hutchinson et al., 2008; Squires et al., 2009) analysiert.

Untersucht wurde auch die Interne-Konsistenz-Reliabilität (Cronbach's α): Sie betrug zwischen 0,65 und 0,92 in der Version Krankenhaus Erwachsene, wobei nur einer von acht Werten unter 0,70 lag (anfänglich war die Dimension „*Organizational Slack*“ noch nicht wie oben beschrieben ausdifferenziert in die Subkonzepte „*Staff*“, „*Space*“, „*Time*“, sondern wurde als ein Konstrukt behandelt) (Estabrooks et al., 2008c). In der Version Krankenhaus Pädiatrie lagen die Werte zwischen 0,48 und 0,90 mit drei Werten unter 0,70 (Hutchinson et al., 2008). Im ACT stationäre Altenpflege lagen die Werte zwischen 0,50 und 0,96 mit nur einem Wert <0.70 .

Tab. 3: Validierung der Instrumente im Rahmen der Pilottests

Version (Quelle)	Sample	Validität	Methoden (M)/Ergebnisse (E)
KH Erwachsene (Estabrooks et al., 2008c)	4 Krankenhäuser; 5 PFK, 2 TP, 4 Ärzte, 5 KE, 4 FK	Response Process	M: 4 Fokusgruppen E: Anpassung verschiedener Formulierungen
	4 Krankenhäuser; 152 PFK, 181 TP, 36 Ärzte, 46 KE, 38 FK	Interne Struktur	M: Hauptkomponentenanalyse, Varimaxro- tation mit Kaisernormalisierung, Extraktions- kriterium: Eigenwerte>1 E: 14-Faktoren-Struktur, 69,97% Varianzauf- klärung
		Beziehung zu anderen Variablen	M: Bivariate Korrelationen der ACT-Skalen mit RU E: Zahlreiche signifikante Korrelationen in vorhergesagter Richtung, allerdings alle <0,3
KH Pädiatrie (Hutchinson et al., 2008)	3 Krankenhäuser (je eine pädiatrische Sta- tion); 249 PFK, 67 TP, 20 Ärzte, 15 KE, 7 FK	Interne Struktur	M: Hauptkomponentenanalyse*, Varimaxro- tation mit Kaisernormalisierung, Extraktions- kriterium: Eigenwerte>1 E: Drei-Faktoren-Struktur, 61,84% Vari- anzaufklärung
		Beziehung zu anderen Variablen	M: Bivariate Korrelationen der ACT-Skalen mit RU E: Zahlreiche signifikante Korrelationen in vorhergesagter Richtung, allerdings alle <0,3
Stationäre Altenpflege (Squires et al., 2009)	1 Wohnbereich; 9 PHK	Response Process	M: Feedback der PHK E: Anpassung verschiedener Formulierungen
	4 Pflegeheime (je ein Wohnbereich für Men- schen mit Demenz); 73 PHK, 18 PFK	Interne Struktur	M: Hauptkomponentenanalyse*, Varimaxro- tation mit Kaisernormalisierung, Extraktions- kriterium: Eigenwerte>1 E: Drei-Faktoren-Struktur, 73,75% Vari- anzaufklärung
		Beziehung zu anderen Variablen	M: Bivariate Korrelationen der ACT-Skalen mit RU E: Nur wenige signifikante Korrelationen, RU-Konzepte nicht für PHK geeignet

FK: Führungskräfte, KE: klinische Experten, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, RU: Research Use, TP: Therapeuten
*nur die drei Konzepte *Leadership*, *Culture* und *Evaluation* wurden untersucht

Das ACT wurde konstruiert, um Kontextfaktoren auf Einrichtung- bzw. Stations-/Wohnbereichsebene zu erfassen. Zu diesem Zweck müssen die individuellen Scores aggregiert werden. Ob diese Aggregation gerechtfertigt ist, überprüften die Entwicklerinnen mit einfaktoriellen Varianzanalysen (*analysis of variance*, ANOVA) mit Institutions- bzw. Stations-/Wohnbereichszugehörigkeit als Faktorvariable (Estabrooks et al., 2008c; Hutchinson et al., 2008; Squires et al., 2009). Auf Basis der ANOVA-Parameter wurden zwei Interclusterkorrelationskoeffizienten (ICC1 und ICC2) sowie die Parameter η^2 und ω^2 berechnet, um die Aggregierbarkeit der Daten zu evaluieren (vgl. Abb. 4 für die Erläuterung der Parameter). Beim Pilottest des ACT Krankenhaus Erwachsene wurden die Scores auf Einrichtungsebene aggregiert. Dort rangierten die ICC1-Werte zwischen 0,000 und 0,007 (keine der ANOVAs war signifikant), alle ICC2-Werte waren <0,6 und die ω^2 -Werte verwiesen auf sehr geringe Effektstärken (Estabrooks et al., 2008c). Dies konnte im Folgebericht (Krankenhaus Pädiatrie, Hutchinson et al., 2008) verbessert werden, in-

dem auf Ebene der Stationen aggregiert wurde. Die ANOVAs für fünf der acht ACT-Scores waren signifikant ($p < 0,05$) mit ICC1-Werten zwischen 0,039 und 0,127, ICC2-Werten zwischen 0,802 und 0,939, durch die Clusterebene erklärten Varianzen (η^2) von 3,1% bis 8,9% und ω^2 -Werten von 0,025 bis 0,083. Auch im Pilottest der stationären Altenpflegeversion (Squires et al., 2009) wurde auf Wohnbereichsebene aggregiert. ANOVAs für sechs der acht ACT-Scores waren signifikant ($p < 0,05$), mit ICC1-Werten von 0,140 bis 0,584, ICC2-Werten zwischen 0,706 und 0,967, η^2 -Werten zwischen 14,6% und 53,1% und ω^2 -Werten von 0,109 bis 0,511.

Interclass Correlation 1 (ICC1): Index für die Clusterung der Werte von Individuen (ACT-Scores) einer Gruppe (einer Einrichtung oder eines Wohnbereichs), bezogen auf den Gruppenmittelwert; Werte reichen von 0 (keinerlei Clusterung) bis 1 (perfekte Übereinstimmung innerhalb der Gruppen); Ab Werten $> 0,1$ gelten die Übereinstimmungen innerhalb der Gruppen als stark.

Formel: $ICC1 = (BMS - WMS) / (BMS + [K - 1] \times WMS)$

Interclass Correlation 2 (ICC2): Index für die Stabilität der aggregierten Daten innerhalb der Gruppe (durchschnittliche Reliabilität der aggregierten Daten); reicht ebenfalls von 0 bis 1; Bei Werten $> 0,6$ gilt die Aggregation der Daten als angemessen.

Formel: $ICC2 = (BMS - WMS) / (BMS + [K - 1] \times WMS)$

η^2 : Anteil der durch die Gruppenzugehörigkeit erklärten Varianz der interessierenden ACT-Variablen

Formel: $\eta^2 = SSB / SST$

ω^2 : Indikator für die Effektstärke (relative Stärke der aggregierten Variablen auf Subgruppenebene); $\omega^2 < 0,06$: geringer/kein Effekt, $0,06 < \omega^2 < 0,15$: mittlerer Effekt, $\omega^2 > 0,15$: großer Effekt

Formel: $\omega^2 = (SSB - [N - 1] \times WMS) / (SST + WMS)$

Legende:
 BMS = *between-group mean square* (Mittlere Summe der Abweichungsquadrate zwischen den Gruppen); K = durchschnittliche Zahl der Individuen pro Gruppe; N = Anzahl der Gruppen; SSB = *sum of squares between groups* (Summe der Abweichungsquadrate zwischen den Gruppen); SST = *sum of squares total* (Gesamtsumme der Abweichungsquadrate); WMS = *within-group mean square* (Mittlere Summe der Abweichungsquadrate innerhalb der Gruppen)

Abb. 4: Parameter zur Evaluation der Aggregierbarkeit der ACT-Scores: Definition und Erläuterung (Shrout and Fleiss, 1979; Estabrooks et al., 2008c; Bortz und Schuster, 2010)

Aufbauend auf den Pilottests, die zu weiteren Modifikationen führten, wurden weitere Validierungen in größeren Stichproben vorgenommen. Fünf Publikationen (Estabrooks et al., 2009b; Cummings et al., 2010; Estabrooks et al., 2011c; Mallidou et al., 2011; Squires et al., 2013b) liegen zur Validierung des ACT Krankenhaus Pädiatrie vor, zwei Artikel (Estabrooks et al., 2011a; Estabrooks et al., 2011b) für das ACT stationäre Altenpflege.

Die Studien aus dem Pädiatriebereich (Tab. 4) beziehen sich alle auf Pflegefachkraftsamples, nur eine (Mallidou et al., 2011) berichtet eine Teilanalyse von Daten des Therapeutenfragebogens (Hauptkomponentenanalyse für die drei *Organizational Slack* Konzepte, bestätigte die angenommene Drei-Faktoren-Struktur). Eine Hauptkomponentenanalyse für das Pflegefachkraftsample ergab eine 13-Faktoren-Struktur (*Informal Interactions* wurde aufgeteilt in zwei, *Structural/Electronic Resources* in drei Konzepte) (Estabrooks et al., 2009b). Die Reliabilität (Cronbach's α)

war akzeptabel ($>0,7$) für neun ACT-Konzepte. Nur die Reliabilität für *Formal Interactions* (0,60), *Structural/Electronic Resources* 2 und 3 (0,60 bzw. 0,54) sowie *Space* (0,63) war geringer. Zwei Studien (Estabrooks et al., 2009b; Squires et al., 2013b) berichten Analysen zur Aggregierbarkeit der Daten auf Wohnbereichsebene – mit guten Ergebnissen (ANOVAs signifikant und ICC1-, ICC2-, η^2 - und ω^2 -Werte überwiegend im akzeptablen Bereich). Deskriptive Vergleiche der ACT-Scores (gruppiert nach Güte des Kontexts, hoch, mäßig hoch, mäßig niedrig und niedrig) mit RU-Scores ließen erkennen, dass Forschungsanwendung mit zunehmender Kontextgüte steigt (Cummings et al., 2010). Bivariate Korrelationen zwischen ACT-Scores und Forschungsanwendung weisen auf signifikante, wenngleich geringe Zusammenhänge hin (Korrelationen zwischen 0,090, *Leadership* und 0,242, *Structural/Electronic Resources*), und auch ANOVAs der ACT-Scores mit Forschungsanwendung als Faktor ergaben signifikante Zusammenhänge (Estabrooks et al., 2009b). Mehrebenenmodelle wurden in zwei pädiatrischen Studien (Estabrooks et al., 2011c; Squires et al., 2013b) eingesetzt. Estabrooks et al. (2011c) untersuchten mittels hierarchisch-linearer Modelle, a) ob die ACT-Scores zwischen verschiedenen Stationen diskriminieren und b) welche Variablen (Level 1: Variablen auf Ebene der Pflegenden, z.B. Arbeitsumfang, Bildung, Arbeitszufriedenheit, ..., Level 2: Merkmale der Station, z.B. Spezialisierung, Größe, Belegung, ...) die Varianz zwischen den Stationen erklären. Die Nullmodelle wie auch Modelle, die Variablen der Pflegenden kontrollierten, ergaben, dass alle ACT-Scores signifikant ($p < 0,01$) zwischen den Stationen variierten. Stationsvariablen erklärten zusätzliche Varianz (15,91%-73,91%) für sieben der zehn ACT-Konzepte (Estabrooks et al., 2011c). Squires et al. (2013b) nutzten *General Estimating Equations* (GEE), um den Einfluss verschiedener Prädiktoren (Level 1: Variablen der Pflegenden, Level 2: ACT-Scores, weitere Stationsmerkmale) auf zwei Formen von Forschungsanwendung zu untersuchen: *Instrumental Research Use* (IRU): Umsetzung von Forschungswissen in Alltagshandeln und *Conceptual Research Use* (CRU): Forschungswissen verändert Denken/Einstellung, nicht aber zwingend Handeln. *Culture* war der einzige signifikante ACT-Prädiktor für IRU. Signifikante Prädiktoren für CRU waren *Leadership*, *Culture*, *Evaluation*, *Formal & Informal Interactions* und *Space* – *Culture* und *Formal Interactions* jedoch mit negativen Assoziationen ($\beta = -0,654$ bzw. $-0,342$), was der Hypothese widerspricht, höhere Kontextscores gingen mit höheren RU-Scores einher. Die Autorinnen empfehlen, diesen überraschenden Befund mittels qualitativer Methoden näher zu untersuchen (Squires et al., 2013b).

Studien zur Validierung des ACT stationäre Altenpflege (Estabrooks et al., 2011a; Estabrooks et al., 2011b) berücksichtigen bislang nur Pflegehilfskraftdaten (Tab. 5). Die interne Struktur des Tools wurde mittels konfirmatorischer Faktorenanalysen untersucht (Estabrooks et al., 2011b). Ein Zehn-Faktoren-Modell (Modell a) aller ACT-Konstrukte wies aus Sicht der Forscherinnen keinen akzeptablen Fit auf: signifikanter χ^2 -Wert und signifikante standardisierte Residuen für Kovarianzen von vier *Structural/Electronic-Resources*-Items sowie einem *Space*-Item mit den restlichen Items der jeweiligen Konzepte. Die misfitproduzierenden Items wurden entfernt

Tab. 4: Studien zur Validierung des ACT Krankenhaus Pädiatrie

Quelle	Sample	Reliabilität	Interne Struktur	Zusammenhänge	Aggregation
Estabrooks et al. (2009b)	7 pädiatrische Krankenhäuser in 6 kanadischen Provinzen; 752 PFK	Cronbach's α 4 Werte <0,7, alle anderen zwischen 0,70 und 0,91	Hauptkomponentenanalyse, Varimaxrotation mit Kaiser-Normalisierung, Extraktionskriterium: Eigenwerte >1, Scree-Test 13-Faktoren-Struktur, 59,26% Varianzaufklärung	Bivariate Korrelationen zwischen ACT-Scores und IRU, ANOVAs der ACT-Scores mit IRU als Faktor Alle ACT-Scores korrelieren signifikant*** mit IRU (jedoch alle Korrelationen <0,3) Zunehmende ACT-Scores gehen mit zunehmender IRU einher (alle ANOVAs signifikant**)	ANOVAs, Faktor: Station, Berechnung von ICC1, ICC2, η^2 , ω^2 Alle ANOVAs signifikant*** ICC1: 0,041-0,251, ICC2: 0,535-0,900, η^2 : 0,075-0,271, ω^2 : 0,057-0,244
Mallidou et al. (2011) (nur <i>Staff, Space, Time</i>)	Wie oben plus 194 TP	Nur für PFK-Sample berichtet (s.o.)	Methoden/Ergebnisse PFK s.o. TP-Sample. 3-Faktoren-Struktur, 65,37% Varianzaufklärung	Nur für PFK-Sample berichtet (s.o.)	Nur für PFK-Sample berichtet (s.o.)
Cummings et al. (2010) (nur <i>Culture, Leadership, Evaluation</i>)	2 Universitätskrankenhäuser; 248 PFK	—	—	Vergleich hoher, mäßig hoher, mäßig niedriger und niedriger Kontextscores mit IRU- und CRU-Scores (Tabellen und Grafiken mit 95%CI) Höhere Kontextscores gehen einher mit höheren RU-Scores	—
Estabrooks et al. (2011c)	32 Stationen in 8 pädiatrischen Krankenhäusern; 844 PFK	—	—	HLM (<i>random intercept, fixed effects</i>); 40 Modelle (4 je ACT-Konzept) wurden in einem schrittweisen Verfahren geschätzt Das ACT diskriminiert trennscharf zwischen verschiedenen Stationen	—
Squires et al. (2013b)	32 Stationen in 8 pädiatrischen Krankenhäusern; 735 PFK	—	—	GEE-Modelle; Zusammenhang der Stationsaggregierten ACT-Scores sowie zahlreicher individueller Variablen mit IRU und CRU Prädiktor für IRU: <i>Culture</i> * Prädiktoren für CRU: <i>Leadership</i> ***, <i>Culture</i> *, <i>Evaluation</i> ***, <i>Formal & Informal Interactions</i> ***, <i>Space</i> *	ANOVAs, Faktor: Station, Berechnung von ICC1, ICC2, η^2 , ω^2 Alle ANOVAs signifikant*** ICC1: 0,052-0,318, ICC2: 0,556-0,913, η^2 : 0,091-0,340, ω^2 : 0,051-0,310

Signifikante Unterschiede: *p < 0,05 **p < 0,01 ***p < 0,001

95%CI: 95%-Konfidenzintervall, ANOVA: *Analysis of Variance*, CRU: *Conceptual Research Use*, GEE: *General Estimating Equation*, ICC: *Interclass Correlation*, IRU: *Instrumental Research Use*, HLM: Hierarchisch-lineare Modelle, PFK: Pflegefachkräfte, TP: Therapeuten

Tab. 5: Studien zur Validierung des ACT stationäre Altenpflege

Quelle	Sample	Reliabilität	Interne Struktur	Zusammenhänge	Aggregation
Estabrooks et al. (2011b)	25 Pflegeheime; 645 PHK	Cronbach's α 2 Werte <0,7, alle anderen zwischen 0,70 und 0,92	Konfirmatorische Faktorenanalysen, 3 Modelle: a) alle 10 Faktoren, b) nur Items mit Zustimmungsskala ¹ (7 Fak- toren), c) nur Items mit Häufigkeits- skala ² (3 Faktoren) Modell a): χ^2 : 4674*** (df: 1550), RMSEA: 0,06, SRMR: 0,07, CFI: 0,91 Modell b): χ^2 : 933*** (df: 474), RMSEA: 0,04, SRMR: 0,04, CFI: 0,98 Modell c): χ^2 : 828*** (df: 176), RMSEA: 0,09, SRMR: 0,08, CFI: 0,85	Bivariate Korrelationen zwischen ACT- Scores und IRU, ANOVAs der ACT- Scores mit IRU als Faktor 8 der 10 ACT-Scores korrelieren signi- fikant** mit IRU (jedoch alle Korrelati- onen <0,2) Zunehmende ACT-Scores gehen mit zunehmender IRU einher (ANOVAs für folgende Konzepte signifikant: <i>Leader- ship**</i> , <i>Culture***</i> , <i>Evaluation*</i> , <i>Social Capital***</i> , <i>Informal Interactions**</i> , <i>Formal Interactions**</i> , <i>Structural/Elec- tronic Resources**</i> , <i>Time***</i>)	ANOVAs, Faktor: a) Einrichtung, b) Wohnbereich, Berechnung von ICC1, ICC2, η^2 , ω^2 a) 8 v. 10 ANOVAs signifikant*** (nicht: <i>Formal & Informal Interac- tions</i>), ICC1: 0,055-0,323, ICC2: 0,599-0,924, η^2 : 0,089-0,339, ω^2 : 0,053-0,313 a) 8 von 10 ANOVAs signifikant ^{2x**} , ^{6x***} (nicht: <i>Formal & Informal Inter- actions</i>), ICC1: 0,062-0,362, ICC2: 0,322-0,804, η^2 : 0,191-0,448, ω^2 : 0,061-0,360
Estabrooks et al. (2011a)	89 Wohnbereiche in 25 Pflegeheimen; 1258 PHK	—	—	2- und 3-Level-HLM; je drei Nullmo- delle pro ACT-Konzept: a) Wohnbe- reich und b) Einrichtung als Level-2- Cluster sowie c) Wohnbereich als Le- vel-2 und Einrichtung als Level-3-Clus- ter 2-Level-Modelle: Mehr erklärte Vari- anz der ACT-Konzepte durch Wohnbe- reich (2,15%-34,45%)* als durch Ein- richtung (1,97%-31,74%)* 3-Level-Modelle erklärten 2,36%- 35,05%* Varianz auf Wohnbereichsebene für alle ACT-Konzepte signifikant*** >0 (auf Einrichtungsebene nur für <i>Leader- ship</i> , <i>Staff</i> , <i>Space</i>)	ANOVAs, Faktor: a) Einrichtung, b) Wohnbereich, Berechnung von ICC1, ICC2, η^2 , ω^2 a) ANOVAs signifikant ^{1x***, 9x***} , ICC1: 0,021-0,317, ICC2: 0,520-0,958, η^2 : 0,039-0,318, ω^2 : 0,021-0,305 a) 9 v. 10 ANOVAs signifikant ^{1x***} , ^{8x***} (nicht: <i>Informal Interactions</i>), ICC1: 0,021-0,328, ICC2: 0,235-0,873, η^2 : 0,091-0,369, ω^2 : 0,021-0,325

Signifikante Unterschiede: *p < 0,05 **p < 0,01 ***p < 0,001

¹Skala 1: (1) strongly disagree, (2) disagree, (3) neither agree or disagree, (4) agree, (5) strongly agree ((6) not accessible), ²Skala 2: (1) never, (2) rarely, (3) occasionally, (4) frequently, (5) almost always

ANOVA: Analysis of Variance, CFI: Comparative Fit Index, HLM: Hierarchisch-lineare Modelle, ICC: Interclass Correlation, IRU: Instrumental Research Use, PHK: Pflegehilfskräfte, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual

Da die zustimmungsskalierten Items (*Leadership, Culture, Evaluation, Social Capital, Staff, Space, Time*) durchgehend deutlich höhere Ladungen aufwiesen als die häufigkeitsskalierten Items (*Formal & Informal Interactions* und *Structural/Electronic Resources*), teilten die Forscherinnen das Instrument und schätzten zwei separate Modelle (Modelle b und c). Während sich die Modellfitparameter für Modell b (verglichen mit Modell a) substantiell verbesserten, verschlechterten sie sich für Modell c (Estabrooks et al., 2011b). Die Reliabilitätswerte (Cronbach's α) waren auch in dieser Untersuchung überwiegend akzeptabel ($>0,7$ für acht Konzepte, 0,37 für *Formal Interactions*, 0,64 für *Space*) (Estabrooks et al., 2011b). Ähnlich wie in den Studien im Pädiatriebereich wurden auch hier bivariate Korrelationen zwischen ACT-Scores und Forschungsanwendung untersucht sowie ANOVAs der ACT-Scores mit Forschungsanwendung als Faktor durchgeführt. Bis auf zwei ACT-Scores (*Staff* und *Space*) sind Korrelationen und ANOVAs signifikant, wobei auch hier die Korrelationen eher gering sind (0,111 *Leadership* bis 0,199 *Time*) (Estabrooks et al., 2011b). Auch die Aggregierbarkeit der Daten (sowohl auf Einrichtungs- als auch auf Wohnbereichsebene) wurde untersucht (Estabrooks et al., 2011a; Estabrooks et al., 2011b) – mit ähnlichen Ergebnissen wie im Pädiatriebereich (Details siehe Tab. 5). Estabrooks et al. (2011a) untersuchten zudem mittels hierarchisch-linearer Modelle, welcher Anteil der ACT-Score-Varianzen durch die Wohnbereichs- sowie die Einrichtungsebenen erklärt wird. Für jedes der ACT-Konzepte schätzten sie je drei Nullmodelle (Modelle ohne Prädiktoren) – a) Level 1: Individuum, Level 2: Wohnbereich, b) Level 1: Individuum, Level 2: Einrichtung und c) Level 1: Individuum, Level 2: Wohnbereich, Level 3: Einrichtung. Von den Zwei-Level-Modellen erklärten die Wohnbereichsmodelle durchgehen mehr Varianz als die Einrichtungsmodelle (Tab. 5). Verglichen mit dem Wohnbereichsmodell erklärte das Drei-Level-Modell nur geringfügig mehr Varianz. Die geschätzte Varianz war in den Wohnbereichsmodellen für alle ACT-Scores signifikant ungleich Null, während dies im Zwei-Level-Einrichtungsmodell nur für *Leadership, Staff* und *Space* zutraf (Estabrooks et al., 2011a).

Das englischsprachige Originalinstrument wurde übersetzt auf Niederländisch, Schwedisch, Chinesisch (Mandarin) und Französisch (für den französischsprachigen Teil Kanadas) und das ACT ist in Kanada, den USA, Schweden, den Niederlanden, dem Vereinigten Königreich, Irland, Australien sowie China im Einsatz (Squires et al., 2013c). Details zum Übersetzungsprozess und zum übersetzten Instrument sind bisher nur für die schwedische Version (Pflegefachkraftbogen ACT stationäre Altenpflege, Eldh et al., 2013) publiziert. Diese Version weist Cronbach's α -Werte zwischen 0,63 und 0,87 auf (*Culture* und *Informal Interactions* 0,63, alle anderen $>0,7$). Die *Response-Process-Validität* wurde in zwei Samples untersucht. Zum einen evaluierten vier Pflegende das Tool während des Übersetzungsprozesses, zum anderen füllten 159 Pflegende das Schwedische ACT aus und bewerteten, in wie weit die Fragen a) einfach zu beantworten (52% Zustimmung), b) relevant (65% Zustimmung) und c) klar formuliert waren (52% waren diesbe-

züglich ambivalent) (Eldh et al., 2013).

Das ACT kann angesichts dieser umfangreichen Befunde als vergleichsweise gut validiert gelten. Allerdings liegen bisher nur Ergebnisse für Pflegefachkräfte und (sehr eingeschränkt) für Therapeuten des Pädiatriebereichs sowie für Pflegehilfskräfte der stationären Altenpflege vor. Berufsgruppenvergleiche wurden bislang nicht angestellt und Validitätsergebnisse für übersetzte Versionen sind (außer den oben referierten) nicht publiziert. Die Befunde zur Internen-Konsistenz-Reliabilität sind akzeptabel, bivariate Korrelationen und ANOVAs weisen, wie angenommen, auf Zusammenhänge zwischen ACT und Forschungsanwendung hin und beobachtete ACT-Scores sind (gemessen an den ANOVA-Parametern) valide und reliabel auf Clusterebene aggregierbar – wobei, wie angenommen, insbesondere die Wohnbereichsebene bedeutsam ist. Die Befunde zur Faktorenstruktur sind nicht eindeutig. Während die Pädiatrieversion für Pflegefachkräfte auf Basis von Hauptkomponentenanalysen 13 Faktoren aufweist, wurde im Pflegehilfskraftsample der stationären Altenpflege ein 10-Faktoren-Modell confirmatorisch untersucht, verworfen und anschließend ein Sieben- und ein Drei-Faktorenmodell separat geschätzt. Werden Zusammenhänge zwischen ACT und RU-Skalen mit multivariaten Methoden untersucht (GEE) entsprechen die Befunde z.T. auch nicht mehr den Hypothesen. Zwar sind einige signifikante Zusammenhänge zu verzeichnen, jedoch in anderer Richtung als angenommen: Höhere (=bessere) *Culture-* und *Formal-Interaction-Scores* scheinen demnach Forschungsanwendung zu reduzieren. Ein Studienprotokoll für die umfangreiche weitere Validierung mittels fortgeschrittener Methoden (z.B. Strukturgleichungs- und *Item-Response-Modelle*) und auf Basis verschiedener Berufsgruppensamples liegt vor (Squires et al., 2013c).

2.3.2 Estabrooks‘ Kinds of Research Use items

Dieses Instrument wurde von Estabrooks (1999a, 1999b) entwickelt. In einer kritischen Auseinandersetzung mit der theoretischen und empirischen Literatur definierte und operationalisierte Estabrooks vier Formen der Forschungsanwendung: a) *instrumental* oder *direct research use* (IRU), b) *conceptual* oder *indirect research use* (CRU), *persuasive* oder *symbolic research use* (PRU) sowie *overall research use* (ORU). IRU beschreibt dabei die Anwendung von Forschungswissen im Praxisalltag, z.B. indem Pflegende den Empfehlungen einer Praxisleitlinie folgen und sturzgefährdeten Bewohnern Hüftprotektoren anlegen. Im Unterschied zur IRU muss CRU nicht zwingend mit einer Handlung einhergehen. Forschungswissen beeinflusst hier primär Denken, Einstellung und Haltung bzw. führt zu neuen Ideen. Im Rahmen von PRU wird Forschungswissen dazu genutzt, um das eigene Handeln zu legitimieren oder um andere Personen davon zu überzeugen, ihre Praxis zu verändern. ORU ist ein übergreifendes Konstrukt und meint jegliche Form der Anwendung von Forschungswissen (Estabrooks, 1998; Estabrooks, 1999a; Estabrooks et al., 2003b). *Estabrooks‘ Kinds of RU items* erfassen jedes dieser vier Konzepte mit einem Einzelitem.

Die jeweilige Form der Forschungsnutzung wird zunächst definiert sowie mit Beispielen veranschaulicht. Daran schließt sich die Frage an, wie oft die befragte Person in der Vergangenheit Forschungswissen auf die zuvor erläuterte Weise genutzt hat. Während die ursprüngliche Instrumentenversion nach den letzten 12 Monaten („*in the past year*“) fragt (Estabrooks, 1999a), bezieht sich die für die stationäre Altenpflege adaptierte Version (Estabrooks et al., 2009c), die im Rahmen dieser Studie genutzt wurde, auf den letzten typischen Arbeitstag („*last typical day of work*“). In der Ursprungsversion (entwickelt für Pflegefachkräfte) wurden diese Fragen anhand einer siebenstufigen Likert-Skala (von (1) *never* bis (7) *nearly every shift*) eingeschätzt. Die hier verwendete Version ist erhältlich für die gleichen Berufsgruppen wie das ACT stationäre Langzeitpflege. Pflegehilfskräfte bewerten die Fragen anhand der Skala: (1) *never*, (2) *rarely*, (3) *occasionally*, (4) *frequently*, (5) *almost always*. Für die anderen Berufsgruppen wird eine andere Skala verwendet: (1) *10% or less of the time*, (2) *about 25% of the time*, (3) *about 50% of the time*, (4) *about 75% of the time*, (5) *almost 100% of the time*. Die vier RU-Konzepte werden als Einzelkonstrukte behandelt, d.h. es wird kein Summenscore oder Mittelwert aus den vier Einzelwerten gebildet (Squires et al., 2011c).

Über die psychometrische Güte des Instruments gibt ein systematisches Review von Squires et al. (2011c) Auskunft. In dieses wurden alle Studien einbezogen, die eine Selbsteinschätzungsskala zur Forschungsnutzung im Gesundheitswesen in irgendeiner Form validierten. Die Forscherinnen identifizierten zehn Studien (acht Forschungsprojekte), die Hinweise zur Güte von *Estabrooks' Kinds of RU items* gaben (Details siehe Tab. 6, in der die Studie von Cobban and Profetto-McGrath (2008) nicht enthalten ist, da es sich um einen für diese Arbeit nicht relevanten Pretest des Tools mit *dental hygienists* handelt). Eine Studie (Estabrooks, 1999a) berichtet den Entwicklungsprozess und gibt Auskunft über die Inhaltsvalidität des Instruments. Dies wurde sichergestellt durch eine intensive und kritische Auseinandersetzung mit der theoretischen und empirischen Literatur zum Thema. Zwei erfahrene Pflegewissenschaftlerinnen evaluierten das Instrument, das auf dieser Basis modifiziert wurde. Das Antwortverhalten der Zielpersonen sowie die Verständlichkeit des Instruments (*response processes*) wurden in drei Studien (Estabrooks, 1999a; Connor, 2006; Estabrooks et al., 2008b) evaluiert. Alle Studien untersuchten in irgendeiner Form Zusammenhänge zwischen verschiedenen Einflussvariablen und den vier RU-Items. Eine der Studien (Estabrooks, 1999a) untersuchte zudem den Zusammenhang zwischen den vier RU-Konstrukten mittels Strukturgleichungsmodellen. Ein reflektives Modell (IRU, CRU & PRU sind Indikatoren von ORU, werden also von ORU beeinflusst) wies einen deutlich schlechteren Modellfit auf als ein formatives Modell (IRU, CRU und PRU sind Prädiktoren für ORU) (Tab. 6). Signifikante Prädiktoren für die Anwendung von Forschungswissen waren in den Strukturgleichungsmodellen betriebsinterne Fort- und Weiterbildungen sowie die Einstellung zu Forschungswissen (Estabrooks, 1999a, 1999b). Zwei Studien (Profetto-McGrath et al., 2003; Profetto-McGrath et al., 2009) berichten bivariate Korrelationen zwischen den vier RU-Scores und der Eigen-

Tab. 6: Studien zur Validierung von *Estabrooks' Kinds of RU items*

Quelle	Sample	Inhalt	Response Processes	Zusammenhänge
Estabrooks (1999a, 1999b)	600 PFK aus verschiedenen Settings (Zufallssample aller in Alberta registrierter PFK)	Konzeption der Items auf Basis theoretischer und empirischer Literatur; vorherige Ansätze wurde kritisch reflektiert Review des Tools durch zwei erfahrene Wissenschaftler	Pretest mit drei Pflegedoktoranden, sowie anschließend mit 23 PFK mit Bachelorabschluss bzw. im Masterstudium	Strukturgleichungsmodelle, um den Zusammenhang der vier RU-Konzepte zu bestimmen (Modell a und b) sowie verschiedene individuelle Prädiktoren für RU (Modell c) Modell a (reflektives Modell): χ^2 : 172,39*** (df: 12), AGF: 0,660 Modell b (formatives Modell): χ^2 : 4,62 (df: 8), p=0,797, AGF: 0,980 Modell c (Einflussfaktoren): χ^2 : 55,91 (df: 50), p=0,0263, AGF: 0,956; signifikante Prädiktoren: betriebsinterne Fort- und Weiterbildungen sowie Einstellung zu Forschungswissen (keine Koeffizienten berichtet)
Kenny (2005)	3 Militärkrankenhäuser (USA); 130 zivile PFK und 160 PFK mit Armeezugehörigkeit	—	—	Regressionsmodelle (Prädiktoren mit p<0,1 für die Regressionskoeffizienten werden berichtet, nicht jedoch die Regressionskoeffizienten selbst) Prädiktoren: <u>IRU</u> : <i>Belief Suspension</i> ¹ , Einstellung zu Forschungswissen, Zugriff auf Forschungswissen, Support durch die Organisation, Zahl der gelesenen Pflegezeitschriften <u>CRU</u> : <i>Belief Suspension</i> ¹ , Merkmale des Forschungswissens <u>PRU</u> : Zugriff auf Forschungswissen, Bewertung der Relevanz des Zugriffs auf Forschungswissen, Vertrauen in Forschungswissen, Merkmale des Forschungswissens <u>ORU</u> : <i>Belief Suspension</i> ¹ , Innovativität der Einr., Interesse, Respekt vor Regeln
Estabrooks et al. (2007)	Samples aus Estabrooks (1999a, 1999b) und Kenny (2005) kombiniert	—	—	Kumulative Logit-Modelle, getrennt für das kanadische Zivillisten-sample und das US-Armysample Prädiktoren: (OR, 95%CI) Zivillisten: <u>IRU</u> : Zeit (1,25, 1,11-1,41), Einstellung zu Forschungswissen (1,17, 1,09-1,25), <u>ORU</u> : Forschungs-Champion vorhanden (1,47, 1,03-2,10), Zeit (1,20, 1,07-1,35), Zahl der Fort- und Weiterbildungen (1,03, 1,01-1,06), Einstellung zu Forschungswissen (1,21, 1,13-1,30), <i>Belief Suspension</i> ¹ (1,07, 1,03-1,11) Army: <u>IRU</u> : <i>Belief Suspension</i> ¹ (1,11, 1,05-1,17), <u>ORU</u> : Einstellung zu Forschungswissen (1,16, 1,06-1,26), <i>Belief Suspension</i> ¹ (1,08, 1,02-1,14)
Profetto-McGrath et al. (2009)	389 PFK (<i>clinical educators</i>) (Zufallssample aller in Alberta registrierter <i>clinical educator</i> PFK)	—	—	Bivariate Korrelationen zwischen IRU, CRU, PRU, ORU und kritischem Denken insgesamt sowie sieben Dispositionen für kritisches Denken Kritisches Denken war signifikant assoziiert mit allen vier RU-Konzepten (Korrelationen zwischen 0,146* und 0,237**)

Tab. 6: Fortsetzung

Quelle	Sample	Inhalt	Response Processes	Zusammenhänge
Milner et al. (2005)	389 PFK (161 <i>staff nurses</i> , 145 <i>managers</i> , 82 <i>clinical educators</i>) aus verschiedenen Settings (Zufallssample aller in Alberta registrierter PFK)	—	—	ANOVAs mit PFK-Rolle als Faktor und Regressionsmodelle für das Gesamtsample Lt. ANOVAs Scores aller vier RU-Konzepte bei <i>clinical educators</i> signifikant** höher als bei <i>staff nurses</i> und <i>managers</i> Prädiktoren (β): <u>IRU</u> : Einstellung zu Forschungswissen (0,120**), <i>Awareness</i> ² (0,037*), Involvement in Forschung (0,142*) <u>CRU</u> : <i>Localite</i> ³ (0,031*) <u>PRU</u> : <i>Awareness</i> ² (0,076*), Nutzung von Massenmedien (Fernsehen, Internet, ...) (0,194**), Involvement in Forschung (0,176*) <u>ORU</u> : Einstellung zu Forschungswissen (0,098**), <i>Awareness</i> ² (0,063**), Involvement in Forschung (0,170**)
Connor (2006)	39 PFK(RN), 31 PFK (LPN) und 73 PHK aus 12 Pflegeheimen	—	Pilotstudie mit je 6 Personen der drei Berufsgruppen; für die RU-Skalen ergab sich kein Modifikationsbedarf	Regressionsmodelle (Prädiktoren mit $p < 0,1$ werden berichtet) Prädiktoren (β): PFK(RN): <u>ORU</u> : Zugriff auf Forschungswissen (0,054), Einstellung zu Forschungswissen (0,117) PFK(LP): <u>IRU</u> : Zahl der Fort- und Weiterbildungen (0,237), <u>PRU</u> : Zugriff auf Forschungswissen (0,122), Support (0,194) PHK: <u>IRU/CRU/ORU</u> : Einstellung zu Forschungswissen (0,070/0,114/0,224), <u>PRU</u> : Zugriff auf Forschungswissen (0,059)
Estabrooks et al. (2008b) Profetto-McGrath et al. (2003)	4 Universitätskrankenhäuser, 2 chirurgische (Erwachsene) und 5 pädiatrische Stationen; 141 PFK	—	<u>Estabrooks et al. (2008b)</u> : Pflegenden wurden 3x nach ORU gefragt. Erster Score im Vergleich zum zweiten signifikant*** höher, dritter signifikant* höher als der zweite. => Lerneffekt im Laufe der Fragebogenbearbeitung	<u>Profetto-McGrath et al. (2003)</u> : Bivariate Korrelationen zwischen IRU, CRU, PRU, ORU und kritischem Denken insgesamt sowie sieben Dispositionen für kritisches Denken Kritisches Denken war signifikant assoziiert mit IRU, CRU und ORU (Korrelationen zwischen 0,24** und 0,35**)

Signifikante Unterschiede: * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$

¹Fähigkeit, bisherige Überzeugungen aufzugeben, um neue Erkenntnisse anzuwenden, ²Häufigkeit, mit der Teilnehmer auf Informationen mit hoher wissenschaftlicher Evidenz zugreift, ³Nutzung lokaler Informationsquellen (z.B. Kollegen)
95%CI: 95%-Konfidenzintervall, AGF: *Adjusted Goodness of Fit*, CRU: *Conceptual Research Use*, IRU: *Instrumental Research Use*, LPN: *Licensed Practical Nurses*, OR: *Odds Ratio*, ORU: *Overall Research Use*, PFK: *Pflegefachkräfte*, PRU: *Persuasive Research Use*, RN: *Registered Nurses*, RU: *Research Use*

schaft kritischen Denkens (signifikante, wenngleich nicht sehr hohe Korrelationen), drei Arbeiten (Kenny, 2005; Milner et al., 2005; Connor, 2006) nutzten lineare Regressionsmodelle und ein Artikel (Estabrooks et al., 2007) berichtet Ergebnisse kumulativer Logit-Modelle. Fort- und Weiterbildungen sowie Einstellung zu Forschungswissen sind auch hier durchgehend signifikante Prädiktoren. Letzteres entspricht den Ergebnissen systematischer Reviews zu individuellen Merkmalen Pflegender als Einflussfaktor für Forschungsanwendung (Estabrooks et al., 2003a; Squires et al., 2011a). Nur für einen der knapp 100 beforschten Individualfaktoren liegen ausreichend Studien angemessener methodischer Güte vor, um eine klare Aussage abzuleiten: Zwischen einer positiven Einstellung zur Nutzung von Forschungswissen und der Anwendung dieses Wissens besteht ein positiver Zusammenhang (Estabrooks et al., 2003a; Squires et al., 2011a).

Auch dieses Instrument kann als umfassend validiert gelten. Zusätzlich zu den hier referierten Studien tragen auch die zur Validität des ACT diskutierten Arbeiten Evidenz zur Güte von *Estabrooks' Kinds of RU items* bei, da letztere in den ACT-Studien zur Erfassung von Forschungsanwendung eingesetzt wurden.

2.3.3 Die *Conceptual Research Use scale*

Die CRU-Skala wurde entwickelt und validiert von Squires et al. (2011b). Dem CRU-Konzept scheint unter den verschiedenen RU-Konzepten eine besondere Bedeutung zuzukommen, so Squires et al. (2011b, 2), denn „[it] is believed to be more reflective of the process of research utilization at the individual practitioner level than are the other (i.e., instrumental, symbolic) kinds of research utilization“. Berufliche Akteure in Gesundheitseinrichtungen wie auch Politiker scheinen Forschungswissen weniger dafür zu nutzen, um in konkreten Situationen zu handeln, als vielmehr dafür, sich allgemein zu informieren. So tritt z.B. in Studien, die verschiedene RU-Formen zugleich erheben, CRU regelmäßig deutlich häufiger auf als andere RU-Formen (Squires et al., 2011b). Vor Entwicklung der CRU-Skala wurde überwiegend das CRU-Item des Estabrooks-Tools genutzt. Beim Einsatz dieses Items mit Pflegehilfskräften stießen die Forscherinnen jedoch auf Validitätsprobleme: Die Hilfskräfte äußerten Schwierigkeiten, das ursprünglich für Pflegefachkräfte konzipierte Item zu verstehen (Squires et al., 2011b). Daher entwickelten die Forscherinnen die CRU-Skala. Wie das Estabrooks-Tool liegt diese in einer Version für Pflegehilfskräfte und in einer zweiten Fassung für alle anderen Berufsgruppen der stationären Altenpflege (siehe ACT) vor. Die Teilnehmer werden gefragt, wie häufig Forschungswissen an ihrem letzten typischen Arbeitstag Verschiedenes bewirkt hat. Es folgen fünf Optionen (z.B. „Help to change your mind about how to care for residents“), die jeweils mit der beim Estabrooks-Tool beschriebenen Skala bewertet werden (Squires et al., 2011b).

In zwei Pflegehilfskraftsamples und einem Expertensample wurde die CRU hinsichtlich Inhaltsvalidität, Antwortprozesse und Verständlichkeit (*response processes*), Faktorstruktur sowie Zusammenhang mit anderen Variablen untersucht (Squires et al., 2011b). Sample 1 bestand aus

zehn Pflegehilfskräften aus zwei Pflegeheimen der Provinz Alberta. Diese füllten die Skala aus und berichteten den Forschenden anschließend ihre Eindrücke in einem informellen Austausch. Die aufgezeichneten Gespräche wurden inhaltsanalytisch ausgewertet. Die Ergebnisse führten zu Umformulierungen einiger Items, ein Item wurde ganz gestrichen (Squires et al., 2011b).

Sample 2 umfasste neun internationale Experten des Bereichs *research utilization in nursing* – drei kanadische, je zwei britische und schwedische und je ein australischer und US-amerikanische Wissenschaftler. Diese erhielten einen Fragebogen mit der Bitte, die Relevanz jedes der fünf Items mit einer vierstufigen Likert-Skala ((1) *not relevant*, (2) *item needs some revision*, (3) *relevant but needs minor revision*, (4) *very relevant*) einzuschätzen. Außerdem war Raum für Freitextkommentare. Übereinstimmungsindizes (von 0, keine Übereinstimmung bis 1 perfekte Übereinstimmung der Rater) wurden für jedes der Einzelitems und für die Gesamtskala berechnet. Die mittlere Itemübereinstimmung betrug 0,84, wobei Item 1 die schlechteste Übereinstimmung aufwies (0,55) und sich drei Rater kritisch zu diesem Item äußerten. Die durchschnittliche Gesamtskalenübereinstimmung mit Item 1 (0,87) konnte nach Entfernen dieses Items deutlich verbessert werden (0,94) (Squires et al., 2011b).

Sample 3 schloss 707 Pflegehilfskräfte aus 30 kanadischen Pflegeheimen ein und diente der statistischen Validierung. Die Interne-Konsistenz-Reliabilität (Cronbach's α) betrug 0,894, die *Split-Half*-Reliabilität nach Gutman war 0,858 und die Spearman-Brown-Reliabilität (ungleiche Länge) ergab 0,894 (Squires et al., 2011b). Eine konfirmatorische Faktorenanalyse bestätigte die Ein-Faktor-Struktur. Das Ausgangsmodell ($\chi^2=69,53$, $df=5$, $p=0,000$, $RMSEA=0,140$, $SRMR=0,030$, $CFI=0,977$) konnte durch die Korrelation von Residualvarianzen (Items 1/2 sowie Items 3/4, theoretisch-sachlogisch wie auch durch Modifikationsindizes begründet) verbessert werden: $\chi^2=6,86$, $df=3$, $p=0,075$, $RMSEA=0,043$, $SRMR=0,009$, $CFI=0,999$. Aufgrund der Befunde der Expertengruppe (Sample 2) sowie der (im Verhältnis zu den anderen Items) geringen Faktorladung wurde Item 1 in einem dritten Modell entfernt, was den Modellfit weiter verbesserte: $\chi^2=2,43$, $df=1$, $p=0,119$, $RMSEA=0,045$, $SRMR=0,007$, $CFI=0,999$ (Squires et al., 2011b). Die bivariate Korrelationsanalyse wurde dennoch mit dem Score aller fünf Items durchgeführt, um alle Validitätsevidenz mit dem gleichen Instrument (Fünf-Item-Skala) zu generieren – vorbehaltlich späterer Modifikationen. Alle Einzelitems wie auch die Gesamtskala waren signifikant ($p<0,01$) positiv korreliert mit IRU (Pearson r von 0,191 bis 0,295), PRU (Pearson r von 0,318 bis 0,406) und ORU (Pearson r von 0,232 bis 0,342) (Squires et al., 2011b). Eine multiple lineare Regression mit ORU als abhängiger Variable und CRU als einem von zehn Prädiktoren wies CRU als signifikanten Prädiktor aus ($\beta=0,109$, $p<0,001$) (Squires et al., 2011b).

Die CRU-Skala, darauf deutet diese starke Studie hin, ist ein reliables und valides Instrument für die Erfassung von CRU in Pflegehilfskraftsamples. Zwar liegt die Annahme nahe, ein Tool, das mit Pflegehilfskräften valide eingesetzt werden kann, sei auch valide in den anderen Berufsgruppen. Empirische Belege hierfür stehen bisher aber noch aus.

3 Zielstellung

Der Organisationskontext, so wurde in Abschnitt 2.1 deutlich, scheint eine wichtige Rolle für den Erfolg der Implementierung evidenzbasierter Neuerungen im stationären Altenpflegebereich zu spielen. Im Detail verstanden sind diese Zusammenhänge jedoch bisher nicht. Insbesondere mangelt es an methodisch hochwertigen Studien, die a) Organisationskontextfaktoren und Forschungsanwendung in großen Stichproben mittels robuster Messinstrumente erfassen und b) Zusammenhänge mit fortgeschrittenen statistischen Methoden (z.B. multiple Regressionen, Mehrebenenmodelle, Strukturgleichungsmodelle) analysieren. International wurde damit begonnen, diese Lücke zu bearbeiten, im deutschsprachigen Raum fehlen bisher aber robuste Forschungsinstrumente – insbesondere zur Erfassung potenziell beeinflussbarer, DI-relevanter Organisationskontextfaktoren, aber auch zur Erhebung abhängiger DI-Variablen (z.B. Ausmaß, in dem Forschungswissen in der Praxis genutzt wird) (vgl. Abschnitte 2.2.1 und 2.2.2). Drei bewährte und gut validierte Instrumente für diesen Zweck wurden in Kanada entwickelt und liegen in englischer Sprache vor: das *Alberta Context Tool (ACT)* (Abschnitt 2.3.1), *Estabrooks' Kinds of RU items* (Abschnitt 2.3.2) und die *Conceptual Research Use (CRU) scale* (Abschnitt 2.3.3). Vor diesem Hintergrund sind die Hauptziele dieser Arbeit:

- a. die deutsche Übersetzung der drei kanadischen Instrumente,
- b. die Adaptation der Instrumente an die kulturellen und strukturellen Gegebenheiten der deutschen stationären Altenpflege sowie
- c. die ausführliche psychometrische Testung der übersetzten Tools.

Bei akzeptablen psychometrischen Eigenschaften sollen zudem erste Einblicke d) in die Ausprägung der Organisationskontext- und RU-Variablen in deutschen Pflegeheimen sowie e) zum Zusammenhang zwischen Organisationskontext und RU in diesen Settings generiert werden. Das übergeordnete Ziel dieser Arbeit ist, robuste Instrumente für die deutschsprachige DIF zur Verfügung zu stellen. Diese, so die Hoffnung, werden zu einem besseren Verständnis des Organisationskontexts führen und so dabei helfen, DI-Prozesse effektiver zu gestalten – eine wichtige Voraussetzung, um die Versorgungsqualität und die Situation der Pflegebedürftigen zu verbessern.

4 Material und Methodik

4.1 Bezugsrahmen der psychometrischen Testung: *The Standards for Educational and Psychological Testing*

The Standards for Educational and Psychological Testing (AERA et al., 1999, im Folgenden kurz: *the Standards*) sind akzeptiert als *best practice* in Sachen psychometrischer Testung (Streiner and Norman, 2008; Moosbrugger und Höfling, 2012). Sie waren Grundlage für die Validierung der hier verwendeten Originalinstrumente (Estabrooks et al., 2011b; Squires et al., 2011b; Squires et al., 2011c; Squires et al., 2013c) wie auch ähnlicher Instrumente für den stati-

onären Altenpflegebereich (z.B. Gagnon et al., 2011; Zúñiga et al., 2013). Es handelt sich bei den *Standards* um einen Bezugsrahmen, der Kriterien für die Evaluation von Forschungsinstrumenten enthält und der sicherstellen will, dass relevante Aspekte beachtet werden (AERA et al., 1999). Die Prinzipien und das Validitätskonzept der *Standards* waren grundlegend für diese Arbeit.

Validität gilt in den *Standards* als vielschichtiges, aber einheitliches Konzept; d.h., es gibt nicht viele verschiedene Validitäten (z.B. eine Kriteriumsvalidität im Unterschied zu einer Konstruktvalidität), sondern nur eine einzige Validität mit verschiedenen Facetten (AERA et al., 1999). Diese Facetten werden sichtbar in unterschiedlichen Evidenzquellen. Diese „may illuminate different aspects of validity, but they do not represent distinct types of validity“ (AERA et al., 1999, 11). Validität ist dann „the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose“ (AERA et al., 1999, 11). Die vier für diese Arbeit relevanten Evidenzquellen für Validität sind: a) Evidenz basierend auf den Inhalten der Instrumente, b) Evidenz basierend auf den Antwortprozessen (*response processes*) der Probanden, c) Evidenz basierend auf der internen Struktur der Instrumente und d) Evidenz basierend auf den Beziehungen zu anderen Variablen (AERA et al., 1999). In den Standards ist noch eine fünfte Evidenzquelle erwähnt: Evidenz basierend auf den Konsequenzen eines Tests (AERA et al., 1999). Dies betrifft z.B. Persönlichkeits- oder Intelligenztests, deren Ergebnis für die Betroffenen sehr belastend sein kann oder die Definition von Cut-Off-Scores oder Normwerten, die dann Basis für evtl. folgenreiche Entscheidungen sind. Dies betrifft die hier einbezogenen Instrumente aktuell noch nicht. (Wohl aber mag dieser Aspekt in der Zukunft relevant werden, wenn z.B. die Instrumentenscores dafür genutzt werden sollen, DI-Prozesse zu planen, zu steuern und zu evaluieren).

Um inhaltsbezogene Validitätsevidenz zu generieren, werden die Instrumenteninhalte (Themen, Formulierungen und Format der Einleitungstexte, Items und Aufgaben) daraufhin überprüft, ob diese die Konstrukte repräsentieren, die zu messen das Instrument intendiert. Theorie- und evidenzgeleitete Instrumentenentwicklung oder Überprüfung der Inhalte durch Experten sind typische Methoden, um diese Evidenzform sicherzustellen bzw. zu prüfen (AERA et al., 1999). Antwortprozesse werden analysiert, um festzustellen, ob die Items von den Probanden verstanden werden, ob die Instrumente insgesamt handhabbar sind und an welchen Stellen und aus welchen Gründen Schwierigkeiten bestehen (ablesbar z.B. an einer Häufung fehlender Werte). Linguistische Validierungsverfahren (z.B. kognitive Interviews mit den Probanden) werden hierfür eingesetzt (AERA et al., 1999). Bei einer Analyse der internen Struktur der Instrumente wird überprüft, wie die Instrumentenitems und -komponenten mit den zugrundeliegenden Konstrukten zusammenhängen. Hauptkomponentenanalysen sowie exploratorische und konfirmatorische Faktorenanalysen sind hier die klassischen Methoden (AERA et al., 1999). Die Beziehung der Instrumentenvariablen zu anderen Parametern kann auf vielfältige Weise überprüft werden – von einfachen bivariaten Korrelationen bis hin zu komplexen *Latent-Variable*-Kausalmodellen. Ziel ist, heraus-

zufinden, ob die Instrumentenkonstrukte bzgl. Umfang und Richtung entsprechend der theoretischen Annahmen mit anderen Konstrukten zusammenhängen (also diese wie angenommen beeinflussen bzw. wie vorhergesagt von diesen beeinflusst werden) (AERA et al., 1999).

4.2 Genereller Aufbau der Studie

Das Projekt war aufgeteilt in drei Phasen, in denen jeweils unterschiedliche Methoden und Samples zum Einsatz kamen und in denen jeweils unterschiedliche Quellen von Validitätsevidenz untersucht wurden. In Phase 1 erfolgte die Übersetzung und kulturelle Adaptation der drei Instrumente. Primärer Fokus war hier, die Qualität der Instrumenteninhalte sicherzustellen und zu überprüfen. Dies erfolgte durch die methodisch rigorose Gestaltung des Übersetzungsprozesses und durch Einbeziehen verschiedener Experten (Abschnitt 4.3). Eine Zeitschriftenpublikation (Hoben et al., 2013) berichtet Methoden und Ergebnisse dieser Studienphase. Phase zwei untersuchte die Antwortprozesse der Probanden mittels kognitiver Interviews, um die Verständlichkeit der Items in allen Berufsgruppen sicherzustellen (Abschnitt 4.4). Auch diese Ergebnisse wurden zur Publikation eingereicht (Hoben et al., submitted). Phase 3 diente der statistischen Validierung der übersetzten Tools. Hier wurden Reliabilität, interne Struktur sowie Beziehungen zu anderen Variablen in einem großen Sample untersucht (Abschnitt 4.5). Zwei Publikationen mit Ergebnissen dieser Phase sind in Vorbereitung. Die Studie wurde von der Ethikkommission der Medizinischen Fakultät, Martin-Luther-Universität Halle-Wittenberg genehmigt (Bearbeitungsnummer: 2011-39).

4.3 Phase 1: Übersetzung und Adaptation der Instrumente

Die Übersetzung von Forschungsinstrumenten erfordert ein methodisch ausgefeiltes Vorgehen, um die Äquivalenz der originalen und der übersetzten Instrumentenversion sicherzustellen (Wild et al., 2005; Acquadro et al., 2008; Squires et al., 2013a). Um Ergebnisse zwischen verschiedenen Kulturen vergleichen zu können, ist konzeptuelle, semantische, operationale und psychometrische Äquivalenz erforderlich (Stewart und Nápoles-Springer, 2000; Acquadro et al., 2008; die von Stewart und Nápoles-Springer (2000) zusätzlich diskutierte Item- und Kriteriumsäquivalenz sind Subkategorien der psychometrischen Äquivalenz). Konzeptuelle Äquivalenz bedeutet, dass Zahl und Definitionen der Konstrukte in beiden Instrumenten übereinstimmen und dass alle Zielpersonen in beiden Kulturen jedes der Konstrukte kennen, als relevant bewerten und akzeptieren. Semantische Äquivalenz meint, dass die Items für die Probanden beider Kulturen die gleiche Bedeutung haben. Operationale Äquivalenz bezieht sich auf die gleichen formalen Instrumentenmerkmale in beiden Kulturen: Art der Verteilung (z.B. online oder in Papierform), Fragebogenformat, Instruktionen, Itemformate und Anforderungen an die Probanden. Psychometrische Äquivalenz ist erreicht, wenn beide Instrumente vergleichbar gute psychometrische Eigenschaften aufweisen (Stewart und Nápoles-Springer, 2000; Acquadro et al., 2008). Verschiedene *Best-Practice*-Leitlinien für die Übersetzung von Forschungsinstrumenten und zur Sicherung kultureller

Äquivalenz sind erhältlich, doch diese unterscheiden sich deutlich (z.B. bzgl. der Definitionen von „Äquivalenz“, „Kultur“ oder „Adaptation“, Art, Gestalt und Reihenfolge der Prozessschritte, einzubeziehende Personen oder der zu berichtenden Aspekte); ein „Goldstandard“ ist nicht verfügbar (Wild et al., 2005; Acquadro et al., 2008; Wild et al., 2009; Koller et al., 2012). Daher gilt es, die methodischen Entscheidungen und die einzelnen Schritte in Übersetzungsprozessen gut zu begründen, transparent darzustellen und detailliert zu dokumentieren (Maneesriwongul und Dixon, 2004; Acquadro et al., 2008).

Das Design dieses Übersetzungsprozesses folgte den *Good-Practice*-Empfehlungen der *Translation and Cultural Adaptation work group* der *International Society of Pharmacoeconomics and Outcomes Research (ISPOR)* (Wild et al., 2005). Diese Empfehlungen basieren auf einem Review existierender Leitlinien für Übersetzung und Adaptation von Assessmentinstrumenten. Ein von McKenna and Doward (2005) kritisch diskutierter Punkt dieser Empfehlungen ist der des Rückübersetzungsschrittes. Dieser, so McKenna and Doward (2005), garantiere keineswegs, wie intendiert, die Äquivalenz der originalen und der übersetzten Instrumentenversion. Ihr eigener Ansatz – der *two-panel approach* – beinhaltet diesen Schritt daher nicht. Um die Qualität während des Übersetzungsprozesses zu sichern, anstatt sie „a posteriori“ (McKenna und Doward, 2005, 89) zu überprüfen, werden die Vorwärtsübersetzungen in einem aufwändigen Prozess von einer Gruppe aus acht bis zwölf professionellen Experten angefertigt und anschließend in einer zweiten Gruppe aus zwei bis acht Laien diskutiert und adaptiert (Hunt et al., 1991; EGQoLHM, 1993). In der vorliegenden Studie wurde der Rückübersetzungsschritt dennoch beibehalten, da dies der einzige Weg war, die englischsprachigen Instrumentenentwicklerinnen in den Prozess einzubeziehen. Daher wurde der Vor- und Rückübersetzungsansatz durch ein zusätzliches Expertenpanel ergänzt. Eine von Acquadro et al. (2008) entwickelte Checkliste war Grundlage für die Qualitätssicherung bei Vorbereitung, Monitoring und Dokumentation des gesamten Prozesses.

Übersetzt wurden die fünf Fragebogenversionen für Pflegehilfskräfte (PHK), Pflegefachkräfte (PFK), Therapie- bzw. Betreuungskräfte (TBK), Experten (Exp.) sowie Führungskräfte (FK). Das Formular für Ärzte wurde nicht übersetzt, da die Probanden in den Einrichtungen angestellt sein müssen, um valide Aussagen zum Organisationskontext zu machen. In Deutschland gibt es jedoch so gut wie keine in Pflegeheimen angestellten Ärzte. Die Fragebögen wurden – beginnend mit der PHK-Version – nacheinander übersetzt. D.h., die PHK-Fragebögen aller drei Instrumente durchliefen zunächst den kompletten Übersetzungsprozess (Abb. 5), der erst mit der Freigabe durch die Entwicklerinnen abgeschlossen war. Anschließend erfolgte die Übersetzung der PFK-, TBK, Experten- und FK-Fragebögen auf die gleiche Weise. Gleiche Wortlaute konnten so von jeweils vorangegangenen Versionen übernommen werden, sodass nur noch die Unterschiede zu übersetzen waren.

1. Vorbereitung: Ein Protokoll des Übersetzungsprozesses und ein Zeitplan wurden an die Entwicklerinnen geschickt mit der Bitte, das Projekt zu genehmigen und zu unterstützen. Diese stimmten zu und stellten Originalfragebögen und Konzeptdefinitionen zur Verfügung. Anschließend erfolgte die Rekrutierung der Übersetzungsteams und Expertenpanels. Zudem wurden standardisierte Vorlagen für die Vor- und Rückübersetzungen sowie die Expertenpanels entwickelt. Die Kooperationsbedingungen wurden vertraglich geregelt (Anhang 1). Der Vertrag legt u.a. fest, dass die Originalfragebögen ausschließlich für diese Studie verwendet und nicht an Dritte weitergegeben werden dürfen. Außerdem bleibt das übersetzte Instrument im Besitz der Entwicklerinnen und unterliegt den gleichen Bedingungen wie die Originalinstrumente. Textteile des ACT dürfen nicht in Publikationen, Forschungsberichten oder Qualifikationsarbeiten veröffentlicht werden. (Eine Ausnahme sind die Beispielimite – jeweils eines pro ACT-Konzept, vgl. Tab. 2) Einblick in die Instrumente wird von Dr. Carole Estabrooks auf Nachfrage gewährt (Kontaktdaten siehe Anhang 1).

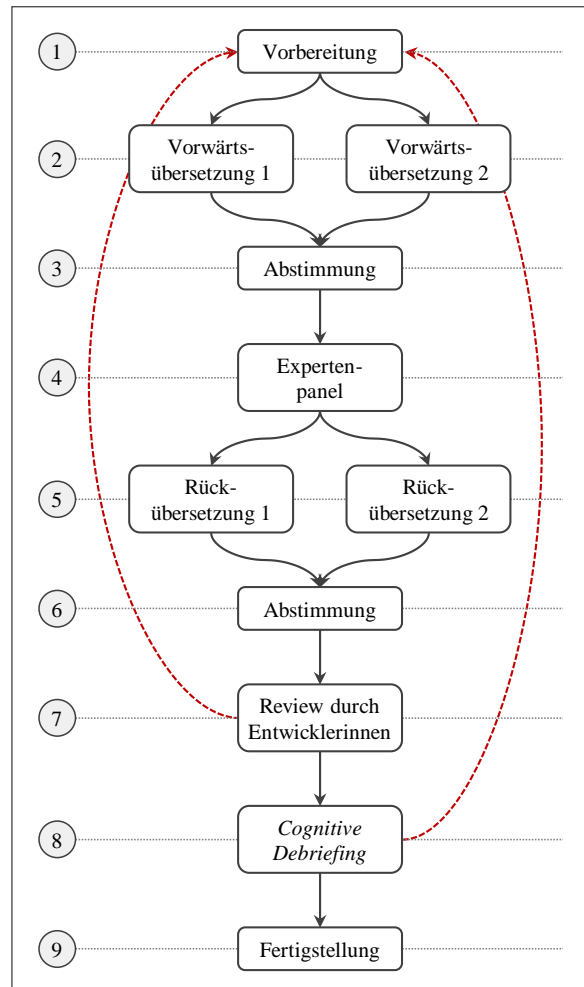


Abb. 5: Schritte des Übersetzungsprozesses (Hoben et al., 2013)

2. Vorwärtsübersetzung: Zwei Personen (eine davon der Promovend) erstellten unabhängig voneinander je eine Vorwärtsübersetzung der Fragebögen. Beide Vorwärtsübersetzer sind deutsche Muttersprachler mit sehr guten Englischkenntnissen. Beide verfügen über langjährige Erfahrung in Praxis, Weiterbildung, Beratung und Forschung in der stationären Altenpflege und sind daher vertraut mit den dortigen Berufsgruppen und Kontextbedingungen sowie mit den Instrumentenkonstrukten. Folgende Elemente waren zu übersetzen: a) Einleitungstexte, die in das Instrument insgesamt einführen, b) Texte, die in die verschiedenen Konstrukte einführen und Erklärungen zur Handhabung der Items beinhalten, c) Itemtexte, d) Anweisungen zum Überspringen bzw. Nichtüberspringen einzelner Items und e) die Skalentexte. Erst wenn beide Übersetzer ihre Version fertiggestellt hatten, wurden die beiden Übersetzungen diskutiert und abgestimmt. Fragen, die während der Übersetzung auftauchten (z.B. Bedeutungen bestimmter Konstrukte, Hintergründe der Instrumentenentwicklung oder Einordnung bestimmter Begriffe) wurden direkt

mit den Entwicklerinnen abgestimmt.

3. Abstimmung der Vorwärtsübersetzungen: Die Vorwärtsübersetzer bestimmten für jedes übersetzte Textelement, ob dieses a) einfach oder schwierig zu übersetzen war, ob sich die beiden Versionen b) nicht, leicht oder stark unterschieden und ob es c) einfach oder schwierig war, eine Einigung zu erzielen. Die Übersetzung eines Textelements wurde als schwierig klassifiziert, wenn dessen Übersetzung viel Zeit in Anspruch nahm, es schwer war, angemessene Formulierungen zu finden oder die Auswahl zwischen verschiedenen Übersetzungsvarianten schwer fiel. Unterschiede der beiden Übersetzungsvarianten wurden als leicht eingeordnet, wenn die inhaltliche Bedeutung ähnlich war und nur der Wortlaut leicht abwich. Unterschiedlich die inhaltliche Bedeutung deutlich, wurden die Diskrepanzen als stark bewertet. Tab. 7 enthält Beispiele für jede der Kombinationen. Die abweichenden Wortlaute wurden so lange diskutiert und ggf. angepasst, bis eine Einigung erzielt war. Entscheidungskriterien waren: Nähe des Wortlauts zum Original, Verständlichkeit für die Zielpersonen, Grammatik und Formulierungen sowie Kürze. Das Abstimmungsergebnis und dessen Begründung wurden dokumentiert.

Tab. 7: Beispiele für leicht/schwierig zu übersetzende Items mit keinen, leichten, starken Diskrepanzen (Hoben et al., 2013)

	Leicht zu übersetzen	Schwierig zu übersetzen
Keine Diskrepanzen	<p><i>ACT: Time, Item 1</i></p> <p><u>Original:</u> How often do you have time to do something extra for residents?</p> <p><u>Übersetzungen 1 und 2:</u> Wie oft haben Sie Zeit, um für Bewohner auch mal etwas zusätzlich zu tun?</p>	<p><i>ACT: Feedback, Einführung</i></p> <p><u>Original:</u> Sometimes you may receive formal information about the care that is being provided to residents.</p> <p><u>Übersetzung 1 und 2:</u> Es kann sein, dass Sie von Zeit zu Zeit formale, also offizielle, Informationen über die von Ihnen geleistete Pflege erhalten.</p>
Leichte Diskrepanzen	<p><i>ACT: Space, Item 1</i></p> <p><u>Original:</u> We have adequate space to provide resident care.</p> <p><u>Übersetzung 1:</u> Wir verfügen über genügend Platz, um die Versorgung der Bewohner zu gewährleisten.</p> <p><u>Übersetzung 2:</u> Wir haben für die Pflege der Bewohner genügend Räumlichkeiten zur Verfügung.</p>	<p><i>ACT: Leadership, Item 5</i></p> <p><u>Original:</u> The leader actively mentors or coaches performance of others.</p> <p><u>Übersetzung 1:</u> Die Führungsperson betreut oder berät die Mitarbeiter hinsichtlich Ihrer Leistung.</p> <p><u>Übersetzung 2:</u> Die Führungsperson berät und betreut aktiv die Arbeitsleistungen anderer.</p>
Starke Diskrepanzen	<p><i>ACT: Connections Among People, Item 4</i></p> <p><u>Original:</u> I am comfortable talking about resident care issues with those in positions of authority.</p> <p><u>Übersetzung 1:</u> Ich fühle mich wohl dabei, mit Personen in verantwortlichen Positionen über die Pflege und Betreuung der Bewohner zu sprechen.</p> <p><u>Übersetzung 2:</u> Ich kann mit höher qualifizierten Mitarbeitern gut über Themen, die die Pflege der Bewohner betreffen, sprechen.</p>	<p><i>Estabrooks' Tool: IRU, Item 1</i></p> <p><u>Original:</u> On your last typical work day, how often did you use this type of best practice knowledge to provide resident care?</p> <p><u>Übersetzung 1:</u> Wenn Sie an Ihren letzten typischen Arbeitstag denken: Wie häufig haben Sie diese Art Wissen zur optimalen Praxisgestaltung bei der Pflege und Betreuung von Bewohnern angewendet?</p> <p><u>Übersetzung 2:</u> Wenn Sie an Ihren letzten typischen Arbeitstag denken: Wie oft haben Sie solche neuartigen Erkenntnisse, Instrumente oder Konzepte in der Pflege der Bewohner eingesetzt?</p>

4. Expertenpanel: In den Expertenpanels wurde jedes Textelement diskutiert, das schwierig

zu übersetzen oder schwierig abzustimmen war oder für das sich die beiden Vorwärtsübersetzungen stark unterschieden. Weitere Textelemente wurden einbezogen, wenn dies mindestens einer der Übersetzer oder eines der Panelmitglieder wünschte. Nur für die PHK- und PFK-Fragebögen fanden Expertenpanels statt. Nahezu alle Formulierungen konnten aus diesen beiden Bögen in die anderen Bögen übernommen werden. Die verbleibenden Unterschiede betrafen lediglich leicht variierte Beispiele oder Fragen nach der Einrichtungs- anstatt der Wohnbereichsebene. Das PHK-Panel umfasst vier Mitglieder, das PFK-Panel drei, wobei eine Expertin Mitglied in beiden Panels war. Eine Übersicht über die Expertise der Panelmitglieder gibt Tab. 8.

Tab. 8: Expertise der Expertenpanelmitglieder (Hoben et al., 2013)

	Aktuelle Rolle	Qualifikation	Berufserfahrung
PHK-Panel	Exp. 1	Pflegewissenschaftlerin mit Schwerpunkt stationäre Altenpflege	Altenpflegerin Dipl. Pflegewirtin (FH) 7 Jahre als Altenpflegerin 7 Jahre als Pflegewissenschaftlerin
	Exp. 2	Universitätsdozentin eines Pflegepädagogikstudiengangs	Gesundheits- und Krankenpflegerin Dipl. Pflegepädagogin (FH) 6 Jahre als Pflegefachkraft in verschiedenen Settings 23 Jahre als Pflegepädagogin in der Altenpflegeausbildung 5 Jahre als Universitätsdozentin und wissenschaftliche Mitarbeiterin mit Schwerpunkt stationäre Altenpflege
	Exp. 3	Pflegepädagogin in der Altenpflegeausbildung	Gesundheits- und Krankenpflegerin Berufspädagogin (Staatsexamen) mit Schwerpunkt Pflege 6 Jahre als Pflegefachkraft in verschiedenen Settings 2 Jahre als Universitätsdozentin eines Pflegepädagogikstudiengangs 8 Jahre als Pflegepädagogin in der Altenpflegeausbildung
	Exp. 4	Pflegefachkraft in der stationären Altenpflege	Gesundheits- und Krankenpflegerin 10 Jahre als Pflegefachkraft in der stationären Altenpflege
PFK-Panel	Exp. 1	s.o.	s.o.
	Exp. 5	Pflegepädagogin in der Altenpflegeausbildung	Gesundheits- und Krankenpflegerin Dipl. Pflegepädagogin 6 Jahre als Pflegefachkraft in verschiedenen Settings 21 Jahre als Pflegepädagogin in der Altenpflegeausbildung
	Exp. 6	Altenpflegerin	Altenpflegerin 5 Jahre als Altenpflegerin

Jeweils eine Woche vor den Paneldiskussionen erhielten die Mitglieder ein Dokument mit dem englischen Wortlaut eines jeden Textelements, der zugehörigen deutschen Vorwärtsübersetzung sowie Kommentaren (z.B. ob es sich um ein obligatorisch oder optional zu diskutierendes Element handelte oder spezifische Fragen der Übersetzer an die Gruppe). Die Mitglieder wurden gebeten, kritisch einzuschätzen, ob das Item a) korrekt übersetzt wurde, b) relevant für die jeweilige Zielgruppe war und c) von diesen Personen verstanden würde. Die beiden Vorwärtsübersetzer nahmen an den Panels teil, beteiligten sich jedoch nicht an den Diskussionen. Sie führten die Instrumente und Hintergründe des Projekts ein, beantworteten Fragen und protokollierten Prozess und Ergebnisse. Die zweite Vorwärtsübersetzerin moderierte die Paneldiskussionen.

5. Rückübersetzung: Die aus den Expertenpanels resultierende Vorwärtsübersetzung wurde

von zwei weiteren Personen unabhängig voneinander ins Englische zurückübersetzt. Beide Rückübersetzerinnen sind erfahrene Pflegekräfte und Pflegewissenschaftlerinnen und vertraut mit der stationären Altenpflege sowie den Instrumentenkonzepten (Organisationkontextfaktoren und RU). Eine Rückübersetzerin ist englische Muttersprachlerin, die andere spricht seit ihrer frühen Kindheit Englisch. Bis zu diesem Zeitpunkt waren beide Personen nicht in den Übersetzungsprozess involviert und kannten die Originalinstrumente nicht. Die beiden Rückübersetzerinnen diskutierten ihre Versionen erst miteinander, wenn diese vollständig fertiggestellt waren.

6. Abstimmung der Rückübersetzungen: Dieser Schritt erfolgte analog zu Schritt 3. Diskrepanzen, Entscheidungen und deren Begründungen wurden dokumentiert.

7. Review durch die Entwicklerinnen: Die Instrumentenentwicklerinnen verglichen die abgestimmte Rückübersetzung mit der Originalversion. Jedes Element wurde daraufhin überprüft, ob die Formulierungen noch die gleiche Bedeutung und den gleichen Fokus reflektierten wie die des Originals. Wenn von den Entwicklerinnen Änderungsbedarf angemeldet wurde, diskutierten die Vorwärtsübersetzer die Gründe für die Diskrepanzen und die entsprechenden Konsequenzen. War die Ursache a) eine unpräzise Vorwärtsübersetzung, wurden die Formulierungen angepasst, erneut zurückübersetzt und wieder an die Entwicklerinnen gesandt. Wenn die Unterschiede b) durch notwendige Adaptationen bedingt waren, wurde dies entsprechend begründet, mit den Entwicklerinnen diskutiert und die Übersetzung wurde nicht modifiziert. Lagen die Gründe c) in einer unpräzisen Rückübersetzung, wurde dies mit den Rückübersetzerinnen diskutiert, die Rückübersetzungen wurden angepasst und wieder an die Entwicklerinnen gesandt. Dieser Kreislauf wurde so lange wiederholt, bis alle Textelemente von den Urheberinnen genehmigt waren.

8. Cognitive Debriefing: In diesem Schritt wurden die übersetzten Tools von den jeweiligen Zielpersonen ausgefüllt und linguistisch validiert. Mittels kognitiver Interviews wurde untersucht, ob die Probanden die Items verstanden wie beabsichtigt und ob sie das Instrument handhaben konnten. Dieser Schritt stellt eigentlich eine ganz eigene Phase dar und ist daher ausführlich in Abschnitt 4.4 dargelegt. Modifikationsbedürftige Items wurden angepasst, zurückübersetzt und wieder an die Entwicklerinnen gesandt. Dies wiederholte sich solange, bis alle Textelemente wie intendiert verstanden und von den Urheberinnen genehmigt wurden.

9. Fertigstellung: In diesem Schritt wurden die Formulare formatiert und abschließend korrektur gelesen.

Der gesamte Prozess (Methoden, Teilnehmer, Ergebnisse, Herausforderungen und getroffene Entscheidungen) wurde entsprechend der eingesetzten Checkliste Acquadro et al. (2008) ausführlich dokumentiert. Auf Basis dieser Protokolle und der Erfahrungen im Übersetzungsprozess diskutierten die Übersetzer die aufgetretenen Herausforderungen und clusterten diese thematisch (Ergebnisse siehe Abschnitt 5.1).

4.4 Phase 2: *Cognitive Debriefing* der übersetzten Instrumente

Cognitive Debriefing ist ein linguistisches Validierungsverfahren, das dazu dient, “to assess the clarity, intelligibility, appropriateness, and cultural relevance of the target language version to the target population” (Mear und GirouDET, 2012, 47). Es handelt sich um einen entscheidenden Schritt im Übersetzungsprozess, da hier überprüft wird, wie die Zielpersonen mit dem Instrument zurechtkommen und ob sie die Textelemente wie vorgesehen verstehen (Willis, 2005; Mear und GirouDET, 2012). Übersetzungsleitlinien unterstreichen relativ einheitlich die Bedeutsamkeit dieses Schrittes, sie unterscheiden sich jedoch substantiell bzgl. der als geeignet erachteten Methoden und des Detaillierungsgrads der Vorgaben (Wild et al., 2005; Acquadro et al., 2008). Für das Design dieser Studienphase waren zusätzliche Informationen erforderlich, da die hier genutzten Übersetzungsleitlinien (Wild et al., 2005; Acquadro et al., 2008) nicht alle methodischen Fragen abdeckten. Die umfassende und detaillierte Übersicht über kognitive Interviewmethoden von Willis (2005) wurde daher als Grundlage genutzt.

4.4.1 *Cognitive-Debriefing-Sample*

Beim *Cognitive Debriefing* handelt es sich um ein Verfahren zur Pretestung und Validierung von Fragebögen mittels qualitativer Interviewmethoden, das typischerweise kleine Stichproben (5-15 Zielpersonen) einschließt (Willis, 2005; Mear und GirouDET, 2012). In dieser Studie wurden 39 Probanden aus fünf Pflegeeinrichtungen des Rhein-Neckarraums eingeschlossen (vgl. Tab. 9 für Ein- und Ausschlusskriterien): 16 PHK, 5 PFK, 7 TBK, 5 Exp. und 6 FK. Zehn der PHK nahmen in der ersten Runde teil. Auf Basis dieser Ergebnisse wurden verschiedene Modifikationen vorgenommen und in einer zweiten Runde mit drei weiteren PHK getestet. Da weitere Modifikationen erforderlich waren, wurde noch eine dritte *Cognitive-Debriefing*-Runde mit drei zusätzlichen PHK durchgeführt. In den anderen Berufsgruppen war jeweils nur eine Runde erforderlich.

Tab. 9: Ein- und Ausschlusskriterien für Probanden der *Cognitive Debriefings* (Hoben et al., submitted)

Einschlusskriterien	Ausschlusskriterien
Angestellt in einer stationären Altenpflegeeinrichtung nach §71 SGB XI	Ehrenamtliche, Zeitarbeiter
Seit >3 Monaten in der Einrichtung	Seit <3 Monaten in der Einrichtung
Stellenumfang mindestens 25% einer Vollzeitstelle	Stellenumfang <25% einer Vollzeitstelle
Ausreichende Deutschkenntnisse, um die Instrumententexte verstehen zu können	Deutschkenntnisse nicht ausreichend, um die Instrumententexte zu verstehen

Um möglichst viele potenzielle Probleme der übersetzten Fragebögen zu entdecken, sollten die Probanden ein breites Spektrum an Merkmalen aufweisen, die das Verständnis potenziell beeinflussen (Willis, 2005; Mear und GirouDET, 2012). Eigenschaften, die sich in diesem Zusammenhang als bedeutsam erwiesen haben, sind: Alter, Geschlecht, Bildung und sozio-ökonomischer Hintergrund (Willis, 2005; Mear und GirouDET, 2012). Bei der Validierung der CRU-Skala (Squires et al., 2011b) wurde zudem deutlich, dass PHK, deren Muttersprache nicht Englisch war,

ein anderes Antwortverhalten zeigten, als englische Muttersprachler. Dies unterstreicht die Bedeutung der Ethnizität und Muttersprache als relevante Einflussfaktoren auf das Itemverständnis. Daher wurden die Probanden anhand der genannten Merkmale (Abb. 6) gesampelt.

Die Leitenden der teilnehmenden Einrichtungen wurden gebeten, potenziell geeignete Probanden zu identifizieren und deren Teilnahmebereitschaft zu eruieren. Beginnend mit den PHK wurde mit Hilfe der Leitenden, zunächst eine Person eingeschlossen mit Eigenschaften, die für das Verständnis der Fragebögen möglichst unvoreilhaft waren (rot umrandete Kästen: Muttersprache nicht Deutsch, geringe Berufserfahrung, möglichst niedriger Bildungsabschluss, keine PHK-Aus-

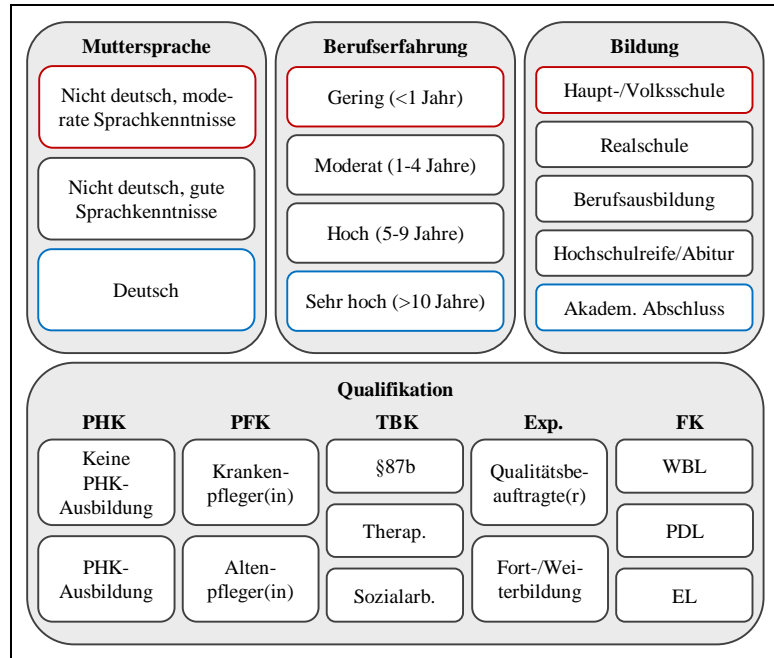


Abb. 6: Samplingkriterien für die Probanden der *Cognitive Debriefings* (Hoben et al., submitted)

(EL: Einrichtungsleitung, Exp.: Experten, FK: Führungskräfte, PDL: Pflegedienstleitung, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, TBK: Therapie-/Betreuungskräfte, WBL: Wohnbereichsleitung)

bildung). Anschließend wurde eine zweite Person mit möglichst optimalen Eigenschaften identifiziert (blau umrandete Kästen: Muttersprache Deutsch, langjährige Berufserfahrung, möglichst hoher Bildungsabschluss, PHK-Ausbildung). Schließlich wurden weitere acht Personen eingeschlossen, deren Merkmalskombinationen sich zwischen diesen beiden Extrempolen bewegten. In Runde zwei und drei des PHK-*Cognitive-Debriefings* sowie in den *Cognitive Debriefings* der anderen Berufsgruppen wurden die Teilnehmer auf die gleiche Weise ausgewählt.

4.4.2 Auswahl der zu testenden Fragebogenitems

Aufgrund der Menge der Fragebogenitems (insgesamt rund 70), der zeitaufwändigen Prozedur (Details siehe Abschnitt 4.4.3) und den knappen Zeitressourcen in der stationären Altenpflege war es nicht möglich, alle drei Instrumente komplett mit allen Probanden zu testen. Daher wurde für jeden Probanden im Vorfeld eine Itemliste erstellt. Die Teilnehmenden füllten den kompletten Fragebogen aus, in den Interviews wurden aber nur bestimmte Items getestet. Sechs ACT-Items (*Feedback 2-4, Space 1-2, Time 3*) und sämtliche RU-Items waren schwierig zu übersetzen und erforderten ausführliche Diskussionen in den Übersetzerteams, Expertenpanels und den Entwicklerinnenreviews. Diese Items wurden daher in alle Listen aufgenommen. Die übrigen PHK-Items wurden per Zufallsauswahl auf die zehn PHK-Listen verteilt, bis a) jedes der Items mindestens

einer der Listen zugeteilt war und b) jede der Listen 20 Items enthielt. Dieses Verfahren ist eine Adaptation der von Schuman (1966) beschriebenen *random probe technique*. In Runde zwei und drei der PHK-*Debriefings* wurden nur die jeweils zuvor modifizierten Items einbezogen. In den *Cognitive Debriefings* mit den anderen Berufsgruppen wurden die Items nicht per Zufall ausgewählt. Auch hier wurden die sechs problematischen ACT-Items sowie alle RU-Items in die Listen aufgenommen. Zusätzlich wurden sämtliche Items hinzugefügt, die sich in den PHK-*Debriefings* als problematisch erwiesen hatten. Außerdem wurden alle Items aufgenommen, deren Wortlaut sich von dem des entsprechenden Items der zuvor übersetzten Fragebögen unterschied. Die resultierenden Listen enthielten 26 Items für die PFK, 25 Items für die TBK, 24 Items für die Exp. und 28 Items für die FK.

4.4.3 Datensammlung

Zu Beginn wurden den Probanden zunächst Hintergründe des Forschungsprojekts, Ziele und Vorgehensweisen des *Cognitive Debriefings* sowie Maßnahmen zum Schutz ihrer Daten erläutert. Personen, die auf dieser Basis ihre informierte Zustimmung zur Teilnahme gaben, erhielten den Fragebogen (alle ACT- und RU-Items) und füllten diesen komplett aus. Der ausgefüllte Fragebogen wurde zunächst auf fehlende Antworten und Ankreuzfehler (z.B. mehr als ein Kreuz bei einem der Items oder zu überspringende Items, die angekreuzt waren) untersucht. Zudem wurden die Teilnehmenden nach Items gefragt, die sie als schwierig zu verstehen oder zu beantworten empfanden. Solch problematische Items wurden zusätzlich in die vordefinierte Liste mit den zu testenden Items aufgenommen. Die Teilnehmenden wurden außerdem gebeten, Struktur und Aufbau der Fragebögen insgesamt zu bewerten.

Anschließend begann das *Cognitive Debriefing*. In qualitativen, halbstrukturierten Einzelinterviews wurde mittels *verbal probing technique* (Willis, 2005) untersucht, wie die Probanden die Textelemente verstanden hatten, was die Hintergründe für ihre Antworten waren und was ihnen dabei durch den Kopf ging. Dabei wurde eine bestimmte Form von Fragen – so genannte *cognitive probes* (Willis, 2005) – eingesetzt, um die Reflexion der Interviewpartner anzuregen. Willis (2005) beschreibt sechs verschiedene *cognitive probes*, die für diese Arbeit adaptiert wurden und in Tab. 10 anhand eines ACT-Beispielitems erläutert sind. Willis (2005) Empfehlungen folgend wurde in dieser Arbeit ein flexibler *Probing*-Ansatz angewendet. D.h., es wurde vor den Interviews

Tab. 10: Sechs *Cognitive-Probe*-Formen nach Willis (2005) (Hoben et al., submitted)

<i>Item: Mein Team unterstützt mich (Bitte kreuzen Sie an, in wie weit Sie zustimmen)</i>	
Cognitive Probe	Beispiel
Verständnis-/Interpretationsprobe	Was bedeutet "Unterstützung durch Ihr Team" für Sie?
Paraphrasierung	Können Sie diese Frage bitte in Ihren eigenen Worten wiedergeben?
Vertrauensprobe	Wie sicher sind Sie, dass Sie Mitglied eines Teams sind, das sie

	unterstützt (bzw. nicht unterstützt)?
Erinnerungsprobe	Wie haben Sie in der letzten Zeit Ihr Team erlebt?
Spezifitätsprobe	Woran machen Sie fest, dass Ihr Team sie unterstützt (bzw. nicht unterstützt)?
Allgemeine Proben	Wie kamen Sie zu dieser Antwort? War diese Frage für Sie leicht oder schwer zu beantworten? An was haben Sie gedacht?

ein Leitfaden mit Beispielfragen für jedes Item und jede der sechs *Probe*-Formen entwickelt. Dieser legte jedoch nicht fest, welche *Probe*-Form wann und bei welchem Item eingesetzt werden sollte, denn:

„(...) the most interesting and productive forms of probing often develop through the course of the interview, as the product of the particular relationship between the interviewer, subject, and survey questionnaire.” (Willis, 2005, 95)

Daher wurden die *Probes* der Gesprächssituation angemessen ausgewählt und ggf. durch neue Fragen, die nicht im Leitfaden vorformuliert waren, ergänzt. Die Gespräche wurden mittels eines digitalen Diktiergeräts aufgezeichnet.

4.4.4 Datenanalyse und Instrumentenmodifikation

Die Interviews wurden transkribiert, paraphrasiert und inhaltsanalytisch (Mayring, 2010) ausgewertet. Die einzelnen Textsegmente wurden mit den Konstruktdefinitionen der Originalinstrumente verglichen und im Übersetzerteam diskutiert. Textelemente wurden modifiziert, wenn die Antworten von mindestens zwei Probanden nicht mit den Konstruktdefinitionen übereinstimmten bzw. deutliche Verständnisprobleme erkennen ließen. Diese modifizierten Textelemente wurden dann in einer weiteren *Cognitive-Debriefing*-Runde mit zusätzlichen Probanden getestet.

4.5 Phase 3: Psychometrische Testung der übersetzten Instrumente

4.5.1 Sample psychometrische Testung

Die psychometrische Testung der übersetzten Fragebögen umfasste a) die Analyse der Faktorenstruktur mittels konfirmatorischer Faktorenanalysen, b) die Analyse der Messinvarianz und der Faktorinvarianz zwischen den verschiedenen Berufsgruppen mittels konfirmatorischer Multigruppenmodelle, c) die Berechnung der Reliabilität der Instrumente auf Basis der Modellparameter der Faktorenmodelle d) die Analyse der Aggregierbarkeit der Daten auf Einrichtungs- und Wohnbereichsebene mittels Mehrebenenfaktorenmodellen und e) die Analyse des Zusammenhangs zwischen Organisationskontext und Anwendung von Forschungswissen mittels Strukturgleichungsmodellen. Die Punkte a) und b) untersuchen Validitätsevidenz bezogen auf die interne Struktur der Instrumente, in Punkt e) wird Validitätsevidenz bzgl. des Zusammenhangs mit anderen Variablen untersucht (vgl. Abschnitt 4.1).

Wie die angemessene Stichprobengröße für konfirmatorische Faktorenmodelle und Strukturgleichungsmodelle zu bestimmen ist, ist umstritten: Es existieren verschiedene Faustregeln

(z.B. absolute Untergrenzen der Stichprobengröße von 100 bis 200 Fälle, 5-10 Fälle pro beobachteter Variablen, 5-20 Fälle pro geschätztem Parameter etc.), aber auch modellbasierte Methoden zur Schätzung der statistischen Power (die bekanntesten sind das Verfahren von Satorra and Saris (1985) sowie der Ansatz von Muthén and Muthén (2002), basierend auf Monte-Carlo-Simulationen) (Brown, 2006; Hancock, 2006; Kline, 2011; Lee et al., 2012; Wang and Wang, 2012). Zwar sind letztere präziser und besser geeignet als Faustregeln, um die erforderliche Samplegröße adäquat zu bestimmen, doch erfordern sie, verschiedene Parameter a priori festzulegen (z.B. auf Basis vorangegangener Studien oder begründeter Annahmen bzgl. des Populationswerts). Suffiziente Informationen für die hier in Frage stehenden Modelle (vgl. Abschnitt 4.5.5) liegen leider nicht vor. Daher wurde die Samplegröße auf 800 Fälle festgelegt. Sie bewegt sich damit für alle Modelle innerhalb der oben angesprochenen Faustregel-Korridore.

Aus der Grundgesamtheit aller Pflegeheime nach §71 SGB XI der Metropolregion Rhein-Neckar wurde eine geschichtete Zufallsstichprobe von 38 Institutionen gezogen. Die Metropolregion umfasst insgesamt 295 Kommunen (acht kreisfreie Städte und sieben Landkreise) im „Drei-Länder-Eck“ Baden-Württemberg, Hessen, Rheinland-Pfalz und beherbergt auf einer Fläche von 5.637 km² eine Bevölkerung von 2,3 Mio. Menschen (Metropolregion Rhein-Neckar, 2013, Stand 2012). Zunächst wurde in einer umfassenden Recherche (Anfragen bei den statistischen Landesämtern der drei Bundesländer und Online-Recherche mittels der drei Portale: www.aok-pflegeheimnavigator.de, www.heimverzeichnis.de und pflegeheim.weisse-liste.de) eine Komplettiliste aller Pflegeheime dieser Region (n=251) erstellt. Die Einrichtungen wurden kategorisiert nach a) Bundesland: Baden-Württemberg (BW), Hessen (HE), Rheinland-Pfalz (RP), b) Größe: klein (K): ≤60 Pflegeplätze, mittel (M): 61-120 Pflegeplätze, groß (G): >120 Pflegeplätze und c) Trägerschaft: frei-gemeinnützig (FG), öffentlich (Ö) und privat (P). Aus der Kombination dieser drei Kriterien ergeben sich 27 mögliche Kategorien, in die die Heime einsortiert wurden (Tab. 11).

Tab. 11: Stationäre Altenpflegeeinrichtungen in der Metropolregion Rhein-Neckar (Schichtung nach den Kriterien Bundesland, Größe und Trägerschaft)

	Baden-Württemberg				Hessen				Rheinland-Pfalz				Gesamt			
	K	M	G	Σ	K	M	G	Σ	K	M	G	Σ	K	M	G	Σ
FG	18	38	15	71	2	6	3	11	5	26	10	41	25	70	28	123
Ö	0	2	1	3	0	0	1	1	1	2	2	5	1	4	4	9
P	32	21	6	59	13	4	3	20	10	15	15	40	55	40	24	119
Σ	50	61	22	133	15	10	7	32	16	43	27	86	81	114	56	251

FG: frei-gemeinnützige Trägerschaft, Ö: öffentliche Trägerschaft, P: private Trägerschaft
 K: kleine Einrichtungen (≤60 Pflegeplätze), M: mittelgroße Einrichtungen (61-120 Pflegeplätze), G: große Einrichtungen (>120 Pflegeplätze)
 Σ: Summe

Aus jeder der Kategorien wurde durch eine unabhängige, nicht anderweitig in die Studie involvierte Person je eine Einrichtung per Zufall ausgewählt. In den Kategorien M-FG und G-FG wurden zusätzlich je zwei weitere Institutionen für jedes der drei Bundesländer gezogen, da Einrichtungen dieser Kategorien auch auf Bundesebene überrepräsentiert sind (Statistisches Bundesamt, 2013). War die Kategorie leer (z.B. BW-K-Ö), wurde ersatzweise eine Einrichtung

aus einer möglichst ähnlichen Kategorie des gleichen Bundeslandes gezogen (z.B. BW-K-FG). Die gezogenen Einrichtungen wurden telefonisch kontaktiert, über die Studie informiert und um Teilnahme gebeten. Als Gegenleistung wurde den Einrichtungen eine anonymisierte Rückmeldung der Studienergebnisse angeboten. Bei Interesse erhielten die Kontaktpersonen schriftliche Informationen per E-Mail und wurden nach einer vereinbarten Bedenkzeit (meist ca. 1-2 Wochen) erneut kontaktiert. Sagte die Einrichtung ab, wurde eine weitere Einrichtung nachgezogen. Wenn sich in einem Bundesland keine Einrichtung einer bestimmten Kategorie (oder einer der Ursprungskategorie ähnlichen Kategorie) fand, wurden entsprechende Einrichtungen eines anderen Bundeslandes kontaktiert. Von 133 kontaktierten Pflegeeinrichtungen sagten insgesamt 41 ihre Teilnahme zu, drei davon sprangen jedoch vor der Erhebung aufgrund von Personalengpässen ab. Dies entspricht einer Beteiligungsquote von 28,6%. Die Absagegründe der anderen Einrichtungen waren: a) andere Projekte (eigene wie z.B. die Einführung einer Pflegedokumentationssoftware oder Beteiligung an anderen Forschungsprojekten) (n=37), b) Personalengpässe (n=27), c) keine entscheidungsbefugte Kontaktperson erreichbar oder verfügbar (z.B. durch Leitungswechsel) (n=11), d) Mitarbeiterbefragung nicht erwünscht (n=6), e) keine Gründe angegeben (n=5), f) Ablehnung durch Betriebsrat oder Mitarbeiter (n=3), g) Teamprobleme (n=2) und h) schlechte Vorerfahrungen mit anderen Forschungsprojekten (n=1).

Die Einschlusskriterien für die in den Einrichtungen tätigen Personen entsprachen denen der *Cognitive Debriefings* (Abschnitt 4.4.1, Tab. 9). Alle PHK, PFK, TBK, Exp. und FK, die diesen Kriterien entsprachen, waren potenziell einschließbar. Zusätzlich wurden in dieser Studienphase auch Altenpflegeschüler (APS) eingeschlossen. Insgesamt wurden 824 Fragebögen ausgefüllt (274 PHK, 197 PFK, 153 TBK, 6 Exp., 129 FK und 65 APS). Der Fragebogenrücklauf in Relation zu allen potenziell einschließbaren Personen, die in der Einrichtung angestellt sind, (Rücklaufquote 1) betrug 37,73% für die Gesamtstichprobe. Gemessen an allen zur Datenerhebung in den Einrichtungen verfügbaren potenziell einschließbaren Personen (Rücklaufquote 2) betrug der Rücklauf in der Gesamtstichprobe 81,50%. Weitere Details zu den Rückläufen zeigt Tab. 12.

Tab. 12: Rücklaufquoten der verteilten Fragebögen

	Fragebögen		Rücklaufquote 1 (%)				Rücklaufquote 2 (%)			
	N	%	Ges.	Min.	Max.	SA	Ges.	Min.	Max.	SA
PHK	274	33,25	33,79	10,00	100,00	20,45	76,11	28,57	77,60	18,95
PFK	197	23,91	31,57	9,09	80,00	14,04	80,74	50,00	82,31	15,47
TBK	153	18,57	50,16	14,29	100,00	23,67	81,38	25,00	83,29	22,51
Exp.	6	0,73	60,00	0,00	100,00	51,64	100,00	100,00	100,00	0,00
FK	129	15,66	73,71	33,33	100,00	20,49	92,81	50,00	93,86	12,27
APS	65	7,89	25,10	0,00	85,71	21,21	87,84	0,00	91,68	21,02
Alle	824	100,00	37,73	23,33	79,31	11,22	81,50	56,67	82,90	11,71

Ges.: Gesamt, Min.: Minimum, Max.: Maximum, SA: Standardabweichung

Rücklaufquote 1 = Zahl der Fragebögen / Zahl der in der Einrichtung beschäftigten Personen × 100

Rücklaufquote 2 = Zahl der Fragebögen / Zahl der zur Datenerhebung verfügbaren Personen × 100

APS: Altenpflegeschüler, Exp.: Experten, FK: Führungskräfte, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, TBK: Therapie-/Betreuungskräfte

4.5.2 Fragebögen und erhobene Variablen

Charakteristika der Pflegeeinrichtungen wurden mittels eines Strukturfragebogens erhoben. Dieser war eine Eigenentwicklung auf Basis a) der Strukturfragebögen vorangegangener Forschungsprojekte im stationären Altenpflegebereich Deutschlands (DEMIAN, Kruse et al., 2010; Implementierung des RAI, Becker et al., 2012) sowie b) der Strukturbögen, die von den Entwicklerinnen des ACT und der RU-Skalen in der kanadischen TREC-Studie (Estabrooks et al., 2009c) eingesetzt wurden. Erhobene Variablen waren: 1) Zahl der Wohnbereiche (WB), 2) Spezialisierung der WB, 3) Zahl der Pflegeplätze jedes WBs, 4) Zahl der Ein- und Zweibettzimmer jedes WBs, 5) Zahl der Bewohner in den einzelnen Pflegestufen (0, I, II, III und Härtefall) jedes WBs, 6) Personalsituation (jeweils Personenzahl und vollzeitäquivalente Stellenzahl aller einbezogenen Berufsgruppen) auf jedem WB sowie auf Einrichtungsebene, 7) Zahl der beschäftigten einschließbaren Personen auf jedem WB und in der Einrichtung insgesamt, 8) Zahl der zur Datenerhebung verfügbaren einschließbaren Personen auf jedem WB und in der Einrichtung insgesamt, 9) Art und Zahl der Implementierungs-/Veränderungsprojekte in den letzten fünf Jahren.

Die Fragebögen für die in den Einrichtungen Beschäftigten lagen in sechs berufsgruppenspezifischen Versionen vor: PHK-, PFK-, TBK-, Exp.-, FK- und APS-Version. Sie enthielten die jeweiligen Versionen des ACT und der RU-Skalen. Da für APS keine eigene Version dieser Instrumente erhältlich ist, wurden für diese Gruppe die PHK-Versionen verwendet. Zusätzlich wurden weitere Variablen einbezogen:

a) Fünf Skalen der deutschen Version des *Copenhagen Psychosocial Questionnaire (COPSOQ)* (Nübling et al., 2005; Nübling et al., 2006): 1) Arbeitszufriedenheit (7 Items, vierstufige Likert-Skala: sehr zufrieden, zufrieden, unzufrieden, sehr unzufrieden), 2) allgemeine Gesundheit (1 Item aus dem EQ-5D-Fragebogen (EuroQoL Group, 1990), elfstufige Likert-Skala von 0, schlechtester denkbarer Gesundheitszustand bis 10, bester denkbarer Gesundheitszustand), 3) verhaltensbezogene Stresssymptome (8 Items, fünfstufige Likert-Skala: trifft genau zu, trifft überwiegend zu, trifft etwas zu, trifft kaum zu, trifft überhaupt nicht zu), 4) kognitive Stresssymptome (4 Items, fünfstufige Likert-Skala: immer, oft, manchmal, selten, fast nie/nie) sowie 5) die *Personal Burnout Scale* des *Copenhagen Burnout Inventory (CBI)* (Borritz and Kristensen, 1999) (6 Items, fünfstufige Likert-Skala: immer, oft, manchmal, selten, fast nie/nie). Diese Skalen wurden hinzugenommen, da es starke Hinweise darauf gibt, dass diese individuellen Faktoren die Anwendung von Forschungswissen beeinflussen (Squires et al., 2011a). Sie sollen Aufschluss über die Validitätsevidenz „Zusammenhang mit anderen Variablen“ der RU-Skalen geben und werden in die Strukturgleichungsmodelle (Abschnitte 4.5.5.5 und 5.3.5) aufgenommen. Die deutsche Version des COPSOQ wurde umfassend validiert (Nübling et al., 2005; Nübling et al., 2006) und verschiedentlich erfolgreich im Pflegeheimbereich eingesetzt (z.B. Simon et al., 2005; Nübling et al., 2010; Kwiatkowski, 2011; Schaeffer et al., 2013). Die Skalen weisen eine akzeptable In-

terne-Konsistenz-Reliabilität (Cronbach's α) auf (Arbeitszufriedenheit: 0,79, Verhaltensbezogener Stress: 0,90, Kognitiver Stress: 0,87, Burnout: 0,91). Untersucht wurden auch die interne Struktur des COPSOQ (Hauptkomponentenanalysen und konfirmatorische Faktorenanalysen) und Zusammenhänge zwischen den COPSOQ-Variablen (bivariate Korrelationen, Regressions- und Strukturgleichungsmodelle) (Nübling et al., 2005; Nübling et al., 2006). Bei den Faktorenanalysen und Strukturgleichungsmodellen wurden jedoch nicht wie üblich die Items der Einzelskalen als Indikatoren einer entsprechenden latenten Konstruktvariablen spezifiziert (z.B. die sieben Arbeitszufriedenheitsitems als Indikatoren der latenten Variablen Arbeitszufriedenheit). Vielmehr wurden die Skalenscores (z.B. Mittelwert der sieben Arbeitszufriedenheitsitems) verschiedener Skalen als Indikatoren bestimmter latenter Konstrukte definiert (z.B. der Arbeitszufriedenheitsscore als einziger Indikator eines latenten „Outcome 1“ sowie die Scores der beiden Stressskalen und der Burnoutskala als Indikatoren eines „Outcome 2“). Die Modellfitwerte des konfirmatorischen Faktorenmodells ($\chi^2=2050$, $df=143$, $p=0,000$, $NFI=0,91$, $TLI=0,89$, $CFI=0,91$, $RMSEA=0,07$, $SRMR=0,05$) sowie des Strukturgleichungsmodells ($\chi^2=1855$, $df=142$, $p=0,000$, $NFI=0,92$, $TLI=0,90$, $CFI=0,91$, $RMSEA=0,07$, $SRMR=0,05$) wurden von den Autoren als akzeptabel bewertet (Nübling et al., 2005), liegen aber größtenteils unter den seit langem geforderten Grenzwerten (NFI , TLI und $CFI > 0,95$, $RMSEA < 0,06$, $SRMR < 0,08$) (Hu and Bentler, 1999).

b) Einstellung gegenüber Forschungswissen: Die in der TREC-Studie (Estabrooks et al., 2009c) verwendete *Attitudes Towards Research Use (ARU) Scale* wurde auf Deutsch übersetzt – allerdings nicht entsprechend des aufwändigen Verfahrens für die anderen Instrumente. Dies war nicht erforderlich, da die kritischen Wortlaute (*best practice* und *research use*) denen der RU-Skalen entsprechen und von diesen übernommen werden konnten. Bei der Skala handelt es sich um eine von Lacey (1994) adaptierte Version des ursprünglich von Champion and Leach (1989) entwickelten Tools. Eine Hauptkomponentenanalyse ergab eine Ein-Faktor-Struktur (48% Varianzaufklärung) und die Interne-Konsistenz-Reliabilität (Cronbach's α) betrug 0,74 (Estabrooks, 1997). Auch diese Skala wurde aufgenommen, um Evidenz zum Zusammenhang mit anderen Variablen der RU-Skalen zu untersuchen. Einstellung gegenüber der Anwendung von Forschungswissen gilt als der bestbelegte individuelle Prädiktor für Forschungsnutzung (Squires et al., 2011a).

c) Soziodemografische Merkmale der Probanden: Auch dieser Fragebogenteil war eine Eigenentwicklung auf Basis der Demografiebögen vorangegangener Studien (TREC-Studie, Estabrooks et al., 2009c; DEMIAN, Kruse et al., 2010; Implementierung des RAI, Becker et al., 2012). Die Variablen waren: 1) Geschlecht, 2) Alter, 3) Muttersprache (und ggf. Zeit, seit der die deutsche Sprache gesprochen wird), 4) Art der beruflichen Qualifikation und Abschlussdatum, 5) Berufserfahrung, 6) Stellenumfang und Arbeitsstunden in den letzten 14 Tagen, 7) Schicht(en) in der (in denen) die meiste Zeit gearbeitet wird sowie 8) höchster Bildungsabschluss.

4.5.3 Datensammlung

Die Strukturdaten der Einrichtungen wurden jeweils in einem Vorgespräch erhoben. Dafür wurde ein separater Termin vor der Befragung der Beschäftigten vereinbart. Der Strukturbogen wurde in Form eines strukturierten Interviews mit den verantwortlichen Leitungspersonen der Einrichtung (meist Pflegedienst- und/oder Heimleitung) ausgefüllt. In diesem Vorgespräch wurde zudem der Erhebungstermin vorbesprochen und organisiert.

An den Erhebungsterminen war der Autor dieser Arbeit ganztags in den Pflegeeinrichtungen anwesend. Er wurde allen anwesenden Mitarbeitern der Früh- und Spätschicht vorgestellt, erläuterte das Forschungsprojekt, verteilte die Fragebögen und stand für Fragen bzw. bei Verständnisproblemen zur Verfügung. Mitarbeiter, die teilnehmen wollten, füllten die Bögen aus und gaben sie bei ihm ab, um zu vermeiden, dass Unbefugte Einblick erhielten.

4.5.4 Datenaufbereitung

Die Eingabe der Fragebogendaten erfolgte mittels der Software *EpiData Entry Client*, Version 1.1.1.1 (EpiData, 2013). Dafür wurde mit dem Programm *EpiData Manager*, Version 1.1.2.0 (EpiData, 2013) eine Eingabemaske erstellt. Für jedes Item wurden die erlaubten Werte einschließlich Labels definiert. Fehlende Werte wurden mit -9 kodiert, Items, die korrekterweise nicht beantwortet wurden (z.B. weil sie zu überspringen waren), erhielten den Kode -8. Die Eingabe nicht erlaubter Werte war somit nicht möglich und produzierte eine Fehlermeldung. Zudem wurde definiert, dass die Items der Reihe nach ausgefüllt und für jedes Item ein Wert eingegeben werden musste. So wurde das Risiko für Eingabefehler minimiert. Nach Eingabe aller Fragebögen erfolgte der Export der Daten in eine SPSS-Datenbank.

Die Analyse der fehlenden Werte, der Normalverteilung der Daten und der deskriptiven Statistiken wurde mit der Statistiksoftware SPSS, Version 20 (IBM, 2011) durchgeführt. Unter den 167 Variablen des Datensatzes war nur eine, die in mehr als 5% aller Fälle fehlte: die Frage nach der Zahl der Arbeitsstunden in den letzten 14 Tagen (Soziodemografiebogen). 203 der 824 Probanden (24,64%) hatten diese Frage nicht beantwortet. Mit 28 fehlenden Werten (3,4%) folgte das COPSOQ-Item zur allgemeinen Gesundheit an zweiter Stelle, Item 5 der CRU-Skala fehlte bei 25 Fällen (3%). Alle anderen Variablen wiesen eine *Missingness* von deutlich unter 3% auf, was als sehr gutes Ergebnis gewertet werden kann und als erster Hinweis darauf, dass keine systematischen Probleme vorliegen (Scheridera, 2007; Graham, 2012). Komplette Datensätze lagen für 445 der 824 Probanden vor (54%). Bei 353 Probanden (42,84%) fehlten weniger als 5% der Variablen (max. 8 fehlende Werte), 16 Probanden (1,94%) wiesen Missings in 5-10% der Variablen auf (zwischen 9 und 16 fehlende Werte) und bei 7 Probanden (0,85%) fehlten 10-17% der Werte (zwischen 17 und 29 Missings). Dies ist ebenfalls als sehr gutes Ergebnis zu werten (Scheridera, 2007; Graham, 2012). Nur in drei Fällen (0,36%) fehlten mehr als 25% der Werte (51, 81 und 86 Missings). Wie der *Missing Completely at Random (MCAR) Test* nach Little (Little

and Rubin, 2002) zeigte ($\chi^2=12426,806$, $df=12434$, $p=0,517$), waren die Missings vollständig zufällig verteilt. Daher wurden die drei Fälle (je ein PHK-, PFK- und TBK-Bogen) mit mehr als 25% fehlenden Werten gelöscht, die restlichen Missings jedoch nicht in irgendeiner Form imputiert. Der Datensatz, der den weiteren Berechnungen zugrunde lag, beinhaltete somit 821 Fälle (273 PHK, 196 PFK, 152 TBK, 6 Exp., 129 FK und 65 APS).

Alle Variablen wurden auf Normalverteilung untersucht. Schiefe- und Wölbungswerte der Verteilungen aller Variablen wichen deutlich von Null ab und bei allen Variablen war der Betrag des Schiefe- und/oder Wölbungswertes deutlich mehr als doppelt so groß wie der zugehörige Standardfehler – ein deutlicher Hinweis darauf, dass die Normalverteilungsannahme nicht erfüllt ist (Miles und Shevlin, 2001). Der Kolmogorov-Smirnov-Anpassungstest (Kolmogorov, 1933; Smirnov, 1948) sowie der Shapiro-Wilk-Test (Shapiro and Wilk, 1965) unterstrichen dieses Ergebnis: die Verteilung aller Variablen wich signifikant ($p<0,0001$) von der Normalverteilung ab. Die deskriptiven Analysen (u.a. Häufigkeiten, Maße der zentralen Tendenz und Dispersion, Boxplots, Stängel-Blatt-Diagramme, M-Schätzer, Ausreißer) ergaben keine extremen bzw. gehäuften Ausreißer, die das Löschen weiterer Fälle erforderten.

Abschließend wurden die Gesamtscores für die einzelnen Skalen entsprechend der Vorgaben der Instrumentenentwicklerinnen berechnet. Tab. 13 listet die englische und deutsche Bezeichnung der in den Fragebögen enthaltenen Konzepte sowie die zugehörigen Items auf. In den statistischen Analysen (Mplus-Syntaxen) wurden durchgehend die Abkürzungen der englischen Konzepte und Items verwendet, um die Kommunikation mit den kanadischen Forscherinnen zu erleichtern. Dies wird in den Ergebnisdarstellungen dieser Arbeit beibehalten.

Die Scores für sieben ACT-Konzepte (Führung, Kultur, Rückmeldung, Miteinander im Team, Personal, Räumlichkeiten, Zeit) wurden berechnet durch Mittelwertbildung der jeweils zugehörigen Items. Dabei enthält das Konzept „Personal“ für PHK, PFK und APS drei Items, für die anderen Berufsgruppen nur zwei. Für die Scoreberechnung der restlichen drei ACT-Konzepte mit Häufigkeitsskala (Offizielle Zusammenkünfte, Informeller Austausch und Strukturelle/Elektronische Ressourcen) wurden die Antworten zunächst umkodiert (1=nie und 2=selten $\rightarrow 0$; 3=gelegentlich $\rightarrow 0,5$; 4=häufig und 5=fast immer $\rightarrow 1$), anschließend wurden die umkodierten Antworten der jeweils zugehörigen Items addiert. Das Konzept „Offizielle Zusammenkünfte“ enthält für PHK und APS ein anderes Item (OZ4) als für die restlichen Berufsgruppen (OZ3), das Konzept „Informeller Austausch“ enthält sieben Items, die in allen Fragebögen enthalten sind und fünf Items, die nur in bestimmten Fragebogenversionen vorkommen (Tab. 13). Aus *Estabrooks Kinds of RU items* wird kein Score gebildet. Das CRU-Item dieser Skala ist nicht in den PHK- und APS-Fragebögen enthalten. Die Scores der restlichen Konzepte (CRU-Skala und COPSOQ-Skalen) wurden durch Mittelwertbildung der zugehörigen Items berechnet. Sowohl Mittelwert- als auch Häufigkeitsscores wurden für Fälle berechnet, bei denen mehr als 75% der zugehörigen Items gültige Werte aufwiesen. Andernfalls wurde der Scorewert als fehlend (Missing) kodiert.

Tab. 13: Englische und deutsche Bezeichnungen der Konzepte und Items in den Fragebögen

	Englische Bezeichnung		Deutsche Bezeichnung		Fragebögen	Score	
	Konzept	Items	Konzept	Items		Min.	Max.
ACT-Konzepte mit Mittelwertscore	Leadership (LEAD)	L1-L6	Führung	F1-F6	Alle	1	5
	Culture (CULT)	C1-C6	Kultur	K1-K6	Alle	1	5
	Feedback (FEED)	F1-F6	Rückmeldung	RM1-RM6	Alle	1	5
	Connections Among People (CAMP)	CAP1-CAP6	Miteinander im Team	MT1-MT6	Alle	1	5
	Staffing (STAF)	ST1-ST2 ST3	Personal	P1-P2 P3	Alle 1, 2, 6	1	5
	Space (SPAC)	SP1-SP3	Räumlichkeiten	R1-R3	Alle	1	5
	Time (TIME)	T1-T4	Zeit	Z1-Z4	Alle	1	5
ACT-Konzepte mit Häufigkeitsscore	Formal Interactions (FINT)	FI1-FI2, FI5 FI3 FI4	Offizielle Zusammenkünfte	OZ1-OZ2, OZ5 OZ3 OZ4	Alle 2-5 1, 6	0	4
	Informal Interactions (IINT)	II1-II2, II4-II6, II10-II11 II3 II7 II8 II9 II12	Informeller Austausch	IA1-IA2, IA4-IA6, IA10-IA11 IA3 IA7 IA8 IA9 IA12	Alle 3-5 2, 3, 5 1, 6 2-5 5	0	8-11 ^a
	Structural/Electronic Resources (STER)	SER1-SER11	Strukturelle/Elektronische Ressourcen	SER1-SER11	Alle	0	11
	Instrumental Research Use (INRU)	IRU	Forschungsanwendung direkt	FAD	Alle	1	5
Estabrooks Kinds of RU items	Conceptual Research Use a) (CORU _a)	CRU _a	Forschungsanwendung konzeptuell a)	FAK _a	2-5	1	5
	Persuasive Research Use (PERU)	PRU	Forschungsanwendung überzeugen	FAÜ	Alle	1	5
	Overall Research Use (ORU)	ORU	Forschungsanwendung allgemein	FAA	Alle	1	5
	Conceptual Research Use b) (CORU _b)	CRU _b 1-CRU _b 6	Forschungsanwendung konzeptuell b)	FAK _b 1-FAK _b 5	Alle	1	5
ARU-Skala	Attitude Towards Research Use (ATRU)	ARU1-ARU6	Einstellung zur Forschungsanwendung	E1-E6	Alle	1	5
COPSOQ-Skalen	Job Satisfaction (JOSA)	JS1-JS7	Arbeitszufriedenheit	AZ1-AZ7	Alle	4	1 ^b
	Gen. Health (GEHE)	GH	Allg. Gesundheit	AG	Alle	0	10
	Behav. Stress (BSTR)	BS1-BS8	Stress – Verhalten	SV1-SV8	Alle	1	5
	Cognitive Stress (CSTR)	CS1-CS4	Stress – kognitiv	SK1-SK4	Alle	1	5
	Burnout (BURN)	B1-B6	Burnout	B1-B6	Alle	1	5

Rot: Nur in bestimmten Fragebogenversionen enthalten

1=Pflegehilfskräfte, 2=Pflegefachkräfte, 3=Therapie-/Betreuungskräfte, 4=Experten, 5=Führungskräfte, 6=Altenpflegeschüler

^aVariiert je nach Zahl der Items im Fragebogen (PHK & APS 8, Exp. 9, PFK & TBK 10, FK 11)

^bInvertierte Skala – im Gegensatz zu allen anderen Scores sind hier höhere Werte schlechter

4.5.5 Datenanalyse

4.5.5.1 Analyse der beobachteten Skalenscores

Als erster Schritt der Datenanalyse wurden die auf Basis der beobachteten Variablenwerte gebildeten Skalenscores analysiert. Untersucht wurde, ob die Faktoren „Einrichtungszugehörigkeit“, „Berufsgruppenzugehörigkeit“ und „Wohnbereichszugehörigkeit innerhalb der Einrichtung“ sowie deren Interaktionen Unterschiede in den Scores bedingen. Aufgrund der deutlich verletzten Normalverteilungsannahmen wurden hierfür nicht die üblicherweise verwendeten varianzanalytischen Verfahren (ANOVA bzw. MANOVA) verwendet. Stattdessen wurde eine nichtparametrische (oder auch permutationsbasierte) Varianzanalyse (PERMANOVA) (Anderson, 2001a, 2001b) durchgeführt. Dieses Verfahren wurde mittels des Pakets „vegan“ (Version 2.0-7) (Oksanen et al., 2013) in die Statistiksoftware „R“ (Version 3.0.2) (CRAN, 2013) implementiert. Jeder der Skalenscores wurde daraufhin überprüft, ob er sich zwischen den a) Pflegeeinrichtungen (E), b) Berufsgruppen (B), c) Wohnbereichen innerhalb der Einrichtung (ExW), d) Berufsgruppen innerhalb der Einrichtung (ExB) sowie e) Wohnbereichen und Berufsgruppen innerhalb der Einrichtung (ExWxB) signifikant unterscheidet.

4.5.5.2 Überprüfung der Faktorenstruktur und der Faktorinvarianz

Um die Faktorenstruktur der übersetzten Instrumente zu untersuchen, wurden konfirmatorische Faktorenanalysen durchgeführt. Überprüft wurde, ob die übersetzten Instrumente die gleiche Faktorenstruktur aufweisen, wie die Originalinstrumente (ACT: zehn Faktoren, CRU-Skala: ein Faktor). Da es sich bei *Estabrooks' Kinds of RU items* um vier Einzelitems ohne gemeinsame Faktorenstruktur handelt, waren diese nicht Bestandteil dieses Analyseschritts. Mittels konfirmatorischer Mehrgruppenfaktorenmodelle wurde zudem überprüft, ob die Assessmentinstrumente zwischen den einbezogenen Berufsgruppen messinvariant waren. Messinvarianz ist Voraussetzung, um Variablen, daraus abgeleitete Scores und Modellparameter zwischen den einbezogenen Berufsgruppen zu vergleichen (Byrne et al., 1989; Brown, 2006; Dimitrov, 2010; Sass, 2011; Wang and Wang, 2012). In der Substichprobe der Experten konnten keine gruppenspezifischen Faktorenmodelle geschätzt werden, da diese mit nur sechs Personen zu klein war. Diese Gruppe wurde daher nur in den Modellen der Gesamtstichprobe berücksichtigt. Die Analysen wurden mit dem Statistikprogramm *Mplus* (Version 7.11, Base Program and Combination Add-On) (Muthén and Muthén, 2013) durchgeführt. Da die Voraussetzung der multivariaten Normalverteilung bei den vorliegenden Daten deutlich verletzt war, wurden alle Modelle mittels eines robusten, standardfehler-, mittelwert- und varianzadjustierten *Maximum-Likelihood*-Schätzers (MLMV) (Asparouhov und Muthén, 2010) geschätzt, der die Daten entsprechend korrigiert. Dieser Schätzer erfordert in *Mplus* den listweisen Ausschluss fehlender Werte (Muthén and Muthén, 2012). Sofern nicht anderweitig berichtet, wurde in allen Modellen aus Gründen der Modellidentifikation bei Faktoren mit mehr als einem Indikator die Ladung des ersten Indikators auf 1,0 fixiert

(=Mplus-StandardEinstellung, Muthén and Muthén, 2012).

Die Evaluation der Modelle erfolgte in Übereinstimmung mit gängigen Empfehlungen (Brown, 2006; Raykov und Marcoulides, 2006; Kline, 2011; Wang und Wang, 2012; West et al., 2012). Als globales, inferenzstatistisches Maß für den Modellfit wird der χ^2 -Wert berichtet. Dieser sollte nicht signifikant sein ($p > 0,05$), ist aber abhängig vom Stichprobenumfang, sodass mit zunehmender Stichprobengröße selbst minimale Fehlspezifikationen eines Modells zu signifikanten χ^2 -Werten führen. Daher wurden zudem deskriptive, nicht stichprobengrößenabhängige Fitindizes herangezogen: a) *Root Mean Square Error of Approximation* (RMSEA), b) *Comparative Fit Index* (CFI), c) *Tucker Lewis Index* (TLI) sowie d) *Standardized Root Mean Square Residual* (SRMR). Richtwerte für die Interpretation dieser Indizes waren die von Hu and Bentler (1999) vorgeschlagenen Cut-Off-Werte akzeptabler Modellpassung: $RMSEA < 0,6$, CFI und TLI $> 0,95$ sowie $SRMR < 0,8$. Neben der globalen Fitevaluation wurden die Modelle auch bzgl. spezifischer Problembereiche analysiert. Die dafür gängigerweise verwendeten standardisierten Varianz-Kovarianz-Residuen (Differenz zwischen Sample- und modellimplizierter Varianz-Kovarianz-Matrix) stehen in Mplus bei Verwendung des MLMV-Schätzers nicht zur Verfügung (dieser liefert nur die unstandardisierten Werte, Muthén und Muthén, 2012). Analysiert wurden daher die geschätzten Modellparameter (Interpretierbarkeit, Größe/Richtung, statistische Signifikanz) und die Modifikationsindizes (Hinweise auf Fehlspezifikationen und Modifikationsoptionen).

Das Prozedere folgte den in einschlägigen Methodentexten (Byrne et al., 1989; Brown, 2006; Dimitrov, 2010; Sass, 2011; Wang and Wang, 2012) beschriebene Schritten (Abb. 7):

1. Separate Schätzung eines gemeinsamen Modells: Um das Modell in allen Gruppen schätzen und später Gruppenvergleiche vornehmen zu können, wurden die acht gruppenspezifischen ACT-Variablen (ST3, FI3, FI4, II3, II7, II8, II9, II12, Tab. 13) ausgeschlossen. Für das ACT wurde zunächst ein Zehn-Faktoren-Modell geschätzt (Modell 1, Abb. 8). Weder in der Gesamtstichprobe noch in einer der Subgruppen erwies sich dieses als passend. Analog zum Vorgehen der kanadischen Studie (Estabrooks et al., 2011b) wurde dieses Modell daher aufgeteilt in ein Sieben-Faktoren-Modell (alle zustimmungsskalierten Konzepte, Modell 2a, Abb. 8) und ein Drei-Faktoren-Modell (alle häufigkeitsskalierten Konzepte, Modell 3a, Abb. 8). Die Modellfitwerte des Modells 2a verbesserten sich gegenüber Modell 1, die Modellanalyse ergab jedoch weiteren Modifikationsbedarf. Alle Modellmodifikationen erfolgten schrittweise (ein Parameter nach dem anderen) und die Modelle wurden jeweils erneut geschätzt. Abb. 8 zeigt das Endergebnis (Modell 2b), das einen akzeptablen Fit aufwies. Vier Items (C2, C3, CAP3 und SP1) wurden entfernt, drei negative, nicht signifikante Residualvarianzen wurden auf null fixiert (ST2 im PHK-Modell, SP2 im FK-Modell und SP3 im PFK-Modell) und die Korrelation verschiedener Residualvarianzen wurden in allen Modellen (schwarze Linien) bzw. in den Modellen einzelner Berufsgruppen (farbige Linien) zugelassen. Die Korrelation von Residualvarianzen erfolgte, wenn a) die Modifikationsindizes dies nahelegten und dies b) aufgrund der Iteminhalte, sachlogischer Erwägungen so-

wie der Erfahrungen in den *Cognitive Debriefings* und Fragebogenerhebungen plausibel erschien. Auch die Analyse des Modells 3a offenbarte deutlichen Modifikationsbedarf. Insbesondere legte diese nahe, die Faktoren *Informal Interactions* und *Structural/Electronic Resources* weiter aufzuteilen und in allen, außer der TBK-Gruppe, eine Querladung des Items II4 auf beide IINT-Faktoren zuzulassen. Im FK-Modell lädt II5 zudem nicht wie in den anderen Modellen auf IINT_b, sondern auf IINT_a. Auch hier wurden Items entfernt (III1 und SER11) und Residualvarianzkorrelationen erlaubt, sofern dies aufgrund der zuvor beschriebenen Kriterien indiziert war. Modell 3b wies ebenfalls einen akzeptablen Fit auf. Ein 13-Faktoren-Modell (Zusammenführung der Modelle 2b und 3b) wies hingegen inakzeptable Fitwerte auf und war in der FK- und APS-Gruppe nicht zuverlässig schätzbar (Ableitungsmatrix nicht positiv definit). Daher erfolgte die Bestimmung der Messinvarianz getrennt für die Modelle 2b und 3b. Im Ein-Faktor-Modell der CRU-Skala (Modell 4, Abb. 9) wurde – wie in der kanadischen Studie (Squires et al., 2011b) – die Korrelation der Residualvarianzen der ersten beiden Items in allen Gruppen erlaubt. Die Korrelation der Residualvarianzen der Items 3 und 4 war hingegen nicht angezeigt. Im FK-Modell wurde zudem eine Korrelation der Residualvarianzen der Items 4 und 5 zugelassen. Dieses Modell wies in allen Gruppen einen sehr guten Fit auf und war Basis für die Mehrgruppenmodelle.

2. Konfigurale Invarianz: Für jedes der drei Modelle (2b, 3b und 4) erfolgte nun jeweils eine simultane Schätzung aller fünf Berufsgruppenvarianten. Alle Modellparameter wurden dabei in der jeweiligen Gruppe frei geschätzt. Die drei Gesamtmodelle wiesen einen akzeptablen Fit auf und waren somit Basis für den nächsten Schritt.

3. Schwache Messinvarianz: In diesem Schritt wurde untersucht, ob die Beziehung zwischen latenten Faktoren und assoziierten Indikatoren (Items) in den Gruppen gleich war. Hierfür wurden die Faktorladungen zwischen den Gruppen gleichgesetzt. Alle anderen Modellspezifikationen blieben gleich wie im Modell der konfiguralen Invarianz. Sind die Faktorladungen für ein Item in den Gruppen gleich, ist dies ein Zeichen dafür, dass eine Änderung des latenten Konstrukts um einen bestimmten Wert das Antwortverhalten (den Itemwert) in den einzelnen Grup-

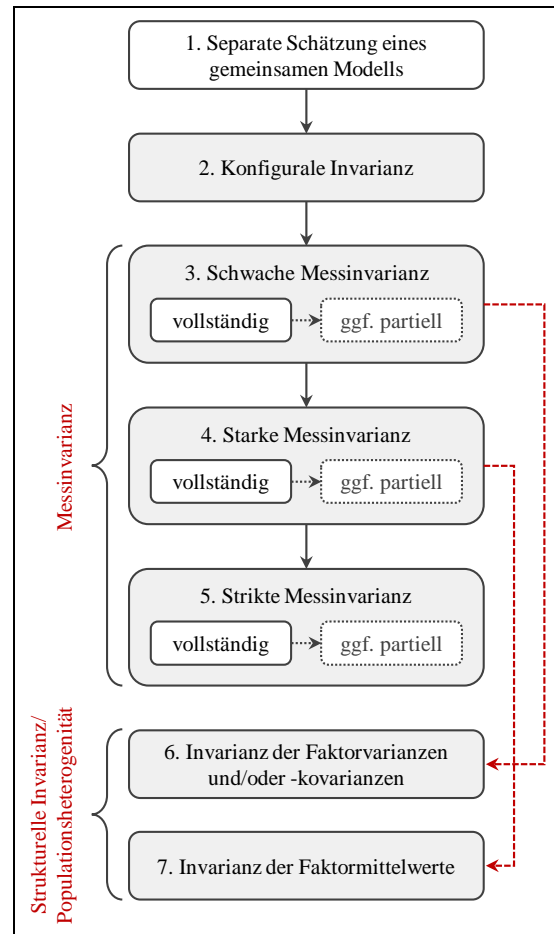


Abb. 7: Schritte der Faktorinvarianzanalyse

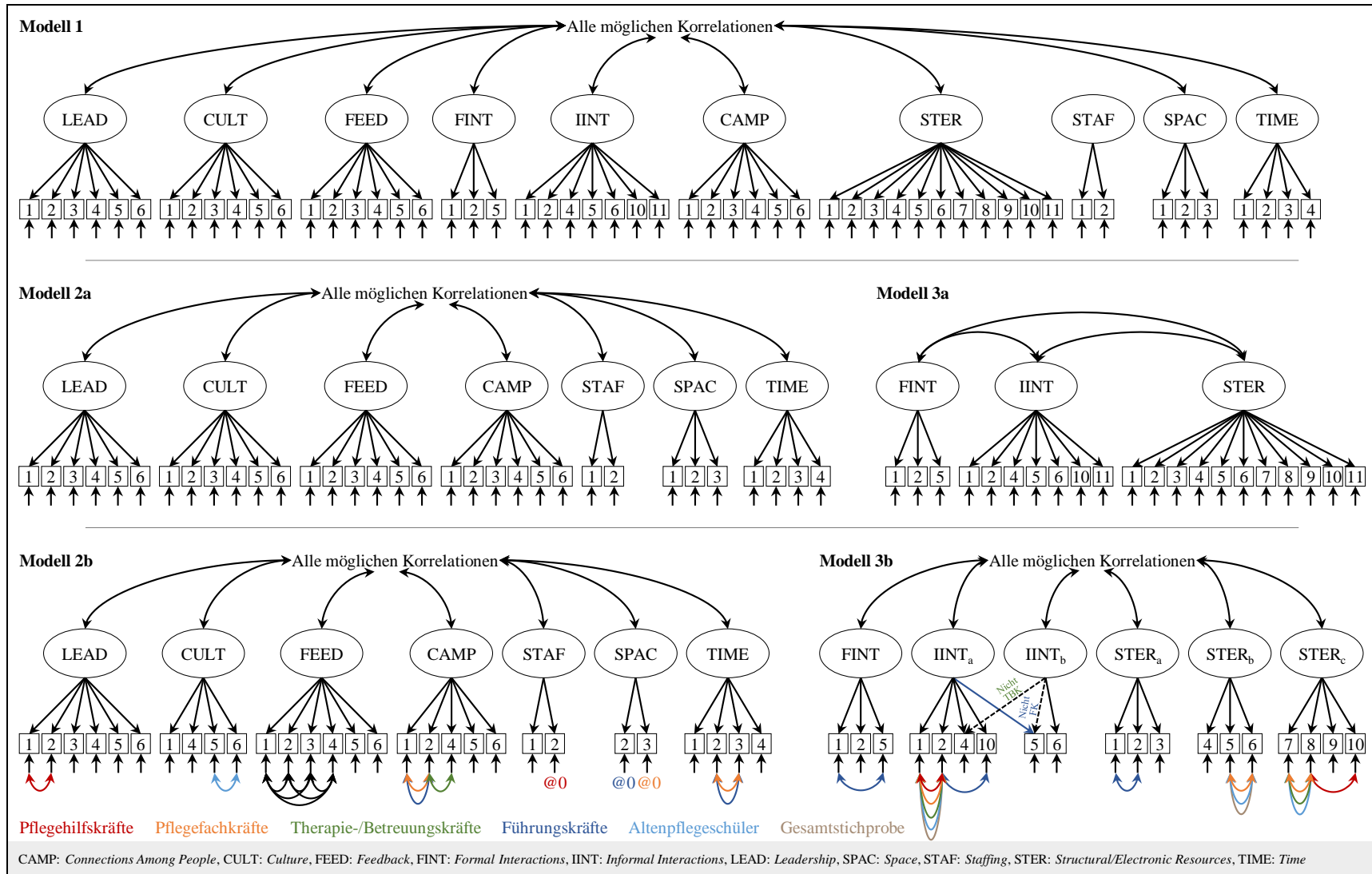


Abb. 8: Faktorenmodelle ACT

pen gleich stark beeinflusst. Schwache (metrische) Messinvarianz erlaubt Gruppenvergleiche der Beziehung zwischen latenten Faktoren und externen Variablen sowie von Faktorvarianzen und -kovarianzen. Nicht vergleichbar sind Strukturkoeffizienten und Faktormittelwerte, da zwar durch die gleichen Ladungen eine gemeinsame Skalierung gegeben ist, nicht jedoch ein gemeinsamer Startpunkt/Ursprung der Skala. Schwache Messinvarianz

wurde konstatiert, wenn das hier geschätzte Modell, verglichen mit dem Modell der konfiguralen Invarianz, keine signifikant schlechtere Modellpassung aufwies (gemessen am χ^2 -Differenztest, resultierender p-Wert $>0,05$). In Mplus erfolgt dieser Modellvergleich bei Verwendung des MLMV-Schätzers mittels der DIFFTEST-Funktion, die einen korrigierten χ^2 -Differenztest liefert (Muthén und Muthén, 2012). In diesem Fall erfolgte Schritt 4. Lag keine schwache Messinvarianz vor, wurden auf Basis der Modifikationsindizes Ladungen nach und nach freigegeben. Begonnen wurde mit der Ladung mit dem größten Modifikationsindexwert. Mussten weniger als 20% der potenziell schätzbaren Ladungen freigesetzt werden, um einen nichtsignifikanten Differenztest zu erhalten, wurde partielle schwache Messinvarianz konstatiert (Byrne et al., 1989; Vandenberg und Lance, 2000; Dimitrov, 2010) und mit Schritt 4 fortgefahren. Ansonsten endete die Kaskade.

4. Starke Messinvarianz: Ausgehend vom Vormodell (Faktorladungen zwischen den Gruppen gleichgesetzt bis auf die evtl. freigegebenen) wurden nun zusätzlich die Indikatorinterzepte (Itemmittelwerte) zwischen den Gruppen gleichgesetzt. Unterscheiden sich die Interzepte eines Items zwischen verschiedenen Gruppen, bedeutet dies, dass Mitglieder unterschiedlicher Gruppen unterschiedlich auf ein Item antworten, selbst wenn der Wert des latenten Faktors (z.B. Einrichtungskultur) in den Gruppen gleich ist (Itembias bzw. *differential item functioning*). Liegt starke Messinvarianz vor, können auch Faktormittelwerte und Strukturkoeffizienten zwischen den Gruppen verglichen werden. Dieses Modell wurde mit dem Modell der (partiellen) schwachen Messinvarianz verglichen. Lag starke Messinvarianz vor, erfolgte Schritt 5. Andernfalls wurden, entsprechend des Vorgehens in Schritt 3, nach und nach Interzepte freigegeben. Bei weniger als 20% freizusetzender Interzepte wurde partielle starke Messinvarianz konstatiert und Schritt 5 erfolgte. Andernfalls endete die Kaskade hier.

5. Strikte Messinvarianz: Analog des Vorgehens in den vorherigen Schritten wurden nun zusätzlich die Residualvarianzen der Items zwischen Gruppen gleichgesetzt und ggf. wieder freigesetzt bis (partielle) strikte Messinvarianz gegeben war oder abgelehnt werden musste. Strikte Messinvarianz ist erforderlich, um die beobachteten Itemwerte oder daraus abgeleitete Scores zwischen den Gruppen zu vergleichen. Liegt diese nicht vor, bedeutet dies, dass die Items in den Gruppen unterschiedlich präzise gemessen werden (mit unterschiedlichen „Fehlern“ behaftet sind). Gruppenunterschiede gehen also nicht ausschließlich auf den „wahren Wert“ (den latenten

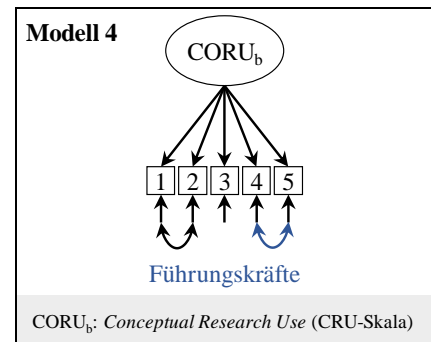


Abb. 9: Faktorenmodell CRU-Skala

Faktor) zurück. Modelle, die von einer „fehlerfreien“ Messung der Variablen ausgehen (z.B. generelle lineare Modelle) sind in diesem Fall für Gruppenvergleiche nicht geeignet. Erforderlich sind Modelle, die diese unterschiedlichen Residuen berücksichtigen (z.B. Strukturgleichungsmodelle) (Dimitrov, 2010; Sass, 2011).

6. Vergleich der Faktorvarianzen: Auf Basis der Modelle aus Schritt 4 wurden zusätzlich die Faktorvarianzen zwischen den Gruppen gleichgesetzt, ggf. nach und nach freigesetzt und das resultierende Modell mit dem Vorgängermodell verglichen.

7. Vergleich der Faktormittelwerte: Auch hier war das aus Schritt 4 resultierende Modell der Ausgangspunkt. Die Faktormittelwerte konnten aus Gründen der Modellidentifikation nicht direkt geschätzt werden. Geschätzt wurde die Abweichung der Mittelwerte von den in der definierten Referenzgruppe (PHK) auf null fixierten Mittelwerten.

4.5.5.3 Bestimmung der Reliabilität

Der wohl am meisten berichtete Reliabilitätskoeffizient ist Cronbach's α (Interne-Konsistenz-Reliabilität) – ungeachtet der Vielzahl an Publikationen, die eine Reihe von Problemen dieses Koeffizienten aufzeigen (vgl. für einen Überblick Green and Yang, 2009a; Sijtsma, 2009; Yang and Green, 2011). Insbesondere zwei Bedingungen, die Cronbach's α voraussetzt, sind in der Praxis selten erfüllt: a) essenzielle Tau-Äquivalenz (alle Items einer Skala hängen gleich stark und in gleicher Richtung mit dem zugrunde liegenden „wahren Wert“ zusammen – haben also die gleiche Ladung auf den jeweiligen Faktor) und b) unkorrelierte Residualvarianzen. Sind diese Voraussetzungen verletzt, sind gebiaste Cronbach's- α -Werte die Folge. Zuverlässiger lässt sich die Reliabilität einer Skala auf Basis von Strukturgleichungsmodellparametern bestimmen (Green and Hershberger, 2000; Raykov and ShROUT, 2002; Green and Yang, 2009b; Yang and Green, 2011). Die Reliabilität (im Folgenden ω) ist dabei der Quotient aus Varianz des wahren Skalenscores (σ_{Wahr}) und Gesamtvarianz ($\sigma_{\text{Ges.}}$) der Skala. $\sigma_{\text{Ges.}}$ ist die Summe aus σ_{Wahr} und der Residualvarianz ($\sigma_{\text{Res.}}$) der Skala. Daraus folgt: $\omega = \frac{\sigma_{\text{Wahr}}}{\sigma_{\text{Wahr}} + \sigma_{\text{Res.}}}$. Da jedes der ACT-Konzepte ein eigenes Konstrukt darstellt, wurde für jedes dieser Konzepte ein eigener Reliabilitätswert berechnet. In diesem Fall (jeweils ein Faktor) entspricht σ_{Wahr} der quadrierten Summe der Ladungen des jeweiligen Faktors und $\sigma_{\text{Res.}}$ der Summe der Residualvarianzen der zugehörigen Indikatoren (Items). Grundlage für die Reliabilitätsberechnungen waren die individuellen Faktorenmodelle 2b und 3b (ACT) sowie 4 (CRU-Skala). In diesen Modellen wurden alle Ladungen frei geschätzt. Um dennoch ein identifiziertes Modell zu erhalten, wurden statt der jeweils ersten Faktorladungen die Faktorvarianzen auf 1,0 fixiert.

4.5.5.4 Aggregierbarkeit der ACT-Daten auf Einrichtungs- bzw. Wohnbereichsebene

Das ACT wurde, wie in Abschnitt 2.3.1 ausgeführt, dafür entwickelt, den Organisationskontext

auf Wohnbereichs- oder Einrichtungsebene abzubilden. Daher müssen die Antworten der Individuen auf der jeweiligen Ebene aggregiert werden. Die Instrumentenentwicklerinnen aggregierten zu diesem Zweck die beobachteten Variablenwerte und überprüften die Güte der aggregierten Scores mittels einfaktoriellen ANOVAS und linearen Mehrebenenregressionsmodellen. Diesem Ansatz wurde in dieser Arbeit aus zwei Gründen nicht gefolgt. Zum einen zeigten die Messinvarianzanalysen, dass die zum gruppenübergreifenden Vergleich beobachteter Werte erforderliche strikte Messinvarianz für die deutsche ACT-Version nicht gegeben ist. Zum anderen konnten Lüdtke et al. (2008) zeigen, dass Mehrebenenansätze, die manifeste (beobachtete) Werte zu Clusterebenenvariablen aggregieren (z.B. der Wohnbereichsdurchschnitt der Führungsvariablen des ACT), oft gebiaste Ergebnisse für die Level-2- bzw. -3-Effekte liefern. Der von Muthén und Kollegen entwickelten *Multilevel-Latent-Covariate-Approach* (Muthén, 1989, 1994; Lüdtke et al., 2008; Asparouhov and Muthén, 2010) vermeidet diesen Bias. Wie Abb. 10 zeigt, werden dabei die Variablen zerlegt in eine *Within*-Komponente (erklärt intraindividuelle Unterschiede auf Level 1 – in diesem Fall individuelle Probanden) und eine *Between*-Komponente (erklärt Unterschiede auf Clusterebene – in diesem Fall Einrichtung oder Wohnbereich). Die *Within*-Komponente der Faktorenmodelle wird exakt so spezifiziert wie die individuellen Faktorenmodelle. Zusätzlich wird auf der *Between*-Ebene für jeden *Within*-Faktor ein *Between*-Pendant spezifiziert. Dessen Indikatoren sind die Clusterebeneninterzepte der Level-1-Indikatoren. Diese Interzepte werden als latente Variablen behandelt und geschätzt. Mplus gibt – neben der Intraklassenkorrelation für jeden Indikator – den Gesamtmodell-

fit aus sowie den SRMR-Wert und die geschätzten Parameter, getrennt nach *Within*- und *Between*-Part des Modells. Die Modelle 2b, 3b (ACT) und 4 (CRU-Skala) wurden als Mehrebenenfaktorenmodelle spezifiziert und für jede der Berufsgruppen separat sowie für die gesamte Stichprobe geschätzt – jeweils mit Einrichtungs- und Wohnbereichsebene als Clustervariable. Als robuster Schätzer wurde für diese Modelle der WLSMV-Schätzer eingesetzt, da der MLMV-Schätzer in Mplus für Mehrebenenmodelle nicht verfügbar ist (Muthén und Muthén, 2012). Die Modelle dienten dazu, zu bestimmen, ob die Einrichtungs-/Wohnbereichszugehörigkeit das Antwortverhalten beeinflusste (ob also die Antworten innerhalb der Cluster ähnlich waren und signifikant zwischen den Clustern variierten). Entgegen der kanadischen Befunde konnten in dieser Arbeit keine signifikanten Clustereffekte nachgewiesen werden. In allen Modellen erwiesen sich die Faktorvarianzen und ein Großteil der Faktorladungen auf der *Between*-Ebene als nicht signifikant.

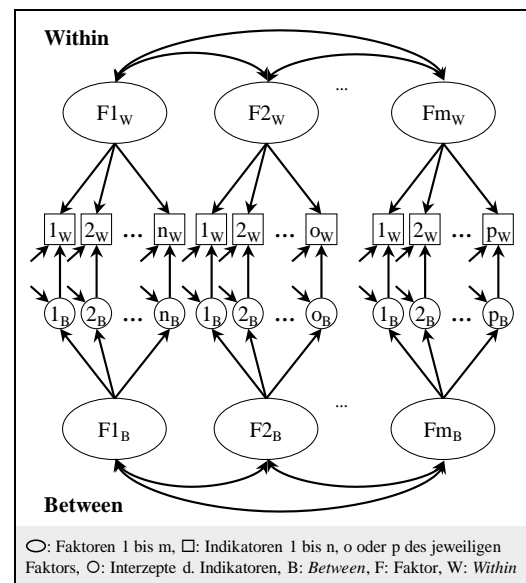


Abb. 10: Mehrebenenfaktorenmodelle

fit aus sowie den SRMR-Wert und die geschätzten Parameter, getrennt nach *Within*- und *Between*-Part des Modells. Die Modelle 2b, 3b (ACT) und 4 (CRU-Skala) wurden als Mehrebenenfaktorenmodelle spezifiziert und für jede der Berufsgruppen separat sowie für die gesamte Stichprobe geschätzt – jeweils mit Einrichtungs- und Wohnbereichsebene als Clustervariable. Als robuster Schätzer wurde für diese Modelle der WLSMV-Schätzer eingesetzt, da der MLMV-Schätzer in Mplus für Mehrebenenmodelle nicht verfügbar ist (Muthén und Muthén, 2012). Die Modelle dienten dazu, zu bestimmen, ob die Einrichtungs-/Wohnbereichszugehörigkeit das Antwortverhalten beeinflusste (ob also die Antworten innerhalb der Cluster ähnlich waren und signifikant zwischen den Clustern variierten). Entgegen der kanadischen Befunde konnten in dieser Arbeit keine signifikanten Clustereffekte nachgewiesen werden. In allen Modellen erwiesen sich die Faktorvarianzen und ein Großteil der Faktorladungen auf der *Between*-Ebene als nicht signifikant.

4.5.5.5 Zusammenhänge zwischen den Variablen: Strukturgleichungsmodelle

Da im vorangegangenen Schritt keine signifikanten Clustereffekte gefunden wurden, wurden in diesem Schritt keine Mehrebenenstrukturgleichungsmodelle spezifiziert. Da zudem einzelne Substichproben zu klein waren, um diese komplexen Modelle separat zu schätzen (APS 65 und FK 129 Probanden), wurden diese nur in der Gesamtstichprobe geschätzt. Dies war möglich, da die Mehrgruppenmodelle eine partielle starke Messinvarianz sowie partielle Faktorvarianz-Invarianz ergaben. Unterschiede zwischen den Gruppen gehen also überwiegend zurück auf Unterschiede des latenten Konstrukts. Ein zusätzliches Mehrebenenfaktorenmodell mit Gruppenzugehörigkeit als Clustervariable bestätigte die Invarianz der Faktorvarianzen zwischen den Gruppen. Die Modellspezifikation und -evaluation entsprach den in Abschnitt 4.5.5.1 beschriebenen Prinzipien.

In einem ersten Schritt wurde ein Gesamtmodell für alle RU-Variablen spezifiziert (Modell 5, Abb. 11). Das CRU-Item der *Estabrooks' Kinds of RU items* wurde nicht be-

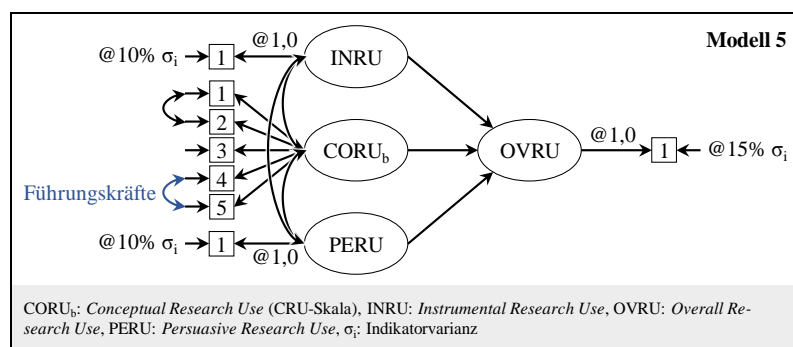


Abb. 11: Gesamtmodell aller RU-Variablen

berücksichtigt, da es a) in den PHK- und APS-Fragebögen nicht enthalten war und b) dieses Konzept durch die CRU-Skala deutlich besser abgebildet wird. Grundlage für die Modellspezifikation waren die Arbeiten von Estabrooks (1999a, 1999b), die zeigten, dass *Instrumental*, *Conceptual* und *Persuasive RU* keine Indikatoren von *Overall RU* sind (also durch *Overall RU* beeinflusst werden), sondern Prädiktoren für *Overall RU*. Die drei Ein-Item-Konzepte wurden auf Basis der Empfehlungen von Hayduk and Littvay (2012) spezifiziert: Um das Modell zu identifizieren und dem jeweiligen latenten Konstrukt (INRU, PERU und OVRU) eine Skala zuzuweisen, wurde die Faktorladung jeweils auf 1,0 fixiert. Die Residualvarianz kann in einem solchen Modell nicht geschätzt werden und muss daher auf Basis vorheriger Forschungsbefunde, theoretischer Erwägungen und der Erfahrungen im Feld festgelegt werden. Aufgrund der Verständnisschwierigkeiten einiger der Probanden (vgl. Abschnitt 5.2.2) erschien die von Estabrooks (1999a) gewählte Festlegung (1% der Itemvarianz) als zu gering. Für IRU und PRU wurde daher eine Residualvarianz von 10% der Itemvarianz festgelegt, für ORU 15%. Entsprechend der Empfehlungen von Hayduk and Littvay (2012) wurden diese Residualvarianzwerte der Reihe nach (immer ein Wert nach dem anderen) zunächst halbiert und anschließend verdoppelt und das Modell jedes Mal erneut geschätzt. Die Modellparameter und -fitwerte blieben nahezu konstant, was darauf hindeutete, dass die Residualvarianzwerte das Modell nicht substantiell beeinflussten.

In diesem Schritt wurden auch die individuellen Variablen (COPSOQ- und ARU-Items) einbezogen. Abb. 12 zeigt das entsprechende Faktorenmodell. Das Ein-Item-Konstrukt Allgemeine

Gesundheit (GEHE) wurde entsprechend der vorangegangenen Ausführungen spezifiziert. Dieses Modell wurde nur in der Gesamtstichprobe geschätzt. Es wies einen akzeptablen Fit auf und wurde daher in die Modelle für die Zusammenhangsberechnungen aufgenommen.

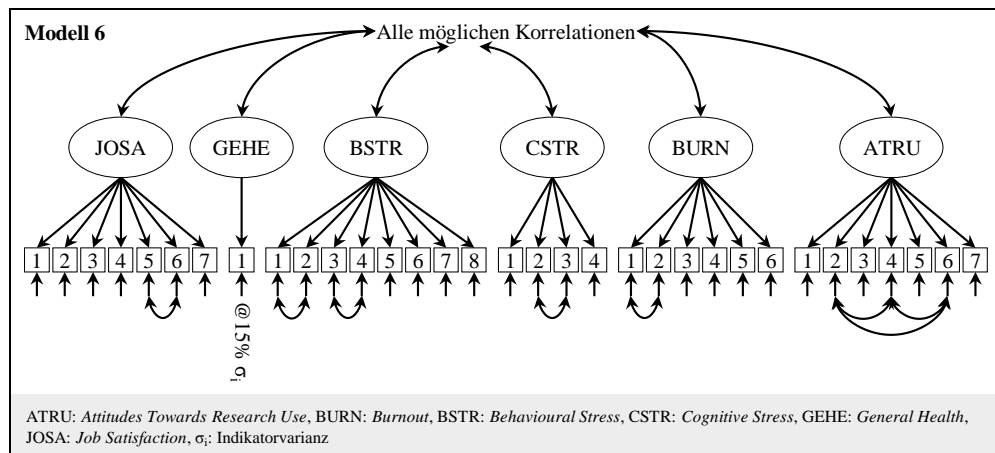


Abb. 12: Faktormodell der individuellen Variablen

Abschließend wurden sechs Strukturgleichungsmodelle geschätzt (Modelle 7a/b, 8a/b und 9a/b, Abb. 13). Die Strukturgleichungsmodelle überprüften folgende Hypothesen:

Modell 7a: Bessere Werte der individuellen Variablen (höhere Arbeitszufriedenheit, bessere allgemeine Gesundheit, weniger verhaltensbezogener und kognitiver Stress, weniger Burnout sowie positivere Einstellung zur Nutzung von Forschungswissen) und höhere Werte instrumenteller, konzeptueller und persuasiver Forschungsnutzung gehen einher mit höheren Werten allgemeiner Forschungsnutzung.

Modell 8a: Bessere Organisationskontextwerte (Führung, Feedback, Kultur, Miteinander im Team, Personalbesetzung, Räumlichkeiten und mehr Zeit) und höhere Werte instrumenteller, konzeptueller und persuasiver Forschungsnutzung gehen einher mit höheren Werten allgemeiner Forschungsnutzung.

Modell 9a: Bessere Organisationskontextwerte (mehr offizielle Zusammenkünfte, mehr informeller Austausch und mehr Nutzung struktureller und elektronischer Ressourcen) und höhere Werte instrumenteller, konzeptueller und persuasiver Forschungsnutzung gehen einher mit höheren Werten allgemeiner Forschungsnutzung.

Modell 7b: Bessere Werte der individuellen Variablen (höhere Arbeitszufriedenheit, bessere allgemeine Gesundheit, weniger verhaltensbezogener und kognitiver Stress, weniger Burnout sowie positivere Einstellung zur Nutzung von Forschungswissen) gehen einher mit höheren Werten instrumenteller, konzeptueller und persuasiver Forschungsnutzung, die wiederum Mediatoren für allgemeine Forschungsnutzung sind.

Modell 8b: Bessere Organisationskontextwerte (Führung, Feedback, Kultur, Miteinander im Team, Personalbesetzung, Räumlichkeiten und mehr Zeit) gehen einher mit höheren Werten instrumenteller, konzeptueller und persuasiver Forschungsnutzung, die wiederum Mediatoren für all-

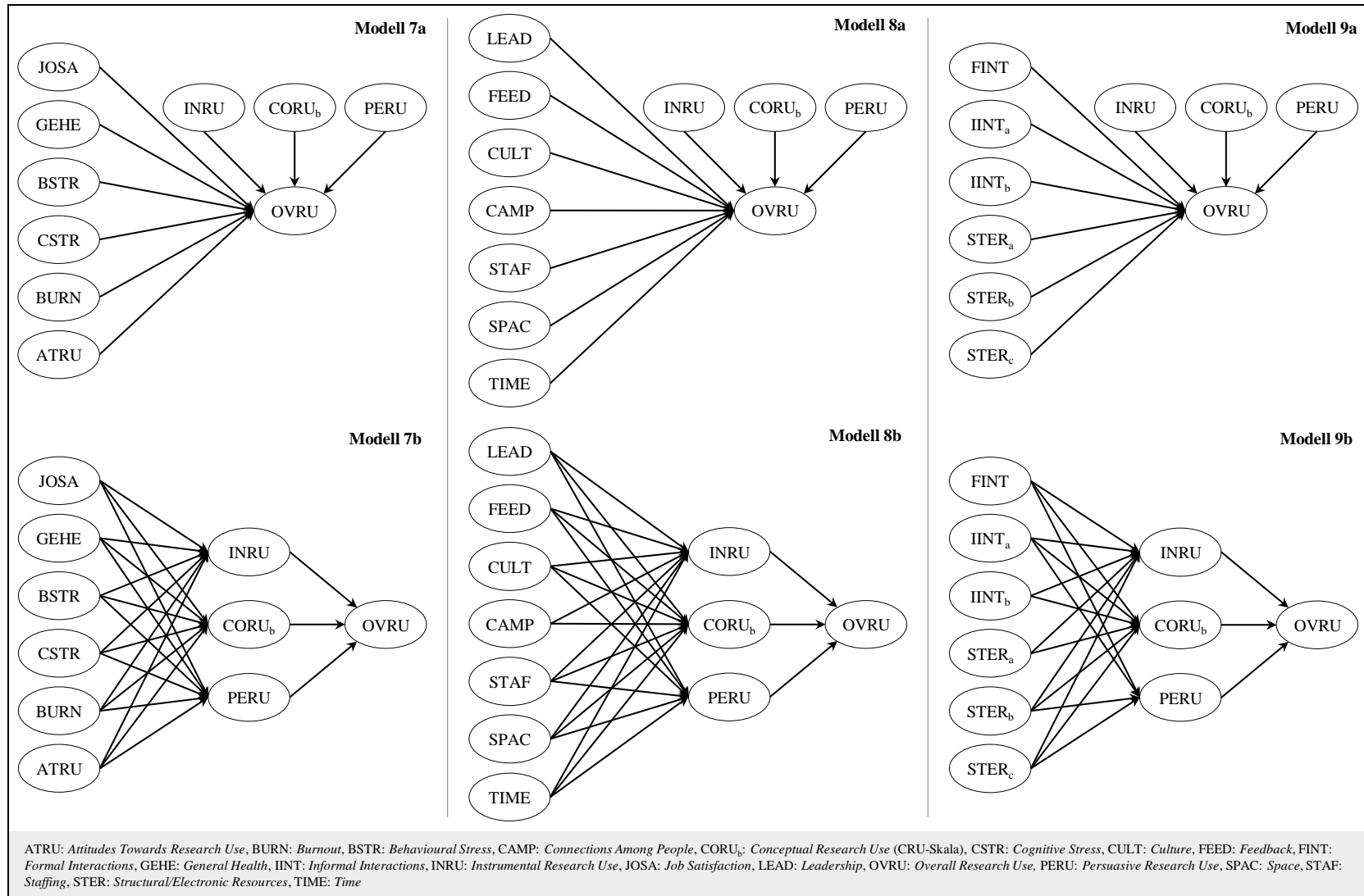


Abb. 13: Strukturgleichungsmodelle (nur Strukturkomponenten; Messmodelle entsprechen denen der zuvor vorgestellten Faktorenmodelle)

gemeine Forschungsnutzung sind.

Modell 9b: Bessere Organisationskontextwerte (mehr offizielle Zusammenkünfte, mehr informeller Austausch und mehr Nutzung struktureller und elektronischer Ressourcen) gehen einher mit höheren Werten instrumenteller, konzeptueller und persuasiver Forschungsnutzung, die wiederum Mediatoren für allgemeine Forschungsnutzung sind.

5 Ergebnisse

5.1 Erkenntnisse aus dem Übersetzungsprozesses

Der Übersetzungsprozess nahm insgesamt 16 Monate in Anspruch, wobei die Übersetzung der PHK-Fragebögen am längsten dauerte (286 Tage vom Beginn des Übersetzungsprozesses bis zur endgültigen Genehmigung aller Textelemente durch die Entwicklerinnen). Drei Abstimmungsrunden mit den Entwicklerinnen (Schritt 7 im Übersetzungsprozess) waren erforderlich, bevor alle Items für die *Cognitive Debriefings* genehmigt waren. In den *Cognitive Debriefings* (Schritt 8) wurden drei Runden benötigt, da verschiedene Items modifiziert und erneut getestet werden mussten, und der resultierende Wortlaut erforderte zwei weitere Abstimmungsrunden mit den Entwicklerinnen. Der Zeitbedarf reduzierte sich mit jeder weiteren Fragebogenversion: 175 Tage PFK (nur noch zwei Runden in Schritt 7 und eine *Cognitive-Debriefing*-Runde), 102 Tage TBK, 32 Tage Exp. und 29 Tage FK (jeweils nur eine Runde in Schritt 7 und 8). Dies lag zum einen daran, dass sich die Zahl der zu übersetzenden Textelemente mit jeder weiteren Version reduzierte. Während die PHK-Fragebögen noch komplett übersetzt werden mussten, konnten die im Wortlaut übereinstimmenden Textelemente später in die anderen Fragebögen übernommen werden. Zum anderen war die Übersetzung der PHK-Fragebögen auch inhaltlich mit den größten Herausforderungen verbunden. Es handelt sich um eine Personengruppe mit durchschnittlich eher niedrigem Bildungsgrad und sehr heterogenen Sprachkompetenzen, so dass es z.T. schwierig war, verständliche Formulierungen zu finden. Schließlich lernten die Übersetzer während des Prozesses kontinuierlich dazu und es gelang ihnen immer besser, die Items angemessen zu formulieren, potenzielle Schwierigkeiten zu antizipieren und kreative Lösungen zu erarbeiten.

Die Herausforderungen im Übersetzungsprozess können in sieben Kategorien unterteilt werden (Abb. 14). Sie resultierten vorwiegend aus einem Spannungsverhältnis zweier z.T. widersprüchlicher Anforderungen: Einerseits sollten die übersetzten Instrumente den kanadischen Originalen so ähnlich wie möglich sein, was Zahl und Inhalte der Konzepte, Items und Skalen angeht, um bestmögliche Vergleichbarkeit zu gewährleisten. Andererseits galt es, die Instrumente an die Gegebenheiten der deutschen stationären Altenpflege anzupassen, um deren Reliabilität und Validität in diesem Kontext sicherzustellen. Sowohl Vergleichbarkeit als auch inhaltliche Güte zu gewährleisten, war nicht immer ganz einfach. Die einzelnen Herausforderungen, die gewählten Lösungsstrategien sowie die vorgenommenen Adaptationen der Tools werden nachfolgend beschrieben.

1. Unterschiedliche Qualifikationsstruktur in der kanadischen und deutschen Altenpflege:

Während Pflege- und Therapieberufe hierzulande überwiegend in konventionellen dualen Ausbildungsgängen erlernt werden, ist dafür in Kanada um ein Studium (mindestens auf Bachelor-, oft auf Masterniveau) erforderlich. Entsprechende grundständige Studiengänge existieren in Deutschland und nehmen mehr und mehr zu, doch die Mehrzahl der in der Praxis tätigen Fachkräfte sind im Moment noch konventionell ausgebildet. Dies wirkte sich in zweifacher Weise auf die Übersetzung der PFK- und TBK-Fragebögen aus: Zum einen war nicht davon auszugehen,

dass die deutschen Fachkräfte ähnlich vertraut mit Konzepten wie „*best practice*“, „Forschung“ oder „wissenschaftliches Wissen“ sind wie ihre kanadischen Kollegen. Daher lehnten sich die deutschen Übersetzungen stärker an den Wortlaut der PHK-Fragebögen an als die kanadischen Originale. Zum anderen existiert in Deutschland kein Pendant zur kanadischen „*Licensed Practical Nurse*“ – einer zweijährigen Pflegefachausbildung. Daher wurde das entsprechende ACT-Item in den deutschen Fragebögen entfernt.

2. Deutsche Übersetzung einfacher englischer Sätze wird lang und kompliziert: Aus unterschiedlichen Gründen sind deutsche Sätze oft bedeutend länger und komplexer als gleichbedeutende englische Sätze. Dies liegt u.a. an a) den deutschen Relativsatzkonstruktionen, b) unterschiedlichen Geschlechterformen zahlreicher deutscher Begriffe, c) der häufigen Notwendigkeit, einfache englische Begriffe mit mehr als einem Wort (z.T. einem ganzen Nebensatz) zu beschreiben, d) dem Umstand, dass schon einzelne deutsche Wörter oft bedeutend länger sind als ihr englisches Äquivalent, etc. Eines der kanadischen ACT-Items lautete z.B.:

„We have *private space* such as a conference room on this unit or floor (*other than* at the bedside, in the hallway or medication room) to discuss resident care plans and share knowledge about resident care and best practices.” (ACT, Space, Item 2, Hervorhebungen im Original)

Die erste Übersetzung war nahezu doppelt so lang, komplex und schwer zu lesen:

„Wir verfügen über einen *nicht öffentlichen Rückzugsraum* wie z.B. einen Besprechungsraum auf diesem Wohnbereich oder Stockwerk (nicht gemeint sind: Zimmer der Bewohnerinnen und Bewohner, Gang, Treppenhaus, Foyer oder Medikamentenlager), in dem wir über die Pflegeplanungen der Bewohnerinnen und Bewohner sprechen und Informationen über die Pflege und Betreuung der Bewohnerinnen und Bewohner sowie über optimale Praxis austauschen können.”

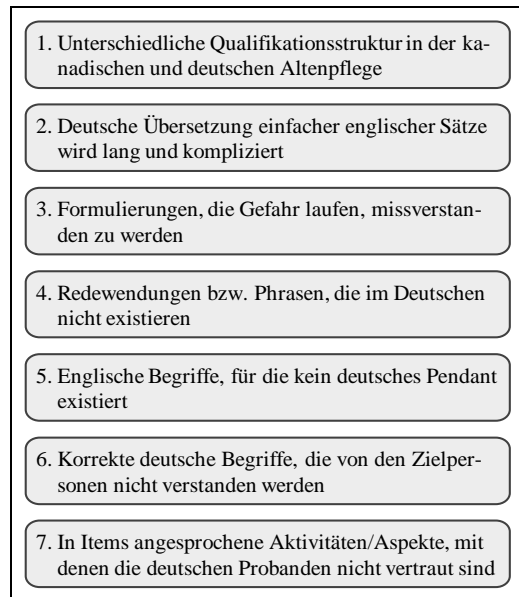


Abb. 14: Herausforderungen im Übersetzungsprozess (Hoben et al., 2013)

Eine Lösung dieses Problems bestand darin, aus einem langen Satz zwei oder mehrere kürzere zu machen. Zum anderen wurde nur die männliche Sprachform verwendet. Auch konnten durch Umformulierung manche Passagen des englischen Originals weggelassen werden. Auf diese Weise gelang es, die abschließende Version bedeutend knapper und einfacher zu halten:

„Wir verfügen über einen *nicht öffentlichen Personal- oder Besprechungsraum* auf diesem Wohnbereich oder Stockwerk. Diesen können wir nutzen, um über die Pflegeplanungen zu sprechen sowie um Wissen über die optimale Pflege und Betreuung der Bewohner auszutauschen.“

3. Formulierungen, die Gefahr laufen, missverstanden zu werden: Dies betraf z.B. die Frage, wie häufig die Probanden etwas „im letzten Jahr“ („*in the last year*“) getan hatten. Wenn ein Teilnehmer dies etwa im Dezember 2012 gefragt würde, kann es sein, dass er an das Jahr 2011 (Januar bis Dezember) denkt, anstatt wie eigentlich beabsichtigt an die letzten 12 Monate (Dezember 2011 bis Dezember 2012). Ein anderes Beispiel ist oben angesprochene Frage nach „*private space*“. Diese enthält Irritationspotenzial, weil diese Räumlichkeiten – wenngleich nicht öffentlich zugänglich – trotz allem dienstlich sind und nicht privat. Solche Formulierungen wurden in der deutschen Übersetzung präzisiert.

4. Redewendungen bzw. Phrasen, die im Deutschen nicht existieren: Eine besondere Herausforderung in dieser Kategorie stellte ein Item dar, das danach fragt, ob genug Personal zur Verfügung steht, um sicherzustellen „*that residents have the best day*“. Eine wörtliche Übersetzung ins Deutsche ergäbe keinen Sinn. Im Austausch mit den Entwicklerinnen wurde die Bedeutung dieser Phrase nach und nach herausgearbeitet. Die deutsche Übersetzung reflektiert schlussendlich, dass es darum geht, „das Wohlbefinden der Bewohner in ihrem Alltag bestmöglich zu gewährleisten“.

5. Englische Begriffe, für die kein deutsches Pendant existiert: Beispiele für diese Kategorie sind die Begriffe „*to coach*“, „*best practice*“ oder „*performance*“. Als besondere Herausforderung erwies sich hier *best practice*. Dieser Begriff wurde in der kanadischen Version speziell für die PHK eingeführt, da diese Schwierigkeiten hatten, „*research use*“ oder „*scientific knowledge*“ zu verstehen. Während „*best practice*“ im Deutschen oft in der englischen Form verwendet wird, wurde der Begriff gerade von den deutschen PHK besonders schlecht verstanden und auch die anderen Formulierungen (Wissenschaftliche Erkenntnisse oder Forschungswissen) waren keine geeigneten Alternativen. Als beste Lösung erwies sich auch hier, die Begriffe so gut wie möglich zu umschreiben. „Optimale Pflege und Betreuung“ bzw. „Fachwissen zur optimalen Pflege und Betreuung“ lautet die abschließende deutsche Übersetzung.

6. Korrekte deutsche Begriffe, die von den Zielpersonen nicht verstanden werden: Für englische Begriffe wie „*formal*“, „*informal*“ oder „*rating scale*“ existieren zwar adäquate deutsche Übersetzungen („formell“, „informell“ oder „Einschätzungsskala“). Insbesondere in der PHK-Gruppe bereiteten diese jedoch Probleme. Neben der Strategie, diese Begriffe zu umschrei-

ben, wurden hier zusätzlich Beispiele zur Illustration eingeführt, die in den kanadischen Originalfragebögen nicht enthalten waren.

7. In Items angesprochene Aktivitäten/Aspekte, mit denen die deutschen Probanden nicht vertraut sind: Ein Beispiel für eine solche Aktivität sind „*family conferences*“ – eine in kanadischen Pflegeheimen gängige Besprechungsform, in der Bewohner und deren Angehörige sowie verschiedene Berufsgruppen (Pfleger, Therapeuten, Führungskräfte, Seelsorger, Sozialarbeiter, Ärzte, Apotheker etc.) integriert sind. Im Gegensatz dazu findet der Austausch mit Familienangehörigen in deutschen Pflegeheimen überwiegend anlassbezogen, spontan und ungeplant statt. Am ehesten planen noch Sozialdienstmitarbeiter, Pflege- oder Einrichtungsleitungen diese Gespräche. Daher wurde vom Originalkonzept abgewichen. In der deutschen Übersetzung wird nach „geplanten Besprechungen mit Angehörigen“ gefragt.

Ein weiteres Beispiel ist die Frage danach, in wie weit die Teilnehmer regelmäßige Informationen über die Qualitätssituation auf ihrem Wohnbereich erhalten. Gemeint sind datenbasierte Informationen, die die Gesamtsituation des Wohnbereichs beschreiben (z.B. Qualitätsindikatoren wie Sturzrate oder Schmerzsituation). In kanadischen Pflegeheimen wird diese Art Informationen routinemäßig erhoben und zurückgemeldet. In fast allen Provinzen ist die Anwendung des Resident Assessment Instrument (RAI) Pflicht – ein standardisiertes Erhebungsinstrument mit dem Bewohner regelmäßig in 18 verschiedenen Bereichen eingeschätzt werden und aus dem verschiedene Qualitätsindikatoren ableitbar sind (Hirdes et al., 2011). Zwar sind auch deutsche Pflegeheime verpflichtet, ähnliche Qualitätsinformationen zu erheben, bestimmte Instrumente sind dafür jedoch nicht vorgeschrieben und so variieren die erhobenen Daten hinsichtlich Umfang und Qualität substantiell. Rückmeldungen an die Teams erfolgen eher bezogen auf individuelle Bewohner, weniger in aggregierter Form für den Wohnbereich und für einen bestimmten Zeitraum. Insofern bestand die Herausforderung darin, die Zielpersonen dazu zu bringen, an die richtige Art von Informationen zu denken und „stimme nicht zu“ oder „stimme überhaupt nicht zu“ anzukreuzen, wenn sie überwiegend individuelle Bewohnerinformationen erhielten. Daher wurde das Wort „Statistik“ eingeführt. Außerdem wurden auch hier Beispiele hinzugefügt (z.B. Zahl der Bewohner, die im letzten halben Jahr auf diesem Wohnbereich gestürzt sind) und es wurde ausdrücklich betont, dass Informationen über einzelne Bewohner nicht gemeint sind.

5.2 Resultate der *Cognitive Debriefings*

5.2.1 Beschreibung des *Cognitive-Debriefing-Samples*

An den *Cognitive Debriefings* nahmen 39 Probanden aus fünf stationären Altenpflegeeinrichtungen der Metropolregion Rhein-Neckar teil. Die mediane Platzzahl der Einrichtungen war 163 (zwischen 82 und 217). Zwei der 16 PHK hatten einen Abschluss als Altenpflegehelfer, eine als Krankenpflegehelfer. Von den fünf PFK hatten drei ein Altenpflege- und zwei ein Gesundheits- und Krankenpflegeexamen. Unter den sechs TBK waren zwei Ergotherapeuten, ein Arbeitserzie-

her, eine Musiktherapeutin, eine Hilfskraft der sozialen Betreuung und zwei Alltagsbegleiter für Menschen mit Demenz nach §87b SGB XI. Die fünf Experten teilten sich auf in drei Qualitätsbeauftragte und zwei Dozenten für Fort- und Weiterbildung. Die Gruppe der Leitenden bestand aus einer Einrichtungsleitung, zwei Pflegedienstleitungen und drei Wohnbereichsleitungen. Im Mittel waren die Probanden 40,26 Jahre alt (Standardabweichung: 10,58 Jahre). Neun der Probanden waren keine deutschen Muttersprachler. Sechs davon gaben polnisch als ihre Muttersprache an, zwei russisch und eine Person spanisch. Im Mittel sprachen diese Personen seit 13,56 Jahren deutsch (zwischen 2 und 39 Jahren, Standardabweichung: 12,76 Jahre). Weitere soziodemografische Details sind in Tab. 14 aufgeführt.

Tab. 14: Soziodemografische Merkmale *Cognitive-Debriefing*-Sample

Merkmal	N
Geschlecht	
Weiblich	35
Männlich	4
Alter	
20-29 Jahre	7
30-39 Jahre	13
40-49 Jahre	9
50-59 Jahre	10
Muttersprache	
Deutsch	30
Nicht Deutsch	9
Berufserfahrung	
<1 Jahre	8
1-4 Jahre	7
5-9 Jahre	8
10-14 Jahre	3
15-20 Jahre	5
20-24 Jahre	3
>24 Jahre	5
Höchster Bildungsabschluss	
Haupt-/Volksschule	4
Realschule	1
Berufsausbildung	23
Gymnasium	7
Akademischer Abschluss	4

5.2.2 *Cognitive Debriefings* mit den Pflegehilfskräften

In der ersten *Cognitive-Debriefing*-Runde mit den Pflegehilfskräften erwiesen sich elf der 58 ACT-Items und alle RU-Items als problematisch – wurden also von mindestens zwei Teilnehmern nicht korrekt verstanden (Tab. 15). Ein Beispiel für ein Item, das keine Verständnisprobleme bereitete ist SP2:

Original: „We have *private space* such as a conference room on this unit or floor (*other than* at the bedside, in the hallway or medication room) to discuss resident care plans and share knowledge about resident care and best practices.”

Übersetzung: „Wir verfügen über einen *nicht öffentlichen Personal- oder Besprechungsraum* auf diesem Wohnbereich oder Stockwerk. Diesen können wir nutzen, um über die Pflegeplanungen zu sprechen sowie um Wissen über die optimale Pflege und Betreuung der Bewohner auszutauschen.“

Die Frage des Interviewers (in diesem Fall eine allgemeine Probe) lautete:

„An welchen Raum haben Sie beim Beantworten dieser Frage gedacht? Können Sie diesen bitte beschreiben?“

Die Antwort der Probandin lässt klar erkennen, dass sie das Item verstand:

„Also wir haben zwei. Wir haben einen, wo das Stationszimmer, also das Dienstzimmer ist, und dann haben wir einen, das ist unser, ich sage mal: der Pausenraum. Also wenn wir wirklich ungestört über pflegerische Dinge reden wollen, und es soll kein Bewohner oder Angehöriger reinkommen, dann gehen wir da hin. Da wird halt geklopft, wenn jemand was will.“

Die Teilnehmerin erkannte, dass a) nach einem ungestörten, für Unbefugte nicht ohne Weiteres zugänglichen Raum gefragt wurde, der b) für Gespräche über Bewohner oder die Pflegepraxis genutzt wird.

Ein Beispiel für ein problematisches Item ist C3 (Tab. 15). Auf die Bitte des Interviewers, den Inhalt des Items in eigenen Worten wiederzugeben, antwortete eine Teilnehmerin:

„Ob wir es hier schaffen, für die Bewohner da zu sein und die Pflege so zu machen, wie die das gerne möchten und brauchen.“

Der Interviewer hakte mit einer Verständnisprobe nach:

„Und was bedeutet „Ausgleich“ für Sie?“

Antwort:

„Dass man dem entgegen kommt, wie es der Bewohner von früher her gewohnt ist, dass man darauf eingeht.“

Es folgte eine weitere Verständnisprobe des Interviewers:

„Und was verstehen Sie unter „Wirtschaftlichkeit“?“

Antwort der Probandin:

„Ich kann mit dem Wort nichts anfangen. Keine Ahnung.“

Das zweite Beispiel zeigt deutlich, dass das Item nicht verstanden wurde. Die Teilnehmerin bezog sich ausschließlich auf den ersten Teil der Frage (Pflege- und Betreuungsqualität) und konnte weder den Aspekt des Ausgleichs bzw. der Balance zwischen zwei Dingen einordnen, noch den Gegenpart zur Pflege- und Betreuungsqualität – die wirtschaftlichen Aspekte.

Als Konsequenz dieser Ergebnisse wurden vier der elf problematischen ACT-Items und alle RU-Items modifiziert. Das *Formal-Interactions*-Item (Fort- und Weiterbildungen außerhalb der Einrichtung) wurde nicht modifiziert. Zwei Teilnehmer hatten den Text nicht aufmerksam genug gelesen und an Schulungen in ihrer Einrichtung gedacht. Dieses Problem war durch eine Umformulierung nicht zu lösen. Die sechs *Feedback*-Items (Frage nach datenbasierten Informationen über die Pflegequalität auf dem Wohnbereich und dem Umgang damit) wurden ebenfalls nicht verändert. Ursache der Verständnisprobleme waren nicht die Items selbst, sondern der Einleitungstext. Die Interviewten dachten überwiegend an Informationen über individuelle Bewohner (anstatt an Daten für den gesamten Wohnbereich wie z.B. die Sturzrate im letzten Halbjahr) und bekundeten demzufolge (fälschlicherweise) bei allen Items, diese Informationen sehr häufig zu erhalten und auch stets Konsequenzen daraus abzuleiten. Die vorgenommenen Veränderungen zielten darauf, dieses Missverständnis zu vermeiden und den Wortlaut zu präzisieren. Daher wurde für Runde zwei der Hinweis hinzugefügt, gefragt werde nach „Statistiken für den gesamten Wohnbereich und für einen bestimmten Zeitraum“. Das Hauptproblem der RU-Items und des ACT-

Tab. 15: Problematische Items in den *Cognitive Debriefings* mit den Pflegehilfskräften und Revisionshistorie (Hoben et al., submitted)

Konzept	Items	Original	Übersetzung Runde 1	Übersetzung Runde 2	Übersetzung Runde 3
Culture	3, 6	Beispielitem: My organization effectively <i>balances</i> best practice and productivity.	Meine Einrichtung schafft erfolgreich <i>den Ausgleich</i> zwischen optimaler Pflege- und Betreuungsqualität und Wirtschaftlichkeit.	Meiner Einrichtung gelingt es, optimale Pflege und Betreuungsqualität und finanzielle Interessen <i>in Einklang</i> zu bringen	Meiner Einrichtung gelingt es, <i>sowohl</i> optimale Pflege- und Betreuungsqualität sicherzustellen <i>als auch</i> wirtschaftlich zu arbeiten (d.h. die zur Verfügung stehenden Mittel - z.B. Zeit, Geld, Personal etc.) optimal einzusetzen).
Feedback	Einleitung	Sometimes you may receive formal information about the care that is being provided to residents. Examples of this kind of information include: <i>number of resident falls and pain control</i> .	Vielleicht erhalten Sie gelegentlich offiziell (schriftlich oder mündlich) Auskunft über die Pflege- und Betreuungsqualität auf Ihrem Wohnbereich. Beispiele für solche Informationen sind: <i>Zahl der Bewohner, die gestürzt sind oder einen Dekubitus haben</i> .	Vielleicht erhalten Sie gelegentlich offiziell (schriftlich oder mündlich) Rückmeldung über die Pflege- und Betreuungsqualität auf Ihrem Wohnbereich. Gemeint sind Statistiken für den gesamten Wohnbereich und für einen bestimmten Zeitraum (z.B. die letzten drei oder sechs Monate). Beispiele für solche Statistiken sind: <i>Zahl der Bewohner, die gestürzt sind oder einen Dekubitus haben</i> .	Vielleicht erhalten Sie gelegentlich offiziell (schriftlich oder mündlich) Rückmeldung über die Pflege- und Betreuungsqualität auf Ihrem Wohnbereich. Gemeint sind ausschließlich Statistiken für den gesamten Wohnbereich und für einen bestimmten Zeitraum (z.B. die letzten drei oder sechs Monate). Nicht gemeint sind Informationen über einzelne Bewohner (z.B. in der Übergabe). Beispiele für solche Statistiken sind: <i>Zahl der Bewohner, die gestürzt sind oder einen Dekubitus haben</i> .
	1-6	Beispielitem: I routinely receive information on my team's performance on data <i>like</i> the examples provided above.	Ich erhalte regelmäßig Informationen über die Arbeitsleistung meines Teams in Form von Daten, <i>ähnlich</i> der oben genannten Beispiele.	[Keine Veränderung]	Ich erhalte regelmäßig Informationen über die Arbeitsleistung meines Teams in Form von Statistiken , <i>ähnlich</i> der oben genannten Beispiele.
Formal Interactions	4	Continuing education (conferences, courses) <i>outside</i> this nursing home?	Fort- und Weiterbildung (z. B. Tagungen/Kongresse oder Schulungen) <i>außerhalb</i> Ihrer Pflegeeinrichtung?	[Keine Veränderung]	[Keine Veränderung]
Structural/ Electronic Resources	5	Policies and procedures for your facility (print/online)?	Prozessvorgaben und Verfahrensanweisungen Ihrer Einrichtung (in gedruckter oder elektronischer Form)?	Verfahrensanweisungen und Vorgaben Ihrer Einrichtung (in gedruckter oder elektronischer Form)?	Verfahrensanweisungen und Vorgaben Ihrer Einrichtung (z.B. Hygienegrundsätze oder Dokumentationsvorgaben in gedruckter oder elektronischer Form)?
Time	3	[How often do you] have time to talk to someone about best practices?	[Wie oft haben Sie] Zeit, sich mit jemandem über den neuesten Erkenntnisstand zur optimalen Pflege und Betreuung zu unterhalten?	[Wie oft haben Sie] Zeit, sich mit jemandem über neues Fachwissen zur optimale Pflege und Betreuung zu unterhalten?	[Keine Veränderung]
Instrumental, Persuasive, Overall RU	1-3	Beispielitem: On your LAST typical work day , how often did you use this type of best practice knowledge <u>to provide resident care</u> ?	Wenn Sie an Ihren LETZTEN typischen Arbeitstag denken: Wie häufig haben Sie diese Art Wissen zur optimalen Gestaltung der <u>Pflege und Betreuung von Bewohnern</u> angewendet?	Wenn Sie an Ihren LETZTEN typischen Arbeitstag denken: Wie häufig haben Sie wissenschaftliche Erkenntnisse der oben genannten Art (so genanntes „best-practice-Wissen“) bei der <u>Pflege und Betreuung von Bewohnern</u> angewendet?	Erinnern Sie sich jetzt bitte an die eben gelesenen Beispiele . Wenn Sie an Ihren LETZTEN typischen Arbeitstag denken: Wie häufig haben Sie solches oder ähnliches Fachwissen (siehe Beispiele) bei der Pflege und Betreuung von Bewohnern angewendet?
Conceptual RU	1-5	Beispielitem: Give you new knowledge about how to care for residents.	Sie haben neues Wissen darüber erlangt, wie Sie die Pflege und Betreuung der Bewohner gestalten können.	Sie wissen jetzt Neues darüber, wie Sie Bewohner pflegen und betreuen können.	[Keine Veränderung]

Zeititem war die Übersetzung des englischen Originalwortlauts „*best practice*“. Mangels eines passenden deutschen Begriffs wird auf Deutsch häufig die englische Bezeichnung verwendet. Daher wurde dieser für Runde zwei in Klammern ergänzt.

In der zweiten *Cognitive-Debriefing*-Runde erwiesen sich zwei der vier modifizierten ACT-Items (C3 und SER5), der *Feedback*-Einleitungstext und alle RU-Items noch immer als problematisch. Im *Feedback*-Einleitungstext wurden daher für Runde 3 a) Textteile optisch hervorgehoben, b) eine Aussage hinzugefügt, die verdeutlichte, was nicht gemeint war (Informationen über einzelne Bewohnersituationen) und das Wort „Statistik“ wurde c) zusätzlich in die Items F1-F4 aufgenommen. Auch die anderen beiden ACT-Items wurden erneut modifiziert. In den RU-Items und dem ACT-Zeititem wurde nun statt der Begriffe „*best practice*“, „Forschungswissen“ oder „wissenschaftliche Erkenntnisse“ die Formulierung „Fachwissen über optimale Pflege und Betreuung“ verwendet. In Runde drei wurden alle Textteile wie beabsichtigt verstanden.

5.2.3 Cognitive Debriefings mit den anderen Berufsgruppen

In die Fragebögen für die anderen Berufsgruppen wurden alle übersetzten Wortlaute der Pflegehilfskraftbögen übernommen, die auch in den kanadischen Originalen übereinstimmten: die beiden *Culture*-Items, der *Feedback*-Einleitungstext und alle sechs zugehörigen Items, und das *Struktural/Electronic-Resources*-Item (Tab. 15). Im kanadischen Original unterscheiden sich die Formulierungen für die RU-Items und das ACT-Zeititem zwischen den Hilfskraft- und den Fachkraftbögen. Während für die Hilfskräfte „*best practice*“ verwendet wird, beinhalten die anderen Bögen den Begriff „*clinical knowledge*“ (ACT-Zeititem) bzw. „*research use*“ (RU-Items). Diese Unterscheidung wurde auch bei der deutschen Übersetzung berücksichtigt („klinische Erkenntnisse“ bzw. „Anwendung wissenschaftlicher Erkenntnisse“). Wie die *Cognitive Debriefings* zeigten, wurden alle untersuchten Items in allen weiteren Berufsgruppen wie intendiert verstanden.

5.3 Psychometrische Güte der übersetzten Instrumente

5.3.1 Beschreibung der Stichprobe und der beobachteten Skalenscores

In die Analysen wurden 821 verwertbare Fragebögen von Probanden aus 38 stationären Altenpflegeeinrichtungen der Metropolregion Rhein-Neckar eingeschlossen. Drei der insgesamt 824 gesammelten Fragebögen wurden aufgrund der hohen Zahl fehlender Werte ausgeschlossen (Details siehe Abschnitt 4.5.4). Einzelheiten zur Grundgesamtheit der Probanden und Einrichtungen sowie zur Rücklaufquote sind in Abschnitt 4.5.1 beschrieben. Tab. 16 zeigt die soziodemografischen Merkmale der eingeschlossenen Teilnehmer (273 PHK, 196 PFK, 152 TBK, 6 Exp., 129 FK und 65 APS). Die Mehrzahl der PHK (N=188) waren nicht oder nur im Rahmen kurzer Weiterbildungen (einige Wochen bis sechs Monate) ausgebildet, 73 Teilnehmer hatten eine Pflegehelferausbildung, zehn hatten eine in Deutschland nicht anerkannte PFK-Ausbildung in ihrem Heimatland absolviert und zwei Personen hatten eine andere tätigkeitsverwandte Aus- oder Wei-

Tab. 16: Soziodemografische Merkmale der Teilnehmer der statistischen Validierung

	PHK (N=273)		PFK (N=196)		TBK (N=152)		Exp. (N=6)		FK (N=129)		APS (N=65)		Alle (N=821)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Geschlecht***														
Weiblich	249	91,21	161	82,14	131	86,18	4	66,67	95	73,64	54	83,08	694	84,53
Männlich	24	8,79	35	17,86	20	13,16	2	33,33	34	26,36	11	16,92	126	15,35
Fehlend	—	—	—	—	1	0,66	—	—	—	—	—	—	1	0,12
Alter***														
≤19 Jahre	15	5,49	—	—	3	1,97	—	—	—	—	12	18,46	30	3,65
20-29 Jahre	33	12,09	54	27,55	18	11,84	2	33,33	13	10,08	33	50,77	153	18,64
30-39 Jahre	52	19,05	47	23,98	17	11,18	2	33,33	23	17,83	9	13,85	150	18,27
40-49 Jahre	81	29,67	39	19,90	46	30,26	2	33,33	47	36,43	7	10,77	222	27,04
50-59 Jahre	75	27,47	39	19,90	55	36,18	—	—	44	34,11	2	3,08	215	26,19
>59 Jahre	12	4,40	10	5,10	11	7,24	—	—	—	—	—	—	33	4,02
Fehlend	5	1,83	7	3,57	2	1,32	—	—	2	1,55	2	3,08	18	2,19
Muttersprache***														
Deutsch	179	65,57	150	76,53	140	92,11	5	83,33	113	87,60	50	76,92	637	77,59
Nicht Deutsch	90	32,97	45	22,96	12	7,89	1	16,67	15	11,63	15	23,08	178	21,68
Fehlend	4	1,47	1	0,51	—	—	—	—	1	0,78	—	—	6	0,73
Berufsqualifikation														
Keine	188	68,86	—	—	50	32,89	—	—	—	—	—	—	238	28,99
APH	45	16,48	—	—	—	—	—	—	—	—	—	—	45	5,48
KPH	28	10,26	—	—	—	—	—	—	—	—	—	—	28	3,41
Nicht anerk. PFK ¹	10	3,66	—	—	—	—	—	—	—	—	—	—	10	1,22
Andere ²	2	0,73	—	—	16	10,53	—	—	—	—	—	—	18	2,19
AP	—	—	144	73,47	—	—	2	33,33	21	16,28	—	—	167	20,34
GKP	—	—	47	23,98	—	—	—	—	6	4,65	—	—	53	6,46
PFK (generalist.) ³	—	—	3	1,53	—	—	—	—	—	—	—	—	3	0,37
HEP	—	—	2	1,02	—	—	—	—	—	—	—	—	2	0,24
Ergotherapeut	—	—	—	—	10	6,58	—	—	—	—	—	—	10	1,22
Physiotherapeut	—	—	—	—	2	1,32	—	—	—	—	—	—	2	0,24
Studium Soz. Arb.	—	—	—	—	9	5,92	—	—	3	2,33	—	—	12	1,46
Ausb. Soz. Arb.	—	—	—	—	1	0,66	—	—	—	—	—	—	1	0,12
§87b ⁴	—	—	—	—	64	42,11	—	—	—	—	—	—	64	7,80
Pflegestudium	—	—	—	—	—	—	1	16,67	6	4,65	—	—	7	0,85
QMB-Weiterb.	—	—	—	—	—	—	3	50,00	—	—	—	—	3	0,37
Leitungsweiterb.	—	—	—	—	—	—	—	—	89	68,99	—	—	89	10,84
BWL-Studium	—	—	—	—	—	—	—	—	3	2,33	—	—	3	0,37
Studium Erz.wiss.	—	—	—	—	—	—	—	—	1	0,78	—	—	1	0,12
1. Lehrjahr	—	—	—	—	—	—	—	—	—	—	23	35,38	23	2,80
2. Lehrjahr	—	—	—	—	—	—	—	—	—	—	21	32,31	21	2,56
3. Lehrjahr	—	—	—	—	—	—	—	—	—	—	20	30,77	20	2,44
Fehlend	—	—	—	—	—	—	—	—	—	—	1	1,54	1	0,12
Stellenumfang***														
25-49%	12	4,4	5	2,55	14	9,21	2	33,33	—	—	—	—	33	4,37
50-75%	52	19,05	25	12,76	57	37,5	1	16,67	2	1,55	—	—	137	18,12
>75%	202	73,99	164	83,67	77	50,66	2	33,33	125	96,9	—	—	570	75,40
Fehlend	7	2,56	2	1,02	4	2,63	1	16,67	2	1,55	—	—	16	2,12
Bildungsabschluss***														
Keiner	5	1,83	—	—	2	1,32	—	—	—	—	1	1,54	8	0,97
Haupt-/Volksschule	49	17,95	—	—	9	5,92	—	—	—	—	13	20,00	71	8,65
Mittlere Reife	29	10,62	—	—	13	8,55	—	—	—	—	27	41,54	69	8,40
Berufsausbildung	158	57,88	160	81,63	95	62,50	5	83,33	79	61,24	18	27,69	515	62,73
Abitur/HS-Reife	15	5,49	26	13,27	19	12,50	—	—	30	23,26	4	6,15	94	11,45
Akad. Abschluss	9	3,30	9	4,59	13	8,55	1	16,67	19	14,73	2	3,08	53	6,46
Fehlend	8	2,93	1	0,51	1	0,66	—	—	1	0,78	0	0,00	11	1,34
	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)	MW (SA)
Erfahrung in der akt. Tätigkeit (Jahre)***	10,82 (9,73)	14,99 (9,62)	5,13 (6,31)	3,33 (1,83)	8,41 (6,66)	3,73 (4,79)	9,78 (9,17)							
Erfahrung in der akt. Einrichtung (Jahre)***	7,07 (7,60)	6,81 (5,95)	5,58 (6,43)	4,12 (3,94)	9,49 (7,97)	2,41 (2,65)	6,71 (6,98)							

***Signifikanter Unterschied (p<0,001) der Verteilungen zwischen den sechs Gruppen (Kruskal-Wallis-Test)

¹Pflegefachkraftausbildung, die in Deutschland nicht anerkannt wird, ²Z.B. Altentherapeut, Arzthelfer, Familienhilfe, Haushaltshilfe, Hauswirtschaft, ³Generalistische, nicht settingspezifische PFK-Ausbildung, ⁴Weiterbildung zum Alltagsbegleiter nach §87b SGB XI

AP: Altenpfleger (examiniert), APH: Altenpflegehelfer, BWL: Betriebswirtschaft, Exp.: Experten, FK: Führungskräfte, GKP: Gesundheits- und Krankenpfleger (examiniert), HEP: Heilerziehungspfleger (examiniert), KPH: Krankenpflegehelfer, MW: Mittelwert, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, QMB: Qualitätsmanagementbeauftragte, SA: Standardabweichung, TBK: Therapie-/Betreuungskräfte

terbildung. Unter den PFK befanden sich 144 Altenpfleger, 47 Gesundheits- und Krankenpfleger,

drei Pflegende mit generalistischem Examen und zwei Heilerziehungspfleger. Von den 152 TBK waren 101 Personen wohnbereichsübergreifend tätig, während die restlichen 51 Teilnehmer fest einem bestimmten Wohnbereichsteam angehörten. Auch in dieser Gruppe hatten 50 Personen keine tätigkeitsverwandte Aus- oder Weiterbildung absolviert, 64 Teilnehmer waren nach §87b SGB XI zum Alltagsbegleiter weitergebildet, 12 Personen hatten eine therapeutische, zehn eine sozialarbeiterische Ausbildung und 16 Teilnehmer hatten eine andere tätigkeitsverwandte Aus- oder Weiterbildung. Vier der Experten waren Qualitätsbeauftragte, zwei hatten eine Stabstellenfunktion mit pflegfachlichem Schwerpunkt. Drei Experten hatten eine Weiterbildung zum Qualitätsbeauftragten, einer hatte ein Pflegestudium absolviert und zwei Probanden gaben nur ihre Altenpflegeausbildung als Qualifikation an. Unter den FK befanden sich 66 Wohnbereichsleitungen, 33 Pflegedienstleitungen, 28 Einrichtungsleitungen und zwei Personen mit einer Assistenz- bzw. Stellvertreterfunktion. Die meisten der Führungskräfte (N=89) hatten sich in einer Leitungsweiterbildung qualifiziert,

sechs hatten ein Pflegestudium und sieben ein anderes Studium (Betriebs-, Sozial- oder Bildungswissenschaft) absolviert, 27 Teilnehmer gaben ihr Pflegeexamen als Qualifikation an. Die Gruppe der Schüler setzte sich zusammen aus 23 Personen im ersten Lehrjahr, 21 Personen im zweiten und 20 Personen im dritten Lehrjahr. Wie Tab. 16 zeigt, unterschieden sich die verschiedenen Berufsgruppen signifikant ($p < 0,001$) hinsichtlich der Verteilungen aller soziodemografischer Merkmale (Geschlecht, Alter, Muttersprache, Stellenumfang und höchster Bildungsabschluss).

Tab. 17 zeigt die Merkmale der eingeschlossenen Pflegeeinrichtungen. Acht der 38 Einrichtungen sind klein (bis zu 60

Tab. 17: Merkmale der Pflegeeinrichtungen

	Klein (N=8)		Mittel (N=21)		Groß (N=9)		Alle (N=38)	
	N	%	N	%	N	%	N	
Bundesland								
BW	4	50,00	9	42,86	3	33,33	16	42,11
HE	—	—	3	14,28	3	33,33	6	15,78
RP	4	50,00	9	42,86	3	33,33	16	42,11
Trägerschaft								
FG	3	37,50	17	80,95	5	55,56	25	65,79
Ö	3	37,50	1	4,76	2	22,22	6	15,79
P	2	25,00	3	14,29	2	22,22	7	18,42
		MD (SW)		MD (SW)		MD (SW)		MD (SW)
WB/Einrichtung		1 (1-2)		3 (1-4)		4 (3-7)		3 (1-7)
Plätze/Einrichtung		51 (24-87)		94 (61-120)		145 (120-181)		97 (24-181)
Plätze/WB		33 (23-60)		25 (20-80)		36 (26-48)		35 (20-80)
		MW (SA)		MW (SA)		MW (SA)		MW (SA)
Personal (Personen)								
PHK***	10,38	(4,98)	20,71	(5,42)	32,56	(9,33)	21,34	(9,81)
PFK***	10,50	(3,07)	15,24	(3,97)	24,44	(6,33)	16,42	(6,56)
TBK	5,00	(1,85)	8,24	(5,16)	10,22	(3,27)	8,03	(4,53)
Exp.	0,13	(0,35)	0,38	(0,50)	0,11	(0,33)	0,26	(0,45)
FK – WBL***	1,00	(0,93)	2,52	(0,98)	4,11	(1,17)	2,58	(1,45)
FK – EL+PDL*	1,88	(0,64)	1,90	(0,30)	2,44	(0,53)	2,03	(0,49)
APS**	3,38	(3,20)	6,67	(3,77)	10,22	(2,28)	6,82	(4,02)
Personal (Min./BT)								
PHK	51,00	(12,00)	54,00	(8,40)	56,40	(12,60)	54,00	(10,20)
PFK*	57,60	(12,00)	46,20	(12,00)	44,40	(9,60)	48,60	(12,00)
TBK	22,20	(14,40)	18,60	(12,00)	16,20	(4,80)	18,60	(11,40)
Exp.	0,00	(0,60)	0,60	(1,20)	0,00	(0,60)	0,60	(1,20)
FK – WBL	6,00	(5,40)	9,60	(3,60)	10,20	(1,80)	9,00	(4,20)
FK – EL+PDL***	13,80	(6,60)	7,20	(1,80)	6,00	(0,60)	8,40	(4,20)
APS	4,80	(5,40)	4,20	(2,40)	4,20	(1,80)	4,80	(3,00)

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$; Unterschiede zwischen kleinen, mittleren und großen Einrichtungen (einfaktorielle ANOVAs); Klein: bis 60 Pflegeplätze, Mittel: 61-120 Pflegeplätze, Groß: >120 Pflegeplätze

APS: Altenpflegeschüler, BW: Baden-Württemberg, EL: Einrichtungsleitungen, Exp.: Experten, FG: Frei-gemeinnützig, FK: Führungskräfte, HE: Hessen, MD: Median, Min./BT: Minuten Versorgungszeit durch die jeweilige Berufsgruppe, die pro Bewohner und Tag zur Verfügung stehen, MW: Mittelwert, Ö: Öffentlich, P: Privat, PDL: Pflegedienstleitungen, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, RP: Rheinland-Pfalz, SA: Standardabweichung, SW: Spannweite, TBK: Therapie-/Betreuungskräfte, WB: Wohnbereiche, WBL: Wohnbereichsleitungen

Pflegeplätze), 21 sind mittelgroß (61-120 Pflegeplätze) und neun sind groß (>120 Pflegeplätze). Je 16 der eingeschlossenen Einrichtungen befinden sich in Baden-Württemberg und Rheinland-Pfalz und sechs in Hessen. 25 Institutionen werden von einem frei-gemeinnützigen Träger betrieben, sechs von einem öffentlichen und sieben von einem privaten Träger. Hinsichtlich der Zahl der angestellten Personen (unabhängig vom Stellenumfang) unterscheiden sich die Einrichtungsgruppen deutlich. Berücksichtigt man die Stellenumfänge und berechnet die Versorgungszeit, die pro Bewohner und Tag durch die jeweilige Berufsgruppe zur Verfügung steht, sind nur noch Unterschiede bei den PFK ($p < 0,05$) und den Führungspersonen auf Einrichtungsebene ($p < 0,001$) festzustellen. Wird als Vergleichsfaktor die Bundeslandzugehörigkeit gewählt, unterscheiden sich die Einrichtungen nur hinsichtlich der durch PFK zur Verfügung stehenden Minuten pro Bewohner und Tag ($p < 0,05$). Die drei Trägertypen unterscheiden sich in keinem der Personalmerkmale signifikant.

In Tab. 18 sind die Ergebnisse der PERMANOVAs der Skalenscores dargestellt. Tab. 19 stellt die Skalenscores der verschiedenen Berufsgruppen und der Gesamtstichprobe gegenüber. Einrichtungs- und Berufsgruppenzugehörigkeit beeinflussen signifikant alle Scores außer den der allgemeinen Gesundheit. Wohnbereichszugehörigkeit innerhalb der Einrichtung beeinflusst alle ACT-Scores, den Score der CRU-Skala und alle COPSOQ-Scores außer GEHE. *Estabrooks' Kinds of RU items* werden von der Wohnbereichsebene nicht beeinflusst. Die Berufsgruppenzugehörigkeit innerhalb einer Einrichtung beeinflusst vier ACT-Scores (SPAC, TIME, FINT und STER), die beiden CORU und den PERU-Score sowie die COPSOQ-Scores JOSA, BSTR und CSTR. Berufsgruppenzugehörigkeit innerhalb eines Wohnbereichs einer Einrichtung beeinflusst nur einen der ACT-Scores (TIME), nur den CORU-Score der *Estabrooks' items* und drei der COPSOQ-Scores (JOSA, BSTR und CSTR). Die beobachteten Scores

Tab. 18: P-Werte der nichtparametrischen Varianzanalysen (PERMANOVA) aller Skalenscores (fett: $p < 0,05$)

	Konzept	Items	N	E	B	ExW	ExB	ExWxB
ACT1	LEAD	6	788	0,003	0,051	0,001	0,757	0,976
	CULT	6	803	0,001	0,002	0,007	0,204	0,391
	FEED	6	801	0,001	0,001	0,005	0,063	0,349
	CAMP	6	807	0,001	0,002	0,002	0,387	0,442
	STAF	3	812	0,001	0,001	0,001	0,886	0,101
	SPAC	3	813	0,001	0,001	0,001	0,042	0,077
	TIME	4	819	0,001	0,001	0,001	0,001	0,016
ACT2	FINT	4	814	0,001	0,001	0,011	0,038	0,398
	IINT	10	788	0,001	0,001	0,001	0,193	0,548
	STER	11	796	0,001	0,001	0,001	0,001	0,734
Estabrooks' RU items	INRU	1	815	0,003	0,001	0,316	0,273	0,916
	CORU _a	1	812	0,001	0,001	0,346	0,001	0,016
	PERU	1	813	0,005	0,001	0,221	0,001	0,129
	OVRU	1	806	0,001	0,001	0,459	0,120	0,383
ARU/CRU	CORU _b	5	791	0,003	0,001	0,013	0,005	0,271
COPSOQ	ATRU	6	801	0,004	0,001	0,567	0,638	0,634
	JOSA	7	791	0,001	0,001	0,001	0,001	0,003
	GEHE	1	794	0,052	0,100	0,077	0,923	0,552
	BSTR	8	806	0,001	0,001	0,008	0,001	0,002
	CSTR	4	809	0,044	0,001	0,036	0,002	0,029
	BURN	6	810	0,003	0,001	0,034	0,649	0,386

ACT1: Mittelwertbasierte ACT-Konzepte, ACT2: Zählbasierte ACT-Konzepte, ATRU: Attitudes Towards RU, B: Berufsgruppe, BSTR: Behavioural Stress, BURN: Burnout, CAMP: Connections Among People, CORU: Conceptual Research Use, CSTR: Cognitive Stress, CULT: Culture, E: Einrichtung, FEED: Feedback, FINT: Formal Interactions, GEHE: Gen. Health, IINT: Informal Interactions, INRU: Instrumental Research Use, JOSA: Job Satisfaction, LEAD: Leadership, OVRU: Overall Research Use, PERU: Persuasive Research Use, SPAC: Space, STAF: Staffing, STER: Structural/Electronic Resources, TIME: Time, W: Wohnbereich

– insbesondere die des ACT – werden demnach entsprechend der Annahmen und der kanadischen Befunde von den drei Faktoren Berufsgruppe, Einrichtung und Wohnbereich innerhalb der Einrichtung beeinflusst.

Tab. 19: Skalenscores nach Berufsgruppen und für die Gesamtstichprobe

	Konzept	PHK	PFK	TBK	Exp.	FK	APS	Alle
Mittelwertbasierte ACT-Skalen	LEAD N	268	195	152	6	127	64	812
	MW (SA)	3,80 (0,73)	3,65 (0,79)	3,88 (0,75)	3,94 (0,90)	3,71 (0,71)	3,73 (0,68)	3,76 (0,75)
	CULT N	272	195	152	6	128	65	818
	MW (SA)	3,88 (0,55)	3,81 (0,58)	4,06 (0,48)	4,36 (0,32)	4,04 (0,51)	3,94 (0,53)	3,93 (0,55)
	FEED N	271	195	152	6	128	65	817
	MW (SA)	2,22 (1,12)	2,52 (1,07)	1,97 (1,09)	2,97 (1,12)	3,17 (1,06)	2,49 (1,17)	2,42 (1,16)
	CAMP N	272	196	152	6	128	65	818
MW (SA)	3,94 (0,59)	3,98 (0,54)	4,02 (0,58)	4,07 (0,67)	4,17 (0,51)	3,96 (0,52)	4,00 (0,57)	
Zählbasierte ACT-Skalen	STAF N	271	196	148	6	127	65	813
	MW (SA)	2,45 (1,00)	2,43 (0,96)	2,83 (1,00)	2,58 (1,07)	2,94 (1,16)	2,53 (0,98)	2,60 (1,03)
	SPAC N	270	194	152	6	128	64	814
	MW (SA)	3,52 (0,96)	3,55 (0,97)	3,57 (0,95)	3,78 (0,86)	3,98 (0,88)	3,54 (0,93)	3,61 (0,95)
	TIME N	273	196	152	6	128	65	820
	MW (SA)	2,37 (0,80)	2,41 (0,72)	2,70 (0,78)	2,71 (1,24)	2,85 (0,87)	2,42 (0,64)	2,52 (0,80)
	FINT N	273	196	152	6	128	65	820
MW (SA)	1,10 (0,82)	0,72 (0,98)	0,48 (0,88)	0,92 (1,43)	1,31 (1,12)	1,37 (0,88)	0,95 (0,98)	
Estabrooks' Kinds of RU-items	IINT N	272	196	151	6	128	65	818
	MW (SA)	3,50 (1,51)	4,56 (1,59)	3,36 (1,85)	3,50 (2,35)	5,54 (1,90)	3,57 (1,77)	4,05 (1,86)
	STER N	273	195	152	6	129	65	820
	MW (SA)	2,14 (1,88)	3,47 (2,21)	3,32 (2,32)	6,17 (1,51)	5,97 (2,42)	2,97 (2,21)	3,37 (2,51)
	INRU N	272	194	151	6	128	65	816
	MW (SA)	3,19 (1,41)	2,65 (1,23)	2,51 (1,39)	2,67 (1,63)	2,48 (1,22)	3,45 (1,03)	2,84 (1,35)
	CORU_a N	—	191	149	6	128	—	474
MW (SA)	—	2,58 (1,23)	2,30 (1,28)	2,00 (1,55)	2,68 (1,23)	—	2,51 (1,26)	
CRU-Skala	PERU N	273	191	151	6	128	65	814
	MW (SA)	2,56 (1,15)	2,57 (1,20)	2,08 (1,30)	2,00 (1,55)	2,40 (1,32)	2,88 (0,98)	2,47 (1,23)
	OVRU N	272	187	151	6	127	64	807
	MW (SA)	2,83 (1,19)	2,46 (1,20)	2,34 (1,30)	2,50 (1,38)	2,46 (1,23)	3,38 (0,95)	2,63 (1,23)
ARU-Skala	CORU_b N	272	187	149	6	126	64	804
	MW (SA)	2,88 (1,17)	2,47 (1,06)	2,38 (1,22)	1,63 (1,20)	2,36 (1,15)	3,59 (0,78)	2,65 (1,18)
COPSOQ-Skalen	ATRU N	267	194	151	6	129	65	812
	MW (SA)	3,65 (0,60)	3,49 (0,58)	3,65 (0,62)	3,17 (0,52)	3,62 (0,56)	3,71 (0,59)	3,61 (0,60)
	JOSA N	271	195	152	6	129	65	818
	MW (SA)	2,09 (0,41)	2,18 (0,40)	1,90 (0,42)	1,64 (0,44)	2,00 (0,51)	2,09 (0,41)	2,06 (0,44)
	GEHE N	266	188	147	6	123	65	795
	MW (SA)	6,65 (2,18)	6,65 (1,93)	7,25 (1,75)	7,50 (2,43)	6,95 (2,14)	6,85 (2,12)	6,83 (2,05)
	BSTR N	271	195	152	6	129	65	818
MW (SA)	4,14 (0,73)	3,78 (0,91)	4,20 (0,77)	4,48 (0,36)	4,04 (0,83)	3,73 (0,89)	4,02 (0,83)	
COPSOQ-Skalen	CSTR N	273	194	151	6	129	64	817
	MW (SA)	4,12 (0,76)	3,83 (0,81)	4,08 (0,75)	4,13 (0,72)	3,87 (0,67)	3,66 (0,88)	3,97 (0,78)
	BURN N	270	192	150	6	128	64	810
MW (SA)	4,12 (0,76)	3,83 (0,81)	4,09 (0,75)	4,13 (0,72)	3,87 (0,67)	3,66 (0,88)	3,97 (0,78)	

N: Zahl der Fälle mit gültigen Werten, M (SA): Mittelwert (Standardabweichung)

APS: Altenpflegeschüler, ATRU: Attitudes Towards RU, BSTR: Behavioural Stress, BURN: Burnout, CAMP: Connections Among People, CORU: Conceptual Research Use, CSTR: Cognitive Stress, CULT: Culture, Exp.: Experten, FEED: Feedback, FINT: Formal Interactions, FK: Führungskräfte, GEHE: Gen. Health, IINT: Informal Interactions, INRU: Instrumental Research Use, JOSA: Job Satisfaction, LEAD: Leadership, MW: Mittelwert, OVRU: Overall Research Use, PERU: Persuasive Research Use, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, SA: Standardabweichung, SPAC: Space, STAF: Staffing, STER: Structural/Electronic Resources, TBK: Therapie-/Betreuungskräfte, TIME: Time

5.3.2 Faktorenstruktur und Reliabilität

Die Fitindizes der verschiedenen ACT-Faktorenmodelle sind in Tab. 20 dargestellt. Die kompletten Mplus-Outputs aller Modelle sind auf Nachfrage beim Autor erhältlich. Modell 1 (Zehn Fak-

Tab. 20: Fitindizes der ACT-Faktorenmodelle

		N	χ^2	df	p	RMSEA (90%CI)	CFI	TLI	SRMR	
Modell 1	PHK	242	1800,981	1333 ¹	0,0000	0,038 (0,033-0,043)	0,762	0,745	0,100	
	PFK	172	1707,796	1332	0,0000	0,041 (0,035-0,046)	0,714	0,693	0,098	
	TBK	136	1678,855	1332	0,0000	0,044 (0,037-0,050)	0,690	0,667	0,110	
	FK	107	1588,735	1333 ²	0,0000	0,042 (0,033-0,050)	0,644	0,618	0,110	
	APS	55	Modell nicht schätzbar							
	Alle	712	3249,087	1332	0,0000	0,045 (0,043-0,047)	0,777	0,671	0,098	
Modell 2a	PHK	249	640,619	475 ¹	0,0000	0,037 (0,030-0,045)	0,910	0,900	0,068	
	PFK	178	594,444	474	0,0001	0,038 (0,027-0,047)	0,896	0,884	0,073	
	TBK	140	616,171	474	0,0000	0,046 (0,035-0,056)	0,851	0,834	0,086	
	FK	115	558,560	475 ²	0,0048	0,039 (0,023-0,052)	0,868	0,853	0,082	
	APS	57	564,000	474	0,0027	0,058 (0,036-0,075)	0,767	0,740	0,094	
	Alle	744	1045,062	474	0,0000	0,040 (0,037-0,044)	0,921	0,911	0,055	
Modell 3a	PHK	263	581,375	186	0,0000	0,090 (0,082-0,098)	0,552	0,495	0,148	
	PFK	183	540,515	186	0,0000	0,102 (0,092-0,112)	0,481	0,414	0,151	
	TBK	147	Modell nicht schätzbar							
	FK	116	428,739	186	0,0000	0,106 (0,093-0,119)	0,509	0,445	0,151	
	APS	62	300,198	186	0,0000	0,100 (0,078-0,120)	0,514	0,451	0,159	
	Alle	771	2108,639	186	0,0000	0,116 (0,111-0,120)	0,590	0,537	0,155	
Modell 2b	PHK	251	419,795	350	0,0061	0,028 (0,016-0,038)	0,962	0,956	0,053	
	PFK	179	405,821	349	0,0193	0,030 (0,013-0,042)	0,952	0,944	0,059	
	TBK	141	401,548	349	0,0273	0,033 (0,012-0,046)	0,945	0,936	0,070	
	FK	118	388,865	349	0,0694	0,031 (0,000-0,047)	0,936	0,926	0,072	
	APS	58	388,725	349	0,0701	0,044 (0,000-0,067)	0,894	0,877	0,083	
	Alle	752	602,051	350	0,0000	0,031 (0,027-0,035)	0,964	0,959	0,041	
Modell 3b	PHK	263	162,869	134	0,0454	0,029 (0,004-0,043)	0,965	0,955	0,055	
	PFK	187	149,882	133	0,1504	0,026 (0,000-0,045)	0,977	0,970	0,045	
	TBK	147	178,228	135	0,0075	0,047 (0,025-0,064)	0,930	0,912	0,075	
	FK	117	172,621	132	0,0101	0,051 (0,026-0,071)	0,918	0,893	0,071	
	APS	63	145,181	133	0,2219	0,038 (0,000-0,074)	0,948	0,934	0,080	
	Alle	777	313,372	134	0,0000	0,042 (0,036-0,047)	0,961	0,950	0,040	
Modell 2b+3b	PHK	244	1148,485	993	0,0004	0,025 (0,018-0,032)	0,920	0,909	0,060	
	PFK	174	1114,214	991	0,0037	0,027 (0,016-0,035)	0,907	0,895	0,060	
	TBK	137	1134,237	993	0,0012	0,032 (0,021-0,041)	0,874	0,857	0,077	
	FK ³	110	1122,841	991	0,0022	0,035 (0,022-0,045)	0,821	0,796	0,079	
	APS ³	63	1201,351	993	0,0000	0,061 (0,047-0,073)	0,618	0,566	0,099	
	Alle	722	1532,170	992	0,0000	0,027 (0,025-0,030)	0,938	0,930	0,045	

¹In diesem Modell wurde eine negative, nicht signifikante Residualvarianz (ST2) auf null fixiert

²In diesem Modell wurde eine negative, nicht signifikante Residualvarianz (SP2) auf null fixiert

³Modell nicht vertrauenswürdig – Fehlermeldung: „Latent variable covariance matrix (PSI) is not positive definite“

90% CI: 90%-Konfidenzintervall, APS: Altenpflegeschüler, CFI: *Comparative Fit Index*, df: *degrees of freedom*, FK: Führungskräfte, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, RMSEA: *Root Mean Square Error of Approximation*, SRMR: *Standardized Root Mean Square Error*, TBK: Therapie-/Betreuungskräfte, TLI: *Tucker Lewis Index*

toren) wies einen sehr schlechten Fit auf und war in der Gruppe der APS (aufgrund der zu kleinen Substichprobe) nicht schätzbar. Daher erfolgte die Zweiteilung des ACT in ein Modell, das die sieben zustimmungsskalierten Konzepte beinhaltet (Modell 2a) und ein Modell mit den drei häu-

figkeitsskalierten Konzepten (Modell 3a). Verglichen mit Modell 1 verbesserten sich die Fitwerte für Modell 2b, vor allem CFI und TLI waren aber noch deutlich kleiner als die empfohlenen Werte ($>0,95$). Die Modellanalyse zeigte, dass vier Items den Fit störten: C2, C3, CAP3 und SP1. In den Erhebungen fragten vor allem die PHK oft nach der Bedeutung von C2 („ich habe die Kontrolle darüber, wie ich meine Arbeit mache“). C3 erwies sich bereits in den *Cognitive Debriefings* als problematisch (Abschnitt 5.2.2). CAP3 weicht deutlich von den anderen fünf Items dieses Konzepts (Miteinander im Team) ab, da es als einziges nicht nach dem Wohnbereichsteam, sondern nach dem Austausch von Informationen mit anderen Teams fragt. SP1 (Frage, ob in der Einrichtung angemessene räumliche Verhältnisse verfügbar seien) wurde schon in der kanadischen Studie entfernt. Auch hier fragten einige Probanden nach, was unter „angemessen“ zu verstehen sei. Diese Probleme spiegelten sich in den Modifikationsindizes wider, die auf diverse Querladungen zu anderen als den erwarteten Faktoren verwiesen. Zudem waren die Faktorladungen dieser Items auf den erwarteten Faktor substantiell kleiner als die der anderen zugehörigen Items. Diese vier Items wurden daher entfernt. Die Residualvarianzen der ersten vier Feedback-Items wurden in allen Gruppen korreliert, da diese Items – wie sich auch in den *Cognitive Debriefings* zeigte (Abschnitt 5.2.2) – eine gemeinsame „Fehlerquelle“ aufweisen: das Wort „Statistiken“ – ein Begriff, den die Befragten in diesem Zusammenhang eher selten verwenden. Zudem wurde in den Berufsgruppen vereinzelt die Korrelation bestimmter Residualvarianzen von Items des gleichen Konzepts zugelassen (Modell 2b, Abb. 8, Details und Begründung siehe Anhang 2). Außer in der Gruppe der APS passte dieses Modell gut zu den Daten.

Die Fitindizes für Modell 3a deuten auf eine sehr schlechte Modellpassung hin. In der Gruppe der TBK war dieses gar nicht schätzbar. Die Analyse des Modells und der Iteminhalte legte nahe, die Faktoren IINT und STER weiter aufzuteilen. Informeller Austausch findet zum einen mit Personen statt, die direkt mit Bewohnern zu tun haben (IINT_a: PHK – Item II1, PFK – Item II2, TBK – Item II4 sowie informelle Gespräche im Allgemeinen – Item II10). Andererseits findet informelle Kommunikation auch mit bewohnerferner tätigen Personen statt (IINT_b: Experten – Items II5 und II6). Die Gruppe der TBK umfasst sowohl bewohnernah tätige Personen (z.B. Alltagsbegleiter, Ergotherapeuten u.ä.), als auch solche, die nicht ganz so bewohnernah tätig sind (z.B. die Sozialdienstleitung). Daher ist Item II4 – außer in der Gruppe der TBL selbst – mit beiden IINT-Faktoren assoziiert. Auch die im Bereich STER abgefragten Ressourcen sind sehr unterschiedlicher Natur. So lassen sich Fachzeitschriften, Lehrbücher und Fachliteraturbibliotheken unter einen Leserressourcenfaktor (STER_a) subsumieren, Pinnwände/schwarze Bretter, Dienst-/Verfahrensweisungen sowie Leitlinien/Standards können einem Regulationenfaktor (STER_b) zugeordnet werden und PCs mit Internetanschluss, Softwarenutzung, E-Mail und Webseiten reflektieren einen EDV-Faktor (STER_c). Die Items II11 (Praxisanleitungen durch Kollegen oder Vorgesetzte) und SER11 (innerbetriebliche Fort- und Weiterbildungen) wurden zunächst dem FINT-Faktor (offizielle Zusammenkünfte) zugeordnet. Dieser beinhaltet bereits das Item zu

betriebsexternen Fort- und Weiterbildungen, was die Zuordnung von SER11 nahelegt. Praxisanleitungen sind zudem hierzulande eher formeller als informeller Natur. Da diese Modelle jedoch auch schlechte Fitwerte aufwiesen, wurden diese beiden Items ganz entfernt. Zudem wurden auch hier gruppenspezifisch Korrelationen bestimmter Residualvarianzen erlaubt (Modell 3b, Abb. 8, Details und Begründung: Anhang 2). Dieses Modell erwies sich als relativ gut passend. Lediglich in der TBK- und FK-Gruppe blieben die CFI- und TLI-Werte unter den empfohlenen Grenzen.

Die erneute Zusammenführung der Modelle 2b und 3b zu einem 13-Faktoren-Modell führte zu einem nichtakzeptablen Modell, das zudem in der FK- und APS-Gruppe keine vertrauenswürdigen Ergebnisse lieferte. Die Ableitungsmatrix war nicht positiv-definit, was vermutlich mit der großen Zahl der zu schätzenden Parameter und den dazu im Verhältnis recht kleinen Substichproben zusammenhängt. Daher wurden die Messinvarianzanalysen separat für Modelle 2b und 3b durchgeführt. Wie Tab. 21 zeigt, wiesen die Modelle für die CRU-Skala (Modell 4, Abb. 9) ausgezeichnete Fitwerte auf und wurde daher ebenfalls in die weiteren Schritte einbezogen. Die Korrelation der Residualvarianzen der ersten beiden CRU-Items wurde in allen Modellen erlaubt. Außerdem wurde in der FK-Gruppe die Korrelation der Residualvarianzen der Items 4 und 5 gestattet (Details und Begründung: Anhang 2).

Tab. 21: Fitindizes des CRU-Skala-Faktorenmodells

		N	χ^2	df	p	RMSEA (90%CI)	CFI	TLI	SRMR
Modell 4	PHK	268	1,519	4	0,8232	0,000 (0,000-0,055)	1,000	1,006	0,003
	PFK	183	4,637	4	0,3266	0,030 (0,000-0,199)	0,999	0,997	0,012
	TBK	147	4,682	4	0,3216	0,034 (0,000-0,133)	0,998	0,995	0,010
	FK	123	3,802	3	0,2837	0,047 (0,000-0,166)	0,998	0,995	0,009
	APS	63	0,297	4	0,9900	0,000 (0,000-0,000)	1,000	1,093	0,005
	Alle	790	6,207	4	0,1842	0,026 (0,000-0,065)	0,999	0,998	0,005

90%CI: 90%-Konfidenzintervall, APS: Altenpflegeschüler, CFI: *Comparative Fit Index*, df: *degrees of freedom*, FK: Führungskräfte, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, RMSEA: *Root Mean Square Error of Approximation*, SRMR: *Standardized Root Mean Square Error*, TBK: Therapie-/Betreuungskräfte, TLI: *Tucker Lewis Index*

Auf Basis der Modellparameter der Modelle 2b, 3b und 4 wurden zudem die Reliabilitätswerte (ω) für die einzelnen Skalen bestimmt. In Tab. 22 sind die Ladungen und Residualvarianzen sowie die sich daraus ergebenden ω -Werte für die ACT-Konzepte und die CRU-Skala dargestellt – getrennt nach Berufsgruppen sowie für die Gesamtstichprobe. Bis auf das CULT-Konzept weisen alle zustimmungsskalierten ACT-Skalen in allen Gruppen eine akzeptable Reliabilität ($\omega > 0,7$) auf. Die Reliabilitätswerte für die häufigkeitsskalierten ACT-Konzepte sind etwas niedriger als die der zustimmungsskalierten Konzepte. IINT_a, STER_a und STER_c weisen in allen Gruppen ω -Werte größer 0,7 auf, STER_b bleibt nur in der APS-Gruppe hinter diesem Wert zurück. Die ω -Werte für FINT sind lediglich in der PHK-Gruppe und der Gesamtstichprobe über 0,7, IINT_b weist allein in der APS-Gruppe einen akzeptablen Wert auf. Die beiden letztgenannten Konzepte stellten sich somit in der vorliegenden Studie als nicht ausreichend zuverlässig heraus. Die Reliabilitätswerte für die CRU-Skala sind in allen Gruppen als sehr gut zu bezeichnen.

Tab. 22: ACT-Modell 2b: Ladungen, Residualvarianzen und Reliabilitätswerte der Skalen

		PHK (N=251)		PFK (N=179)		TBK (N=141)		FK (N=118)		APS (N=58)		Alle (N=752)			
Faktor	Item	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ		
Zustimmungskalierte ACT-Konzepte	LEAD	L1	0,620	0,535	0,740	0,564	0,584	0,740	0,638	0,464	0,517	0,614	0,649	0,574	
		L2	0,396	0,850	0,415	0,889	0,653	0,674	0,468	0,896	0,305	1,087	0,459	0,865	
		L3	0,609	0,703	0,722	0,690	0,595	0,527	0,647	0,654	0,581	0,670	0,648	0,661	
		L4	0,709	0,253	0,735	0,414	0,651	0,305	0,712	0,305	0,699	0,200	0,709	0,310	
		L5	0,765	0,371	0,988	0,343	0,906	0,254	0,797	0,275	0,920	0,089	0,867	0,311	
		L6	0,726	0,326	0,852	0,400	0,759	0,322	0,734	0,279	0,728	0,441	0,774	0,360	
		Reliabilität (ω)	0,828		0,857		0,859		0,848		0,819		0,845		
		CULT	C1	0,463	0,523	0,750	0,413	0,548	0,584	0,513	0,475	0,646	0,410	0,587	0,494
			C4	0,547	0,629	0,408	0,821	0,449	0,837	0,550	0,644	0,627	0,509	0,510	0,711
			C5	0,361	0,412	0,278	0,363	0,231	0,379	0,224	0,471	0,194	0,267	0,292	0,394
			C6	0,632	0,268	0,625	0,412	0,374	0,275	0,301	0,470	0,453	0,210	0,534	0,349
		Reliabilität (ω)	0,687		0,679		0,553		0,550		0,725		0,655		
		FEED	F1	1,061	0,364	0,966	0,566	1,027	0,257	0,701	1,040	1,168	0,402	1,097	0,569
			F2	1,093	0,325	0,939	0,474	1,047	0,280	0,950	0,691	1,151	0,268	1,107	0,421
			F3	1,034	0,265	0,979	0,286	1,010	0,089	0,995	0,586	1,094	0,240	1,083	0,294
			F4	1,096	0,234	1,056	0,251	1,091	0,048	1,137	0,303	1,042	0,114	1,147	0,212
			F5	1,186	0,052	1,155	0,055	1,096	0,009	1,202	0,119	1,158	0,116	1,227	0,061
			F6	0,903	0,569	0,851	0,574	0,828	0,340	0,710	1,074	0,881	0,568	0,903	0,621
		Reliabilität (ω)	0,957		0,941		0,973		0,895		0,961		0,952		
		CAMP	CAP1	0,464	0,352	0,341	0,291	0,506	0,310	0,389	0,359	0,425	0,226	0,443	0,311
			CAP2	0,599	0,351	0,367	0,459	0,492	0,500	0,376	0,376	0,488	0,406	0,527	0,382
			CAP4	0,492	0,314	0,441	0,463	0,277	0,579	0,352	0,436	0,476	0,318	0,446	0,415
			CAP5	0,590	0,254	0,528	0,248	0,590	0,244	0,380	0,190	0,344	0,193	0,517	0,255
		CAP6	0,697	0,332	0,655	0,377	0,682	0,409	0,605	0,471	0,608	0,290	0,646	0,404	
	Reliabilität (ω)	0,834		0,747		0,761		0,707		0,793		0,790			
	STAF	ST1	0,749	0,786	0,895	0,722	0,780	0,530	1,170	0,155	0,728	0,705	0,904	0,604	
		ST2	1,080	0,000	0,849	0,506	0,944	0,373	0,944	0,608	0,957	0,222	0,951	0,374	
	Reliabilität (ω)	0,810		0,712		0,767		0,854		0,754		0,779			
	SPAC	SP2	0,933	0,289	0,783	0,522	0,745	0,381	1,073	0,000	1,009	0,155	0,929	0,281	
		SP3	1,358	0,615	1,626	0,000	1,276	0,569	1,050	1,010	1,407	0,608	1,317	0,729	
	Reliabilität (ω)	0,853		0,917		0,811		0,817		0,884		0,833			
	TIME	T1	0,551	0,753	0,610	0,369	0,505	0,648	0,807	0,334	0,459	0,452	0,591	0,643	
		T2	0,684	0,417	0,681	0,366	0,648	0,737	0,670	0,720	0,449	0,309	0,753	0,454	
		T3	0,754	0,304	0,611	0,484	0,671	0,644	0,715	0,620	0,635	0,295	0,692	0,483	
		T4	0,836	0,505	0,740	0,564	0,584	0,740	0,638	0,464	0,517	0,614	0,794	0,511	
	Reliabilität (ω)	0,801		0,805		0,735		0,802		0,790		0,793			
Häufigkeitsskalierte ACT-Konzepte	FINT	FI1	0,789	0,527	0,596	0,484	0,788	0,781	0,493	0,964	0,665	0,590	0,682	0,706	
		FI2	0,737	0,227	0,833	0,559	0,832	0,116	0,849	0,737	0,823	0,180	0,920	0,348	
		FI5	0,489	0,628	0,405	0,990	0,356	1,041	0,523	0,863	0,424	1,184	0,545	0,897	
		Reliabilität (ω)	0,746		0,623		0,668		0,576		0,652		0,703		
		HINT _a	II1	0,740	0,450	0,561	0,337	0,740	0,881	0,923	0,047	0,846	0,282	0,662	0,671
			II2	0,681	0,494	0,506	0,536	0,869	0,596	0,544	0,194	0,691	0,497	0,639	0,617
			II4	0,790	0,828	0,604	0,796	0,901	0,717	0,675	0,763	0,769	0,553	0,655	0,956
			II5	—	—	—	—	—	—	0,205	0,878	—	—	—	—
			II10	1,030	0,586	0,999	0,840	0,965	0,715	0,464	1,025	1,141	0,341	1,035	0,588
		Reliabilität (ω)	0,817		0,740		0,806		0,731		0,877		0,760		
		HINT _b	II4	0,415	0,828	0,306	0,796	—	—	0,288	0,763	0,363	0,553	0,326	0,956
			II5	0,735	0,318	0,617	0,542	0,373	0,409	0,344	0,878	0,780	0,242	0,549	0,561
			II6	0,626	0,230	0,800	0,358	0,838	0,277	0,908	0,619	0,776	0,314	0,831	0,327
		Reliabilität (ω)	0,696		0,636		0,681		0,512		0,769		0,612		
	STER _a	SER1	0,222	0,153	0,419	0,377	0,399	0,660	0,708	0,740	0,326	0,265	0,471	0,411	
		SER2	0,622	0,293	0,745	0,416	0,897	0,683	0,776	0,635	0,825	0,599	0,795	0,510	
		SER3	0,776	0,376	0,824	0,381	0,820	0,445	1,096	0,348	0,871	0,183	0,913	0,409	
	Reliabilität (ω)	0,761		0,771		0,715		0,794		0,796		0,781			

Tab. 22: Fortsetzung

		PHK (N=251)		PFK (N=179)		TBK (N=141)		FK (N=118)		APS (N=58)		Alle (N=752)		
Faktor	Item	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ	
Häufigkeitsskalierte ACT-Konzepte	STER _b	SER4	0,538	1,362	0,621	1,347	0,680	1,270	0,428	1,475	0,773	0,889	0,547	1,394
		SER5	1,282	0,357	0,848	0,680	1,073	1,175	1,004	0,391	0,566	1,281	1,158	0,575
		SER6	1,105	0,452	0,816	0,783	1,377	0,512	0,999	0,516	0,950	0,839	1,215	0,468
	Reliabilität (ω)		0,798		0,650		0,768		0,713		0,635		0,778	
	STER _c	SER7	1,217	0,204	1,290	1,140	1,376	1,335	1,225	0,730	1,112	1,231	1,413	1,073
		SER8	0,945	0,287	0,885	1,191	0,784	2,038	1,101	1,800	0,670	1,593	0,968	1,359
		SER9	0,508	0,411	1,225	0,358	1,320	0,653	1,471	0,758	0,634	0,444	1,337	0,450
		SER10	0,277	0,218	0,946	0,466	1,230	0,356	1,253	0,946	0,633	0,579	1,094	0,479
		Reliabilität (ω)		0,886		0,857		0,835		0,858		0,707		0,873
	CRU-Skala	CORU _b	CRU _{b1}	1,076	0,403	0,897	0,447	1,093	0,375	1,053	0,387	0,670	0,340	1,074
CRU _{b2}			1,154	0,220	0,907	0,398	1,180	0,229	1,144	0,218	0,660	0,288	1,133	0,282
CRU _{b3}			1,187	0,184	1,064	0,297	1,132	0,362	1,134	0,120	0,829	0,335	1,181	0,255
CRU _{b4}			1,129	0,196	1,056	0,292	1,233	0,168	1,149	0,360	0,786	0,233	1,154	0,239
CRU _{b5}			1,146	0,230	1,111	0,140	1,240	0,331	1,111	0,249	0,683	0,321	1,186	0,222
Reliabilität (ω)		0,963		0,942		0,959		0,959		0,897		0,959		

λ : Faktorladungen (alle $p < 0,000$), δ : Residualvarianzen, APS: Altenpflegeschüler, CAMP: *Connections Among People*, CULT: *Culture*, CORU: *Conceptual Research Use*, FEED: *Feedback*, FINT: *Formal Interactions*, FK: *Führungskräfte*, IINT: *Informal Interactions*, LEAD: *Leadership*, PFK: *Pflegefachkräfte*, PHK: *Pflegehilfskräfte*, SPAC: *Space*, STAF: *Staffing*, STER: *Structural/Electronic Resources*, TBK: *Therapie-/Betreuungskräfte*

5.3.3 Faktorinvarianz

Die Ergebnisse der Messinvarianzuntersuchung für das ACT-Modell 2b sind in Tab. 23 dargestellt. Die Mplus-Outputs aller Modelle sind auf Nachfrage beim Autor erhältlich. Vollständige schwache Messinvarianz konnte nicht erzielt werden. Nach Freisetzen von drei der insgesamt 110 potenziell frei zu schätzenden Ladungen (von 29 Ladungen pro Modell waren sieben auf 1,0 fixiert; in jedem der fünf Modelle – PHK, PFK, TBK, FK und APS – waren somit 22 Ladungen potenziell schätzbar) konnte jedoch partielle schwache Messinvarianz festgestellt werden. Auch vollständige starke Messinvarianz war nicht gegeben. Nach freisetzen von acht der 145 potenziell schätzbaren Interzepte erwies sich das Modell aber als partiell stark messinvariant. Strikte Messinvarianz war in diesem Modell weder vollständig noch partiell erfüllt. Dies überraschte nicht, da aufgrund der Heterogenität der verschiedenen Berufsgruppen, insbesondere auch was Qualifikation und Bildungsgrad angeht (siehe Abschnitt 5.3.1), nicht davon auszugehen war, dass die Items mit derselben Präzision bzw. „Fehlerquote“ beantwortet würden. Die Faktorvarianzen waren ebenfalls nicht vollständig invariant. Nach Freisetzung zweier der 35 potenziell schätzbaren Faktorvarianzen konnte partielle Faktorvarianz-Invarianz verzeichnet werden.

Auch in der Messinvarianzanalyse des Modells 3b (Tab. 23) ergab sich keine vollständige schwache Messinvarianz. Nach Freisetzen von acht der 69 potenziell schätzbaren Ladungen war jedoch partielle schwache Messinvarianz erfüllt. Vollständige starke Messinvarianz wurde ebenfalls nicht erreicht. Nachdem 13 der 95 potenziell schätzbaren Interzepte freigesetzt worden waren, war jedoch partielle starke Messinvarianz erreicht. Wie Modell 2b erwies sich auch dieses Modell weder als vollständig noch als partiell strikt messinvariant, und auch hier waren zwei Faktorvarianzen frei zusetzen (von 30 potenziell schätzbaren Varianzen), um partielle Faktorva-

rianz-Invarianz zu erzielen.

Die CRU-Skala (Tab. 24) erwies sich als vollständig schwach messinvariant. Starke Messinvarianz konnte partiell erzielt werden, indem einer der 25 potenziell schätzbaren Interzepte freigesetzt wurde. Wie das ACT erwies sich auch diese Skala als nicht strikt messinvariant, allerdings konnte nach Freisetzen einer der 25 potenziell schätzbaren Residualvarianzen partielle strikte Messinvarianz festgestellt werden. Auch Faktorvarianz-Invarianz wurde nur partiell erreicht. Hierfür war die Freisetzung einer der fünf potenziell schätzbaren Faktorvarianzen erforderlich.

Tab. 23: ACT-Modelle 2b und 3b: Ergebnisse der Messinvarianzuntersuchung

Modell	Form der Invarianz	χ^2 (df)	p	RMSEA (90%CI)	CFI	TLI	SRMR	Vgl. mit	χ^2_{diff} (df _{diff})	p _{diff}
2b ₁	Konfigurale	1998,108 (1746)	0,0000	0,031 (0,024-0,037)	0,947	0,938	0,064	—	—	—
2b ₂	Schwache	2123,086 (1835)	0,0000	0,032 (0,025-0,039)	0,939	0,933	0,075	2b ₁₁	169,854 (89)	0,0000
2b ₃	Partielle schwache ¹	2087,454 (1833)	0,0000	0,030 (0,03-0,037)	0,946	0,941	0,069	2b ₁₁	102,468 (87)	0,1231
2b ₄	Starke	2256,038 (1921)	0,0000	0,034 (0,028-0,040)	0,930	0,926	0,074	2b ₃	299,021 (88)	0,0000
2b ₅	Partielle starke ²	2171,729 (1913)	0,0000	0,030 (0,023-0,036)	0,946	0,942	0,071	2b ₃	99,212 (80)	0,7160
2b ₆	Strikte	2470,612 (2048)	0,0000	0,037 (0,031-0,042)	0,911	0,912	0,074	2b ₅	360,091 (135)	0,0000
2b ₈	Faktorvarianz	2209,636 (1939)	0,0000	0,031 (0,023-0,037)	0,943	0,940	0,084	2b ₅	52,966 (26)	0,0014
2b ₉	Part. Faktorvarianz ³	2198,742 (1937)	0,0000	0,030 (0,023-0,036)	0,945	0,942	0,082	2b ₅	32,432 (24)	0,1166
3b ₁	Konfigurale	804,527 (667)	0,0002	0,036 (0,026-0,045)	0,952	0,939	0,062	—	—	—
3b ₂	Schwache	994,059 (722)	0,0000	0,049 (0,042-0,057)	0,905	0,888	0,089	3b ₁₁	239,697 (55)	0,0000
3b ₃	Partielle schwache ⁴	857,915 (714)	0,0002	0,036 (0,026-0,045)	0,950	0,940	0,068	3b ₁₁	61,745 (47)	0,0730
3b ₄	Starke	1071,458 (766)	0,0000	0,051 (0,043-0,058)	0,894	0,881	0,085	3b ₃	359,936 (52)	0,0000
3b ₅	Partielle starke ⁵	899,960 (753)	0,0002	0,035 (0,025-0,044)	0,949	0,942	0,069	3b ₃	49,057 (39)	0,1298
3b ₆	Strikte	1297,781 (829)	0,0000	0,060 (0,054-0,067)	0,837	0,832	0,115	3b ₅	489,513 (76)	0,0000
3b ₈	Faktorvarianz	943,047 (777)	0,0000	0,037 (0,028-0,045)	0,942	0,936	0,093	3b ₅	58,397 (24)	0,0001
3b ₉	Part. Faktorvarianz ⁶	923,800 (775)	0,0002	0,035 (0,025-0,044)	0,948	0,942	0,079	3b ₅	30,094 (22)	0,1162

¹Ladungen für drei Items freigesetzt: C1 bei PHK, ST2 und SP3 bei FK

²Interzepte für acht Items freigesetzt: T1 und T3 bei PHK, T3 bei PFK, SP2 und T1 in TBK, F1 bei FK, F4 und T3 bei APS

³Varianzen für zwei Faktoren freigesetzt: CULT bei TBK, STAF bei MA

⁴Ladungen für acht Items freigesetzt: II6, SER9 und SER10 bei PHK; II2 bei TBK; III1, III10 und SER1 bei FK, SER5 bei APS

⁵Interzepte für 13 Items freigesetzt: F11, F15, SER3, SER4 und SER9 bei PHK; III10 und SER3 bei PFK; III1, III4, III5 und III10 bei TBK; III1 und SER3 bei MA

⁶Varianzen für zwei Faktoren freigesetzt: IINT_b bei PHK, STER_c bei APS

90%CI: 90%-Konfidenzintervall, CFI: Comparative Fit Index, df: degrees of freedom, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Error, TLI: Tucker Lewis Index

Nachdem für das ACT weder vollständige noch partielle strikte Messinvarianz festgestellt werden konnte, müssen die Gruppenvergleiche der beobachteten Itemwerte bzw. der daraus abgeleiteten Skalenscores (Abschnitt 5.3.1) als nicht vertrauenswürdig eingestuft werden. Zudem ergab sich für die häufigkeitsskalierten ACT-Konzepte eine andere Faktorenstruktur als ange-

nommen. Daher wurden – auf Basis der Modelle 2b₅, 3b₅ und 4₅ (partielle schwache Messinvarianz) – die Faktorvarianzen für das Gesamtsample bestimmt sowie die latenten Faktormittelwerte verglichen (Tab. 25). Das PHK-Sample wurde als Referenzgruppe mit Faktormittelwerten=0 definiert. Deutliche Unterschiede zwischen den Berufsgruppen sind sichtbar. Am geringsten unter-

Tab. 24: CRU-Skala, Modell 4: Ergebnisse der Messinvarianzuntersuchung

Modell	Form der Invarianz	χ^2 (df)	p	RMSEA (90%CI)	CFI	TLI	SRMR	Vgl. mit	χ^2_{diff} (df _{diff})	p _{diff}
4 ₁	Konfigurale	15,286 (19)	0,7042	0,000 (0,000-0,054)	1,000	1,004	0,008	—	—	—
4 ₂	Schwache	34,011 (35)	0,5157	0,000 (0,000-0,055)	1,000	1,001	0,043	4 ₁	26,271 (16)	0,0503
4 ₄	Starke	58,041 (51)	0,2318	0,030 (0,000-0,061)	0,997	0,997	0,051	4 ₂	31,482 (16)	0,0117
4 ₅	Partielle starke ¹	53,366 (50)	0,3462	0,021 (0,000-0,057)	0,999	0,999	0,049	4 ₂	23,714 (15)	0,0701
4 ₆	Strikte	94,710 (74)	0,0527	0,042 (0,000-0,066)	0,991	0,994	0,046	4 ₅	41,155 (24)	0,0160
4 ₇	Partielle strikte	85,957 (73)	0,1425	0,034 (0,000-0,059)	0,994	0,996	0,049	4 ₅	32,919 (23)	0,0825
4 ₈	Faktorvarianz	65,137 (54)	0,1426	0,036 (0,000-0,065)	0,995	0,996	0,221	4 ₅	18,449 (4)	0,0010
4 ₉	Part. Faktorvarianz ³	57,340 (53)	0,3174	0,023 (0,000-0,057)	0,998	0,998	0,088	4 ₅	5,917 (83)	0,1157

¹Interzept für ein Item freigesetzt: CRU_b4 bei FK

²Residualvarianz für ein Item freigesetzt: CRU_b2 bei PFK

³Faktorvarianz freigesetzt bei APS

90%CI: 90%-Konfidenzintervall, CFI: Comparative Fit Index, df: degrees of freedom, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Error, TLI: Tucker Lewis Index

Tab. 25: Gruppenübergreifende Faktorvarianzen und gruppenspezifische Faktormittelwertabweichungen (Referenzgruppe: PHK, fett: p<0,05)

		Alle		PFK		TBK		FK		APS	
	Faktor	Varianz	p	MW _{diff}	p	MW _{diff}	p	MW _{diff}	p	MW _{diff}	p
Modell 2b ₅	LEAD	0,409	0,000	-0,181	0,007	0,036	0,584	-0,131	0,057	-0,069	0,392
	FEED	0,474	0,000	-0,180	0,017	0,185	0,009	0,148	0,046	0,004	0,968
	CULT	1,100 ¹	0,000	0,246	0,014	-0,336	0,001	0,850	0,000	0,199	0,151
	CAMP	0,187	0,000	0,001	0,985	0,022	0,657	0,134	0,005	0,018	0,757
	STAF	0,588 ²	0,000	0,064	0,386	0,325	0,000	0,510	0,000	0,072	0,441
	SPAC	0,739	0,000	0,054	0,525	0,002	0,977	0,474	0,000	0,116	0,324
	TIME	0,360	0,000	0,216	0,001	0,352	0,000	0,608	0,000	0,119	0,132
Modell 3b ₅	FINT	0,450	0,000	0,357	0,000	0,450	0,000	0,992	0,000	0,208	0,026
	IINT _a	0,464	0,000	0,163	0,031	0,464	0,000	0,268	0,003	-0,202	0,065
	IINT _b	0,215 ³	0,000	0,206	0,000	0,215	0,000	0,449	0,000	0,065	0,332
	STER _a	0,084	0,000	0,227	0,000	0,084	0,000	0,330	0,000	0,197	0,000
	STER _b	0,321	0,000	0,291	0,000	0,321	0,000	0,580	0,000	0,265	0,004
	STER _c	1,644 ⁴	0,000	0,761	0,000	1,644	0,000	2,461	0,000	0,400	0,000
4 ₅	CORU _b	1,112 ⁵	0,000	-0,359	0,000	-0,454	0,000	-0,517	0,000	0,679	0,000

¹Varianz unterscheidet sich in der TBK-Gruppe: 0,254 (p=0,000)

²Varianz unterscheidet sich in der FK-Gruppe: 1,341 (p=0,000)

³Varianz unterscheidet sich in der PHK-Gruppe: 0,548 (p=0,000)

⁴Varianz unterscheidet sich in der APS-Gruppe: 0,541 (p=0,000)

⁵Varianz unterscheidet sich in der APS-Gruppe: 0,465 (p=0,000)

APS: Altenpflegeschüler, FK: Führungskräfte, MW_{diff}: Mittelwertdifferenz zur Referenzgruppe (PHK), PFK: Pflegefachkräfte, TBK: Therapie-/Betreuungskräfte

scheidet sich die Gruppe der APS von den PHK. Bei den zustimmungsskalierten ACT-Konzepten

sowie beim informellen Austausch (IINT-Konzepte) besteht kein Unterschied, die formalen Zusammenkünfte (FINT) sowie die drei Ressourcen-Konzepte (STER) werden von dieser Gruppe höher eingeschätzt als von den PHK. LEAD unterscheidet sich nur zwischen PFK und PHK, wobei die PFK die Führungsqualität etwas niedriger bewerten. Auch das Miteinander im Team (CAMP) sowie die räumlichen Verhältnisse (SPAC) werden nur von einer Gruppe, den FK, anders eingeschätzt als von den PHK. Die FK beurteilen diese Faktoren etwas höher. STAF und TIME fallen bei TBK und FK höher aus als bei PHK. PFK, TBK und FK bewerten alle häufigkeitsskalierten ACT-Konzepte höher als die PHK. CORU wird bemerkenswerter Weise von den besser ausgebildeten Berufsgruppen (PFK, TBK und FK) niedriger eingeschätzt als von den PHK, von den APS jedoch etwas höher.

5.3.4 Aggregierbarkeit der ACT-Daten

Die Fitwerte der Mehrebenenmodelle (Tab. 26, Mplus-Outputs auf Nachfrage beim Autor) verweisen auf eine sehr gute Modellpassung in den einzelnen Subgruppen (CFI, RMSEA und TLI)

Tab. 26: Fitindizes der ACT-Mehrebenenmodelle (Cluster: Einrichtung bzw. Wohnbereich)

		N_{L1}	N_{L2}	$\frac{N_{L1}}{N_{L2}}$	χ^2	df	p	RMSEA	CFI	TLI	SRMR _w	SRMR _B
Modell 2b (Cluster: Einrichtung)	PHK	251	38	6,605	757,135	716	0,1759	0,014	0,962	0,957	0,054	0,251
	PFK	179	38	4,711	565,988	574	0,5862	0,000	1,000	1,016	0,059	0,360
	TBK	141	38	3,711	551,771	541	0,3648	0,012	0,967	0,962	0,071	0,411
	FK	118	38	3,105	548,227	567	0,7067	0,000	1,000	1,085	0,072	0,411
	APS	58	27	2,231	446,613	477	0,8374	0,000	1,000	1,285	0,091	0,549
	Alle	753	38	19,816	922,497	696	0,0000	0,021	0,924	0,914	0,040	0,181
Modell 2b (Cluster: Wohnbereich)	PHK	251	91	2,758	773,003	714	0,0620	0,018	0,932	0,922	0,063	0,173
	PFK	179	91	1,967	629,905	666	0,8522	0,000	1,000	1,113	0,064	0,368
	TBK	141	73	1,932	673,266	663	0,3826	0,010	0,962	0,957	0,086	0,416
	FK	118	93	1,269	536,641	548	0,6279	0,000	1,000	1,065	0,108	0,379
	APS	58	44	1,318	550,564	579	0,9132	0,000	1,000	2,496	0,122	0,428
	Alle	753	143	5,266	953,279	718	0,0000	0,021	0,920	0,910	0,041	0,122
Modell 3b (Cluster: Einrichtung)	PHK	263	38	6,921	295,464	276	0,2010	0,016	0,968	0,960	0,051	0,219
	PFK	187	38	4,921	237,446	241	0,5526	0,000	1,000	1,013	0,048	0,241
	TBK	147	38	3,868	235,143	226	0,3243	0,017	0,949	0,935	0,078	0,469
	FK	117	38	3,079	151,096	147	0,3915	0,015	0,975	0,967	0,071	0,316
	APS	63	27	2,333	182,435	190	0,6402	0,000	1,000	1,127	0,100	0,368
	Alle	777	38	20,447	426,422	275	0,0000	0,027	0,932	0,915	0,038	0,166
Modell 3b (Cluster: Wohnbereich)	PHK	263	94	2,798	297,356	258	0,0464	0,024	0,937	0,921	0,058	0,222
	PFK	187	94	1,989	208,219	227	0,8093	0,000	1,000	1,083	0,059	0,309
	TBK	147	74	1,986	246,013	231	0,2374	0,021	0,922	0,901	0,084	0,327
	FK	117	94	1,245	153,092	168	0,7887	0,000	1,000	1,128	0,082	0,374
	APS	63	47	1,340	174,362	201	0,9129	0,000	1,000	1,994	0,135	0,287
	Alle	777	143	5,434	437,489	276	0,0000	0,027	0,924	0,906	0,034	0,167

APS: Altenpflegeschüler, CFI: *Comparative Fit Index*, df: *degrees of freedom*, FK: Führungskräfte, N_{L1} : Zahl der Level-1-Einheiten (individuelle Probanden), N_{L2} : Zahl der Level-2-Einheiten (Einrichtungen oder Wohnbereiche), PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, RMSEA: *Root Mean Square Error of Approximation*, SRMR: *Standardized Root Mean Square Error*, TLI: *Tucker Lewis Index*, TBK: Therapie-/Betreuungskräfte

sowie des *Within*-Teils der Modelle (SRMR_w). Die CFI- und TLI-Werte der Gesamtstichproben-

modelle liegen jedoch unter 0,95. Die Passung der *Between*-Teile aller Modelle mit SRMR_B-Werten weit über 0,8 (zwischen 0,122 und 0,549) ist dagegen nur suboptimal. Vor allem in den Subgruppenmodellen ist die durchschnittliche Clustergröße sehr gering – je nach Modell und Subgruppe zwischen 2,231 und 6,921 Individuen pro Einrichtung bzw. zwischen 1,245 und 2,798 Individuen pro Wohnbereich. Auch die durchschnittliche Clustergröße der Modelle für die Gesamtstichprobe ist mit 19,816 bzw. 20,447 Individuen pro Einrichtung und 5,266 bzw. 5,434 Individuen pro Wohnbereich nicht ideal. Ein Großteil der *Between*-Faktorvarianzen (Faktorvarianzen zwischen den Einrichtungen bzw. Wohnbereichen, Tab. 27) ist dann auch statistisch nicht signifikant ($p > 0,05$). Zwar weisen einige der ACT-Items Intraklassenkorrelationen (ICC) von deut-

Tab. 27: *Between*-Faktorvarianzen der ACT-Mehrebenenmodelle (fett: $p < 0,05$)

		PHK				PFK				TBK			
		Einrichtung		Wohnbereich		Einrichtung		Wohnbereich		Einrichtung		Wohnbereich	
Konzept		Var.	p	Var.	p	Var.	p	Var.	p	Var.	p	Var.	p
Zustimmungsskalierte ACT-Konzepte	LEAD	0,009	0,491	0,070	0,167	0,021	0,595	0,107	0,393	0,156	0,182	0,156	0,058
	CULT	0,058	0,142	0,021	0,310	0,038	0,470	0,019	0,676	-0,035	0,827	0,124	0,109
	FEED	0,350	0,018	0,401	0,007	0,033	0,564	0,057	0,530	0,243	0,324	-0,062	0,460
	CAMP	0,029	0,246	0,074	0,182	0,035	0,196	0,055	0,264	0,048	0,237	0,054	0,160
	STAF	0,122	0,238	0,221	0,074	0,183	0,149	0,105	0,396	-0,020	0,677	0,004	0,752
	SPAC	0,067	0,513	0,284	0,012	-0,527	0,000	0,159	0,192	—	—	0,076	0,306
	TIME	0,076	0,192	0,070	0,304	0,052	0,451	0,029	0,619	—	—	—	—
Häufigkeitsskalierte ACT-Konzepte	FINT	0,127	0,186	0,406	0,000	0,016	0,608	0,074	0,316	0,112	0,435	0,096	0,236
	IICP	0,023	0,480	0,399	0,000	0,023	0,396	0,010	0,739	-0,006	0,934	0,010	0,811
	IISP	0,018	0,590	0,176	0,011	0,080	0,429	0,089	0,428	-0,048	0,560	0,021	0,539
	SERW	0,002	0,576	0,075	0,000	0,034	0,344	0,001	0,919	-0,743	0,000	—	—
	SERR	0,178	0,214	0,235	0,008	0,138	0,291	0,020	0,627	0,076	0,566	0,031	0,667
	SERE	0,265	0,318	0,804	0,000	1,297	0,279	1,503	0,156	0,246	0,633	0,363	0,557
		FK				APS				Alle			
		Einrichtung		Wohnbereich		Einrichtung		Wohnbereich		Einrichtung		Wohnbereich	
Konzept		Var.	p	Var.	p	Var.	p	Var.	p	Var.	p	Var.	p
Zustimmungsskalierte ACT-Konzepte	LEAD	0,035	0,636	-0,064	0,785	0,007	0,851	0,028	0,762	0,021	0,149	0,131	0,008
	CULT	0,046	0,519	0,202	0,070	0,004	0,962	0,132	0,566	0,044	0,169	0,044	0,072
	FEED	0,007	0,815	0,330	0,415	0,140	0,603	0,112	0,627	0,114	0,115	0,238	0,002
	CAMP	0,051	0,338	0,171	0,163	0,070	0,483	0,061	0,498	0,008	0,247	0,045	0,016
	STAF	0,172	0,305	0,780	0,015	0,045	0,843	0,109	0,661	0,117	0,048	0,224	0,007
	SPAC	0,119	0,642	0,619	0,016	0,027	0,910	0,030	0,915	0,014	0,518	0,161	0,009
	TIME	0,105	0,420	0,257	0,037	0,127	0,424	0,194	0,242	0,051	0,138	0,128	0,012
Häufigkeitsskalierte ACT-Konzepte	FINT	—	—	0,129	0,673	0,184	0,508	-0,416	0,428	0,064	0,114	0,040	0,086
	IICP	0,014	0,829	0,041	0,681	0,517	0,135	0,786	0,088	0,019	0,281	0,106	0,042
	IISP	0,000	0,998	—	—	0,357	0,107	0,034	0,710	0,008	0,535	0,108	0,079
	SERW	0,000	0,999	0,001	0,978	0,035	0,631	-0,005	0,985	0,008	0,261	0,045	0,012
	SERR	0,004	0,979	—	—	0,325	0,259	-0,160	0,904	0,062	0,258	0,037	0,274
	SERE	0,120	0,675	1,815	0,243	-0,010	0,920	0,962	0,291	0,221	0,371	0,991	0,133

APS: Altenpflegeschüler, CAMP: *Connections Among People*, CULT: *Culture*, FEED: *Feedback*, FINT: *Formal Interactions*, FK: Führungskräfte, IICP: *Informal Interactions*, LEAD: *Leadership*, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, SPAC: *Space*, STAF: *Staffing*, STER: *Structural/Electronic Resources*, TBK: *Therapie-/Betreuungskräfte*, TIME: *Time*, Var.: Varianz

lich $> 0,1$ auf – insbesondere bei auf Wohnbereichsebene geclusterten Daten (Tab. 28) –, es finden sich aber auch zahlreiche Items mit sehr viel kleineren ICC-Werten bis hin zu null. Die Mehrzahl

Tab. 28: Intraklassenkorrelationskoeffizienten (ICC) der ACT-Items (fett: ICC>0,1)

Konzept	Item	PHK		PFK		TBK		FK		APS		Alle			
		Einr.	WB	Einr.	WB	Einr.	WB	Einr.	WB	Einr.	WB	Einr.	WB		
Zustimmungsskalierte ACT-Konzepte	LEAD	L1	0,009	0,104	0,013	0,132	0,192	0,348	0,000	0,000	0,051	0,042	0,029	0,123	
		L2	0,011	0,118	0,000	0,000	0,084	0,117	0,000	0,000	0,085	0,000	0,000	0,078	
		L3	0,130	0,238	0,060	0,126	0,000	0,000	0,072	0,085	0,447	0,369	0,060	0,158	
		L4	0,084	0,232	0,028	0,095	0,000	0,043	0,062	0,034	0,155	0,135	0,036	0,132	
		L5	0,053	0,103	0,000	0,077	0,025	0,055	0,063	0,000	0,001	0,000	0,016	0,069	
		L6	0,101	0,244	0,008	0,068	0,087	0,192	0,000	0,000	0,020	0,124	0,021	0,127	
	CULT	C1	0,095	0,113	0,050	0,139	0,074	0,186	0,317	0,371	0,000	0,000	0,073	0,090	
		C4	0,088	0,301	0,000	0,000	0,041	0,456	0,243	0,185	0,000	0,222	0,049	0,085	
		C5	0,077	0,184	0,117	0,178	0,000	0,006	0,261	0,339	0,000	0,102	0,027	0,058	
		C6	0,142	0,254	0,046	0,065	0,111	0,176	0,028	0,188	0,010	0,090	0,045	0,103	
	FEED	F1	0,228	0,284	0,029	0,031	0,240	0,293	0,010	0,225	0,093	0,061	0,083	0,130	
		F2	0,262	0,334	0,065	0,131	0,202	0,259	0,000	0,000	0,025	0,098	0,108	0,160	
		F3	0,218	0,352	0,027	0,104	0,209	0,292	0,014	0,044	0,000	0,036	0,094	0,168	
		F4	0,244	0,336	0,031	0,088	0,238	0,325	0,000	0,000	0,139	0,351	0,115	0,168	
		F5	0,299	0,391	0,001	0,003	0,210	0,328	0,000	0,129	0,101	0,212	0,123	0,187	
		F6	0,364	0,342	0,001	0,061	0,237	0,302	0,150	0,124	0,000	0,000	0,152	0,191	
	CAMP	CAP1	0,051	0,108	0,075	0,113	0,065	0,128	0,000	0,000	0,177	0,191	0,025	0,089	
		CAP2	0,078	0,213	0,111	0,180	0,253	0,245	0,000	0,000	0,012	0,270	0,037	0,129	
		CAP4	0,117	0,243	0,000	0,000	0,300	0,464	0,117	0,511	0,000	0,114	0,090	0,193	
		CAP5	0,110	0,286	0,000	0,137	0,111	0,193	0,122	0,098	0,351	0,288	0,028	0,103	
		CAP6	0,057	0,204	0,076	0,206	0,136	0,188	0,211	0,200	0,152	0,220	0,042	0,131	
		STAF	ST1	0,075	0,144	0,158	0,206	0,014	0,027	0,121	0,472	0,000	0,116	0,106	0,213
	SPAC	ST2	0,221	0,284	0,211	0,194	0,107	0,227	0,285	0,406	0,040	0,308	0,200	0,266	
		SP2	0,090	0,223	0,053	0,111	0,000	0,379	0,158	0,493	0,023	0,026	0,046	0,179	
		SP3	0,017	0,254	0,000	0,165	0,000	0,277	0,048	0,104	0,000	0,000	0,015	0,118	
	TIME	T1	0,098	0,216	0,138	0,220	0,000	0,071	0,208	0,418	0,191	0,682	0,083	0,200	
		T2	0,192	0,261	0,028	0,011	0,000	0,000	0,045	0,221	0,000	0,124	0,048	0,142	
		T3	0,045	0,154	0,019	0,081	0,000	0,017	0,088	0,082	0,000	0,209	0,026	0,071	
		T4	0,090	0,151	0,000	0,011	0,000	0,218	0,017	0,282	0,000	0,064	0,019	0,130	
	Häufigkeitsskalierte ACT-Konzepte	FINT	FI1	0,121	0,149	0,037	0,103	0,098	0,120	0,000	0,000	0,174	0,105	0,069	0,100
			FI2	0,059	0,085	0,096	0,195	0,072	0,268	0,000	0,089	0,063	0,000	0,062	0,097
			FI5	0,062	0,030	0,048	0,000	0,255	0,477	0,000	0,000	0,003	0,000	0,041	0,079
IINT _a		II1	0,048	0,125	0,054	0,000	0,000	0,013	0,042	0,172	0,450	0,787	0,027	0,118	
		II2	0,038	0,128	0,035	0,010	0,058	0,078	0,073	0,068	0,368	0,180	0,026	0,086	
		II10	0,163	0,257	0,073	0,070	0,114	0,140	0,000	0,000	0,145	0,341	0,090	0,153	
IINT _b		II4	0,075	0,173	0,000	0,130	0,009	0,002	0,000	0,000	0,000	0,276	0,057	0,122	
		II5	0,160	0,233	0,192	0,129	0,069	0,070	0,000	0,011	0,412	0,284	0,078	0,102	
		II6	0,215	0,219	0,161	0,104	0,152	0,285	0,000	0,000	0,213	0,230	0,135	0,212	
STER _a		SER1	0,008	0,000	0,080	0,000	0,000	0,000	0,003	0,000	0,095	0,240	0,011	0,069	
		SER2	0,133	0,177	0,082	0,021	0,004	0,000	0,007	0,042	0,000	0,077	0,037	0,093	
		SER3	0,111	0,270	0,157	0,139	0,000	0,000	0,000	0,122	0,000	0,000	0,054	0,132	
		SER4	0,204	0,185	0,000	0,029	0,132	0,174	0,000	0,000	0,216	0,166	0,100	0,112	
		SER5	0,336	0,426	0,092	0,089	0,005	0,085	0,000	0,000	0,000	0,251	0,127	0,179	
		SER6	0,261	0,368	0,159	0,245	0,035	0,071	0,002	0,000	0,000	0,000	0,066	0,096	
STER _c		SER7	0,386	0,370	0,591	0,606	0,168	0,169	0,131	0,948	0,474	0,275	0,242	0,389	
		SER8	0,389	0,365	0,495	0,483	0,283	0,212	0,217	0,318	0,618	0,500	0,239	0,311	
		SER9	0,217	0,338	0,590	0,538	0,077	0,103	0,000	0,798	0,000	0,000	0,155	0,366	
	SER10	0,160	0,121	0,412	0,380	0,033	0,037	0,000	0,478	0,198	0,624	0,120	0,352		

APS: Altenpflegeschüler, CAMP: *Connections Among People*, CULT: *Culture*, Einr.: Einrichtung, FEED: *Feedback*, FINT: *Formal Interactions*, FK: Führungskräfte, IINT: *Informal Interactions*, LEAD: *Leadership*, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, SPAC: *Space*, STAF: *Staffing*, STER: *Structural/Electronic Resources*, TBK: Therapie-/Betreuungskräfte, TIME: *Time*, WB: Wohnbereich

der Faktorladungen in den *Between*-Modellteilen weisen ebenfalls nichtsignifikante p-Werte auf.

Entgegen der kanadischen Befunde kann eine Clusterung der ACT-Daten auf Einrichtungs- bzw. Wohnbereichsebene in dieser Studie nicht bestätigt werden. Versuchsweise wurden die beobachteten ACT-Scores zusätzlich entsprechend der kanadischen Methode auf Aggregierbarkeit überprüft (Berechnung der ICC1-, ICC2, η^2 - und ω^2 -Werte auf Basis permutationsbasierter ANOVAs). Die ANOVAs waren in der Gesamtstichprobe für alle ACT-Konzepte sowohl für die Einrichtungs- als auch die Wohnbereichsebene signifikant und die meisten ICC- und/oder ω^2 -Werte stützten die Aggregierbarkeitsannahme. Aufgrund der robusteren Methodik werden die Befunde der Faktorenmodelle als vertrauenswürdiger angesehen als die der ANOVAs. Daher wurden die weiteren Modelle nicht als Mehrebenenmodelle spezifiziert.

Um die Invarianz der Faktorvarianzen sicherzustellen wurden abschließend Mehrebenenmodelle spezifiziert, in denen die Berufsgruppenzugehörigkeit als Clustervariable definiert wurde (Tab. 29). In allen drei Modellen (ACT 2b und 3b sowie CRU 4) waren die Faktorvarianzen der *Between*-Ebene statistisch nicht signifikant ($p > 0,05$), in den beiden ACT-Modellen zeigten die SRMR_B-Werte eine schlechte Passung des *Between*-Teils auf. Die Invarianz der Faktorvarianzen zwischen den Berufsgruppen kann somit als bestätigt angesehen werden.

Tab. 29: Fitindizes der ACT- und RU-Mehrebenenmodelle (Cluster: Berufsgruppe)

Modell	N _{L1}	N _{L2}	$\frac{N_{L1}}{N_{L2}}$	χ^2	df	p	RMSEA	CFI	TLI	SRMR _w	SRMR _B
2b	753	6	125,500	671,457	692	0,7053	0,000	1,000	1,040	0,042	0,483
3b	777	5	155,400	253,928	275	0,8143	0,000	1,000	1,157	0,045	0,194
4	791	6	131,833	10,383	11	0,4963	0,000	1,000	1,002	0,006	0,001

CFI: Comparative Fit Index, df: degrees of freedom, N_{L1}: Zahl der Level-1-Einheiten (individuelle Probanden), N_{L2}: Zahl der Level-2-Einheiten (Berufsgruppen), RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Error, TLI: Tucker Lewis Index, TBK: Therapie-/Betreuungskräfte

5.3.5 Zusammenhänge zwischen den Variablen

Das Gesamtmodell aller RU-Items (Abb. 11) weist eine sehr gute Modellpassung auf (Tab. 30, Mplus-Outputs auf Nachfrage beim Autor). Lediglich der RMSEA-Wert des PFK-Modells bleibt über dem empfohlenen Wert von 0,06. Die Modelle erklären zwischen 60,4% und 80,6%

Tab. 30: Fitindizes RU-Gesamtmodell

	N	χ^2	df	p	RMSEA (90%CI)	CFI	TLI	SRMR	R ²	PR ²
PHK	267	17,849	16	0,3328	0,021 (0,000-0,062)	0,999	0,998	0,016	0,806	0,000
PFK	181	34,752	16	0,0043	0,080 (0,043-0,117)	0,974	0,955	0,034	0,651	0,000
TBK	147	19,627	16	0,2375	0,039 (0,000-0,090)	0,993	0,988	0,020	0,782	0,000
FK	123	18,391	15	0,2427	0,043 (0,000-0,100)	0,994	0,989	0,023	0,612	0,000
APS	62	10,344	16	0,8480	0,000 (0,000-0,067)	1,000	1,061	0,037	0,604	0,000
Alle	786	43,496	16	0,0002	0,047 (0,030-0,064)	0,993	0,988	0,016	0,744	0,000

90%CI: 90%-Konfidenzintervall, APS: Altenpflegeschüler, CFI: Comparative Fit Index, df: degrees of freedom, FK: Führungskräfte, PFK: Pflegefachkräfte, PHK: Pflegehilfskräfte, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Error, TBK: Therapie-/Betreuungskräfte, TLI: Tucker Lewis Index

der Varianz abhängigen Variablen „Forschungsnutzung im Allgemeinen“ (OVRU). Auch das Faktorenmodell der individuellen Variablen in der Gesamtstichprobe (N=735 gültige Fälle) (Abb.

12) weist eine akzeptable Modellpassung auf: $\chi^2=732,985$, $df=422$, $p=0,0000$, $RMSEA=0,030$ (90% CI=0,026-0,034), $CFI=0,952$, $TLI=0,946$, $SRMR=0,053$. Daher konnten neben den ACT-Skalen auch diese Variablen in die Gesamtstrukturgleichungsmodelle aufgenommen werden.

Die Fitindizes für die Strukturgleichungsmodelle (Tab. 31) weisen auf eine akzeptable Modellpassung hin. Lediglich die TLI-Werte der Modelle 7b, 8b, 9b (Strukturbilder siehe Abb. 13) liegen knapp unter dem empfohlenen Wert von 0,95. Die Ergebnisse dieser Modelle sind somit als vertrauenswürdig einzuordnen.

Tab. 31: Fitindizes der Strukturgleichungsmodelle

Modell	N	χ^2	df	p	RMSEA (90%CI)	CFI	TLI	SRMR
7a	702	986,543	690	0,0000	0,025 (0,021-0,028)	0,963	0,958	0,048
7b	702	1079,700	699	0,0000	0,028 (0,025-0,031)	0,952	0,946	0,050
8a	727	863,892	570	0,0000	0,027 (0,023-0,030)	0,969	0,963	0,039
8b	727	998,073	581	0,0000	0,031 (0,028-0,035)	0,956	0,949	0,045
9a	751	499,235	278	0,0000	0,033 (0,028-0,037)	0,972	0,964	0,041
9b	727	680,816	289	0,0000	0,042 (0,038-0,048)	0,950	0,939	0,049

90%CI: 90%-Konfidenzintervall, CFI: *Comparative Fit Index*, df: *degrees of freedom*, RMSEA: *Root Mean Square Error of Approximation*, SRMR: *Standardized Root Mean Square Error*, TLI: *Tucker Lewis Index*

Tab. 32 zeigt die Zusammenhänge (Strukturkoeffizienten) zwischen den latenten Variablen der Strukturgleichungsmodelle. In den drei ersten Modellen (7a, 8a und 9a) sind die drei Konzepte instrumentelle, konzeptuelle und persuasive Forschungsnutzung (INRU, CORU, PERU) jeweils signifikant und in vorhergesagter Richtung mit Forschungsanwendung im Allgemeinen (OVRU) assoziiert. In Modell 7a sind zudem die individuellen Merkmale verhaltensbezogener Stress (BSTR) und Einstellung gegenüber der Anwendung von Forschungswissen (ATRU) signifikante (positiv assoziierte) Prädiktoren für OVRU. In Modell 8a beeinflusst die verfügbare Zeit (TIME) als einziges zustimmungsskaliertes ACT-Konzept OVRU signifikant. Keines der häufigkeitsskalierten ACT-Konzepte (Modell 9a) hängt signifikant mit OVRU zusammen.

Auch in den Modellen 7b, 8b und 9b sind die drei RU-Konzepte signifikant positiv assoziierte Prädiktoren für OVRU. Sie fungieren zudem als Mediatoren für einige der anderen Konzepte. So sind in Modell 7b ATRU und kognitiver Stress (CSTR) signifikant und in vorhergesagter Richtung mit jedem der drei RU-Konzepte assoziiert – vor allem CSTR mit sehr hohen Strukturkoeffizienten (γ). Auch BSTR weist hohe und signifikante Zusammenhänge mit den drei RU-Konzepten auf, allerdings in anderer Richtung als angenommen. Das gleiche gilt für Burnout (BURN), das INRU und CORU negativ beeinflusst. Auch einige der zustimmungsskalierten ACT-Konzepte sind in unvorhergesehener Richtung mit RU assoziiert (Modell 8b): Führung (LEAD), Personalbesetzung (STAF) sowie Zeit (TIME) mit allen drei RU-Konzepten, Miteinander im Team (CAMP) nur mit CORU. Kultur (CULT) beeinflusst alle drei RU-Konzepte in vorhergesagter Richtung. Unter den häufigkeitsskalierten ACT-Konzepten hängen zwei der drei Ressourcenkonzepte (STER_b: regulative Ressourcen und STER_c: EDV-basierte Ressourcen) in unvorhergesehener Richtung mit allen drei RU-Konzepten zusammen. Der informelle Austausch mit eher be-

wohnerfern tätigen Personen (IINT_b) beeinflusst INRU und CORU negativ. In vorhergesehener Richtung sind die offiziellen Zusammenkünfte (FINT) und der informelle Austausch mit bewohnernah tätigen Personen (IINT_a) mit allen drei RU-Konzepten assoziiert.

Tab. 32: Strukturkoeffizienten und erklärte Varianz der Strukturgleichungsmodelle (fett: p<0,05)

Abh. Var.:	Variante a		Variante b								
	OVRU		INRU		CORU _b		PERU		OVRU		
	γ	p	γ	p	γ	p	γ	p	β	p	
Prädiktoren Modell 7	JOSA	-0,102	0,299	-3,446	0,063	-3,903	0,050	-3,503	0,069	—	—
	GEHE	0,014	0,528	0,116	0,636	0,071	0,790	0,033	0,897	—	—
	BSTR	0,226	0,038	-11,006	0,035	-12,048	0,032	-12,005	0,027	—	—
	CSTR	-0,091	0,288	12,159	0,027	13,180	0,026	12,805	0,025	—	—
	BURN	-0,131	0,171	-3,645	0,038	-3,834	0,043	-3,404	0,064	—	—
	ATRU	0,355	0,000	2,825	0,006	2,802	0,010	2,888	0,007	—	—
	INRU	0,303	0,000	—	—	—	—	—	—	0,339	0,000
	CORU _b	0,209	0,000	—	—	—	—	—	—	0,249	0,000
	PERU	0,373	0,000	—	—	—	—	—	—	0,370	0,000
	<i>R² (p)</i>	0,760 (0,000)		0,514 (0,000)		0,753 (0,000)		0,641 (0,000)		0,739 (0,000)	
Prädiktoren Modell 8	LEAD	0,100	0,196	-5,167	0,002	-5,423	0,002	-5,293	0,002	—	—
	CULT	-0,216	0,189	9,700	0,001	10,421	0,001	9,753	0,001	—	—
	FEED	-0,029	0,305	-0,132	0,552	-0,110	0,642	-0,068	0,760	—	—
	CAMP	0,275	0,098	-0,165	0,302	-0,334	0,011	—	—	—	—
	STAF	-0,069	0,101	-0,932	0,012	-1,029	0,008	-1,061	0,004	—	—
	SPAC	0,022	0,532	-0,330	0,263	-0,441	0,160	-0,492	0,101	—	—
	TIME	0,180	0,024	-1,911	0,035	-2,037	0,032	-1,787	0,049	—	—
	INRU	0,341	0,000	—	—	—	—	—	—	0,351	0,000
	CORU _b	0,264	0,000	—	—	—	—	—	—	0,261	0,000
	PERU	0,343	0,000	—	—	—	—	—	—	0,355	0,000
<i>R² (p)</i>	0,756 (0,000)		0,480 (0,000)		0,775 (0,000)		0,633 (0,000)		0,744 (0,000)		
Prädiktoren Modell 9	FINT	0,100	0,302	7,653	0,000	8,372	0,000	7,193	0,000	—	—
	IINT _a	0,100	0,058	1,186	0,001	1,226	0,001	1,128	0,001	—	—
	IINT _b	0,063	0,779	-0,927	0,001	-0,945	0,000	—	—	—	—
	STER _a	-0,087	0,431	-0,045	0,810	-0,285	0,077	—	—	—	—
	STER _b	0,001	0,987	-4,217	0,003	-4,654	0,003	-4,211	0,002	—	—
	STER _c	0,010	0,719	-1,106	0,000	-1,158	0,000	-1,048	0,000	—	—
	INRU	0,340	0,000	—	—	—	—	—	—	0,339	0,000
	CORU _b	0,279	0,000	—	—	—	—	—	—	0,274	0,000
	PERU	0,333	0,000	—	—	—	—	—	—	0,354	0,000
<i>R² (p)</i>	0,759 (0,000)		0,516 (0,000)		0,792 (0,000)		0,613 (0,000)		0,752 (0,000)		

β : Strukturkoeffizienten zwischen endogenen Variablen, γ : Strukturkoeffizienten zwischen exogenen und endogenen Variablen
 Abh. Var.: Abhängige Variable, ATRU: Attitudes Towards RU, BSTR: Behavioural Stress, BURN: Burnout, CAMP: Connections Among People, CORU: Conceptual Research Use, CSTR: Cognitive Stress, CULT: Culture, FEED: Feedback, FINT: Formal Interactions, GEHE: General Health, IINT: Informal Interactions, INRU: Instrumental Research Use, JOSA: Job Satisfaction, LEAD: Leadership, OVRU: Overall Research Use, PERU: Persuasive Research Use, SPAC: Space, STAF: Staffing, STER: Structural/Electronic Resources, TIME: Time

6 Diskussion

In dieser Arbeit wurde das bisher einzige Erhebungsinstrument zur Erfassung DI-relevanter, potenziell modifizierbarer Organisationskontextfaktoren in deutschen Pflegeheimen kreiert und umfassend validiert. Zu diesem Zweck wurde ein bewährtes kanadisches Instrument – das ACT –

auf Basis internationaler *Best-Practice*-Leitlinien übersetzt und an den kulturellen und strukturellen Kontext der deutschen stationären Altenpflege angepasst. Übersetzt und validiert wurden weiterhin zwei RU-Tools – *Estabrooks' Kinds of RU items* und die CRU-Skala. Diese beiden RU-Skalen sind bisher die einzigen, deren deutsche Versionen umfassend statistisch überprüft wurden. Zudem wurden in dieser Arbeit auch international erstmals Reliabilitäts- und Validitätsbefunde der Instrumente für andere Berufsgruppen als PHK und PFK generiert. Die statistische Validierung folgte den international als *best practice* in Sachen psychometrische Testung anerkannten *Standards* (AERA et al., 1999). In drei Studienphasen mit je unterschiedlichen Methoden und Stichproben wurden vier Evidenzquellen für Validität untersucht: a) Instrumenteninhalte (Phase 1, Übersetzungsprozess: rigoroses Design des Übersetzungsprozesses, Abstimmung mit den Instrumentenentwicklerinnen und Expertenpanels), b) Antwortverhalten der Zielpersonen (*response processes*) (Phase 2, *Cognitive Debriefings*) sowie c) interne Struktur der Instrumente und d) Zusammenhänge zwischen den Variablen (Phase 3, psychometrische Testung: konfirmatorische Faktorenanalysen, konfirmatorische Mehrgruppenfaktorenmodelle zur Faktorinvarianzanalyse, konfirmatorische Mehrebenenfaktorenmodelle zur Bestimmung der Aggregierbarkeit der Daten auf Einrichtungs- bzw. Wohnbereichsebene und Strukturgleichungsmodelle).

6.1 Übersetzung und Adaptation der Instrumente

6.1.1 Einordnung der Erkenntnisse des Übersetzungsprozesses

Werden Erhebungsinstrumente in eine andere Sprache übersetzt, sind detaillierte Informationen über den Übersetzungsprozess erforderlich, um a) Möglichkeiten und Grenzen bzw. Stärken und Schwächen der übersetzten Instrumente abschätzen zu können, b) die Instrumente möglichst optimal einsetzen zu können, c) die Qualität von Studien, in denen die Instrumente eingesetzt wurden, bewerten zu können und d) die erhobenen Daten interpretieren zu können (Maneesriwongul and Dixon, 2004; Acquadro et al., 2008). Leider werden diese Informationen selten ausreichend detailliert berichtet (Maneesriwongul and Dixon, 2004; Acquadro et al., 2008). Methoden und Erkenntnisse des Übersetzungsprozesses – insbesondere Herausforderungen und entsprechende Lösungsstrategien – wurden daher im Rahmen dieser Arbeit ausführlich dokumentiert und international publiziert (Hoben et al., 2013). Übersetzungen der hier verwendeten Tools in andere Sprachen bzw. Übersetzungen ähnlicher Instrumente auf Deutsch können auf Basis dieser Erfahrungen optimiert werden.

Die primäre Herausforderung bestand darin, die kanadischen Instrumente einerseits an den Kontext der deutschen stationären Altenpflege anzupassen, dabei aber andererseits so nah wie möglich an den Originalfragebögen zu bleiben, um eine Vergleichbarkeit der Daten sicherzustellen. Durch die Anwendung rigoroser Übersetzungsmethoden gelang es, diese kulturelle Äquivalenz zu erzielen. Ein solcher Übersetzungsprozess ist ressourcenaufwändig und anforderungsreich (Acquadro et al., 2008; Wild et al., 2009). Die Herausforderungen, die im Rahmen dieser Arbeit

zu bewältigen waren, entsprechen denen, die auch im Rahmen anderer Übersetzungsprojekte berichtet werden (Bravo et al., 1993; Parent-Vachon et al., 2008; Mahler et al., 2009; Menon et al., 2012; Eldh et al., 2013).

Spezifisch für diese Arbeit war insbesondere die Herausforderung, angemessene Formulierungen für die Gruppe der PHK zu finden. PHK sind, was ihre Merkmale angeht, die heterogenste Berufsgruppe in Pflegeheimen. Das Spektrum reichte in dieser Arbeit von Personen ohne Schulabschluss und Berufsausbildung bis hin zu PHK mit einem akademischen Abschluss (z.B. ein russisches Diplom in Erziehungswissenschaft, das hierzulande nicht anerkannt wird). Tendenziell finden sich in dieser Gruppe jedoch überwiegend Personen mit eher niedrigem Bildungs- und Qualifikationsgrad. PHK waren daher mit Begriffen wie „*best practice*“ oder „Einschätzungsskala“ kaum vertraut und die Relevanz von Forschung und daraus resultierender Erkenntnisse für ihre Arbeit war ihnen oft unklar. Bisweilen fehlten deutsche Alternativen zu englischen Begriffen (z.B. *best practice*), z.T. verstanden die PHK die deutschen Begriffe nicht (z.B. Einschätzungsskala). Für PHK verständliche Formulierungen zu finden, die a) die Bedeutung des Originalwortlauts nicht unangemessen veränderten und b) nicht zu lang und komplex waren, erforderte einigen Aufwand. Wie auch in der schwedischen Übersetzung des ACT (Eldh et al., 2013) waren hier die Diskussionen mit den Instrumentenentwicklerinnen von großer Bedeutung. Sie halfen dabei, die Bedeutung der in den Instrumenten enthaltenen Konzepte zu klären, Schwierigkeiten frühzeitig zu erkennen und ggf. zu vermeiden und kulturelle Äquivalenz sicherzustellen. Auch die Expertenpanels waren essenziell. Mit ihrer Hilfe konnten Adaptationsbedarfe frühzeitig erkannt, verständliche Formulierungen gefunden und Inhaltsvalidität sichergestellt werden.

Für die Übersetzung der PHK-Fragebögen waren einige Diskussions-, Abstimmungs-, *Cognitive-Debriefing*- und Modifikationsrunden erforderlich. Dieser hohe Aufwand zahlte sich jedoch aus: Die Übersetzung der weiteren Fragebogenversionen erforderte deutlich weniger Zeit. Die übernommenen Formulierungen wurden von den anderen Berufsgruppen ohne Schwierigkeiten verstanden und die Zahl der Abstimmungs- und Modifikationsrunden reduzierte sich von Gruppe zu Gruppe.

6.1.2 Limitationen des Übersetzungsprozesses

Eine Limitation verfügbarer Übersetzungsleitlinien besteht darin, dass diese sich nahezu ausnahmslos auf Instrumente zur Erfassung patientenbezogener Outcomes (z.B. subjektiv berichtete Lebensqualität) beziehen. Zum Zeitpunkt der Konzeption dieser Arbeit konnte keine Leitlinie gefunden werden, die sich auf die Übersetzung von Instrumenten bezog, mit denen personalbezogene Outcomes (z.B. Arbeitszufriedenheit) oder organisationsbezogene Outcomes (z.B. Organisationskultur) erfasst werden. Inzwischen wurde von Squires et al. (2013a) ein Verfahren zur simultanen Übersetzung und Harmonisierung letztgenannter Instrumente in mehrere Sprachen vorgelegt. Es bezieht sich auf Instrumente zur Erfassung organisationsbezogener Outcomes in

Institutionen des Gesundheitswesens. Personen, die beruflich im Gesundheitswesen tätig sind, agieren innerhalb professioneller und organisationaler Kontexte. Diese Subkulturen beeinflussen die Übersetzung ebenso, wie der übergreifende kulturelle Kontext des jeweiligen Landes. Konsequenzen dieser Subkulturen für die Übersetzung von Instrumenten werden von Übersetzungsleitlinien, die sich auf patientenbezogene Instrumente beziehen, nicht reflektiert und methodisch berücksichtigt. So werden sich vermutlich Qualifikation und professionelle Rollen zwischen Ursprungs- und Zielkultur unterscheiden, möglicherweise existieren bestimmte Berufsgruppen in einer der Kulturen gar nicht, Fachtermini werden evtl. unterschiedlich verwendet etc. Bereits geringfügig unterschiedliche Formulierungen können jedoch, wie Squires et al. (2013a) verdeutlichen, die Bedeutung eines Fragebogenitems, ja die gesamte Itemstruktur grundlegend verändern. Mitglieder von Übersetzungsteams müssen daher sowohl mit den Konzepten der zu übersetzenden Instrumente vertraut sein als auch mit den kulturellen Kontexten des jeweiligen Gesundheitssettings beider Länder. Die hier verwendeten Übersetzungsleitlinien für Instrumente zur Erfassung patientenbezogener Outcomes (Wild et al., 2005; Acquadro et al., 2008) waren wertvolle Instrumente für die Konzeption und Durchführung des Übersetzungsprozesses. Künftige Übersetzungen von Instrumenten zur Erfassung personal- bzw. organisationsbezogener Outcomes werden aber sicher vom Einsatz spezifischerer Verfahren, wie etwa dem von Squires et al. (2013a), profitieren.

6.2 *Cognitive Debriefing* der übersetzten Instrumente

6.2.1 Einordnung der *Cognitive-Debriefing*-Ergebnisse

Cognitive Debriefing ist ein entscheidender Schritt im Übersetzungsprozess, um das Antwortverhalten der Zielpersonen eines übersetzten Fragebogens zu untersuchen (*response process validity*) (AERA et al., 1999; Willis, 2005; Mear und GirouDET, 2012). Leider wird auch dieser Schritt oft nicht bzw. methodisch insuffizient durchgeführt oder nicht ausreichend berichtet (Willis, 2005; Maer and GirouDET, 2012). Methoden und Ergebnisse der *Cognitive Debriefings* im Rahmen dieser Arbeit sind ebenfalls zur internationalen Publikation eingereicht (Hoben et al., submitted).

Mithilfe der qualitativen Interviews konnten verschiedene Verständnisschwierigkeiten der PHK aufgedeckt und verstanden sowie entsprechende Lösungen gefunden werden. Problematische Items, die die Validität gefährdeten, wurden identifiziert und angepasst. Wie bereits angesprochen, bestand die größte Herausforderung darin, eine Übersetzung für den Begriff „*best practice*“ zu finden (alle acht RU-Items sowie ACT-Items *Culture 3* und *Time 3*). Diese Formulierung wird von den englischsprachigen PHK häufig verwendet und besser verstanden als der in den anderen Fragebogenversionen verwendete Begriff „*research*“ bzw. „*research use*“ (Squires et al., 2008; Squires et al., 2009; Estabrooks et al., 2011d; Squires et al., 2011b). Die deutschen PHK verstanden jedoch weder den englischen Begriff (der mangels eines exakten Pendant häufig auch im Deutschen verwendet wird), noch waren ihnen die Begriffe „Forschung“, „Forschungswis-

sen“, „wissenschaftliche Erkenntnisse“, „Evidenz“ o.ä. im Kontext ihrer Tätigkeit geläufig. Ähnlich wie die kanadischen und schwedischen Probanden (Estabrooks et al., 2011d) benannten auch die deutschen PHK oft eher Barrieren für die Anwendung fachlich angemessener Pflege anstatt auf die Fachlichkeit und Qualität der geleisteten Pflege selbst einzugehen. Gleichwohl waren auch die deutschen PHK der Ansicht, ein gewisser Grad an Wissen sei erforderlich für ihre Tätigkeit. Die eigene Erfahrung oder das Fachwissen von Kollegen standen dabei im Vordergrund. Dies stimmt mit den internationalen Befunden überein, nach denen selbst PFK informelle, interaktive oder erfahrungsbasierte Wissensformen formalen Quellen (z.B. Fachzeitschriften oder Lehrbücher) vorziehen (Estabrooks et al., 2005a; Estabrooks et al., 2005b; Spenceley et al., 2008).

Dennoch wurde das Ziel, eine für PHK verständliche Formulierung zu finden, als bedeutsam angesehen und weiterverfolgt. Etwa 40% der Pflegenden in der stationären Altenpflege Deutschlands sind PHK (Statistisches Bundesamt, 2013). Sie unterstützen Bewohner bei der Nahrungsaufnahme, mobilisieren Bewohner, lagern sie, um Druckgeschwüre zu vermeiden, führen Mund- und Zahnpflege durch, interagieren mit Personen mit Demenz etc. Jede dieser Tätigkeiten kann Bewohner massiv gefährden und schädigen, wenn sie unangemessen ausgeführt wird. Es ist daher – so die hier vertretene Ansicht – bedeutsam, herauszufinden, a) wie PHK ihr eigenes Tun einschätzen, in wie weit sie also der Ansicht sind, entsprechend des aktuellen Wissensstands zu pflegen und was ihnen b) dabei hilft bzw. sie dabei behindert. Nur wenige Studien zu diesem Thema bezogen bisher PHK ein (Boström et al., 2006; Connor, 2006; Boström et al., 2007; Squires et al., 2011b) – vermutlich auch mangels geeigneter Assessmentinstrumente. Der rigorose Übersetzungsprozess (Hoben et al., 2013) und insbesondere die *Cognitive Debriefings* waren entscheidend, um reliable und valide Tools für die Befragung von PHK in der stationären Altenpflege Deutschlands zu kreieren. Dies gelang vor allem, indem auf den ersten Blick naheliegende Begriffe („Forschung“, „Wissenschaft“, „best practice“) vermieden und durch einfache Umschreibungen und Beispiele ersetzt wurden. „Fachwissen über optimale Pflege und Betreuung“ stellte sich in den PHK-Bögen als die geeignete Lösung heraus.

Bei den anderen Berufsgruppen waren diese Schwierigkeiten nicht in solch ausgeprägter Form zu erwarten. Für einige Items, bei denen auch in diesen Gruppen Probleme antizipiert wurden (z.B. die Frage nach Statistiken in den *Feedback*-Items des ACT), wurden die PHK-Formulierungen übernommen. Für andere Formulierungen, wie z.B. „research use“, wurden die spezifischen deutschen Begriffe gewählt (Anwendung wissenschaftlicher Erkenntnisse). Alle Items, die zuvor von den PHK verstanden wurden, erwiesen sich auch in den anderen Gruppen als verständlich, was den Aufwand für die PHK-Übersetzung rechtfertigt und als notwendig erweist. Auch die berufsgruppenspezifischen Formulierungen wurden in allen Gruppen verstanden, so dass sich keine Hinweise auf weitere Modifikationsbedarfe ergaben.

6.2.2 Limitationen des *Cognitive-Debriefing*-Teils

Auch für diesen Studienteil sind einige Limitationen zu erwähnen. *Cognitive Debriefings* laufen zum einen generell Gefahr, die Probleme eines Fragebogens zu unterschätzen (Willis, 2005): Erstens sind Personen, die freiwillig an *Cognitive Debriefings* teilnehmen, eher bereit, sich intensiv mit den Items auseinanderzusetzen und Zeit dafür aufzuwenden, als andere Fragebognutzer. Erstere sind zudem tendenziell besser gebildet und vertrauen eher darauf, in der Lage zu sein, die Fragebogeninhalte zu verstehen. Zweitens handelt es sich um eine Testsituation, in der die Teilnehmer bestrebt sind, eine gute Leistung zu erbringen, höflich zu sein und in der sie ggf. bereit sind, über umständliche Formulierungen hinwegzusehen und sich bemühen, diese dennoch zu verstehen. Eine weitere Limitation ist darin zu sehen, dass nicht alle Items mit allen Probanden getestet wurden. Dieser Umstand wurde in Kauf genommen, um die Durchführbarkeit des Projekts und die Kooperationsbereitschaft der Teilnehmer zu gewährleisten. Einige problematische Items wurden mit allen Teilnehmern getestet. Die anderen Items wurden den Probanden zufällig zugeteilt. Auf diese Weise wurde sichergestellt, dass a) alle Items in irgendeiner Form getestet wurden, b) problematische Items intensiver untersucht wurden und c) eine Verzerrung durch die willkürliche Auswahl von Items durch den Promovenden (Selektionsbias) vermieden wurde. Schließlich ist die relativ kleine Stichprobengröße, bedingt durch das qualitative Design, anzusprechen. Sie bewegt sich innerhalb genereller Empfehlungen für *Cognitive Debriefings* (Willis, 2005; Maer and Giroudet, 2012), erlaubt aber keine verallgemeinerbaren Aussagen. Die Einrichtungen und Teilnehmer wurden aufgrund ihrer Kooperationsbereitschaft und Merkmalskombinationen ausgewählt, um ein möglichst breites Spektrum potenzieller Fragebogenprobleme zu identifizieren. Statistische Repräsentativität war nicht das Ziel.

Trotz dieser Limitationen und obgleich einige Probleme möglicherweise unentdeckt blieben, konnten die Fragebögen so weit optimiert werden, dass die Erhebung in einer großen Stichprobe möglich war, ohne überzufällig viele fehlende Werte oder unlogische Wertekombinationen zu produzieren. Die *Cognitive Debriefings* liefern somit Evidenz für eine akzeptable *Response-Process-Validität* der übersetzten Instrumente.

6.3 Psychometrische Testung der übersetzten Instrumente

6.3.1 Faktorenstruktur und Reliabilität

Die zunächst angenommene Zehn-Faktoren-Struktur des deutschen ACT (Modell 1) konnte in den konfirmatorischen Faktorenanalysen nicht bestätigt werden. Dies entspricht den kanadischen Befunden auf Basis von PHK-Fragebögen (Estabrooks et al., 2011b). Die ACT-Items wurden bewusst so konzipiert, dass sie zehn verschiedene Konzepte reflektieren. Allerdings wurden alle Items eines Bereichs jeweils als nichtredundante Items konzipiert. D.h., die Items ähneln sich inhaltlich, sind aber bewusst so formuliert, dass sie sich nicht exakt gleichen. Damit stellt das ACT von vorne herein kein ideales Faktorenmodell dar, in dem die latenten Konstrukte perfekte

Prädiktoren für die zugehörigen Indikatoren (Items) sind. Anders ausgedrückt: Den Items, die einer Domäne (einem Konzept) zugeordnet sind, muss nicht zwingend eine einzige gemeinsame Ursache (das latente Konstrukt) zugrunde liegen (Estabrooks et al., 2011b). Die Fitindizes weisen dann auch auf eine schlechte Passung des Zehn-Faktoren-Modells hin.

Entsprechend des Vorgehens der kanadischen Forscherinnen erfolgte daraufhin die Aufteilung des ACT in ein Sieben-Faktoren-Modell (zustimmungsskalierte Items, Modell 2a) und ein Drei-Faktoren-Modell (häufigkeitsskalierte Items, Modell 3a). Nach Entfernen von vier Items und Korrelation der Residualvarianzen verschiedener, jeweils zum gleichen Konzept gehörender Items (z.T. in allen Modellen, z.T. nur in den Modellen spezifischer Berufsgruppen) erwies sich das Sieben-Faktoren-Modell (Modell 2b) als akzeptabel. Das Drei-Faktoren-Modell wies extrem schlechte Fitwerte auf. Erst die Aufteilung zweier Faktoren (IINT und STER) in zwei bzw. drei Faktoren, die Streichung zweier Items und die Korrelation einiger Residualvarianzen führten zu akzeptabel passenden Faktorenmodellen (Modell 3b). Auch Estabrooks et al. (2011b) erwähnen in ihrer Diskussion, ihr erstes Modell (zehn Faktoren) deute darauf hin, dass die STER-Items möglicherweise nicht nur einen Faktor reflektierten, und auch dort wies das Drei-Faktoren-Modell inakzeptable Fitwerte auf. Die Forscherinnen nahmen jedoch keine weiteren Modellmodifikationen vor. Die exploratorische Faktorenanalyse der Pädiatrieversion des ACT (Estabrooks et al., 2009b) ergab eine 13-Faktoren-Struktur, bei der die häufigkeitsskalierten ACT-Konzepte so aufgeteilt waren, wie in Modell 3b dieser Arbeit (ein Konzept formaler Zusammenkünfte, zwei Konzepte informellen Austauschs sowie drei Konzepte struktureller/elektronischer Ressourcen). Die Faktorenmodelle der CRU-Skala passten sehr gut zu den in dieser Arbeit erhobenen Daten. Die Ergebnisse entsprechen denen der kanadischen Validationsstudie (Squires et al., 2011b).

Aufgrund dieser Befunde sollten nicht zehn, sondern 13 Scores aus den beobachteten ACT-Items gebildet werden, und auch der Reliabilitätsbestimmung sollten die 13 Konzepte zugrunde liegen statt der angenommenen zehn. Dies wurde bei den Reliabilitätsberechnungen berücksichtigt, die zudem nicht wie in den kanadischen Arbeiten auf den beobachteten Werten basierten (Cronbach's α , *Split-Half*-Reliabilität, Spearman-Brown-Reliabilität) (Estabrooks et al., 2009b; Estabrooks et al., 2011b; Squires et al., 2011b), sondern auf den Modellparametern angemessen passender Faktorenmodelle (Green and Hershberger, 2000; Raykov and Shrout, 2002; Green and Yang, 2009b; Yang and Green, 2011). Die Reliabilitätswerte sind mehrheitlich akzeptabel ($>0,7$), wobei die häufigkeitsskalierten Konzepte tendenziell niedrigere Werte aufweisen. Auch dies entspricht den kanadischen Befunden (Estabrooks et al., 2009b; Estabrooks et al., 2011b; Squires et al., 2011b). Ein direkter Vergleich der Reliabilitätswerte dieser Arbeit mit den kanadischen Ergebnissen ist allerdings nur für die zustimmungsbasierten Konzepte der PHK (vgl. mit Estabrooks et al., 2011b) und für die Werte der PFK (vgl. mit Estabrooks et al., 2009b) möglich.

6.3.2 Faktorinvarianz

In dieser Arbeit wurde erstmalig die Faktorinvarianz des ACT stationäre Altenpflege und der CRU-Skala zwischen den verschiedenen Berufsgruppen bestimmt. Die beiden ACT-Modelle erwiesen sich als partiell stark messinvariant, die CRU-Skala sogar als partiell strikt messinvariant. In allen Modellen sind zudem die Faktorvarianzen partiell invariant. Dass das ACT weder vollständige noch partielle strikte Messinvarianz aufweist, ist nicht überraschend. Die verschiedenen Berufsgruppen unterscheiden sich deutlich bzgl. verschiedener Merkmale, die das Fragebogenverständnis beeinflussen (Alter, Sprachkompetenz, Qualifikation, Berufserfahrung und Bildung). Es war daher nicht zu erwarten, dass zwischen den Berufsgruppen gleiche oder ähnliche Residualvarianzen auftreten würden. Damit ist allerdings ein Berufsgruppenvergleich der beobachteten Itemwerte und daraus abgeleiteter Skalenscores des ACT nicht zu empfehlen (Byrne et al., 1989; Brown, 2006; Dimitrov, 2010; Sass, 2011; Wang and Wang, 2012). Alle weiteren Analysen basierten daher auf Modellen, die eine Messfehlertheorie beinhalten (Strukturgleichungsmodelle) und die damit den unterschiedlichen Residualvarianzen Rechnung tragen.

Während die Faktorvarianzen der ACT-Konzepte und CORU sich bis auf wenige Ausnahmen in den Berufsgruppen gleichen, gibt es deutliche Unterschiede der Faktormittelwerte. Bemerkenswert ist vor allem der Vergleich des CORU-Mittelwerts. Die höher qualifizierten Berufsgruppen (PFK, TBK, FK) weisen hier signifikant niedrigere Werte auf als die PHK, während der APS-Mittelwert signifikant höher ist als der der PHK. Befassen sich also PHK und APS intensiver mit aktuellen Erkenntnissen als die Fachkräfte bzw. verändert neues Wissen ihr Denken und ihre Einstellung stärker als bei Fachkräften? Der hohe Wert der APS leuchtet intuitiv ein, sind diese doch im Rahmen ihrer Ausbildung häufiger und intensiver mit neuem Wissen konfrontiert als die anderen Berufsgruppen. Weshalb aber übersteigt der PHK-Wert den der Fachkräfte? Eine naheliegende Erklärung betrifft die in den PHK- und APS-Bögen abweichenden Formulierungen („aktuelles Fachwissen zur optimalen Pflege und Betreuung“ statt „wissenschaftliche Erkenntnisse“). Wahrscheinlich reflektieren diese beiden Varianten nicht das gleiche Konzept. Fachwissen muss nicht zwingend wissenschaftlich fundiert sein und ist ein breiterer Begriff. Eine vergleichende Analyse des Verständnisses beider Formulierungen in allen Berufsgruppen (z.B. mittels Fokusgruppen oder semistrukturierter Einzelinterviews) sowie ggf. alternativer Formulierungen ist hier als künftiger Forschungs- bzw. Weiterentwicklungsbedarf festzuhalten.

6.3.3 Aggregierbarkeit der ACT-Daten

Obgleich mit dem ACT zunächst Einschätzungen von Individuen erhoben werden, wurde es gezielt dafür konzipiert, verschiedene Merkmale einer Einrichtung bzw. eines Wohnbereichs abzubilden. Dafür müssen die Einschätzungen der Individuen auf Einrichtungs- bzw. Wohnbereichsebene aggregiert werden. Die Einschätzungen von Individuen eines bestimmten Clusters (Einrichtung oder Wohnbereich), so die Annahme, variieren weniger (sind sich also ähnlicher) als die

Einschätzungen aller einbezogener Teilnehmer. Die Varianzen der aggregierten Antworten sollten sich demnach zwischen den Clustern signifikant unterscheiden. Dies wurde in den kanadischen Studien stets auf Basis der beobachteten Skalenscores untersucht, die auf Clusterebene aggregiert und mittels ANOVAs und sich daraus ergebender Parameter sowie Mehrebenenregressionsmodellen evaluiert wurden (Estabrooks et al., 2009b; Estabrooks et al., 2011a; Estabrooks et al., 2011b; Estabrooks et al., 2011c; Squires et al., 2013b). Die aus den ANOVA-Parametern abgeleiteten Indizes stützten in diesen Arbeiten mehrheitlich die Annahme der Aggregierbarkeit ($ICC1 > 0,1$, $ICC2 > 0,6$, $\omega^2 > 0,06$ und hohe Anteile der durch die Clusterzugehörigkeit erklärten Varianz η^2). Auch die Mehrebenenmodelle untermauerten diese These (Varianzen aggregierter ACT-Scores unterscheiden sich signifikant zwischen den Clustern, aggregierte ACT-Scores sind signifikante Prädiktoren für RU und Clusterebene erklärt substantielle Anteile der RU-Varianz). Dabei erwies sich die Wohnbereichsebene, verglichen mit der Einrichtungsebene, als erklärungskräftiger.

In der vorliegenden Arbeit wurde eine abweichende Methode gewählt, um die Aggregierbarkeit zu evaluieren. Zum einen waren die beobachteten Werte als nicht zuverlässig zwischen den Berufsgruppen vergleichbar (deutlich unterschiedliche Residualvarianzanteile in den einbezogenen Berufsgruppen und damit keine strikte Messinvarianz). Zum anderen bergen Mehrebenenansätze auf Basis manifester Werte die Gefahr gebieter Clusterlevelleffekte (Lüdtke et al., 2008). Daher wurden konfirmatorische Mehrebenenfaktorenmodelle auf Grundlage des *Multilevel-Latent-Covariate-Approach* (Muthén, 1989, 1994; Lüdtke et al., 2008; Asparouhov and Muthén, 2010) geschätzt. Im Gegensatz zu den kanadischen Befunden und den in dieser Arbeit versuchsweise durchgeführten permutationsbasierten ANOVAs stützen die Ergebnisse der latenten Mehrebenenmodelle die Aggregierbarkeitsthese nicht. Die meisten ACT-Konzepte auf Clustersebene variierten nicht signifikant zwischen den Clustern (nichtsignifikante *Between*-Varianzen) und die Passung der *Between*-Modellteile war schlecht ($SRMR_B$ zwischen 0,122 und 0,549). Am ehesten zeigten sich signifikante Clustervarianzen auf Wohnbereichsebene (Modelle 2b bei den PHK, FK und in der Gesamtstichprobe, Modelle 3b bei den PHK und in der Gesamtstichprobe), wobei die CFI- und TLI-Werte in den Gesamtstichprobenmodellen auf eine inakzeptable Modellpassung hinwiesen (zwischen 0,906 und 0,924). Die vorliegenden Daten lassen somit nicht erkennen, dass der Kontext der Einrichtungen bzw. Wohnbereiche die Einschätzungen der dort arbeitenden Individuen angleicht.

6.3.4 Zusammenhänge zwischen den Variablen

Die strukturelle Beziehung der RU-Variablen zueinander entspricht den Annahmen, die sich aus den kanadischen Arbeiten (Estabrooks, 1999a, 1999b) ergeben: Instrumentelle, konzeptuelle und persuasive Forschungsanwendung (INRU, CORU und PERU) sind signifikante Prädiktoren für Forschungsanwendung im Allgemeinen (OVRU) – sowohl im Gesamtmodell aller RU-Items

(Modell 5) als auch in allen anderen RU-Strukturgleichungsmodellen (Modelle 7a/b, 8a/b und 9a/b).

Unter den individuellen Merkmalen waren einzig die Einstellung zur Anwendung von Forschungswissen (ATRU) und verhaltensbezogener Stress (BSTR) signifikante Prädiktoren für OVRU (Modell 7a). Dies entspricht den internationalen Befunden, zusammengefasst von Estabrooks et al. (2003a) und Squires et al. (2011a) in ihren systematischen Reviews. Die besondere Relevanz der ATRU bestätigte sich auch in den anderen Modellen mit Individualmerkmalen als Prädiktoren (Modelle 8a und 9a), in denen sie sich als signifikanter Prädiktor für INRU, CORU und PERU erwies. Dies galt auch für kognitiven Stress (CSTR). Bemerkenswertweise waren in diesen Modellen verhaltensbezogener Stress (BSTR) und Burnout (BURN), entgegen der Hypothesen, negativ mit den drei RU-Konzepten assoziiert. Je niedriger also die BSTR- und BURN-Scores (je mehr Stress und Burnout die Teilnehmer angaben), desto höher die RU-Scores (desto häufiger gaben die Teilnehmer an, Forschungswissen zu nutzen). Dieser Befund überrascht und ist aus der internationalen Literatur nicht bekannt. Möglicherweise liegt verhaltensbezogenen Stresssymptomen (z.B. sich zurückziehen, keine Zeit haben, sich zu entspannen oder zu vergnügen oder keine Energie haben, sich mit anderen Menschen zu beschäftigen), Burnoutsymptomen (z.B. Müdigkeit, emotionale oder körperliche Erschöpfung) und der Anwendung von Forschungswissen eine gemeinsame Ursache zugrunde – die Motivation, die Arbeit besonders gut zu machen bzw. die Tendenz, sich dafür zu verausgaben. Hier besteht weiterer Forschungsbedarf. Qualitative Studien könnten hier Aufschluss darüber geben, wie und warum erhöhte Stress- und Burnoutsymptome mit vermehrter Forschungsnutzung zusammenhängen. Außerdem sollten in künftigen Studien zusätzliche Motivations-, Verhaltens- und Dispositionsvariablen erhoben und in Modelle integriert werden.

Zusammenhänge der ACT-Konzepte wurden bisher nur für INRU und CORU untersucht (Estabrooks et al., 2009b; Estabrooks et al., 2011b; Squires et al., 2013b). Beziehungen der ACT-Variablen zu OVRU wurden bisher noch nicht publiziert. In den hier geschätzten Modellen erwies sich – entgegen der Annahmen – einzig der Faktor Zeit (TIME) als signifikanter Prädiktor für OVRU. Während in den bivariaten Korrelationen und ANOVAs der kanadischen Studien (Estabrooks et al., 2009b; Estabrooks et al., 2011b) nahezu alle ACT-Konzepte signifikant und in erwarteter Richtung mit INRU assoziiert sind, trifft dies in den Modellen dieser Arbeit nicht zu. Feedback (FEED), Miteinander im Team (CAMP) und schriftbasierte Ressourcen (STER_a) sind mit keinem der drei RU-Konzepte assoziiert. Führung (LEAD), Personalbesetzung (STAF), Zeit (TIME), informeller Austausch mit eher bewohnerfern tätigen Kollegen (IINT_b), regulative Ressourcen (STER_b) sowie EDV-basierte Ressourcen (STER_c) hängen – im Widerspruch zu den Annahmen – negativ mit den drei RU-Konzepten zusammen. Auch Squires et al. (2013b) berichten in ihren Mehrebenenmodellen (GEE) von negativ mit RU (in diesem Fall CRU) assoziierten ACT-Scores (Kultur und formale Zusammenkünfte). Die Ursachen können lediglich vermutet

werden. Möglicherweise bewerten Personen, die sich viel mit Forschung beschäftigen und diese umzusetzen versuchen, den Organisationskontext kritischer als ihre Kollegen. Vielleicht stehen aber auch Faktoren wie Zeit und Austausch deshalb weniger zur Verfügung, weil die Zeit für die Beschäftigung mit und Anwendung von Forschungswissen aufgewendet wird. Eine Führungsperson könnte vielleicht auch deshalb kritischer bewertet werden, weil sie regelmäßig darauf besteht, dass die neuesten Erkenntnisse umgesetzt werden. Und schließlich ist denkbar, dass Personen sich gerade deshalb selbst mit Forschungswissen auseinandersetzen, weil keine entsprechenden Vorgaben und Ressourcen in der eigenen Einrichtung vorhanden sind. Auch was diese Fragen angeht, besteht weiterer Forschungsbedarf.

6.4 Limitationen der psychometrischen Testung

6.4.1 Stichprobe

Mit 38 teilnehmenden Pflegeeinrichtungen und insgesamt 821 einbezogenen Fragebögen wurde die angestrebte Stichprobengröße erreicht. Die Methode der Stichprobenziehung (stratifizierte Zufallsauswahl aus dem Pool aller Pflegeeinrichtungen der Metropolregion Rhein-Neckar) ist als stark zu bewerten. In Anbetracht der verfügbaren Ressourcen und des Hintergrunds dieser Arbeit (Dissertationsprojekt, das nicht in ein größeres Forschungsprojekt oder -programm eingebunden war), bewegt sich das erzielte Sample zudem im Bereich des maximal Machbaren. Dennoch sind einige stichprobenbedingte Limitationen zu konstatieren. Zum einen handelt es sich bei den teilnehmenden Pflegeeinrichtungen trotz der Zufallsziehung sehr wahrscheinlich um eine Positivselektion. 95 kontaktierte Einrichtungen lehnten eine Teilnahme ab – mehrheitlich aufgrund von Personalengpässen. So mussten immer wieder Einrichtungen nachgezogen und z.T. auf andere als die vorgesehenen Kategorien ausgewichen werden, weil alle gelisteten Einrichtungen bereits kontaktiert waren. Dies konterkarierte das Verfahren der Zufallsauswahl etwas. Es ist anzunehmen, dass hauptsächlich Einrichtungen zusagten, die a) motiviert waren, die anonymisiert zurückgemeldeten Studienergebnisse für die interne Qualitätsentwicklung zu nutzen, b) sich als gut organisiert einschätzten und c) sich aufgrund der verfügbaren Ressourcen in der Lage sahen, teilzunehmen. Zudem war die Stichprobe regional begrenzt. Zwar umfasst die Metropolregion Rhein-Neckar Teile dreier Bundesländer (Baden-Württemberg, Hessen und Rheinland-Pfalz), doch die Repräsentativität dieser Heime für das gesamte Bundesgebiet ist fraglich. Da diese Arbeit primär auf die Validierung der übersetzten Instrumente zielte und nicht auf die repräsentative Darstellung der Merkmalsausprägungen (RU und Organisationskontext), ist letztgenannte Limitation nicht zu stark zu gewichten.

Stärker ins Gewicht fällt jedoch die Stichprobengröße. Diese folgte nicht einer eigentlich als ideal angesehenen modellbasierten Powerschätzung (Satorra and Saris, 1985; Muthén and Muthén, 2002; Brown, 2006; Hancock, 2006; Kline, 2011; Lee et al., 2012; Wang and Wang, 2012). Mangels ausreichender Informationen aus Vorstudien wurde die Stichprobengröße auf Ba-

sis gängiger Faustregeln festgelegt. Anhand der Informationen dieser Arbeit sollten künftig präzisere, modellbasierte Stichprobenkalkulationen erfolgen. Die Substichproben der Mehrgruppenmodelle waren – insbesondere bei den APS (N=65) und den FK (N=129) – relativ klein, angesichts der großen und komplexen ACT-Faktorenmodelle. Zwar besaßen auch die Modelle mit kleinen Substichproben ausreichend Power um schlechte Modellpassung anzuzeigen, künftige Arbeiten können jedoch auf Basis größerer Substichproben sicher zuverlässigere Ergebnisse liefern. Mit durchschnittlichen Clustergrößen von 2,231 bis 6,921 Individuen pro Einrichtung bzw. 1,245 bis 2,798 Individuen pro Wohnbereich waren die berufsgruppenspezifischen Substichproben deutlich zu klein für die Mehrebenenmodelle. Vermutlich sind größere Stichproben in künftigen Studien eher in der Lage, Clustereffekte der ACT-Antworten abzubilden.

6.4.2 Datenerhebung

Die in dieser Arbeit gewählte Datenerhebungsmethode – von den Probanden selbst auszufüllende, papier- und stiftbasierte Fragebögen – reduzierte vermutlich die Validität der erhobenen Daten etwas. Auch in den ersten kanadischen Studien, in denen die hier verwendeten Instrumente eingesetzt wurden, füllten die Probanden die Fragebögen selbst aus – allerdings als Online-Version (Estabrooks et al., 2008c; Hutchinson et al., 2008; Estabrooks et al., 2009b). Insbesondere bei den PHK, aber auch in den anderen Berufsgruppen erwies es sich jedoch als hilfreich, die Befragung als standardisierte Interviews durchzuführen, da die Teilnehmer die Fragen oft nicht aufmerksam genug bzw. nicht vollständig lasen oder falsch verstanden (persönliche Kommunikation mit den Instrumentenurheberinnen sowie Squires et al., 2012). Das standardisierte Befragungsverfahren in Kombination mit einem rigorosen Verfahren zur Sicherung der Datenqualität erhöhte die Datenqualität in den kanadischen Studien substantiell (persönliche Kommunikation mit den Instrumentenurheberinnen sowie Squires et al., 2012). Obgleich sich die Items der deutschen Version in den *Cognitive Debriefings* als verhältnismäßig gut verständlich erwiesen und die Zahl fehlender Antworten in der Fragebogenerhebung relativ gering war, ist es doch wahrscheinlich, dass einige Probanden – ähnlich wie in den kanadischen Studien – mit bestimmten Items Probleme hatten oder die Fragen flüchtig bzw. unaufmerksam lasen. Ressourcenbedingt waren standardisierte Interviews mit über 800 Teilnehmern in dieser Arbeit nicht möglich. In künftigen Studien sollte dies aber in Erwägung gezogen werden. In wie weit sich die Erhebungsmethode auf die Validität auswirkt, könnte untersucht werden, indem beide Methoden angewendet und die Ergebnisse mit Multitrait-Multimethod-Modellen (MTMM) (Raykov and Marcoulides, 2011; Schermelleh-Engel und Schweizer, 2012) ausgewertet würden.

6.4.3 Modellfitevaluation

Der Stellenwert des χ^2 -Tests und relativer Modellfitindizes für die Evaluation von Strukturgleichungsmodellen ist ein umstrittenes Thema. Die radikalsten Forderungen lauten, relative Fitindi-

zes komplett zu verbannen (Barrett, 2007) bzw. primär dem χ^2 -Test Bedeutung beizumessen und dabei erst ab p-Werten $>0,75$ von akzeptabler Modellpassung auszugehen (Hayduk, 1996; Hayduk and Glaser, 2000; Hayduk et al., 2007; McIntosh, 2007). Nur der χ^2 -Test, so die Vertreter dieser Position, sei ein inferenzstatistischer Hypothesentest – nämlich der Nullhypothese, die Samplekovarianzmatrix entspreche exakt der modellimplizierten Kovarianzmatrix. Der p-Wert des χ^2 -Tests zeigt an, mit welcher Wahrscheinlichkeit die Samplekovarianzmatrix zu beobachten wäre, entspräche die modellimplizierte Kovarianzmatrix der wahren Populationsmatrix. Die relativen Fitindizes werden hingegen aus drei Gründen abgelehnt: a) testeten diese – im Gegensatz zum χ^2 -Test – nicht die Theorie, die das Modell reflektiere, sondern lediglich die ungefähre Passung des Modells (die auch bei fehlspezifizierten Modellen akzeptabel sein könne), b) weichten diese Indizes die strengen Kriterien des χ^2 -Tests auf, indem die χ^2 -Statistik nach Stichprobengröße, Variablenzahl, Freiheitsgraden o.ä. adjustiert werde und c) verführten die Indizes dazu, arbiträren Cut-Off-Werten zu folgen, die auf unzureichenden bzw. widersprüchlichen empirischen Grundlagen beruhten (Hayduk, 1996; Hayduk and Glaser, 2000; Barrett, 2007; Hayduk et al., 2007; McIntosh, 2007).

Folgte man dieser extremen Position, wären fast alle in dieser Arbeit geschätzten Modelle abzulehnen. Die Beiträge eines Sonderhefts der Zeitschrift *Personality and Individual Differences* (Heft 5, 2007) zum Thema Modellfitevaluation machen allerdings deutlich, dass diese Position von zahlreichen namhaften Methodikern (Bentler, 2007; Goffin, 2007; Markland, 2007; Miles und Shevlin, 2007; Millsap, 2007; Mulaik, 2007; Steiger, 2007) nicht geteilt wird. Sie arbeiten zahlreiche Probleme des χ^2 -Tests heraus: a) Die Cut-Off-Werte für den p-Wert des χ^2 -Tests sind ebenso arbiträr wie die der relativen Fitindizes, b) wie Simulationsstudien zeigten, setzt ein nicht-signifikanter χ^2 -Test nicht zwingend ein korrekt spezifiziertes Modell voraus (und umgekehrt), c) dürfte die Nullhypothese, Sample- und modellimplizierte Kovarianzmatrizen seien identisch (Kovarianzresiduen=0), so gut wie nie erfüllt sein, da Modelle immer nur eine komplexitätsreduzierte, ungefähre Annäherung an die Realität darstellen. Ist d) die Nullhypothese nicht erfüllt, testet der χ^2 -Test lediglich, ob das Modell genügend Power hat, diese abzulehnen. Da der χ^2 -Test abhängig ist von der Stichprobengröße, führen mit zunehmender Stichprobengröße selbst trivial kleine Kovarianzresiduen zu signifikanten χ^2 -Werten. Daher empfehlen die Autoren (Bentler, 2007; Goffin, 2007; Markland, 2007; Miles and Shevlin, 2007; Millsap, 2007; Mulaik, 2007; Steiger, 2007), sowohl χ^2 -Wert, Freiheitsgrade und p-Wert zu berichten als auch verschiedene etablierte relative Fitindizes. Die empfohlenen Grenzwerte sollten dabei als Interpretationshilfe und Richtwerte, nicht als definitive Cut-Off-Werte behandelt werden. Außerdem muss das Modell, neben der globalen Evaluation, zusätzlich hinsichtlich verschiedener lokaler Hinweise auf Missspezifikation und Missfit untersucht werden. Selbst bei einem signifikanten χ^2 -Test kann ein Modell dann als akzeptabel (als theoretische wie statistisch ausreichende Annäherung an die Daten) angesehen werden. Dieser letzten Position wurde in der vorliegenden Arbeit gefolgt.

6.4.4 Modellmodifikation

Die Post-Hoc-Modifikation von Strukturgleichungsmodellen, insbesondere wenn sie überwiegend auf Basis von Modifikationsindizes statt substantieller Theorie und Sachlogik erfolgt, ist problematisch (Brown, 2006; Raykov and Marcoulides, 2006; Hoyle, 2011; Kline, 2011; Chou and Huh, 2012; Wang and Wang, 2012): Modellmodifikationen auf Basis von Modifikationsindizes verbessern zwar den Modellfit, führen aber vielfach zu Fehlspezifikationen, die nicht dem wahren Modell entsprechen (MacCallum, 1986). Erfolgt die Modifikation primär theorie- und erst in zweiter Linie datengeleitet, verbessern sich die Chancen, sich dem wahren Modell anzunähern substantiell (Silvia and MacCallum, 1988). Dennoch bleibt das Problem bestehen, dass mit jeder Modifikation (bzw. mit jeder neuen Schätzung eines Modells) das Risiko für Typ-I-Fehler steigt (statistischer Test eines inakzeptablen Modells weist aufgrund zufallsbedingter Schwankungen auf akzeptable Passung hin = „*capitalization on chance*“) (MacCallum, 1986). Mit jeder Modifikation verschiebt sich das Vorgehen tendenziell von einem confirmatorischen zu einem eher exploratorischen Prozedere. Die Verteilung der Kovarianzresiduen (Differenz zwischen Samplekovarianzmatrix und modellimplizierter Kovarianzmatrix) entspricht dann nicht mehr einer χ^2 -Verteilung, Parameterschätzer und Modellfittests sind damit gebiast (MacCallum et al., 1992). Um sicherzustellen, dass das modifizierte Modell nicht nur Eigenschaften des Studiensamples repräsentiert, sondern tatsächlich auf die Population zutrifft, muss es in einem neuen, unabhängigen Sample erneut getestet werden (Brown, 2006; Raykov and Marcoulides, 2006; Hoyle, 2011; Kline, 2011; Chou and Huh, Chou und Huh, 2012; Wang and Wang, 2012). Dies war im Rahmen der vorliegenden Arbeit nicht möglich und muss zukünftigen Studien vorbehalten bleiben. Die Modifikationen der Modelle dieser Arbeit erfolgten überwiegend auf Basis sachlogischer Überlegungen. Substantielle theoretische oder empirische Befunde als Basis für Modellmodifikationen liegen kaum vor. Modifikationen, die zwar durch Modifikationsindizes angezeigt, jedoch nicht sachlogisch, theoretisch oder empirisch zu begründen waren, wurden nicht vorgenommen. Insofern wurden die Modelle so wenig wie möglich verändert. Die Ergebnisse dieser Arbeit sind dennoch mit Vorsicht und unter dem Vorbehalt der Gegenvalidierung in einer separaten Stichprobe zu interpretieren.

6.4.5 Partielle Faktorinvarianz

Wenn ein Modell in einem oder mehreren Bereichen keine vollständige Invarianz erzielt, ist dies aus zwei Gründen problematisch: Zum einen bedeutet dies, dass bestimmte Modellparameter (je nach Invarianzbereich) nicht perfekt zwischen den Gruppen vergleichbar sind (Byrne et al., 1989; Brown, 2006; Dimitrov, 2010; Sass, 2011; Wang and Wang, 2012). Diese Limitation kann akzeptiert werden, wenn die Zahl der freigesetzten (also nicht invarianten) Parameter „gering“ ist, wobei hier keine anerkannte Regel existiert, was als „gering“ zu bezeichnen ist und die Entscheidung und deren Begründung dem Forschenden obliegen (Vandenberg and Lance, 2000; Brown,

2006; Schmitt and Kuljanin, 2008; Dimitrov, 2010; Sass, 2011). In dieser Arbeit wurde die von Dimitrov (2010) vorgeschlagene Grenze von max. 20% freigesetzter Parameter (gemessen an der Gesamtzahl der restringierten, potenziell frei schätzbaren Parameter) gewählt, wobei der Anteil der freizusetzenden Parameter (abgesehen von der strikten Messinvarianz der ACT-Modelle) in den hier getesteten Modellen deutlich darunter lag. Die zweite Problematik betrifft die mit der Befreiung verschiedener restringierter Parameter verbundene multiple Schätzung von Modellen. Wie oben bereits angesprochen erhöht dies die Gefahr der „*capitalization on chance*“ (Byrne et al., 1989; Vandenberg and Lance, 2000; Brown, 2006; Schmitt and Kuljanin, 2008; Dimitrov, 2010; Sass, 2011; Wang and Wang, 2012). Auch hier sind die Modelle daher unter Vorbehalt zu interpretieren und in unabhängigen Stichproben erneut zu überprüfen.

6.4.6 Zusammenhänge zwischen den Variablen

Zwischen den ACT- bzw. COPSQ-/ARU- und den RU-Konzepten konnten zahlreiche signifikante Beziehungen festgestellt werden. Darin ist die Stärke dieses Studienteils zu sehen. Allerdings wird vor allem OVRU von weit weniger Faktoren beeinflusst als angenommen, und verschiedene Konzepte sind mit den drei anderen RU-Konzepten (INRU, CORU, PERU) in anderer Richtung als erwartet assoziiert. Insofern sind die Ergebnisse, was diese Quelle für Validitätsevidenz betrifft, ambivalent. Möglicherweise sind die Modelle korrekt und die Instrumente erfassen die Konstrukte valide (worauf die Faktorenmodelle hindeuten) – dann müssen die theoretischen Annahmen modifiziert werden. Hierfür ist weitere Forschung erforderlich, um die genauen Ursachen und Mechanismen für die überraschenden Befunde zu klären. Möglicherweise müssen aber auch die Instrumente weiter optimiert und die Modelle überdacht werden. Eine eindeutige Klärung dieser Frage ist im Rahmen dieser Arbeit nicht möglich.

Eine grundsätzliche Problematik betrifft das Konzept der RU. In Relation zu dem eingangszitierten umfassenden Evidenzverständnis (vgl. Einleitung) erscheint das in den RU-Tools reflektierte Evidenzverständnis reduktionistisch. Gefragt wird lediglich nach der Anwendung wissenschaftlicher Evidenz. Die anderen Wissensquellen bleiben außen vor. Außerdem fokussieren die Instrumente nicht die Situationsangemessenheit und Güte der Umsetzung sowie die Qualität des genutzten Wissens, sondern ausschließlich die Nutzungshäufigkeit. Peters et al. (2013) listen diverse Konstrukte auf, die als abhängige Variablen zur Evaluation von DI-Prozessen sehr viel besser geeignet erscheinen als die Häufigkeit der Nutzung wissenschaftlicher Evidenz: a) *acceptability* (Wahrnehmung, dass eine Innovation akzeptabel und sinnvoll ist), b) *adoption* (Anwendung der Neuerung), c) *appropriateness* (Wahrnehmung, dass die Innovation relevant und angemessen ist), d) *feasibility* (Machbarkeit/Umsetzbarkeit der Innovation), e) *fidelity* (Grad, in dem die Innovation so umgesetzt wird, wie vorgesehen), f) *implementation cost* (Kosten der Implementierung), g) *coverage* (Grad, in dem die potenziellen Rezipienten der Innovation diese empfangen) sowie h) *sustainability* (Verstetigungs-/Nachhaltigkeitsgrade der Innovation). Bedauerlicher-

weise sind diese Konzepte oft nicht trennscharf definiert und bisher auch nicht so präzise operationalisiert, dass sie in standardisierte Erhebungsinstrumente überführt werden konnten. Ein entscheidender Entwicklungsbedarf ist daher die Konstruktion und Validierung standardisierter Erhebungsinstrumente, um künftig alternative und besser geeignete DI-Outcomes erfassen zu können.

7 Zusammenfassung

Ausgangspunkt dieser Arbeit sind die z.T. gravierenden Diskrepanzen zwischen wissenschaftlichen Erkenntnissen und dem tatsächlichen Handeln beruflicher Akteure des Gesundheitswesens. Berufsgruppen-, setting- und länderübergreifend werden statt nachgewiesenermaßen effektiver Interventionen regelmäßig suboptimale oder gar schädigende Maßnahmen durchgeführt – mit schwerwiegenden Konsequenzen für die hilfe- und pflegebedürftigen Personen. Bewohner stationärer Altenpflegeeinrichtungen sind eine höchst vulnerable Personengruppe mit komplexen Pflegebedarfen und multiplen Risiken, und daher von diesen Defiziten besonders bedroht. Evidenzbasierte Neuerungen in Einrichtungen des Gesundheitswesens zu implementieren ist jedoch eine Herausforderung, die auf unzählige, komplex zusammenwirkende Barrieren trifft. Die Disseminations- und Implementierungsforschung (DIF) ist daher bestrebt, diese komplexen Prozesse zu erforschen, zu verstehen und effektiv zu beeinflussen. Besondere Aufmerksamkeit erfährt in diesem Kontext seit einiger Zeit der Zusammenhang zwischen Organisationskontextfaktoren und der Anwendung wissenschaftlicher Erkenntnisse. Gerade im stationären Altenpflegebereich sind diese Zusammenhänge bisher jedoch wenig erforscht. International erfährt diese Thematik allerdings mehr und mehr Aufmerksamkeit. Das *Alberta Context Tool* (ACT), *Estabrooks' Kinds of Research Use (RU) items* und die *Conceptual Research Use (CRU) scale* sind drei robuste, international bewährte Forschungsinstrumente, die zur Erforschung dieser Zusammenhänge entwickelt wurden. Auf Deutsch liegen keine Instrumente mit vergleichbarer Qualität und Eignung vor und entsprechende Forschungsaktivitäten sind hierzulande eher die Ausnahme.

Mit dieser Studie sollten die Voraussetzungen dafür geschaffen werden, diese Zusammenhänge mittels statistischer Methoden in stationären Altenpflegeeinrichtungen Deutschlands zu untersuchen. Ein besseres Verständnis dieser Prozesse, so die These dieser Arbeit, trägt dazu bei, diese künftig effektiver zu beeinflussen und damit die Versorgungsqualität und schlussendlich die Lebensqualität hilfe- und pflegebedürftiger Menschen zu verbessern. Daher wurden die drei Instrumente auf Basis internationaler *Best-Practice*-Leitlinien (Wild et al., 2005; Acquadro et al., 2008) in einem mehrstufigen, iterativen Prozess a) auf Deutsch übersetzt und b) an den Kontext der stationären Altenpflege Deutschlands angepasst. Die übersetzten Instrumente wurden zudem c) auf Basis der *Standards for Educational and Psychological Testing* (AERA et al., 1999) umfassend validiert.

Die Übersetzung und Adaptation der kanadischen Instrumente erwies sich als komplexer,

anforderungsreicher und ressourcenaufwändiger Prozess – insbesondere in der Gruppe der Pflegehilfskräfte (PHK). Folgende Herausforderungen waren zu meistern: a) Unterschiedliche Qualifikationsstruktur in der kanadischen und deutschen Altenpflege, b) einfache englische Sätze, die in der deutschen Übersetzung lang und kompliziert wurden, c) potenziell missverständliche Formulierungen, d) englische Idiome und Phrasen, e) fehlende deutsche Pendanten zu englischen Begriffen, f) für Zielpersonen unverständliche deutsche Begriffe sowie g) in Items angesprochene Aktivitäten/Aspekte, mit denen die deutschen Probanden nicht vertraut waren. Die in den Übersetzungsprozess involvierten Experten (insbesondere die Mitglieder der Expertenpanels und die Instrumentenentwicklerinnen) waren eine wertvolle Ressource beim Bewältigen dieser Herausforderungen. Sie bescheinigten den deutschen Instrumentenversionen schließlich eine akzeptable inhaltliche Validität.

Die *Cognitive Debriefings* mit den PHK zeigten umfangreiche Modifikationsbedarfe auf. Insbesondere erwies es sich als schwierig, geeignete Übersetzungen für die Begriffe „*best practice*“ bzw. „*research use*“ zu finden. Drei Runden waren erforderlich, bis alle Items verstanden wurden. Der hohe Aufwand bei Übersetzung und *Cognitive Debriefings* der PHK-Versionen zahlte sich jedoch aus: Der Zeitbedarf für die anderen Berufsgruppen reduzierte sich substantiell und alle Items wurden von diesen wie beabsichtigt verstanden.

In den statistischen Analysen wurden verschiedene Evidenzquellen für die Validität der deutschen Instrumentenversionen untersucht. In Einklang mit den kanadischen Befunden, weist das ACT keine Zehn-Faktoren-Struktur aller Items auf. Vielmehr beinhaltet es zwei verschiedene Arten von Konzepten – zustimmungsskalierte und häufigkeitsskalierte. Erstere haben, wie in der kanadischen Version, eine Sieben-Faktoren-Struktur. Letztere besitzen in der deutschen Version eine Sechs-Faktoren-Struktur, während die kanadischen Analysen von drei Faktoren ausgehen. (Allerdings sind die Modellfitwerte der kanadischen Drei-Faktoren-Lösung inakzeptabel.) Die akzeptable Modellpassung der beiden ACT-Modelle wie auch die der Ein-Faktor-Lösung der CRU-Skala verweisen auf Evidenz bzgl. der internen Struktur der Instrumente. Die Reliabilitätswerte auf Basis der Faktormodellparameter bewegten sich ebenfalls überwiegend auf akzeptablem Niveau. In dieser Arbeit wurde international erstmalig die Faktorinvarianz der übersetzten Instrumente untersucht. Mit partieller starker Messinvarianz (ACT) und partieller strikter Messinvarianz (CRU-Skala) sind damit Vergleiche der Faktorvarianzen, Faktormittelwerte und Strukturparameter zwischen den Berufsgruppen möglich. Mangels (partieller) strikter Messinvarianz sind allerdings die beobachteten Itemwerte bzw. -scores des ACT nicht vergleichbar. Im Gegensatz zu den kanadischen Befunden auf Basis der manifesten Itemscores ergaben die konfirmatorischen Mehrebenenfaktorenmodelle dieser Arbeit keine Aggregierbarkeit der Daten auf Einrichtungs-/Wohnbereichsebene. Die Evidenz auf Basis der Strukturgleichungsmodelle (Zusammenhänge zwischen den Variablen) ist ambivalent. Die Struktur der RU-Konzepte entspricht den Hypothesen und auch einige der COPSOQ-/ARU sowie ACT-Konzepte sind, wie erwartet, signifi-

kante Prädiktoren der RU-Konzepte. Allerdings waren auch zahlreiche COPSOQ-/ARU sowie ACT-Konzepte nicht wie erwartet mit RU assoziiert (insbesondere *Overall RU* war hier betroffen) bzw. wiesen – unerwarteterweise – Zusammenhänge in negativer Richtung mit den anderen drei RU-Konzepten auf.

Zusammenfassend ist also festzuhalten: Mit den deutschen Versionen des ACT, der *Estabrooks-Items* und der CRU-Skala liegen drei verhältnismäßig robuste Forschungsinstrumente vor, um den Zusammenhang zwischen Organisationskontext und Anwendung wissenschaftlicher Erkenntnisse in der stationären Altenpflege Deutschlands zu erforschen. Diese Arbeit zeigt starke Evidenz auf für die Validität der deutschen Instrumentenversionen. Allerdings sind einige Befunde ambivalent bzw. widersprechen den Hypothesen. Daher gilt es, die Ergebnisse dieser Arbeit in künftigen Studien zu überprüfen.

Literaturverzeichnis

- Aarons GA, Glisson C, Green PD, Hoagwood K, Kelleher KJ, Landsverk JA (2012) The organizational social context of mental health services and clinician attitudes toward evidence-based practice: A United States national study. *Implement Sci* 7:56.
- Acquadro C, Conway K, Hareendran A, Aaronson N (2008) Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Health* 11:509-521.
- AERA, APA, NCME – American Educational Research Association, American Psychological Association, National Council on Measurement in Education: Standards for Educational and Psychological Testing. 5th ed. AERA, Washington, DC, 1999.
- Aiken LH, Clarke SP, Sloane DM, Sochalski JA, Busse R, Clarke H, Giovannetti P, Hunt J, Rafferty AM, Shamian J (2001) Nurses' reports on hospital care in five countries. *Health Aff (Millwood)* 20:43-53.
- Anderson MJ (2001a) A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26:32-46.
- Anderson MJ: PERMANOVA: a FORTRAN computer program for permutational multivariate analysis of variance. Department of Statistics, University of Auckland, New Zealand, 2001b.
- Asparouhov T, Muthén B: Simple Second Order Chi-Square Correction: Technical Appendix Related to New Features in Mplus Version 6. Muthén & Muthén, Los Angeles, CA, 2010.
- Balzer K, Butz S, Bentzel J, Boulkhemair D, Lüthmann D: Beschreibung und Bewertung der fachärztlichen Versorgung von Pflegeheimbewohnern in Deutschland. 2013.
- Barrett P (2007) Structural equation modelling: Adjudging model fit. *Pers Individ Dif* 42:815-824.
- Bartholomeyczik S (2008) Institut für Forschung und Transfer in der Pflege und Behandlung von Menschen mit Demenz: Konzept. *Pflege & Gesellschaft* 13:337-349.
- Becker C, Behncke A, Börner M, Hoben M, Kuß O, Selinger Y, Targan K, Zimmermann M, Behrens J: Implementierung des Resident Assessment Instruments (RAI) als Qualitätsentwicklungs- und Steuerungsinstrument in der stationären Langzeitpflege (DRKS-ID: DRKS00000418). Fachlicher Abschlussbericht ans Bundesministerium für Bildung und Forschung (BMBF). Institut für Gesundheits- und Pflegewissenschaften, Medizinische Fakultät, Martin-Luther-Universität Halle-Wittenberg, Halle/Saale, 2012.
- Behrens J, Langer G: Evidence-based Nursing and Caring: Methoden und Ethik der Pflegepraxis und Versorgungsforschung. 3., überarb. u. erg. Aufl. Huber, Bern u.a., 2010.
- Bentler PM (2007) On tests and indices for evaluating structural models. *Pers Individ Dif* 42:825-829.
- Bero L, Grilli R, Grimshaw JM, Mowatt G, Oxman A, Zwarenstein M (2001) Cochrane effective practice and organization of care review group. The Cochrane Library.
- Berta W, Teare GF, Gilbert E, Ginsburg LS, Lemieux-Charles L, Davis D, Rappolt S (2010) Spanning the know-do gap: Understanding knowledge application and capacity in long-term care homes. *Soc Sci Med* 70:1326-1334.
- Borritz M, Kristensen TS: Copenhagen Burnout Inventory. National Institute of Occupational Health, Copenhagen, 1999.
- Bortz J, Schuster C: Statistik für Human- und Sozialwissenschaftler. 7., vollst. überarb. u. erw. Aufl. Springer, Berlin-Heidelberg, 2010.
- Boström A-M, Slaughter SE, Chojecki D, Estabrooks CA (2012) What do we know about knowledge translation in the care of older adults? A scoping review. *J Am Med Dir Assoc* 13:210-219.
- Boström AM, Wallin L, Nordström G (2006) Research use in the care of older people: A survey among healthcare staff. *Int J Older People Nurs* 1:131-140.
- Boström AM, Wallin L, Nordström G (2007) Evidence-based practice and determinants of research use in elderly care in Sweden. *J Eval Clin Pract* 13:665-673.
- Bravo M, Woodbury-Farina M, Canino GJ, Rubio-Stipec M (1993) The Spanish translation and cultural adaptation of the Diagnostic Interview Schedule for Children (DISC) in Puerto Rico. *Cult Med Psychiatry* 17:329-344.
- Bredthauer D, Becker C, Eichner B, Koczy P, Nikolaus T (2005) Factors relating to the use of physical restraints in psychogeriatric care: A paradigm for elder abuse. *Z Gerontol Geriatr* 38:10-18.
- Breimaier HE, Halfens RJG, Lohrmann C (2011) Nurses' wishes, knowledge, attitudes and perceived barriers on implementing research findings into practice among graduate nurses in Austria. *J Clin Nurs* 20:1744-1756.
- Brown TA: Confirmatory Factor Analysis for Applied Research. Guildford, New York, 2006.
- Byrne BM, Shavelson RJ, Muthén B (1989) Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychol Bull* 105:456-466.

- Capezuti E, Taylor J, Brown H, Strothers HS, 3rd, Ouslander JG (2007) Challenges to implementing an APN-facilitated falls management program in long-term care. *Appl Nurs Res* 20:2-9.
- Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S (2007) A conceptual framework for implementation fidelity. *Implement Sci* 2:40.
- Champion VL, Leach A (1989) Variables related to research utilization in nursing: An empirical investigation. *J Adv Nurs* 14:705-710.
- Chaudoir SR, Dugan AG, Barr CH (2013) Measuring factors affecting implementation of health innovations: A systematic review of structural, organizational, provider, patient, and innovation level measures. *Implement Sci* 8:22.
- Chou C-P, Huh J: Model Fit and Model Selection in Structural Equation Modeling. In: Hoyle RH (ed): *Handbook of Structural Equation Modeling*. The Guildford Press, New York, 2012, pp. 232-246.
- Cobban SJ, Profetto-McGrath J (2008) A pilot study of research utilization practices and critical thinking dispositions of Alberta dental hygienists. *Int J Dent Hygiene* 6:229-237.
- Colón-Emeric C, Schenck A, Gorospe J, McArdle J, Dobson L, DePorter C, McConnell E (2006) Translating evidence-based falls prevention into clinical practice in nursing facilities: Results and lessons from a quality improvement collaborative. *J Am Geriatr Soc* 54:1414-1418.
- Connor N: The relationship between organizational culture and research utilization practices among nursing home departmental staff. PhD Thesis. Dalhousie University, 2006.
- CRAN: The Comprehensive R Archive Network. <http://cran.r-project.org/> [Letzter Zugriff: 14.06.2013].
- Cranley LA, Norton PG, Cummings GG, Barnard D, Estabrooks CA (2011) SCOPE: Safer care for older persons (in residential) environments: A study protocol. *Implement Sci* 6:71.
- Cummings GG, Hutchinson AM, Scott SD, Norton PG, Estabrooks CA (2010) The relationship between characteristics of context and research utilization in a pediatric setting. *BMC Health Serv Res* 10:168.
- DGP – Deutsche Gesellschaft für Pflegewissenschaft e.V.: Sektion Dissemination und Implementierung – SDI. <http://www.dg-pflegewissenschaft.de/2011DGP/sektionen/pflege-und-gesellschaft/disseminations-und-implementierungswissenschaft-di-w> [Letzter Zugriff: 16.11.2013].
- Dijkstra R, Wensing M, Thomas R, Akkermans R, Braspenning J, Grimshaw J, Grol R (2006) The relationship between organisational characteristics and the effects of clinical guidelines on medical performance in hospitals, a meta-analysis. *BMC Health Serv Res* 6:53.
- Dimitrov DM (2010) Testing for Factorial Invariance in the Context of Construct Validation. *Measurement and Evaluation in Counseling and Development* 43:121-149.
- Dopson S, Fitzgerald L (eds): *Knowledge to Action? Evidence-Based Health Care in Context*. Oxford University Press, New York, 2005.
- Eccles MP, Armstrong D, Baker R, Cleary K, Davies H, Davies S, Glasziou P, Ilott I, Kinmonth A-L, Leng G, Logan S, Marteau T, Michie S, Rogers H, Rycroft-Malone J, Sibbald B (2009) An implementation research agenda. *Implement Sci* 4:18.
- Edwards H, Chapman H, Forster E, Gaskill D, Morrison P, Sanders F (2003) Challenges associated with implementing an education program in a residential aged care setting. *Aust Health Rev* 26:107-115.
- EGQoLHM – European Group for Quality of Life and Health Measurement: *European Guide to the Nottingham Health Profile*. Brookwood Medical Publications, Brookwood, Surrey, 1993.
- Eldh AC, Ehrenberg A, Squires JE, Estabrooks CA, Wallin L (2013) Translating and testing the Alberta context tool for use among nurses in Swedish elder care. *BMC Health Serv Res* 13:68.
- EpiData: EpiData Software. <http://www.epidata.dk/index.htm> [Letzter Zugriff: 04.12.2013].
- Estabrooks CA: *Research Utilization in Nursing: An Examination of formal Structure and Influencing Factors*. PhD Thesis. University of Alberta, Faculty of Nursing, Edmonton, AB, 1997.
- Estabrooks CA (1998) Will evidence-based nursing practice make practice perfect? *Can J Nurs Res* 30:15-36.
- Estabrooks CA (1999a) The conceptual structure of research utilization. *Res Nurs Health* 22:203-216.
- Estabrooks CA (1999b) Modeling the individual determinants of research utilization. *West J Nurs Res* 21:758-772.
- Estabrooks CA, Chong H, Brigidear K, Profetto-McGrath J (2005a) Profiling Canadian nurses' preferred knowledge sources for clinical practice. *Can J Nurs Res* 37:118-140.
- Estabrooks CA, Derksen L, Winther C, Lavis JN, Scott SD, Wallin L, Profetto-McGrath J (2008a) The intellectual structure and substance of the knowledge utilization field: A longitudinal author co-citation analysis, 1945 to 2004. *Implement Sci* 3:49.
- Estabrooks CA, Floyd JA, Scott-Findlay S, O'Leary KA, Gushta M (2003a) Individual determinants of research utilization: A systematic review. *J Adv Nurs* 43:506-520.
- Estabrooks CA, Hutchinson AM, Squires JE, Birdsell J, Cummings GG, Degner L, Morgan D, Norton PG

- (2009a) Translating research in elder care: An introduction to a study protocol series. *Implement Sci* 4:51.
- Estabrooks CA, Kenny DJ, Adewale AJ, Cummings GG, Mallidou AA (2007) A comparison of research utilization among nurses working in Canadian civilian and United States Army healthcare settings. *Res Nurs Health* 30:282-296.
- Estabrooks CA, Morgan DG, Squires JE, Bostrom AM, Slaughter SE, Cummings GG, Norton PG (2011a) The care unit in nursing home research: Evidence in support of a definition. *BMC Med Res Methodol* 11:46.
- Estabrooks CA, Rutakumwa W, O'Leary KA, Profetto-McGrath J, Milner M, Levers MJ, Scott-Findlay S (2005b) Sources of practice knowledge among nurses. *Qual Health Res* 15:460-476.
- Estabrooks CA, Scott SD, Squires JE, Stevens B, O'Brien-Pallas L, Watt-Watson J, Profetto-McGrath J, McGilton K, Golden-Biddle K, Lander J, Donner G, Boschma G, Humphrey CK, Williams J (2008b) Patterns of research utilization on patient care units. *Implement Sci* 3:31.
- Estabrooks CA, Squires JE, Adachi AM, Kong L, Norton PG: Utilization of Health Research in Acute Care Settings in Alberta (ACUTE): ACUTE Technical Report. University of Alberta, Faculty of Nursing, Edmonton, 2008c.
- Estabrooks CA, Squires JE, Cummings GG, Birdsell JM, Norton PG (2009b) Development and assessment of the Alberta Context Tool. *BMC Health Serv Res* 9:234.
- Estabrooks CA, Squires JE, Cummings GG, Teare GF, Norton PG (2009c) Study protocol for the translating research in elder care (TREC): Building context - An organizational monitoring program in long-term care project (project one). *Implement Sci* 4:52.
- Estabrooks CA, Squires JE, Hayduk LA, Cummings GG, Norton PG (2011b) Advancing the argument for validity of the Alberta Context Tool with healthcare aides in residential long-term care. *BMC Med Res Methodol* 11:107.
- Estabrooks CA, Squires JE, Hutchinson AM, Scott S, Cummings GG, Kang SH, Midodzi WK, Stevens B (2011c) Assessment of variation in the Alberta Context Tool: The contribution of unit level contextual factors and specialty in Canadian pediatric acute care settings. *BMC Health Serv Res* 11:251.
- Estabrooks CA, Squires JE, Strandberg E, Nilsson-Kajermo K, Scott SD, Profetto-McGrath J, Harley D, Wallin L (2011d) Towards better measures of research utilization: A collaborative study in Canada and Sweden. *J Adv Nurs* 67:1705-1718.
- Estabrooks CA, Wallin L, Milner M (2003b) Measuring knowledge utilization in health care. *Int J Policy Eval Manag* 1:3-36.
- EuroQoL Group (1990) EuroQoL - an new facility for the measurement of health-related quality of life. *Health Policy* 16:199-208.
- Fixsen DL, Naoom SF, Blase KA, Friedman RM, Wallace F: Implementation Research: A Synthesis of the Literature. University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network, Tampa, 2005.
- Fleuren M, Wiefferink K, Paulussen T (2004) Determinants of innovation within health care organizations: Literature review and Delphi study. *Int J Qual Health Care* 16:107-123.
- Flodgren G, Rojas-Reyes MX, Cole N, Foxcroft DR (2012) Effectiveness of organisational infrastructures to promote evidence-based nursing practice. *Cochrane Database Syst Rev Issue 2*.
- Foxcroft D, Cole N (2000) Organisational infrastructures to promote evidence based nursing practice. *Cochrane Database Syst Rev Issue 3*.
- Foxcroft D, Cole N (2003) Organisational infrastructures to promote evidence based nursing practice. *Cochrane Database Syst Rev Issue 4*.
- Funk SG, Champagne MT, Wiese RA, Tornquist EM (1991) BARRIERS: The Barriers to Research Utilization Scale. *Appl Nurs Res* 4:39-45.
- Gagnon M-P, Labarthe J, Legare F, Ouimet M, Estabrooks C, Roch G, Ghandour EK, Grimshaw J (2011) Measuring organizational readiness for knowledge translation in chronic care. *Implement Sci* 6:72.
- Gitlin LN, Jacobs M, Earland TV (2010) Translation of a dementia caregiver intervention for delivery in homecare as a reimbursable medicare service: Outcomes and lessons learned. *Gerontologist* 50:847-854.
- Glaser J, Lampert B, Weigl M: Arbeit in der stationären Altenpflege: Analyse und Förderung von Arbeitsbedingungen, Interaktion, Gesundheit und Qualität. Wirtschaftsverlag NW, Bremerhaven, 2008.
- Glisson C (2002) The organizational context of children's mental health services. *Clin Child Fam Psychol Rev* 5:233-253.
- Goffin RD (2007) Assessing the adequacy of structural equation models: Golden rules and editorial policies. *Pers Individ Dif* 42:831-839.

- Graham ID, Bick D, Tetroe J, Straus SE, Harrison MB: Measuring Outcomes of evidence-based Practice: Distinguishing between Knowledge Use and its Impact. In: Bick D, Graham ID (eds): *Evaluating the Impact of implementing evidence-based Practice*. Wiley-Blackwell, Chichester, 2010.
- Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, Robinson N (2006) Lost in knowledge translation: Time for a map? *J Contin Educ Health Prof* 26:13-24.
- Graham JW: *Missing data: analysis and design*. Springer, New York, 2012.
- Green LW, Ottoson JM, García C, Hiatt RA (2009) Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annu Rev Public Health* 30:151-174.
- Green SB, Hershberger SL (2000) Correlated errors in true score models and their effect on Coefficient Alpha. *Struct Equ Modeling* 7:251-270.
- Green SB, Yang Y (2009a) Commentary on Coefficient Alpha: a cautionary tale. *Psychometrika* 74:121-135.
- Green SB, Yang Y (2009b) Reliability of summed item scores using structural equation modeling: An alternative to Coefficient Alpha. *Psychometrika* 74:155-167.
- Greenhalgh T, Glenn R, Bate P, Macfarlane F, Kyriakidou O: *Diffusion of Innovations in Health Service Organisations: A systematic Literature Review*. Blackwell, Massachusetts u.a., 2005.
- Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O (2004) Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Q* 82:581-629.
- Greif S, Runde B, Seeberg I: *Erfolge und Misserfolge beim Change Management*. Hogrefe, Göttingen u.a., 2004.
- Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE (2012) Knowledge translation of research findings. *Implement Sci* 7:50.
- Hancock GR: In: Hancock GR, Mueller RO (eds): *Structural Equation Modeling: A second Course*. Information Age Publishing, Greenwich, CT, 2006, pp. 69-115.
- Hayduk LA: *LISREL: Issues, Debates, and Strategies*. The Johns Hopkins University Press, Baltimore-London, 1996.
- Hayduk LA, Cummings G, Boadu K, Pazderka-Robinson H, Boulianne S (2007) Testing! testing! one, two, three – Testing the theory in structural equation models! *Pers Individ Dif* 42:841-850.
- Hayduk LA, Glaser DN (2000) Jiving the four-step, waltzing around factor analysis, and other serious fun. *Struct Equ Modeling* 7:1-35.
- Hayduk LA, Littvay L (2012) Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Med Res Methodol* 12:159.
- Hirdes JP, Mitchell L, Maxwell CJ, White N (2011) Beyond the 'iron lungs of gerontology': Using evidence to shape the future of nursing homes in Canada. *Can J Aging* 30:371-390.
- Hoben M: *Implementierungsprozesse in der stationären Langzeitpflege am Beispiel des Resident Assessment Instrument (RAI): Interpretativ-rekonstruktive Analyse der Perspektive von Führungspersonen*. Masterarbeit, Martin-Luther-Universität Halle-Wittenberg, Medizinische Fakultät, Institut für Gesundheits- und Pflegewissenschaft, Halle/Saale, 2010.
- Hoben M, Bär M, Mahler C, Berger S, Squires JE, Estabrooks CA, Kruse A, Behrens J (submitted) Linguistic validation of the Alberta Context Tool and two measures of research use, for German residential long term care. *BMC Res Notes*.
- Hoben M, Mahler C, Bär M, Berger S, Squires JE, Estabrooks CA, Behrens J (2013) German translation of the Alberta Context Tool and two measures of research use: methods, challenges and lessons learned. *BMC Health Serv Res* 13:478.
- Howlett B, Rogo EJ, Shelton TG: *Evidence-based Practice for Health Professionals: An interprofessional Approach*. Jones & Bartlett Learning, Burlington, MA, 2013.
- Hoyle RH: *Structural Equation Modeling for Social and Personality Psychology*. Sage, Los Angeles u.a., 2011.
- Hu L-t, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 6:1-55.
- Hunt SM, Alonso J, Bucquet D, Niero M, Wiklund I, McKenna S (1991) Cross-cultural adaptation of health measures: European Group for Health Management and Quality of Life Assessment. *Health Policy* 19:33-44.
- Hutchinson AM, Kong L, Adachi AM, Estabrooks CA, Stevens B: *Context and Research Use in the Care of Children: A Pilot Study – Project 2 CIHR Team in Children's Pain, Technical Report*. University of Alberta, Faculty of Nursing, Edmonton, 2008.
- IBM: *IBM SPSS Statistics, Release 20.0.0*. Chicago, 2011.
- Ishii S, Weintraub N, Mervis JR (2009) Apathy: A common psychiatric syndrome in the elderly. *J Am Med*

- Dir Assoc 10:381-393.
- Jordan R, Sirsch E, Gesch D, Zimmer S, Bartholomeyczik S (2012) Verbesserung der zahnmedizinischen Betreuung in der Altenpflege durch Schulungen von Pflegekräften. *Pflege* 25:97-105.
- Kajermo KN, Boström AM, Thompson DS, Hutchinson AM, Estabrooks CA, Wallin L (2010) The BARRIERS scale – The barriers to research utilization scale: A systematic review. *Implement Sci* 5:32.
- Kaplan HC, Brady PW, Dritz MC, Hooper DK, Linam WM, Froehle CM, Margolis P (2010) The influence of context on quality improvement success in health care: A systematic review of the literature. *Milbank Q* 88:500-559.
- Kenny DJ (2005) Nurses' use of research in practice at three US Army hospitals. *Nurs Leadersh* 18:45-67.
- Kitson A, Harvey G, McCormack B (1998) Enabling the implementation of evidence based practice: A conceptual framework. *Qual Health Care* 7:149-158.
- Kitson AL (2009) The need for systems change: Reflections on knowledge translation and organizational change. *J Adv Nurs* 65:217-228.
- Kline RB: *Principles and Practice of Structural Equation Modeling*. 3rd ed. The Guilford Press, New York, London, 2011.
- Koller M, Kantzer V, Mear I, Zarzar K, Martin M, Greimel E, Bottomley A, Arnott M, Kulis D, Isoqol T-S (2012) The process of reconciliation: Evaluation of guidelines for translating quality-of-life questionnaires. *Expert Rev Pharmacoecon Outcomes Res* 12:189-197.
- Kolmogorov AN (1933) Sulla determinazione empirica di una legge di distribuzione (On the empirical determination of a distribution law). *Giorn Ist Ital Attuar* 4:83-91.
- Köpke S, Koch F, Behncke A, Balzer K (2013) Einstellungen Pflegender in deutschen Krankenhäusern zu einer evidenzbasierten Pflegepraxis. *Pflege* 26:163-175.
- Kruse A, Berendonk C, Stanek S, Schönit M, Kaspar R, Bär M, Ehret S, Motruk M, Böggemann M: *DEmenzkranken Menschen in Individuell bedeutsamen Alltagssituationen – Handlungskompetenzen in der Betreuung demenzkranker Menschen fördern*. Schlussbericht Projekt B3 des Pflegeforschungsverbands Nordrhein-Westfalen. Ruprecht-Karls-Universität Heidelberg, Institut für Gerontologie, Heidelberg, 2010.
- Kuske B, Luck T, Hanns S, Matschinger H, Angermeyer MC, Behrens J, Riedel-Heller SG (2009) Training in dementia care: A cluster-randomized controlled trial of a training program for nursing home staff in Germany. *Int Psychogeriatr* 21:295-308.
- Kwiatkowski B (2011) Belastungssituation bei Betreuungskräften gemäß § 87b SGB XI in der stationären Altenpflege: Eine Pilotstudie mit Verlaufsmessung über sechs Monate. *Pflege Z* 64:286-290.
- Lacey EA (1994) Research utilization in nursing practice: A pilot study. *J Adv Nurs* 19:987-995.
- Lahmann NA, Dassen T, Poehler A, Kottner J (2010) Pressure ulcer prevalence rates from 2002 to 2008 in German long-term care facilities. *Aging Clin Exp Res* 22:152-156.
- Lee T, Cai L, MacCallum RC: *Power Analysis for Tests of Structural Equation Models*. In: Hoyle RH (ed): *Handbook of Structural Equation Modeling*. The Guildford Press, New York, 2012, pp. 181-194.
- Little RJA, Rubin DB: *Statistical Analysis with Missing Data*. 2nd ed. Wiley, New York, NY, 2002.
- Lüdtke O, Marsh HW, Robitzsch A, Trautwein U, Asparouhov T, Muthen B (2008) The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychol Methods* 13:203-229.
- MacCallum R (1986) Specification searches in covariance structure modeling. *Psychol Bull* 100:107-120.
- MacCallum RC, Roznowski M, Necowitz LB (1992) Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychol Bull* 111:490-504.
- Mahler C, Jank S, Reuschenbach B, Szecsenyi J (2009) „Komm, lass uns doch schnell mal den Fragebogen übersetzen“: Richtlinien zur Übersetzung und Implementierung englischsprachiger Assessment-Instrumente. *Pflegewissenschaft* 11:5-12.
- Majic T, Pluta JP, Mell T, Aichberger MC, Treusch Y, Gutzmann H, Heinz A, Rapp MA (2010) The pharmacotherapy of neuropsychiatric symptoms of dementia: A cross-sectional study in 18 homes for the elderly in Berlin. *Dtsch Ärztebl Int* 107:320-327.
- Mallidou AA, Cummings GG, Ginsburg LR, Chuang YT, Kang S, Norton PG, Estabrooks CA (2011) Staff, space, and time as dimensions of organizational slack: A psychometric assessment. *Health Care Manage Rev* 36:252-264.
- Maneesriwongul W, Dixon JK (2004) Instrument translation process: a methods review. *J Adv Nurs* 48:175-186.
- Markland D (2007) The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Pers Individ Dif* 42:851-858.
- Masso M, McCarthy G (2009) Literature review to identify factors that support implementation of evi-

- dence-based practice in residential aged care. *Int J Evid Based Healthc* 7:145-156.
- Mayring P: *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. 11. aktual. u. überarb. Aufl. Beltz, Weinheim–Basel, 2010.
- McCormack B, Kitson A, Harvey G, Rycroft-Malone J, Titchen A, Seers K (2002) Getting evidence into practice: The meaning of 'context'. *J Adv Nurs* 38:94-104.
- McIntosh CN (2007) Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007). *Pers Individ Dif* 42:859-867.
- McKenna SP, Doward LC (2005) The translation and cultural adaptation of patient-reported outcome measures. *Value Health* 8:89-91.
- McKibbon KA, Lokker C, Wilczynski NL, Ciliska D, Dobbins M, Davis DA, Haynes RB, Straus SE (2010) A cross-sectional study of the number and frequency of terms used to refer to knowledge translation in a body of health literature in 2006: A Tower of Babel? *Implement Sci* 5:16.
- Mear I, GirouDET C: *Linguistic Validation Procedures*. In: Acquadro C, Conway K, GirouDET C, Mear I (Hrsg): *Linguistic Validation Manual for Health Outcome Assessments*. 2. ed. MAPI Institute, Lyon, 2012, S. 33-117.
- Meijers JMM, Janssen MAP, Cummings GG, Wallin L, Estabrooks CA, Halfens RYG (2006) Assessing the relationships between contextual factors and research utilization in nursing: Systematic literature review. *J Adv Nurs* 55:622-635.
- Menon B, Cherkil S, Aswathy S, Unnikrishnan AG, Rajani G (2012) The process and challenges in the translation of World Health Organization Quality of Life (WHOQOL- BREF) to a regional language; Malayalam. *Indian J Psychol Med* 34:149-152.
- Metropolregion Rhein-Neckar: Rhein-Neckar in Zahlen. <http://www.m-r-n.com/start/forschen-und-studieren/rhein-neckar-in-zahlen.html> [Letzter Zugriff: 03.12.2013].
- Meyer G, Köpke S (2012) Wie kann der beste pflegewissenschaftliche Kenntnisstand in die Pflegepraxis gelangen? *Pflege & Gesellschaft* 17:36-44.
- Meyer G, Köpke S, Haastert B, Mülhauser I (2009) Restraint use among nursing home residents: Cross-sectional study and prospective cohort study. *J Clin Nurs* 18:981-990.
- Miles J, Shevlin M: *Applying Regression & Correlation: A Guide for Students and Researchers*. Sage, London u.a., 2001.
- Miles J, Shevlin M (2007) A time and a place for incremental fit indices. *Pers Individ Dif* 42:869-874.
- Millsap RE (2007) Structural equation modeling made difficult. *Pers Individ Dif* 42:875-881.
- Milner FM, Estabrooks CA, Humphrey C (2005) Clinical nurse educators as agents for change: increasing research utilization. *Int J Nurs Stud* 42:899-914.
- Moosbrugger H, Höfling V: *Standards für psychologisches Testen*. In: Moosbrugger H, Kelava A (Hrsg): *Testtheorie und Fragebogenkonstruktion*. 2., aktual. u. überarb. Aufl. Springer Medizin, Heidelberg, 2012, S. 203-224.
- Mowatt G, Thomson MA, Grimshaw J, Grant A (1998) Implementing early warning messages on emerging health technologies. *Int J Technol Assess Health Care* 14:663-670.
- Mulaik S (2007) There is a place for approximate fit in structural equation modelling. *Pers Individ Dif* 42:883-891.
- Müller-Hergl C: „...und dann klappte die Umsetzung nicht...“: Implementierung neuer Pflegepraxis bei Demenz. Universität Witten-Herdecke, Institut für Pflegewissenschaft, Dialogzentrum Demenz, Witten-Herdecke, 2006. <http://www.uni-wh.de/fileadmin/media/g/pflege/dzd/Downloads/Arbeitspapiere/Implementierung.pdf> [Letzter Zugriff: 15.11.2013].
- Muthén BO (1989) Latent variable modeling in heterogeneous populations. *Psychometrika* 54:557-585.
- Muthén BO (1994) Multilevel covariance structure analysis. *Sociol Methods Res* 22:376-398.
- Muthén BO, Muthén LK: *Mplus Version 7.11 – Base Program and Combination Add-On*. Los Angeles, CA, 2013.
- Muthén LK, Muthén BO (2002) How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equ Modeling* 9:599-620.
- Muthén LK, Muthén BO: *Mplus User's Guide: Statistical Analysis With Latent Variables*. 7th. ed. Muthén & Muthén, Los Angeles, CA, 2012.
- NIH – National Institutes of Health, Office of Behavioral and Social Sciences Research: *Dissemination and Implementation*. http://obssr.od.nih.gov/scientific_areas/translation/dissemination_and_implementation/index.aspx [Letzter Zugriff: 16.11.2013].
- Nübling M, Stöbel U, Hasselhorn H-M, Michaelis M, Hofmann F: *Methoden zur Erfassung psychischer*

- Belastungen: Erprobung eines Messinstrumentes (COPSOQ). Wirtschaftsverlag, Bremerhaven, 2005.
- Nübling M, Stössel U, Hasselhorn HM, Michaelis M, Hofmann F (2006) Measuring psychological stress and strain at work - Evaluation of the COPSOQ Questionnaire in Germany. *Psychosoc Med* 3:Doc05.
- Nübling M, Vomstein M, Schmidt SG, Gregersen S, Dulon M, Nienhaus A (2010) Psychosocial work load and stress in the geriatric care. *BMC Public Health* 10:428.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H: Package 'vegan': Community Ecology Package. Version 2.0-7, March 19, 2013. 2013.
- Parahoo K (1998) Research utilization and research related activities of nurses in Northern Ireland. *Int J Nurs Stud* 35:283-291.
- Parent-Vachon M, Parnell LK, Rachelska G, Lasalle L, Nedelec B (2008) Cross-cultural adaptation and validation of the Questionnaire for Pruritus Assessment for use in the French Canadian burn survivor population. *Burns* 34:71-92.
- Peters DH, Adam T, Alonge O, Agyepong IA, Tran N (2013) Implementation research: what it is and how to do it. *BMJ* 347:f6753.
- Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M (2011) Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health* 37:65-76.
- Proctor EK, Brownson RC: Measurement Issues in Dissemination and Implementation Research. In: Brownson RC, Colditz GA, Proctor EK (eds): *Dissemination and Implementation Research in Health: Translating Science to Practice*. Oxford University Press, Oxford, 2012, pp. 261-280.
- Profetto-McGrath J, Hesketh KL, Lang S, Estabrooks CA (2003) A study of critical thinking and research utilization among nurses. *West J Nurs Res* 25:322-337.
- Profetto-McGrath J, Smith KB, Hugo K, Patel A, Dussault B (2009) Nurse educators' critical thinking dispositions and research utilization. *Nurse Educ Pract* 9:199-208.
- Quasdorf T, Hoben M, Riesner C, Dichter MN, Halek M (2013) Einflussfaktoren in Disseminations- und Implementierungsprozessen. *Pflege & Gesellschaft* 18:235-252.
- Rabin BA, Brownson RC: Developing the terminology for dissemination and implementation research. In: Brownson RC, Colditz GA, Proctor EK (eds): *Dissemination and implementation research in health: Translating science to practice*. Oxford University Press, Oxford, 2012, pp. 23-51.
- Raykov T, Marcoulides GA: *A First Course in Structural Equation Modeling*. 2nd ed. Psychology Press, New York, NY u.a., 2006.
- Raykov T, Marcoulides GA: *Introduction to Psychometric Theory*. Routledge, New York-Hove, 2011.
- Raykov T, Shrout PE (2002) Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Struct Equ Modeling* 9:195-212.
- Reuther S, van Nie N, Meijers J, Halfens R, Bartholomeyczik S (2013) Mangelernährung und Demenz bei Bewohnern in Einrichtungen der stationären Altenpflege in Deutschland: Ergebnisse von Prävalenzerhebungen aus den Jahren 2008 und 2009. *Z Gerontol Geriatr* 46:260-267.
- Roes M, De Jong A, Wulff I (2013) Implementierungs- und Disseminationsforschung: Ein notwendiger Diskurs. *Pflege & Gesellschaft* 18:197-213.
- Rycroft-Malone J, Kitson A, Harvey G, McCormack B, Seers K, Titchen A, Estabrooks C (2002) Ingredients for change: Revisiting a conceptual framework. *Qual Saf Health Care* 11:174-180.
- Sarges W, Wottawa H, Roos C (Hrsg): *Handbuch wirtschaftspsychologischer Testverfahren, Band II: Organisationspsychologische Instrumente*. Pabst, Lengerich u.a., 2010.
- Sass DA (2011) Testing Measurement Invariance and Comparing Latent Factor Means Within a Confirmatory Factor Analysis Framework. *J Psychoeduc Assess* 29:347-363.
- Satorra A, Saris W (1985) Power of the likelihood ratio test in covariance structure analysis. *Psychometrika* 50:83-90.
- Saxer S (2002) Transfer von Forschungsergebnissen in die Praxis: Hemmende und fördernde Faktoren. *Printernet* 4:17-23.
- Schaeffer D, Wingenfeld K, Kleina T, Brause M, Horn A: *Qualität und Gesundheit in der stationären Altenhilfe: Eine empirische Bestandsaufnahme – ZQP-Abschlussbericht*. Universität Bielefeld, Zentrum für Qualität in der Pflege, Bielefeld, 2013.
- Scheridera CFG: *Datenqualität mit SPSS*. Oldenbourg, München, 2007.
- Schermelleh-Engel K, Schweizer K: *Multitrait-Multimethod-Analysen*. In: Moosbrugger H, Kelava A (Hrsg): *Testtheorie und Fragebogenkonstruktion*. 2., aktual. u. überarb. Aufl. Springer Medizin, Heidelberg, 2012, S. 345-362.
- Schmitt N, Kuljanin G (2008) Measurement invariance: Review of practice and implications. *Hum Resour*

Manage R 18:210-222.

- Schnittger T, Hilgefort M, Hauken T (2012) Status quo des Evidence-based Nursing: „Interesse, Wissen und Bereitschaft für den Wissenschaftstransfer der Pflegefachkräfte in drei deutschen Kliniken“ – Eine Querschnittsstudie im deskriptiven Design. *Pflegewissenschaft* 14:140-150.
- Schubert B, Wrobel M (2009) Identifizierung von Hindernissen, die die Implementierung von Forschungswissen in die Pflegepraxis hemmen. *Pflegewissenschaft* 12:343-351.
- Schubert M, Glass TR, Clarke SP, Aiken LH, Schaffert-Witvliet B, Sloane DM, De Geest S (2008) Rationing of nursing care and its relationship to patient outcomes: the Swiss extension of the International Hospital Outcomes Study. *Int J Qual Health Care* 20:227-237.
- Schuman H (1966) The random probe: A technique for evaluating validity of closed questions. *Am Sociol Rev* 31:218-222.
- Seers K, Cox K, Crichton NJ, Edwards RT, Eldh AC, Estabrooks CA, Harvey G, Hawkes C, Kitson A, Linck P, McCarthy G, McCormack B, Mockford C, Rycroft-Malone J, Titchen A, Wallin L (2012) FIRE (Facilitating Implementation of Research Evidence): A study protocol. *Implement Sci* 7:25.
- Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52:591-611.
- Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 420-428.
- Sijtsma K (2009) On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika* 74:107-120.
- Silvia ESM, MacCallum RC (1988) Some factors affecting the success of specification searches in covariance structure modeling. *Multivariate Behav Res* 23:297-326.
- Simon M, Tackenberg P, Hasselhorn H-M, Kümmerling A, Büscher A, Müller BH: Auswertung der ersten befragung der NEXT-Studie in Deutschland. Universität Wuppertal, 2005.
- Smirnov N (1948) Table for estimating the goodness of fit of empirical distributions. *Ann Math Stat* 19:279-281.
- Sommer S, Marckmann G, Pentzek M, Wegscheider K, Abholz HH, in der Schmitt J (2012) Advance directives in nursing homes: prevalence, validity, significance, and nursing staff adherence. *Dtsch Arztebl Int* 109:577-583.
- Spenceley SM, O'Leary KA, Chizawsky LL, Ross AJ, Estabrooks CA (2008) Sources of information used by nurses to inform practice: An integrative review. *Int J Nurs Stud* 45:954-970.
- Squires A, Aiken LH, van den Heede K, Sermeus W, Bruyneel L, Lindqvist R, Schoonhoven L, Stromseng I, Busse R, Brzostek T, Ensio A, Moreno-Casbas M, Rafferty AM, Schubert M, Zikos D, Matthews A (2013a) A systematic survey instrument translation process for multi-country, comparative health workforce studies. *Int J Nurs Stud* 50:264-273.
- Squires JE, Estabrooks CA, Gustavsson P, Wallin L (2011a) Individual determinants of research utilization by nurses: A systematic review update. *Implement Sci* 6:1.
- Squires JE, Adachi AM, Estabrooks CA: Developing a Valid and Reliable Measure of Research Utilization: Measurement Technical Report. University of Alberta, Faculty of Nursing, Edmonton, 2008.
- Squires JE, Estabrooks CA, Newburn-Cook CV, Gierl M (2011b) Validation of the conceptual research utilization scale: An application of the standards for educational and psychological testing in healthcare. *BMC Health Serv Res* 11:107.
- Squires JE, Estabrooks CA, O'Rourke HM, Gustavsson P, Newburn-Cook CV, Wallin L (2011c) A systematic review of the psychometric properties of self-report research utilization measures used in healthcare. *Implement Sci* 6:83.
- Squires JE, Estabrooks CA, Scott SD, Cummings GG, Hayduk L, Kang SH, Stevens B (2013b) The influence of organizational context on the use of research by nurses in Canadian pediatric hospitals. *BMC Health Serv Res* 13:351.
- Squires JE, Hayduk L, Hutchinson AM, Cranley LA, Gierl M, Cummings GG, Norton PG, Estabrooks CA (2013c) A protocol for advanced psychometric assessment of surveys. *Nurs Res Pract* 2013:Article ID 156782.
- Squires JE, Hutchinson AM, Bostrom AM, Deis K, Norton PG, Cummings GG, Estabrooks CA (2012) A data quality control program for computer-assisted personal interviews. *Nurs Res Pract* 2012:Article ID 303816.
- Squires JE, Kong L, Brooker S, Mitchell A, Sales AE, Estabrooks CA: Examining the Role of Context in Alzheimer Care Centers: A Pilot Study, Technical Report. University of Alberta, Faculty of Nursing, Edmonton, 2009.
- Statistisches Bundesamt (Hrsg): *Pflegestatistik 2011: Pflege im Rahmen der Pflegeversicherung – Deutschlandergebnisse*. Erschienen am 18. Januar 2013. Statistisches Bundesamt, Wiesbaden, 2013.

- Steiger JH (2007) Understanding the limitations of global fit assessment in structural equation modeling. *Pers Individ Dif* 42:893-898.
- Sterns S, Miller SC, Allen S (2010) The complexity of implementing culture change practices in nursing homes. *J Am Med Dir Assoc* 11:511-518.
- Stewart AL, Nápoles-Springer A (2000) Health-related quality-of-life assessments in diverse population groups in the United States. *Med Care* 38:II102-II124.
- Straus SE, Richardson WS, Glasziou P, Haynes RB: *Evidence-based Medicine: How to practice and teach it*. 4. ed., reprinted. Churchill Livingstone–Elsevier, Edinburgh u.a., 2011.
- Streiner DL, Norman GR: *Health Measurement Scales: A practical Guide to their Development and Use*. 4th ed. Oxford University Press, Oxford, 2008.
- Treusch Y, Jerosch D, Majic T, Heinz A, Gutzmann H, Rapp MA (2010) Wie können demenzkranke Pflegeheimbewohner mit Apathie besser versorgt werden? *Psychiatr Prax* 37:84-88.
- Vandenberg RJ, Lance CE (2000) A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organ Res Methods* 3:4-70.
- Volkert D, Pauly L, Stehle P, Sieber CC (2011) Prevalence of malnutrition in orally and tube-fed elderly nursing home residents in Germany and Its relation to health complaints and dietary intake. *Gastroenterol Res Pract* 2011:Article ID 247315.
- Walker Schlaefli Y: *Einflussfaktoren bei der Anwendung von pflegewissenschaftlichen Erkenntnissen: Eine Untersuchung bei HochschulabsolventInnen in Pflegeberufen in der deutschsprachigen Schweiz*. Unpublizierte Master Thesis zur Erlangung des akademischen Titels Master in Nursing Science, September 2005. Fachbereich Pflegewissenschaft, Fakultät der Gesundheitswissenschaften, Universität Maastricht; Weiterbildungszentrum für Gesundheitsberufe (WE'G) Aarau, Maastricht, Aarau, 2005.
- Wang J, Wang X: *Structural Equation Modeling: Applications Using Mplus*. Wiley, Chichester, 2012.
- West SG, Taylor AB, Wu W: *Model Fit and Model Selection in Structural Equation Modeling*. In: Hoyle RH (ed): *Handbook of Structural Equation Modeling*. The Guildford Press, New York, 2012, pp. 209-231.
- Wilborn D, Dassen T (2010) Pressure ulcer prevention in German healthcare facilities: Adherence to National Expert Standard? *J Nurs Care Qual* 25:151-159.
- Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, Hareendran A, Wiklund I, Chong LY, von Maltzahn R, Cohen L, Molsen E (2009) Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: The ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. *Value Health* 12:430-440.
- Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P (2005) Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 8:94-104.
- Willis GB: *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage, Thousand Oaks, 2005.
- Yang Y, Green SB (2011) Coefficient Alpha: a reliability coefficient for the 21st century? *J Psychoeduc Assess* 29:377-392.
- Zúñiga F, Schwappach D, De Geest S, Schwendimann R (2013) Psychometric properties of the Swiss version of the Nursing Home Survey on Patient Safety Culture. *Safety Sci* 55:88-118.

Thesen

(1) Organisationskontextfaktoren scheint in Disseminations- und Implementierungsprozessen eine besondere Bedeutung zuzukommen. Eine signifikante Forschungslücke sind methodisch suffiziente Überprüfungen dieser Hypothesen mittels statistischer Methoden: Lange Zeit konzentrierte sich die Disseminations- und Implementierungsforschung (DIF) primär auf die Frage, wie und welche Eigenschaften individueller Akteure DI-Prozesse beeinflussen. Diese individuellen Merkmale haben jedoch nur begrenzte Erklärungskraft. Einzig die Einstellung gegenüber der Anwendung wissenschaftlicher Erkenntnisse erwies sich durchgehend als zuverlässiger Prädiktor. Organisationskontextfaktoren erfahren daher aus drei Gründen zunehmend Aufmerksamkeit: a) Der Kontext einer Organisation beeinflusst die Situation ihrer Akteure wie auch ihre Einstellung und ihr Handeln. b) DI-Prozesse sind in erster Linie soziale Prozesse. Daher beeinflusst der Organisationskontext nicht nur vereinzelte Individuen, sondern ganze Teams. c) Zudem sind viele dieser Organisationskontextfaktoren potenziell modifizierbar – im Gegensatz zu manch anderen in der DIF populären Einflussfaktoren (z.B. bestimmte Eigenschaften von Individuen oder der Innovation). Insbesondere quantitative Studien zum Zusammenhang von Organisationskontext und Anwendung von Forschungswissen leiden unter methodischen Limitationen. Unzureichende theoretisch fundierte und definierte Konzepte sowie Erhebungsinstrumente unzureichender Güte sind die Hauptursachen dafür. Narrative Reviews und qualitative Studien betonen die Bedeutung des Organisationskontexts, methodisch suffiziente Überprüfungen dieser Hypothesen mittels statistischer Methoden sind bislang jedoch rar. Insgesamt fokussieren die Studien vor allem den Krankenhausbereich. Im stationären Altenpflegebereich ist die Studienlage sehr dünn.

(2) Voraussetzung für solche Studien sind standardisierte Erhebungsinstrumente hoher Güte. Mit dem *Alberta Context Tool (ACT)*, *Estabrooks' Kinds of Research Use (RU) items* und der *Conceptual Research Use (CRU) scale* liegen drei international breit genutzte Instrumente vor, für die weder national noch international Alternativen mit vergleichbarer Eignung und Güte verfügbar sind: Die drei Instrumente wurden a) auf Basis robuster DI-Theorie und -Evidenz sowie anerkannter Instrumentenentwicklungsmethoden konstruiert, erfassen b) spezifisch potenziell modifizierbare Organisationskontextfaktoren des stationären Altenpflegebereichs bzw. verschiedene Formen der Anwendung wissenschaftlicher Erkenntnisse in diesen Settings, liegen c) für verschiedene Berufsgruppen in stationären Altenpflegeeinrichtungen vor, konnten sich d) in internationalen Studien bewähren und weisen e) eine akzeptable psychometrische Güte auf. Die auf Deutsch vorliegenden Alternativen sind zu unspezifisch, liegen nicht in Versionen für die stationäre Altenpflege vor und/oder sind unzureichend validiert. Auch keines der verfügbaren englischsprachigen Alternativinstrumente weist vergleichbare Eigenschaften auf.

(3) Validitätsevidenz bezogen auf die Instrumenteninhalte – deutsche Instrumentenversionen sind akzeptabel valide: Die Übersetzung eines Erhebungsinstruments in eine andere

Sprache bewegt sich in einem Spannungsfeld: Einerseits soll die übersetzte Version strukturell und inhaltlich so nahe wie möglich am Original bleiben (um Vergleichbarkeit zu gewährleisten), andererseits soll es im Zielkontext verständlich, relevant, reliabel und valide sein, was ggf. Anpassungen erfordert. Durch den anforderungsreichen und ressourcenaufwändigen Übersetzungsprozess gelang es, die vielfältigen Herausforderungen zu meistern und sowohl eine angemessene kulturelle Adaptation als auch inhaltliche Validität sicherzustellen. Dies bescheinigten insbesondere die Expertenpanelmitglieder wie auch Instrumentenentwicklerinnen.

(4) Validitätsevidenz bezogen auf die Antwortprozesse der Zielpersonen – deutsche Instrumentenversionen sind akzeptabel valide: Die *Cognitive Debriefings* mit den Pflegehilfskräften (PHK) deckten eine Vielzahl von Problemen in den Fragebögen auf. Problematische Items wurden modifiziert, erneut getestet usw., bis nach drei Runden alle Items von den PHK verstanden wurden. Die *Cognitive Debriefings* mit den anderen Berufsgruppen – Pflegefachkräfte (PFK), Therapie-/Betreuungskräfte (TBK), Experten (Exp.) und Führungskräfte (FK) – ergaben keine weiteren Modifikationsbedarfe. Alle Items wurden wie intendiert verstanden.

(5) Validitätsevidenz bezogen auf die interne Struktur der Instrumente – deutsche Instrumentenversionen sind akzeptabel valide, deutsche Befunde weichen z.T. von den kanadischen ab bzw. gehen darüber hinaus: Die vorliegende Studie untersuchte erstmalig konfirmatorisch die Faktorenstruktur der Fragebögen für die stationäre Altenpflege bei PHK, PFK, TBK, FK und Altenpflegeschülern (APS). Wie in der kanadischen Studie (nur PHK) zeigte sich, dass das ACT keine einheitliche Zehn-Faktoren-Struktur aufweist. Vielmehr handelt es sich bei den zustimmungs- und den häufigkeitsskalierten Konzepten wohl um unterschiedliche Konstrukte. Erstere weisen, wie in der kanadischen Arbeit, eine Sieben-Faktoren-Struktur mit akzeptablem Fit auf. Eine Drei-Faktoren-Struktur führte, wie auch in der kanadischen Studie, zu inakzeptablen Fitwerten. Im Gegensatz zur kanadischen Arbeit wurden die Konstrukte in dieser Studie daher weiter aufgeteilt. Die Sechs-Faktoren-Struktur weist akzeptable Fitwerte auf. Die häufigkeitsskalierten ACT-Konzepte sind dabei aufgeteilt, wie in der Pädiatrieversion des ACT für PFK (Hauptkomponentenanalysen). Die Ein-Faktor-Struktur der CRU-Skala weist in der vorliegenden, wie auch in der kanadischen Studie eine akzeptable Modellpassung auf.

(6) Die Reliabilität der deutschen Instrumentenversionen ist akzeptabel, deutsche Befunde gehen über die kanadischen hinaus und beruhen auf robusteren Methoden: Während die kanadischen Reliabilitätsbefunde auf beobachteten Werten basierten, wurden in dieser Arbeit die Parameter angemessen passender Faktorenmodelle zugrunde gelegt – eine robustere, weniger biasanfällige Methode. Die Reliabilität der ACT-Konzepte ist mehrheitlich akzeptabel ($>0,7$), die der CRU-Skala sehr gut. Aufgrund der unterschiedlichen Faktorenstruktur und der je unterschiedlichen einbezogenen Berufsgruppen, ist allerdings ein direkter Vergleich der Reliabilitätswerte dieser Arbeit mit den kanadischen Ergebnissen nur eingeschränkt möglich.

(7) Das deutsche ACT ist partiell stark, die deutsche CRU-Skala partiell strikt messinvariant zwischen den Berufsgruppen: International erstmalig wurde die Faktorinvarianz des ACT und der CRU-Skala zwischen den Berufsgruppen bestimmt. Da das ACT keine (partielle) strikte Messinvarianz erzielte, sind die beobachteten Itemwerte und daraus abgeleitete Skalenscores nicht zwischen den Berufsgruppen vergleichbar – wohl aber Varianzen, Mittelwerte und Strukturparameter der latenten Konstrukte. Modelle, die von „unabhängigen“ Residuen ausgehen (z.B. die in den kanadischen Studien eingesetzten Regressionsmodelle oder generellen lineare Modelle), waren in diesem Fall für Gruppenvergleiche nicht geeignet. Erforderlich waren Modelle, die invariante Residualvarianzen berücksichtigen (z.B. Strukturgleichungsmodelle).

(8) Die Individualantworten des deutschen ACT sind – im Gegensatz zu den kanadischen Befunden – nicht auf Einrichtungs- bzw. Wohnbereichsebene aggregierbar, deutsche Befunde beruhen auf robusteren Methoden als die kanadischen: Da die beobachteten Werte nicht zuverlässig zwischen den Berufsgruppen vergleichbar waren und Mehrebenenansätze auf Basis manifester Werte die Gefahr gebieter Clusterlevelleffekte bergen, wurde hier ein anderer Ansatz gewählt, als in den kanadischen Arbeiten: konfirmatorische Mehrebenenfaktorenmodelle. Im Gegensatz zu den kanadischen Befunden und den in dieser Arbeit versuchsweise durchgeführten permutationsbasierten ANOVAs (auf Basis der beobachteten Werte), stützen die Ergebnisse der latenten Mehrebenenmodelle die Aggregierbarkeitsthese nicht. Die meisten ACT-Konzepte auf Clusterebene variierten nicht signifikant zwischen den Clustern (nichtsignifikante *Between*-Varianzen) und die Passung der *Between*-Modellteile war schlecht.

(9) Validitätsevidenz bezogen auf die Zusammenhänge zwischen den Variablen – deutsche wie auch die kanadische Befunde sind hier ambivalent, deutsche Befunde beruhen auf robusteren Methoden: Die strukturelle Beziehung der RU-Variablen zueinander entspricht den Annahmen, die sich aus den kanadischen Arbeiten ergeben: Instrumentelle, konzeptuelle und persuasive Forschungsanwendung sind signifikante Prädiktoren für Forschungsanwendung im Allgemeinen – sowohl im Gesamtmodell aller RU-Items als auch in allen anderen RU-Strukturgleichungsmodellen. Auch einige der COPSOQ-/ARU sowie ACT-Konzepte sind, wie erwartet, signifikante Prädiktoren der RU-Konzepte. Allerdings waren auch zahlreiche Prädiktoren nicht wie erwartet mit RU assoziiert (insbesondere *Overall RU* war hier betroffen) bzw. wiesen – unerwarteter Weise – Zusammenhänge in negativer Richtung mit den anderen drei RU-Konzepten auf.

(10) Die Befunde dieser Arbeit sind vorbehaltlich ihrer Überprüfung in unabhängigen Stichproben bzw. mittels anderer Methoden zu interpretieren. Mit dieser Einschränkung sind die übersetzten Instrumente vergleichsweise reliabel und valide: Modellmodifikationen müssen in neuen, unabhängigen Stichproben erneut überprüft werden. Dies war im Rahmen dieser Arbeit nicht möglich und muss daher künftigen Studien vorbehalten bleiben. Die Mehrebenenmodelle erfordern größere Stichproben, als in dieser Arbeit – insbesondere, was die durchschnittliche Clustergröße angeht. Auch hier müssen künftige Studien die Befunde erneut überprüfen.

Die Gründe für die überraschenden Ergebnisse der Strukturgleichungsmodelle (negative Zusammenhänge zwischen Prädiktoren und RU) sollten mit qualitativen Methoden untersucht werden. Auch Modelle, in die zusätzlich Motivations-, Verhaltens- und Dispositionsvariablen integriert werden, könnten aufschlussreiche Erkenntnisse liefern. Tendenziell sprechen die Befunde dafür, dass die deutschen Instrumentenversionen akzeptabel reliabel und valide sind – insbesondere aufgrund der sehr robusten Methoden und im Vergleich zu alternativen Instrumenten. Limitationen und zukünftig nötige klärende Untersuchungen wurden in dieser Arbeit skizziert. Mit den deutschen Versionen des ACT, der *Estabrooks-Items* und der CRU-Skala liegen somit drei verhältnismäßig robuste Forschungsinstrumente vor, um den Zusammenhang zwischen Organisationskontext und Anwendung wissenschaftlicher Erkenntnisse in der stationären Altenpflege Deutschlands zu erforschen.

Anhang

Anhang 1: Vertrag über die Nutzung des ACT (*Full Use Permission Agreement*)



Knowledge Utilization Studies Program

Complete and return by fax or post to:

Dr. Carole A. Estabrooks
5-112 Clinical Sciences Building
Faculty of Nursing
University of Alberta
Edmonton, AB T6G 2G3 CANADA
Telephone: (780) 492-6187; Fax: (780) 492-6186
Email: kusp@ualberta.ca

Alberta Context Tool (ACT): Full Use Permission Agreement

Full Name Matthias Hoben
Full Mailing Address Network Aging Research (NAR), Ruprecht-Karls-University Heidelberg, Bergheimer Str. 20, 69115 Heidelberg, Germany
Telephone +49 (0)6221 / 726 58 33 Fax +49 (0)6221 / 54 81 00
Email Address hoben@nar.uni-heidelberg.de

The following constitutes an agreement between Matthias Hoben
(Name, please print) Of ^{(1) Institute for Health and Nursing Science, Martin-Luther-University Halle-Wittenberg, Magdeburger Str. 8, 06112 Halle/Saale, Germany}
^{(2) Network Aging Research (NAR), mailing address: see above}
(Name and mailing address of organization) hereinafter called **Researcher**
and
the Knowledge Utilization Studies Program of the University of Alberta, Edmonton, AB, Canada,
hereinafter called **KUSP**.

Conditions of use are located at the end of this document.

Permission is granted for the project described in this agreement only, as outlined below:

Name of research project or thesis: Translation of scientific knowledge and implementation into nursing care of people with dementia

Anticipated start and completion date of project: Translation, adaptation and psychometric testing of the Alberta Context Tool (ACT)

Projected size of research sample: 500

Number of surveys to be administered: 1500

Signatures


Researcher  Date March, 28., 2011

Matthias Hoben
Martin-Luther-Universität Halle-Wittenberg
Medizinische Fakultät
Institut für Gesundheits- und Pflegewissenschaft
Direktions- & Supervisionsambulanz
Mendelsburger Str. 8 • 06109 Halle (Saale)
Prof. Dr. Johann Behrens

Students' Supervisor  Date March, 25., 2011

KUSP  Date Mar 28, 2011

The researcher also gives KUSP permission to list their project, in the ACT manual and/or on the future ACT website, with other projects and researchers that have used the ACT.

Researcher  Date March, 28., 2011

ACT survey to be sent:

Please indicate the survey **version** that you require:

- Acute Care (Adults)
- Acute Care (Pediatrics)
- Long -Term Care
- Home Care

Please indicate the survey **form** that you require:

- Nurses (RNs/LPNs)
- ~~Nurses~~ Health care aides
- Managers
- Practice Specialists (e.g., Clinical Educator, Quality Improvement Specialist)
- Allied Health Care Providers

The Knowledge Utilization Studies Program (KUSP) will provide the researcher with a copy of the Alberta Context Tool (ACT). The researcher is responsible for the reproduction of ACT, the distribution of the survey, and the collection of data.

The researcher will retain full rights to the data for publication. On completion of the study the researcher will forward a digital copy of the ACT and demographic data from their study. These data will be used to assess the psychometric properties of the ACT and to build the ACT's normative record on an ongoing basis. KUSP will retain rights to use these data within analyses of its larger ACT data set but will not publish analyses based on these data alone.

The data should be received within one year of project completion and submitted as follows:

- in Excel format
- with documentation (i.e., codebook)
- by secure courier on a DVD (DVD-R format) **OR** uploaded to the KUSP secure data site (by arrangement with the KUSP Data Manager)

The researcher will not distribute ACT to any other party. The text will not be copied in any publication, research reports, or theses arising from the research.

The researcher will not adapt or modify the ACT without permission.

Permission to use ACT is granted solely for the project described in the *Full Use Permission Agreement* between KUSP and the researcher and is not transferrable to other researchers or projects.

If the ACT will be distributed in a language other than English, professional translation and back translation from English to the second language is required. Consultation with Dr. Estabrooks during and following completion of the back translation must precede use of the tool. All costs associated with translation and back translation are the responsibility of the requesting researcher. The translated version of ACT will become the property of Dr Estabrooks who will provide it, where requested, to other researchers under the same conditions as have been outlined above.

All copies of ACT must include the following text:

<p>© Carole A. Estabrooks, 2007 All rights reserved. No part of this instrument may be produced, stored in a retrieval system, or transmitted in any form or by any means without the prior written permission of the copyright owner.</p>
--

Anhang 2: Begründung für die Korrelation von Residualvarianzen in den ACT- und CRU-Faktorenmodellen der einzelnen Berufsgruppen

Gruppe	Modell	Korrelierte Residualvarianzen	Begründung (gemeinsame „Fehlerquelle“)
PHK	ACT 2b	L1 mit L2	L1 fragt, ob die Führungsperson Rückmeldung sucht, selbst wenn diese schwer anzunehmen sei. L2 fragt, ob die Führungsperson ihr Augenmerk eher auf Erfolge als auf Misserfolge lege. Bei beiden Items meldeten viele der PHK zurück, sie hätten sich darüber noch nie Gedanken gemacht und wüssten nicht so genau, was sie ankreuzen sollten.
	ACT 3b	II1 mit II2	II1 fragt nach dem Austausch mit PHK, II2 nach dem Austausch mit PFK. Insbesondere die bewohnernah tätigen Berufsgruppen (PHK, PFK, TBK und APS) schienen hier oft nicht genau genug zu lesen und beide Items in einen Topf zu werfen.
		II8 mit III0	II8 fragt nach dem Austausch mit Personen, die neue Ideen in den Wohnbereich einbringen, III0 fragt informellen Gesprächen allgemein (personunenabhängig). Der Unterschied dieser beiden Items war vielen PHL nicht klar. Vermutlich dachten sie jeweils an die gleichen Personen.
PFK	ACT 2b	CAP1 mit CAP2	CAP1 fragt, ob die Teammitglieder Informationen untereinander austauschen, CAP2 fragt, ob eigene Beobachtungen von Personen in verantwortlichen Positionen ernstgenommen werden. Sowohl PFK als auch FK (von denen die meisten eine PFK-Ausbildung und -sozialisierung aufweisen) brachten diese beiden Items in engen Zusammenhang.
		T2 mit T3	T2 fragt nach der Zeit, etwas in der Fachliteratur nachzuschlagen, T3 fragt nach der Zeit, sich mit Kollegen über neue klinische Erkenntnisse auszutauschen. Sowohl PFK als auch FK (von denen die meisten eine PFK-Ausbildung und -sozialisierung aufweisen) brachten diese beiden Items in engen Zusammenhang.
	ACT 3b	III1 mit II2	III1 fragt nach dem Austausch mit PHK, II2 nach dem Austausch mit PFK. Insbesondere die bewohnernah tätigen Berufsgruppen (PHK, PFK, TBK und APS) schienen hier oft nicht genau genug zu lesen und beide Items in einen Topf zu werfen.
		SER5 mit SER6	SER5 fragt nach Verfahrens- oder Dienstanweisungen, SER6 fragt nach fachlichen Standards (z.B. pflegerische Expertenstandards). Insbesondere PFK und APS fiel es oft schwer, zwischen diesen beiden Vorgabenarten zu unterscheiden.
		SER7 mit SER8	SER7 fragt nach einem PC mit Internetanschluss, SER8 fragt nach pflegespezifischer Software auf diesen PCs (z.B. Pflegedokumentationssoftware). Insbesondere für bewohnernah tätige Berufsgruppen, die viel mit der Dokumentation ihrer Tätigkeiten befasst sind (PFK, TBK und APS) war die Nutzung eines PCs während ihrer Arbeitszeit identisch mit der Nutzung dieser Software.
TBK	ACT 2b	CAP2 mit CAP4	CAP2 fragt, ob eigene Beobachtungen von Personen in verantwortlichen Positionen ernstgenommen werden, CAP4 fragt, ob man sich dabei wohlfühlt, mit Personen in verantwortlichen Positionen über die Versorgung der Bewohner zu sprechen. In der TBK-Gruppe waren diese „Personen in verantwortlichen Positionen“ der beiden Items oft identisch.
	ACT 3b	III1 mit II2	III1 fragt nach dem Austausch mit PHK, II2 nach dem Austausch mit PFK. Insbesondere die bewohnernah tätigen Berufsgruppen (PHK, PFK, TBK und APS) schienen hier oft nicht genau genug zu lesen und beide Items in einen Topf zu werfen.
		SER7 mit SER8	SER7 fragt nach einem PC mit Internetanschluss, SER8 fragt nach pflegespezifischer Software auf diesen PCs (z.B. Pflegedokumentationssoftware). Insbesondere für bewohnernah tätige Berufsgruppen, die viel mit der Dokumentation ihrer Tätigkeiten befasst sind (PFK, TBK und APS) war die Nutzung eines PCs während ihrer Arbeitszeit identisch mit der Nutzung dieser Software.

Gruppe	Modell	Korrelierte Residualvarianzen	Begründung (gemeinsame „Fehlerquelle“)
FK	ACT 2b	CAP1 mit CAP2	CAP1 fragt, ob die Teammitglieder Informationen untereinander austauschen, CAP2 fragt, ob eigene Beobachtungen von Personen in verantwortlichen Positionen ernstgenommen werden. Sowohl PFK als auch FK (von denen die meisten eine PFK-Ausbildung und -sozialisierung aufweisen) brachten diese beiden Items in engen Zusammenhang.
		T2 mit T3	T2 fragt nach der Zeit, etwas in der Fachliteratur nachzuschlagen, T3 fragt nach der Zeit, sich mit Kollegen über neue klinische Erkenntnisse auszutauschen. Sowohl PFK als auch FK (von denen die meisten eine PFK-Ausbildung und -sozialisierung aufweisen) brachten diese beiden Items in engen Zusammenhang.
	ACT 3b	FI1 mit FI5	FI1 fragt nach der Teilnahme an Fallbesprechungen, FI5 fragt nach der Teilnahme an Fort- und Weiterbildungen (außerhalb der Einrichtung). In beiden Fällen dachten FK oft eher an sich als Organisierende oder Leitende dieser Aktivitäten, nicht als Teilnehmer.
		II2 mit II10	II2 fragt nach dem Austausch mit PFK, II10 fragt nach informellen Gesprächen allgemein (personenunabhängig). PFK waren oft die Personengruppe, mit der sich FK am häufigsten informell austauschten.
		SER1 mit SER2	SER1 fragt nach der Nutzung von Lehrbüchern, SER2 nach der Nutzung einer Fachbibliothek (in der Einrichtung). Sehr oft wurde von den FK der eigene Lehrbuchbestand als die „Fachbibliothek“ der Einrichtung angesehen.
	CRU 4	CRU _b 4 mit CRU _b 5	CRU _b 4 fragt nach neuen Ideen (ausgelöst durch Forschungswissen), Bewohner zu versorgen, CRU _b 5 fragt nach Forschungswissen, das dazu führte, dass Dinge, die bisher getan wurden, nun besser einleuchten. Der Unterschied dieser beiden Items war FK oft nicht klar.
APS	ACT 2b	C5 mit C6	C5 fragt, ob sich alle Akteure in der Einrichtung anstrengen, den Bewohnern das zukommen zulassen, was sie benötigen, C6 fragt, ob man selbst vom Team unterstützt wird. Für APS hingen diese beiden Aspekte oft eng zusammen oder wurden von diesen als identisch betrachtet.
	ACT 3b	II1 mit II2	II1 fragt nach dem Austausch mit PHK, II2 nach dem Austausch mit PFK. Insbesondere die bewohnernah tätigen Berufsgruppen (PHK, PFK, TBK und APS) schienen hier oft nicht genau genug zu lesen und beide Items in einen Topf zu werfen.
		SER5 mit SER6	SER5 fragt nach Verfahrens- oder Dienstanweisungen, SER6 fragt nach fachlichen Standards (z.B. pflegerische Expertenstandards). Insbesondere PFK und APS fiel es oft schwer, zwischen diesen beiden Vorgabenarten zu unterscheiden.
		SER7 mit SER8	SER7 fragt nach einem PC mit Internetanschluss, SER8 fragt nach pflegespezifischer Software auf diesen PCs (z.B. Pflegedokumentationssoftware). Insbesondere für bewohnernah tätige Berufsgruppen, die viel mit der Dokumentation ihrer Tätigkeiten befasst sind (PFK, TBK und APS) war die Nutzung eines PCs während ihrer Arbeitszeit identisch mit der Nutzung dieser Software.
Gesamtstichprobe	ACT3b	II1 mit II2	II1 fragt nach dem Austausch mit PHK, II2 nach dem Austausch mit PFK. Insbesondere die bewohnernah tätigen Berufsgruppen (PHK, PFK, TBK und APS) schienen hier oft nicht genau genug zu lesen und beide Items in einen Topf zu werfen.
		SER5 mit SER6	SER5 fragt nach Verfahrens- oder Dienstanweisungen, SER6 fragt nach fachlichen Standards (z.B. pflegerische Expertenstandards). Insbesondere PFK und APS fiel es oft schwer, zwischen diesen beiden Vorgabenarten zu unterscheiden.

Lebenslauf

PERSÖNLICHE ANGABEN

Geburtsdatum: 16. Oktober 1976
Geburtsort: Mutlangen (Ostalbkreis)

WISSENSCHAFTLICHER WERDEGANG

seit 07/2010 Promotion, Martin-Luther-Universität Halle-Wittenberg (MLU), Medizinische Fakultät, Institut für Gesundheits- und Pflegewissenschaft (IGPW)
Stipendiat im Graduiertenkolleg Demenz, Netzwerk Altersforschung (NAR), Ruprecht-Karls-Universität (RKU) Heidelberg, gefördert von der Robert Bosch Stiftung

09/2008-09/2010 MSc Gesundheits- und Pflegewissenschaft, IGPW, Medizinische Fakultät, MLU Halle-Wittenberg

09/2003-07/2007 Dipl. Pflegewirt (FH), Hochschule Esslingen/Neckar, Fakultät für Soziale Arbeit, Gesundheit und Pflege, Studiengang Pflege/Pflegemanagement

UNIVERSITÄRE BERUFSTÄTIGKEIT

12/2008-06/2010 Wissenschaftlicher Mitarbeiter, IGPW, Medizinische Fakultät, MLU Halle-Wittenberg

07/2007-11/2008 Wissenschaftlicher Mitarbeiter, Institut für Gerontologie, Fakultät für Verhaltens- und empirische Kulturwissenschaften, RKU Heidelberg

AUßERUNIVERSITÄRE BERUFSTÄTIGKEIT

Seit 06/2005 Freiberufliche Tätigkeit als Dozent und Berater für Wissenstranslation, Qualitätssicherung und -entwicklung in der stationären Altenpflege, der geriatrischen Rehabilitation und im Akutkrankenhausbereich

04/2001-06/2004 Tätigkeit als Gesundheits- und Krankenpfleger in verschiedenen Settings und Bereichen

BERUFLICHER WERDEGANG

04/1998-03/2001 Ausbildung zum Krankenpfleger, Kliniken Ludwigsburg-Bietigheim gGmbH

08/1996-03/1998 12 Monate Zivildienst, Rettungshelfer, Deutsches Rotes Kreuz Esslingen/Neckar, danach sieben Monate Pflegehelfer im Altenheim Obertor, Esslingen/Neckar

08/1993-07/1996 Fachgebundene Hochschulreife und Ausbildung zum Physikalisch-Technischen-Assistenten, naturwissenschaftlich-technisches Gymnasium der Kerschensteinerschule Stuttgart

MITGLIEDSCHAFTEN

Seit 10/2011 Gründungsmitglied der Sektion Dissemination und Implementierung (SDI) der Deutschen Gesellschaft für Pflegewissenschaft (DGP)

Seit 04/2008 Mitglied im Deutschen Berufsverband für Pflegeberufe (DBfK)

seit 10/2007 Mitglied in der AG Pflegeforschung Rhein-Neckar, Heidelberg

seit 04/2005 Mitglied der Deutschen Gesellschaft für Pflegewissenschaft (DGP)

Heidelberg/Neckar,

(Datum)

(Unterschrift)

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Ich versichere, dass ich für die inhaltliche Erstellung der vorliegenden Arbeit nicht die entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen habe. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Heidelberg/Neckar, den

(Datum)

(Unterschrift)

Erklärung über frühere Promotionsversuche

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Frühere Promotionsversuche wurden von mir nicht unter-
nommen.

Heidelberg/Neckar, den

(Datum)

(Unterschrift)

Danksagung

Mein erster und ganz besonderer Dank geht an alle Pflegeeinrichtungen und die dort tätigen Personen, die sich an der Studie beteiligt haben. Die Unterstützung der Führungspersonen beim Vorbereiten der Datenerhebung, bei Rekrutierung und Kontaktaufnahme mit den Beschäftigten und die Bereitschaft so vieler Beteiligter, die Fragebögen auszufüllen und in den kognitiven Interviews Rede und Antwort zu stehen – während der ohnehin bereits knapp bemessenen Arbeitszeit – all dies war unschätzbar wertvoll und für diese Arbeit fundamental. Vielen Dank dafür.

Danken möchte ich auch den Betreuern dieser Arbeit, Prof. Dr. Behrens und Prof. Dr. h.c. Kruse. Ihre zugewandte und verlässliche Unterstützung, ihre Offenheit für das in Deutschland bislang noch eher wenig beachtete Thema sowie die Freiheit, die sie mir bei der Konzeption und Umsetzung des Projekts ließen, machten das Projekt zu dem, was es schlussendlich geworden ist.

Auch Prof. Dr. Carole Estabrooks und Dr. Janet Squires – den beiden maßgeblichen Urheberinnen der verwendeten Assessmentinstrumente – hat diese Arbeit vieles zu verdanken. Sie genehmigten die Übersetzung der Instrumente und deren Einsatz in der Validierungsstudie und die Kooperation mit ihnen im Rahmen des Übersetzungsprozesses war eine große Bereicherung. Dafür und für manche fachliche und moralische Unterstützung *my sincere gratitude*.

Unterstützung im Übersetzungsprozess erhielt ich von zahlreichen weiteren Personen: Dr. Marion Bär (zweite Vorwärtsübersetzung) sowie Dr. Cornelia Mahler und Sarah Berger (Rückübersetzungen) waren nicht nur bei der Konzeption des Übersetzungsprozesses eine große Hilfe, sie leisteten auch unschätzbare Dienste bei der zeit- und manchmal nervenintensiven Vor- und Rückübersetzungsarbeit. Danke dafür. Marion Bär danke ich zudem für die Moderation der Expertenpanels. Schließlich danke ich auch den Mitgliedern der Expertenpanels. Ihre Unterstützung war Gold wert. Zweien meiner Kolleginnen gilt zudem mein herzlicher Dank für die Zufallsauswahl der Items für die *Cognitive Debriefings* und der zu rekrutierenden Pflegeheime in der Validierungsstudie (Stefanie Schmidt) sowie für die Unterstützung bei der Übersetzung der *Attitudes Towards Research Use scale* (Birgit Kramer).

Ohne eine finanzielle Basis kann kein Projekt wie das vorliegende zustande kommen. Für diese sorgte im vorliegenden Fall die Robert Bosch Stiftung mit ihrem großzügigen Promotionsstipendium. Auch dafür mein ausdrücklicher Dank. Für die Bereitstellung der infrastrukturellen Rahmenbedingungen und die logistische Unterstützung geht mein Dank ans Netzwerk AlternsfoRschung (NAR). Insbesondere bin ich Herrn Prof. Dr. Beyreuther zutiefst zu Dank verpflichtet: für ein offenes Ohr, das Vertrauen in mein Projekt und die großzügige und unbürokratische Unterstützung in unzähligen Situationen. Danke schließlich an die Kollegiaten des Graduiertenkollegs Demenz und des NAR-Kollegs, mit denen ich in der Zeit am NAR bereichernde fachliche Gespräche führen und äußerst erinnerungswürdige Momente teilen durfte.

Damit eine Promotion gelingen kann, sind nicht nur die fachlichen und beruflichen Rahmenbedingungen entscheidend, sondern auch die privaten. Zahlreiche nahestehende Personen haben

mich bei der Erstellung dieser Arbeit begleitet, unterstützt und – was manchmal fast wichtiger war – bisweilen von dieser abgelenkt. Ein besonderer Dank geht hier an meine Eltern, meinen Schwiegervater, meine Schwester Miriam, meine Schwägerin Susanne und ihren Freund Jochen (danke auch fürs Gegenlesen) sowie an Amelie und Matthias. Vor allem aber gilt meine kaum in Worte zu fassende Dankbarkeit meiner Frau, Charlotte. Selbst mit ihrer Dissertation befasst, gelang es ihr stets auf bewundernswerte Weise, den Alltag nicht aus dem Blick zu verlieren, uns (und insbesondere mich) zu erden und Raum zu schaffen für die kleinen, schönen Dinge im Leben. Danke für Deine Liebe und Deinen Beistand!