Suchtmonitoring Schweiz
Monitorage suisse des addictions
Monitoraggio svizzero delle dipendenze
Addiction Monitoring in Switzerland

Juli 2017

Entwicklung einer Kurzform der Compulsive Internet Use Scale (CIUS)

Dieser Bericht wurde vom Bundesamt für Gesundheit in Auftrag gegeben.
Vertragsnummer 16.924108 / 204.0001 - 1511/1

SUCHT | SCHWEIZ

Zitiervorschlag:

Gmel G. (2017). Entwicklung einer Kurzform der Compulsive Internet Use Scale (CIUS). Sucht Schweiz, Lausanne, Schweiz

Impressum

Auskunft:	suchtmonitoring@bag.admin.ch
Bearbeitung:	Sucht Schweiz: Gerhard Gmel, Christiane Gmel
Vertrieb:	Bundesamt für Gesundheit, Direktionsbereich Öffentliche Gesundheit, Nationale Präventionsprogramme
Copyright:	© Bundesamt für Gesundheit, Bern 2017
ISBN:	978-2-88183-219-2

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Zusammenfassung

Das Ziel dieses Berichtes war es im Auftrag des Bundesamtes für Gesundheit abzuschätzen, inwieweit man die *Compulsive Internet Use Scale* (CIUS) kürzen kann. Vorgabe dabei ist, eine zumindest einstellige Itemzahl (Im Orginal: 14 Items) zu erreichen. Dazu wurden im ersten Schritt Verfahren der klassischen Testtheorie eingesetzt. Diese haben die Annahme, dass die fünfstufigen Ratingskalen der Items mit intervallskalierten Verfahren analysiert werden können.

In einem zweiten Schritt wurde diese Annahme aufgehoben und mit Verfahren, die nur ordinales Datenniveau annehmen, gearbeitet. Dabei wurden Verfahren der probabilistischen Testtheorie sowie konfirmatorische Faktorenanalysen unter der Annahme ordinalskalierter Items durchgeführt. Im Vordergrund stehen dabei auch Verfahren zur Prüfung des *differential item functioning*, d.h. eine Prüfung der Testfairness in verschiedenen Populationssegmenten (nach Alter, Geschlecht und Sprachregion), sowie der Prüfung verschiedener Ausprägungen der Messinvarianz in diesen verschiedenen Populationssegmenten. Dies ist wichtig, damit der Test gesamtschweizerisch einsetzbar ist, also z.B. nicht in verschiedenen Sprachregionen unterschiedlich interpretiert wird.

Allgemein kann gesagt werden, dass beide Varianten zu ähnlichen Lösungen kommen. Es konnte eine Skala bestehend aus 9 Items (sowie eine noch kürzere Skala mit 8 Items) gefunden werden, die hervorragende psychometrische Eigenschaften hat und nahezu starke (skalare) Invarianz über verschiedene Populationssegmente aufweist.

- Lösung mit 9 Items: Item 1, Item 4, Item 5, Item 7, Item 9, Item 10, Item 11, Item 12, Item 14 der originalen 14 Items umfassenden Skala

- Lösung mit 8 Items: zusätzlich ohne Item 10

Die Skalenlösung beinhält weiterhin Items aller 5 Unterdimensionen des CIUS (Konflikt, Coping, Kontrollverlust, Vertieftsein (preoccupation), Entzugserscheinungen). Die Skalenentwicklung wurde an einer Lernstichprobe durchgeführt und anschliessend erneut an einer zweiten Validierungsstichprobe getestet. Auch in der Validierungsstichprobe ergaben sich hervorragende psychometrische Eigenschaften.

Die Skala bestehend aus 9 Items weist ein Cronbachs Alpha von über 0.8 auf. Ebenso weisen die Modelanpassungen mit der konfirmatorischen Faktorenanalyse unter der Annahme ordinalen Skalenniveaus hervorragende Werte auf (RMSEA < 0.05 und CFI > 0.95). Analysen mittels *graded response models* der *item response theory* (IRT) zeigen, dass die Kurzform ein vergleichbares, wenn nicht höheres Informationspotenzial hat.

Vergleiche mit der Originalskala mit dem vorgeschlagenen Schwellenwert von 28 Punkten ergeben eine Fläche unter der *receiver operating characteristic* (ROC) *curve* von über 0.98 bei Sensitivitäten und Spezifitäten von über 0.9. Sowohl IRT als auch ROC Analysen deuten auf einen Schwellenwert für problematischen Internetgebrauch von 15 Punkten bei der Kurzform hin.

Es lässt sich folgendes Fazit ziehen: Unter der Annahme, dass der CIUS im Original ein valides Instrument für die Messung des problematischen oder abhängigen Internetgebrauches ist, so sind es auch die hier untersuchten Kurzformen, wenn sie nicht sogar besser sind. Die Kurzform ist sowohl einsetzbar in unterschiedlichen Sprachregionen als auch in verschiedenen Altersgruppen und bei beiden Geschlechtern.

Résumé

À la demande de l'Office fédéral de la santé, l'objectif de ce rapport était d'évaluer dans quelle mesure il est possible de raccourcir la *Compulsive Internet Use Scale* (CIUS). Il s'agissait d'obtenir au moins un nombre d'items à un seul chiffre (la version originale contient 14 items). Dans un premier temps, les opérations de la théorie des tests classique ont été appliquées. Ces dernières utilisent l'hypothèse que les échelles de notation des items à cinq niveaux peuvent être analysées au moyen d'opérations des échelles d'intervalle.

Dans un deuxième temps, cette hypothèse a été écartée et seules les méthodes qui acceptent un niveau ordinal de mesure des données ont été employées. À cet effet, les opérations de la théorie de tests probabilistes ainsi que les analyses factorielles confirmatoires ont été exécutées dans l'hypothèse d'items sur une échelle ordinale. L'accent a également été mis sur les opérations de contrôle du *differential item functioning*, c'est-à-dire le contrôle de l'équité des tests (test fairness) dans divers segments de la population (en fonction de l'âge, du sexe et de la région linguistique), ainsi que le contrôle de diverses manifestations de l'invariance des mesures dans ces divers segments de la population. C'est un élément important car il permet d'appliquer ce test à l'ensemble de la Suisse, ce qui, par exemple, évite des interprétations différentes d'une région linguistique à l'autre.

D'un point de vue général, on peut dire que les deux variantes conduisent à des solutions similaires. On a pu trouver une échelle composée de 9 items (ainsi qu'une échelle encore plus courte, de 8 items) qui présente d'excellentes propriétés psychométriques et une invariance (scalaire) quasi forte entre les divers segments de la population.

- Solution à 9 items : item 1, item 4, item 5, item 7, item 9, item 10, item 11, item 12, item 14 de l'échelle originale couvrant les 14 items

- Solution à 8 items : idem, l'item 10 en moins

Cette solution d'échelle contient en outre les items de chacune des 5 sous-dimensions du CIUS (conflit, coping, perte de contrôle, préoccupation, symptômes de manque). Le développement de l'échelle a été effectué sur un échantillon d'apprentissage, puis à nouveau testé sur un deuxième échantillon de validation. Ce dernier échantillon comportait lui aussi d'excellentes propriétés psychométriques.

L'échelle à 9 items présente un alpha de Cronbach de plus de 0.8. De même, les adaptations du modèle à l'aide de l'analyse factorielle confirmatoire dans l'hypothèse de niveaux d'échelle ordinale ont, elles aussi, présenté des valeurs exceptionnelles (RMSEA < 0.05 et CFI > 0.95). Les analyses à l'aide des *graded response models* de la théorie des réponses aux items (*item response theory*, IRT) montrent que la version raccourcie a un potentiel d'information comparable, voire plus important.

Les comparaisons avec l'échelle d'origine sur la base de la valeur seuil proposée à 28 points donnent une surface en-dessous de la courbe ROC (*receiver operating characteristic*) supérieure à 0.98 en matière de sensibilité et supérieure à 0.9 en matière de spécificité. Tant les analyses IRT que les analyses ROC indiquent une valeur seuil de l'utilisation problématique d'Internet de 15 points dans la version raccourcie.

On peut en conclure ce qui suit : en supposant que le CIUS est un instrument valide pour mesurer l'utilisation problématique ou addictive d'Internet, les versions raccourcies examinées ici le sont également, voire elles sont plus performantes. La version raccourcie s'applique à la fois aux diverses régions linguistiques, aux divers groupes d'âge et aux deux sexes.

1. Einleitung

Das Ziel dieses Berichtes ist es abzuschätzen, inwieweit man die *Compulsive Internet Use Scale* (CIUS) kürzen kann. Vorgabe ist dabei, eine zumindest einstellige Itemzahl (Im Orginal: 14 Items) zu erreichen. Dazu werden im ersten Schritt Verfahren der klassischen Testtheorie eingesetzt. Diese haben die Annahme, dass die fünfstufigen Ratingskalen der Items mit intervallskalierten Verfahren analysiert werden können. Eine solche Annahme wird von mehreren Autoren vertreten (z.B. Norman, 2010).

Die folgenden Fragestellungen sollen beantwortet werden:

a) Ist der CIUS in der Schweiz ein eindimensionales Messinstrument?
b) Welche Items bieten sich aufgrund der klassischen Testtheorie (Trennschärfe, Schwierigkeit) zur Kürzung an?
c) Wie sehen bestehende, international verwendete Kurzformen aus und wie verhalten sich diese im Vergleich mit für die Schweiz vorgeschlagene Kürzungsvarianten sowie im Vergleich mit der Originalversion?

Im zweiten Teil werden weitere Verfahren der probabilistischen Testtheorie sowie konfirmatorische Faktorenanalyse unter der Annahme ordinalskalierter Items durchgeführt. Im Vordergrund stehen dabei auch Verfahren zur Prüfung des *differential item functioning*, d.h. eine Prüfung der Testfairness in verschiedenen Populationssegmenten (nach Alter, Geschlecht und Sprachregion). Zusätzlich werden verschiedene Ausprägungen der Messinvarianz in diesen Populationssegmenten geprüft. In diesem Zusammenhang ist es wichtig zu betonen, dass in der Schweiz die Fragen in der jeweiligen Landessprache der befragten Person gestellt werden. Dabei können die Fragen in den Formulierungen leicht unterschiedlich sein, da sich Übersetzungen nicht immer exakt 1:1 übertragen lassen. Im Suchtmonitoring wurde als deutsche Version jene der Pinta-Studie (Bischof et al., 2013) verwendet. In der französischsprachigen Schweiz kam die validierte Version von Khazaal und Kollegen (2012) zum Einsatz. Beide Versionen orientieren sich sehr wortgetreu an der englischen Version von Meerkerk (2007), wobei dieser vermutlich eine niederländische Version gebraucht hat und somit die englische Version auch nur eine Übersetzung für die Dissertation darstellt. In Diskussionen mit Muttersprachlern, wurden von Sucht Schweiz beide Versionen bei einzelnen Wörtern aufeinander abgestimmt. Dies waren aber minimale Korrekturen im Vergleich zu den Originalversionen. Gemäss der Vorlage der englischen, deutschen und französischen Version wurde die Skala von Sucht Schweiz von einem Muttersprachler ins Italienische übersetzt, wobei bei Rückfragen die Bedeutung der jeweiligen Frage mit mehrsprachigen anderen Wissenschaftlern von Sucht Schweiz diskutiert worden ist. Aus praktischen Gründen und Kostengründen konnten keine Verfahren der Rückübersetzung eingesetzt werden. Dies könnte einen Schwachpunkt der hier berichteten Studie darstellen, da "differential item functioning" nicht eindeutig auf kulturelle Unterschiede zwischen den Sprachregionen zurückgeführt werden kann, sondern auch an Unterschieden in den Übersetzungen liegen könnte. Die aus unserer Sicht starke Übereinstimmung der Fragenformulierungen (siehe Anhang Fragebogen) macht Übersetzungsschwächen aus unserer Sicht jedoch sehr unwahrscheinlich.

Bei den Analysen wurden die Daten der Befragung des Suchtmonitorings von Januar bis Juni im Jahr 2013 als Lernstichprobe herangezogen. Die Daten im gleichen Zeitraum des Jahres 2015 wurden abschliessend zur Validierung der Kurzform verwendet. Im Jahr 2013 füllten 1371 Personen die Daten zum CIUS aus, im Jahr 2015 waren es 1550 Personen. Es wurden Unterstichproben nach Alter, Geschlecht und Sprachregion gebildet. Wegen der geringen Fallzahlen in der italienischsprachigen Schweiz wurden deren Daten mit jener der französischsprachigen Schweiz zusammengezogen. Dies könnte einen Schwachpunkt der vorliegenden Arbeit darstellen. Alternativ hätten wir nur die deutsche und französische Sprachregion auswählen können. Die italienische Sprachregion liefert einfach zu geringe Fallzahlen für eine selbstständige Betrachtung. Wir haben uns für den vorliegenden Bericht entschieden, auf diese Variante zu verzichten, um a) eine gesamtschweizerische Untersuchung durchzuführen und b) mit möglichst grossen Fallzahlen zu arbeiten.

Die Altersgruppen 15-19 Jahre, 20-34 Jahre und über 34 Jahre wurden so gewählt, dass es in den beiden jüngeren Altersgruppen ausreichende Fallzahlen gab (vgl. Tabelle 1). Ab 35 Jahre benutzen zwar noch viele Menschen das Internet, aber kaum jemand erreicht einen problematischen Gebrauch (Schwellenwert > 27) auf der Originalskala des CIUS. Bei der Prüfung der Testfairness im Hinblick auf unterschiedliche Altersgruppen wird deshalb das Augenmerk vorrangig auf die beiden jüngeren Altersgruppen gelegt. In der zur Zeit einzigen, in einer peer-reviewten Zeitschrift veröffentlichten Kurzform von Cartierre und Kollegen (2011) wurden beispielsweise 269 Schülern der "classe de quatrième", das ist die dritte Klasse der französischen Sekundarstufe (vergleichbar mit der "Quarta" in deutschen Gymnasien), aus 4 französischen *collèges* mit einem Durchschnittsalter von 13.8 Jahren herangezogen. Die hier verwendeten Stichproben sind also in jeder Altersgruppe um einiges grösser.

Tabelle 1: Fallzahlen in der Lern- und Validierungsstichprobe

Alter	Lernstichprobe	Validierungsstichprobe
15-19 Jahre	469	348
20-34 Jahre	514	425
35 und mehr Jahre	388	777
Total	1371	1550

2. Klassische Testtheorie zur Findung einer Kurzform unter Annahme intervalskalierter Items

2.1 Dimensionalität des CIUS

In einem ersten Schritt wird gemäss dem Kaiser-Meyer-Olkin-Kriterium (KMO) festgestellt, ob ein Datensatz prinzipiell für eine explorative Faktorenanalyse geeignet ist. Bei Werten unter 0.5 ist der Datensatz ungeeignet, bei Werten über 0.6 brauchbar, jedoch deutet sich an, dass Modifikationen vorgenommen werden müssen, die zu einer Verbesserung führen sollten. Bei Werten über 0.8 (1 ist der Höchstwert) gilt der Datensatz als gut geeignet für explorative Faktorenanalysen. Der KMO-Wert setzt sich im Wesentlichen aus Werten der "*Measure Sampling Adaequacy (MSA)*" für jedes einzelne Item zusammen. Nach Kaiser (1974) gelten folgende Daumenregeln für den KMO-Wert.

- 0.00 to 0.49: unacceptable (nicht akzeptabel).
- 0.50 to 0.59: miserable (schlecht).
- 0.60 to 0.69: mediocre (mittelmässig, unbedeutend)
- 0.70 to 0.79: middling (mittelmässig, leidlich ok).
- 0.80 to 0.89: meritorious (Annerkennung verdienend, lobenswert)
- 0.90 to 1.00: marvelous (fantastisch, ausgezeichnet).

Tabelle 2 gibt einen Überblick zu den KMO und MSA-Werten.

Tabelle 2: KMO- und MSA-Werte in der Gesamtstichprobe und Unterstichproben nach Alter, Geschlecht und Sprachregion

	Total	Alter			Geschlecht		Sprachregion	
	Total	**15-19**	**20-34**	**35+**	**Männlich**	**Weiblich**	**Deutsch**	**Lateinisch**
KMO	0.918	0.869	0.894	0.869	0.897	0.925	0.912	0.907
Bereich der MSA	.861-.959	.804-.924	.830-.943		.841-.946		.843-.963	.873-.944
Items mit MSA <0.9	12, 13	Alle ausser 4, 5, 14	1, 2, 8, 9, 12, 13	Alle ausser 4, 5, 7, 10; mit 12, 13 < .8	1, 2, 7, 8, 9, 12, 13	12, 13	9, 12, 13	1, 2, 9, 12, 13

Es zeigt sich, dass der Datensatz hervorragend für eine Faktorenanalyse geeignet ist, da alle KMO Werte, sogar in den Unterstichproben, deutlich über 0.8 liegen. Die MSA Werte der einzelnen Items liegen, bis auf 2 Ausnahmen (Item 12, und Item 13) bei den weniger relevanten über 34-Jährigen (weniger relevant, da dort kaum noch problematischer Internetgebrauch anzutreffen ist), bei über 0.8 und in der Regel sogar über 0.9. Auffallend ist jedoch, dass die Itempaare 1 und 2, 8 und 9 sowie 12 und 13 deutlich häufiger als die anderen Items unter denjenigen mit den geringsten MSA Werten liegen. Es ist auch zu bemerken, dass bei den Iteminterkorrelationen (nicht tabelliert) jeweils die Itempaare 1 und 2, 8 und 9 sowie 12 und 13 die höchsten Interkorrelationen aufweisen, also darauf hinweisen wechselseitig recht redundant zu sein.

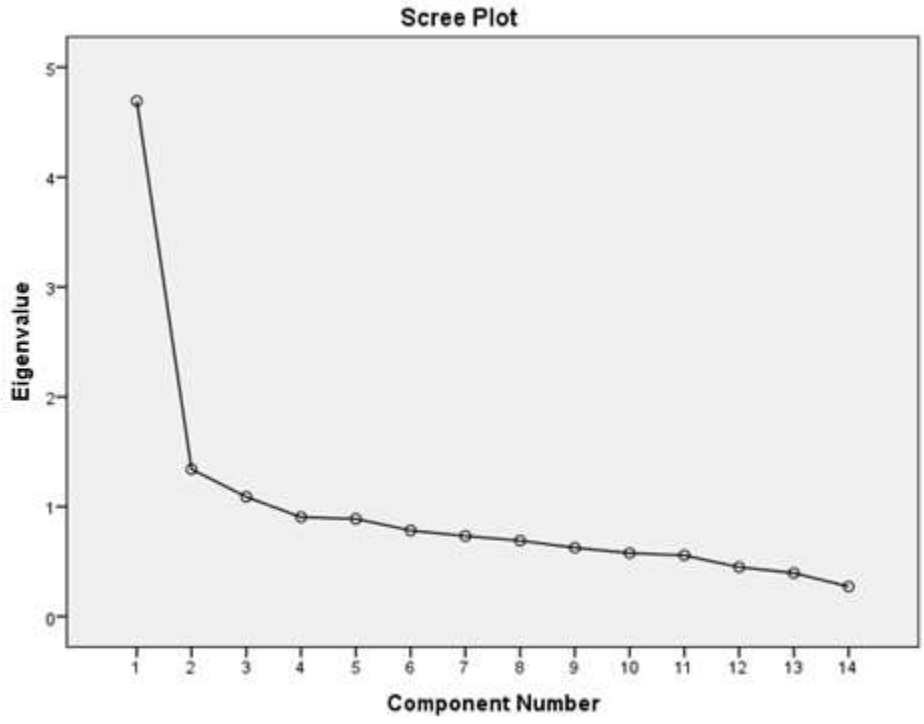
Die explorative Faktorenanalyse (Tabelle 3) deutet an, dass es im Wesentlichen einen Hauptfaktor gibt, der um die 40% der Gesamtvarianz erklärt. Gemäss dem Kaiser-Guttman-Kriterium sollten auch Faktoren mit einem Eigenwert über 1 eventuell herangezogen werden (ein Eigenwert von 1 erklärte im vorliegenden Fall mit 14 Items (1/14) 7.1% der Gesamtvarianz). Allerdings hat sich gezeigt, dass das Kaiser-Guttman-Kriterium die Faktorenzahl erheblich überschätzt (Zwick & Velicer, 1986). Eine erste graphische Annäherung bietet Catells *Scree Plot*, der sich als akkurat erwiesen hat. Darüber hinaus wurde die *parallel analysis* nach Horn (1965) und der *minimum average partial* (MAP) Test (Velicer, 1976) durchgeführt.

Tabelle 3: Erklärte Varianz des 1. Faktors und Eigenwert des zweiten Faktors.

	Total	Alter			Geschlecht		Sprachregion	
		15-19	20-34	35+	Männlich	Weiblich	Deutsch	Lateinisch
Erklärte Varianz des ersten Faktors	40.728	33.617	37.779	33.515	37.249	44.426	40.983	40.076
Eigenwert des zweiten Faktors	1.070	1.216	1.207	1.341	1.147	1.047	1.026	1.208

Insgesamt deuten die graphischen Analysen für die gesamte Stichprobe, aber auch in Unterstichproben auf ein eindimensionales Konstrukt des CIUS hin. Zur Verdeutlichung zeigen wir den Scree Plot für den höchsten zweiten Eigenwert über alle Unterstichproben hinweg (Abbildung 1).

Abbildung 1: Scree-Plot für die Stichprobe der über 35-Jährigen mit dem höchsten 2. Eigenwert.



Es ist eindeutig, dass selbst in der im Hinblick auf das Eigenwertekriterium schlechtesten Substichprobe der über 34-Jährigen die Eigenwerte von der 14-Faktorenlösung bis zur 2-Faktorenlösung praktisch linear ansteigen und dass der Knick eindeutig auf eine eindimensionale Lösung hinweist. Dies bestätigt die bisherige Literatur zum CIUS, in der in der Regel eindimensionale Lösungen gefunden worden sind. Sowohl der *parallel test* nach Horn (1965) als auch Velicer's MAP Tests weisen für die Gesamtstichprobe aber auch für alle Unterstichproben eine eindimensionale Lösung aus.

Cronbachs Alpha ist ein Mass der Reliabilität und gibt die interne Konsistenz an. Folgende Richtwerte für Cronbachs Alpha werden gegeben (z.B. DeVellis, 2012):

- $\alpha \geq 0.9$ exzellent
- $0.9 > \alpha \geq 0.8$ gut
- $0.8 > \alpha \geq 0.7$ akzeptabel
- $0.7 > \alpha \geq 0.6$ fragwürdig
- $0.6 > \alpha \geq 0.5$ schlecht
- $0.5 > \alpha$ unakzeptabel

Die folgenden Alpha-Werte wurden in der Stichprobe ermittelt:

- Total: $\alpha = .887$
- 15-19 Jahre: $\alpha = .847$
- 20-34 Jahre: $\alpha = .872$
- 35+ Jahre: $\alpha = .842$
- deutsche Sprachregion: $\alpha = .888$
- lateinische Sprachregion: $\alpha = .884$
- Männer: $\alpha = .869$
- Frauen: $\alpha = .903$

2.1.1 Fazit

Insgesamt ist also der CIUS als eindimensionales Konstrukt mit guter, fast exzellenter interner Konsistenz anzusehen.

2.2 Kürzungspotential

Bei potentiellen Kürzungen ist zu bedenken, ob man dabei ganze Grundkriterien (theoretische Unterdimensionen) des CIUS kürzt. Nach Meerkerk (2007) bildet der CIUS Kriterien der Abhängigkeit ab. Diese sind:

- Konflikt: Item 3, Item 8, Item 10, Item 11
- Coping: Item 12, Item 13
- Kontrollverlust: Item 1, Item 2, Item 5, Item 9
- Vertieftsein (preoccupation): Item 4, Item 6, Item 7
- Entzugserscheinungen: Item 14

Wie bereits bei der Faktorenanalyse erklärt, korrelieren beispielsweise die Items 1 und 2 sehr hoch miteinander und haben auch die schlechteren MSA-Werte. Man könnte also eines dieser Items weglassen, da noch 3 weitere Items das Kriterium "Kontrollverlust" messen würden. Noch schlechter erwiesen sich die Kennwerte für die Items 12 und 13. Würde man beide Items in einer Kurzform nicht mehr berücksichtigen, so wäre das Kriterium "Coping" nicht mehr repräsentiert.

Es gibt zur Zeit zwei Kurzskalen: eine von Cartierre und Kollegen (2011) mit 9 Items und eine Kurzform von Bischof und Kollegen (2016) in Deutschland mit 5 Items (vgl. Tabelle 4).

- Cartierre: Item 1, Item 3, Item 4, Item 5, Item 7, Item 9, Item 11, Item 12, Item 14
- Bischof: Item 1, Item 3, Item 5, Item 11, Item 12 (bei einer 7-Itemlösung kämen noch die Items 7 und 9 dazu).

Man kann erkennen, dass bei Cartierre und Kollegen (2011), alle ursprünglichen Kriterien abgedeckt bleiben. Interessanterweise wurde in der Version auch jeweils ein Item unserer "verdächtigen Paare" (1-2; 8-9; 12-13) gekürzt. Anders sieht es bei Bischof und Kollegen (2016) aus. Deren Kurzversionen bestehen nur noch aus Items, die Konflikt, Coping und Kontrollverlust beinhalten, d.h. keines der Items der Kurzform repräsentiert noch die ursprünglichen Konzepte "Vertieftsein" und "Entzugserscheinungen".

Tabelle 4: Items des CIUS und der entsprechenden Kurzformen (französische und italienische Versionen im Anhang)

Item-nummer	Di-men-sion	Wortlaut der Frage	Voll-stän-dige Ver-sion von Meer-kerk	Kurz-form von Car-tierre et al.	Kurz-form von 5 (7) Items nach Bi-schof et al.
Item 1	KV	Wie häufig finden Sie es schwierig, mit dem Internetgebrauch aufzuhören, wenn Sie online sind?	X	X	X
Item 2	KV	Wie häufig setzen Sie Ihren Internetgebrauch fort, obwohl Sie eigentlich aufhören wollten?	X		
Item 3	K	Wie häufig sagen Ihnen andere Menschen (z.B. Partner, Kinder, Eltern, Freunde), dass Sie das Internet weniger nutzen sollten?	X	X	X
Item 4	V	Wie häufig bevorzugen Sie das Internet, statt Zeit mit anderen zu verbringen (z.B. Partner, Kinder, Eltern, Freunde)?	X	X	
Item 5	KV	Wie häufig schlafen Sie zu wenig wegen des Internets?	X	X	X
Item 6	V	Wie häufig denken Sie an das Internet, auch wenn Sie gerade nicht online sind?	X		
Item 7	V	Wie oft freuen Sie sich bereits auf Ihre nächste Internetsitzung?	X	X	(X)
Item 8	K	Wie häufig denken Sie darüber nach, dass Sie weniger Zeit im Internet verbringen sollten?	X		
Item 9	KV	Wie häufig haben Sie erfolglos versucht, weniger Zeit im Internet zu verbringen?	X	X	(X)
Item 10	K	Wie häufig erledigen Sie Ihre Aufgaben (zuhause oder auf der Arbeit) hastig, damit Sie früher ins Internet können?	X		
Item 11	K	Wie häufig vernachlässigen Sie Ihre Alltagsverpflichtungen (Arbeit, Schule, Familienleben), weil Sie lieber ins Internet gehen?	X	X	X
Item 12	C	Wie häufig gehen Sie ins Internet, wenn Sie sich niedergeschlagen fühlen	X	X	X
Item 13	C	Wie häufig nutzen Sie das Internet, um Ihren Sorgen zu entkommen oder um sich von einer negativen Stimmung zu entlasten?	X		
Item 14	E	Wie häufig fühlen Sie sich unruhig, frustriert oder gereizt, wenn Sie das Internet nicht nutzen können?	X	X	

Bemerkung: K = Konflikt, C=Coping, KV= Kontrollverlust, V=Vertieftsein, E=Entzugserscheinung

2.2.1 Itemschwierigkeiten und Itemtrennschärfen

Itemschwierigkeit bezeichnet bei dichotomen Items den Prozentsatz an Personen, welche die Kategorie "richtig" (bzw. "zutreffend", "ja" etc.) ankreuzen. Itemschwierigkeiten sollten zwischen 20 und 80 liegen, also nicht zu schwer oder nicht zu leicht sein (Bortz & Döring, 2005).

Bei mehrstufigen Items, wie im Falle des CIUS berechnet man die Schwierigkeit eines Items i mit p_i = Mittelwert des Items i / maximalen Wert*100. Im Falle des CIUS liegen die Itemschwierigkeiten

zwischen 3.8 (Item 10 bei über 35-Jährigen) und 38.0 (Item 1 bei 15- bis 19-Jährigen). Die Items sind im Prinzip also alle "schwer". Wir haben uns deshalb für eine strengeres potentielles Ausschlusskriterium entschieden, d.h. einem Kriterium mit einer Itemschwierigkeit < 15 (vgl. Tabelle 5).

Der Trennschärfe eines Items ist zu entnehmen, wie gut das gesamte Testergebnis aufgrund der Beantwortung eines einzelnen Items vorhersagbar ist (Bortz & Döring, 2005). Dabei wird als Daumenregel die Korrelation des Items mit der Summe der übrigen Items herangezogen (*corrected item-total correlation*). Als Daumenregel gelten Items mit einer Korrelation < 0.5 als ungenügend trennscharf. Trennschärfe ist eines der bedeutsamsten Kennwerte der klassischen Testtheorie.

Tabelle 5: Itemschwierigkeiten und Trennschärfen im der Gesamtstichprobe und nach Alter, Geschlecht und Sprachregion

		Items mit Schwierigkeit < 15	Items mit Trennschärfe < 0.5
Total	Total	3, 5, 9, 10, 11, 14	3, 6, 7, 14
Alter	15-19 Jahre	keines	3, 4, 6, 7, 9, 11, 13, 14 (indes 9, 11, 13 sind dicht an 0.5)
	20-34 Jahre	3, 5, 9, 10, 14	3, 5, 6, 14 (mit 5 und 6 dicht an 0.5)
	35+	alle ausser 1 und 2	3, 5, 6, 7, 10, 11, 14, (mit 7 dicht an 0.5)
Geschlecht	männlich	9, 10, 11, 14	3, 4, 6, 7, 9, 14 (7 und 14 dicht an 0.5)
	weiblich	3, 5, 9, 14	14
Sprachregion	deutsch	3, 5, 9, 10, 11, 14	3, 4, 6, 14 (6 und 14 dicht an 0.5)
	lateinisch	5, 9,	3, 5, 6, 7, 14 (3 und 5 dicht an 0.5)

2.2.2 Fazit

Es lässt sich sagen (Tabelle 3), dass insbesondere die Items 3, 6, und 14 (und eingeschränkt Item 7) konsistent schlechte Trennschärfen über die Substichproben aufweisen. Die Items 3 und 14 weisen zudem auch sehr hohe Schwierigkeiten auf. Zusätzlich sind aufgrund der Schwierigkeiten auch noch die Items 5, 9 und ggf. 10 kritisch zu betrachten.

2.3 Mögliche Kurzformen gemäss der klassischen Testtheorie

Aus den bisher besprochenen Ergebnissen ergeben sich mögliche Kandidaten für Kürzungen:

- Entweder Item 1 oder Item 2 (hohe Interkorrelationen und schlechteste, obwohl noch gute MSA). Da Item 1 sowohl in der Cartierre-Skala als auch in der Bischof-Skala enthalten ist, jedoch nicht Item 2, wählen wir Item 2 als Ausschlusskandidaten.
- Entweder Item 8 oder Item 9 (hohe Interkorrelationen und schlechteste (obwohl noch gute) MSA-Werte). Da Item 9 auch durch eine hohe Itemschwierigkeit auffällt, wählen wir Item 9.
- Entweder Item 12 oder Item 13 (hohe Interkorrelationen und schlechteste (obwohl noch gute) MSA). Da Item 12 in beiden Kurzformen enthalten ist und Item 13 zumindest bei der jüngsten Altersgruppe eine schlechte Trennschärfe aufweist, entscheiden wir uns für Item 13 als Ausschlussmöglichkeit.
- Item 3, 6 werden wegen schlechter Trennschärfen ausgeschlossen.
- Item 14 ist eine Ausschlussoption, ist aber das einzige Item für Entzugserscheinungen.
- Zu berücksichtigen sind zudem Item 7 (Trennschärfe) und Item 10 (Itemschwierigkeit).

Folgende Varianten werden über Korrelationen mit der Orginalsummenskala nach Meerkerk, der Cartierre-Skala sowie der Bischof-Skala verglichen.

1) Maximale Kürzung: Item 1, Item 4, Item 5, Item 8, Item11, Item12 (Ausschluss: Item 2, Item 3, Item 6, Item 7, Item 9, Item10, Item 13, Item 14). Diese Skala würde alle Kriterien ausser Entzugserscheinungen beinhalten.
2) Maximal Kürzung ohne Ausschluss von Entzugserscheinung: wie 1) aber inklusive Item 14.
3) Unklare Kürzung: wie 1 aber inklusive Item 7 und Item 10.
4) Unklare Kürzung ohne Ausschluss von Entzugserscheinung: wie 3 aber mit Item14.

In Tabelle 6 werden die Pearson Korrelationen der vier vorgeschlagenen Skalen mit dem originalen CIUS sowie der deutschen Kurzform und der Kurzform nach Cartierre et al. dargestellt.

Tabelle 6: Korrelationen der 4 Versionen mit der Original-Skala und der deutschen 5-Item-Skala und der Skala nach Cartièrre et al., gesamte Stichprobe.

	Original CIUS	Deutsche 5-Item Kurzform	Cartierre 9-Item Kurzform
Schweiz Version 1 (6 Items)	0.950	0.941	0.942
Schweiz Version 2 (=Version 1 plus Entzug, 7 Items)	0.961	0.932	0.961
Schweiz Version 3 (8 Items)	0.969	0.923	0.955
Schweiz Version 4 (= Version 3 plus Entzug; 9 Items)	0.975	0.915	0.968
Deutsche 5-Item Kurzform	0.930	1	0.946
Cartierre 9-Item Kurzform	0.975	0.946	1

Wie Tabelle 6 belegt, korrelieren alle Skalen (einschliesslich der deutschen Kurzform und der Kurzform nach Cartierre et al.) in der Gesamtstichprobe sehr hoch miteinander. Ähnlich hohe Korrelationen (> 0.9) werden auch bei beiden Geschlechtern, in den hier gewählten zwei Sprachregionen und in den drei hier gewählten Altersgruppen gefunden.

Die Originalskala von Meerkerk et al. (2007) ist extrem homogen, d.h., dass praktisch jede Auswahl von Items sehr hoch mit der Gesamtskala korreliert. Es zeigt sich, dass insbesondere die Anzahl an Items und nicht so sehr, welche Items konkret ausgewählt worden sind, die Korrelation mit der Gesamtskala bestimmen. Je mehr Items ausgeschlossen werden, desto geringer wird die Korrelation mit der Gesamtskala. Beispielsweise korreliert die von uns vorgeschlagene Skala mit 9 Items genauso hoch mit der Gesamtskala wie die 9-Itemskala nach Cartierre et al. (2011), obwohl teilweise unterschiedliche Items eingehen.

2.3.1 Fazit

Es kann Folgendes festgehalten werden: Unter der Annahme, dass der CIUS im Original ein valides Instrument für die Messung des problematischen oder abhängigen Internetgebrauches ist, so sind es auch die hier untersuchten Kurzformen. Die hohen Korrelationen mit der Originalskala zeigen, dass die Kurzformen sehr Ähnliches messen wie die Originalskala. Generell muss betont werden, dass es zur Zeit keinen "Goldstandard" gibt gegen den die Langform oder eine der Kurzformen getestet worden sind. Dies gilt aber generell auch für andere Skalen der Internetabhängigkeit.

3. Probabilistische Testtheorie und ordinale konfirmatorische Faktorenanalyse

In diesem Kapitel geht es darum, Modelle der probabilistischen Testtheorie heranzuziehen, um eine Itemselektion für eine Verkürzung der *Compulsive Internet Use Scale* (CIUS) auf unter 10 Items zu erreichen. Im Wesentlichen werden hierbei zwei Grundansätze gewählt: a) die konfirmatorische, ordinale Faktorenanalyse und b) Modelle der *item response theory* (IRT), nämlich "*graduated response models*", die ebenfalls bei mehrstufigen ordinalen Items verwendet werden.

3.1 Konfirmatorische Faktorenanalyse und Messinvarianz

Ein Hauptaugenmerk für die Itemselektion bei der konfirmatorischen Faktorenanalyse (CFA) liegt, neben der Modellanpassung, auf stark korrelierenden Fehlern der Items, da diese Korrelationen auf redundante Informationen hinweisen. Merkeerk (2007) fand z.B., dass sich nur dann eine befriedigende Skala finden lässt, wenn man die fünf Itempaare (bzw. deren jeweilige Fehler) 1 und 2, 6 und 7, 8 und 9, 10 und 11 sowie 12 und 13 korrelieren lässt. Diese Fehlerkorrelationen bei Meerkerk stellten den Ausgangspunkt der 9-Itemversion nach Cartierre et al. (2011) dar, die jeweils eines der Items aus den fünf Paaren wegliessen, wobei unklar blieb, warum welches Item aus den jeweiligen Paaren weggelassen worden ist. Ausserdem wurde von Cartierre und Kollegen (2011) die Skala nur an Jugendlichen der 4. Gymnasialklasse (Durchschnittsalter 13.8 Jahre) nicht jedoch in der Allgemeinbevölkerung psychometrisch getestet. Hohe Fehlerkorrelationen wurden auch von Khazaal und Kollegen (2012) festgestellt. Sie betrafen die Itempaare 1 und 2 sowie 12 und 13.

Ein weiteres Kriterium für den Ausschluss von Items bilden geringe Trennschärfen (Ladungen) der Items auf dem Faktor.

3.1.1 Invarianztestung über konfirmatorische Faktorenanalysen

Mittels multipler Gruppenvergleiche wird festgestellt, inwieweit die finale Lösung der Kurzform weitestgehend invariant über verschiedene Bevölkerungsgruppen, d.h. über Sprachregionen (deutsch- versus lateinischsprachig (italienisch und französisch)), Alter (15-19, 20-34, 35+) und Geschlecht, sind. Bei der Invarianztestung werden in steigender Folge stärkere Restriktionen ins Model zur Prüfung a) der konfiguralen Invarianz (gleiche Faktorzahl über die Gruppen), b) der metrischen Invarianz (*weak invariance*, gleiche Trennschärfen (Ladungen) auf den Faktoren) und c) der skalaren oder starken Messinvarianz (gleiche Itemschwierigkeiten (Schwellenwerte)) eingeführt. Die Güte der Modellanpassung kann über den *comparative fit index* (CFI; Daumenregel: Werte zwischen .90 und .95 sind akzeptabel, Werte über 0.95 sind gut) und den *root mean square error of approximation* (RMSEA: Werte zwischen 0.08 und 0.05 sind akzeptabel, Werte kleiner 0.05 sind gut) eingeschätzt werden.

Unterschiede zwischen diesen hierarchisch geschachtelten Modellen können über $\chi 2$-Verfahren getestet werden, wobei diese sehr stark auf hohe Stichproben reagieren, d.h. häufig signifikante Unterschiede ausweisen, obwohl die Unterschiede praktisch kaum relevant sind. Deshalb werden zusätzlich noch die Unterschiede zwischen den Modellanpassungen über Differenzen in den RMSEA- und CFI-Werten, also ΔRMSEA und ΔCFI, herangezogen. Signifikante Unterschiede beim $\chi 2$-Test gelten dann als praktisch unbedeutend, wenn ΔRMSEA und ΔCFI < 0.01 (Elosua, 2011) bzw. < 0.03 sind (< 0.03 für schwache /metrische und < 0.01 für starke/skalare Messinvarianz vgl. Chen, 2007). Die verschiedenen CFAs wurden in Mplus berechnet, wobei von ordinalem Datenniveau ausgegangen wurde.

3.2 IRT mit graded response models (GRM)

Im Rahmen der probabilistischen Testtheorie werden *graded response models* (GRM) eingesetzt. GRM stellen die Erweiterung von *item response models* mit dichotomen Items auf ordinale Skalen mit mehr als zwei Ausprägungen (der CIUS besteht aus fünfstufigen Antwortskalen) dar. Diese Modelle liefern, Eindimensionalität vorausgesetzt, ebenso Kennwerte der Itemschwierigkeit (*threshold parameters/*Schwellenwerte*)* und der Trennschärfe (*discrimination parameter*) wie in der ordinalen konfirmatorischen Faktorenanalyse. Erwünscht sind Items mit Schwierigkeiten (*thresholds*) über die gesamte Skalenbreite hinweg bei möglichst hohem Diskriminationspotenzial (hoher Trennschärfe). Gemäss Baker (2001) gilt folgende Daumenregel für die Trennschärfe:

- keine: 0
- sehr gering: 0.01 - 0.34
- gering: 0.35 - 0.64
- moderat: 0.65 - 1.34
- hoch: 1.35 - 1.69
- sehr hoch: > 1.70

Diese Kennwerte, insbesondere jene der Trennschärfe, liefern ein weiteres Kriterium für den Ausschluss von Items in der Kurzform. Probabilistischen Verfahren liefern mit der *item information function* (IIF) sowie der *test information function* (TIF) und der *test characteristic curve* (TCC) interessante Abbildungen. Insbesondere die letzten beiden können zur Einschätzung herangezogen werden, wieviel der Gesamttest an Information verliert, wenn einzelne Items weggelassen werden. Diese Funktionen stellen also ein weiteres Kriterium dafür dar, um welche Items die vollständige Form des CIUS reduziert werden kann, ohne dabei einen zu grossen Informationsverlust einzugehen (An & Yung, 2014).

Probabilistische Testmodelle wie das GRM sind ideal, um sog. *differential item functioning* (DIF) bei der Validierung von Skalen zu ermitteln. Die Ermittlung von DIFs dient der Findung von Items, die in verschiedenen Gruppen (z.B. Sprachregionen) unterschiedlich "funktionieren", also unterschiedlich zu den Gesamtskalenwerten beitragen. DIFs ermitteln Gruppenunterschiede, nachdem für Unterschiede zwischen Gruppen auf dem Konstrukt (hier kompulsive Internetnutzung) kontrolliert worden ist. Die Kontrolle des Gesamttestwerts ist wichtig, da sich ja beispielsweise Romands von Deutschschweizern in ihrer kompulsiven Internetnutzung unterscheiden können, was sich natürlich auch auf Unterschiede bei einzelnen Items auswirken würde. Es sollen mit DIFs aber nicht Unterschiede zwischen Gruppen gefunden werden, sondern unterschiedliche Wirkungsweisen der Items auf die Gesamtskala. Es werden also die Auswirkung einzelner Items untersucht, wobei der jeweilige Gesamttestwert bereinigt wird. DIFs messen somit unabhängig vom Gesamttestwert, ob sich Itemschwierigkeiten und Trennschärfen unterschiedlich auf den Gesamttestwert auswirken, also sich im Prinzip die Gesamtskala unterschiedlich zusammensetzt. DIFs stellen somit ein weiteres Kriterium für die Itemselektion im Hinblick auf eine Kurzform dar, die gesamtschweizerisch valide sein soll.

3.3 Differential Item Functioning (DIF)

Beim DIF unterscheidet man, ob dieses gleichförmig (uniform) oder ungleichförmig ist. Gleichförmig bedeutet, dass das Diskriminationspotential des Items in den zu vergleichenden Gruppen gleich ist, aber die Schwellenwerte (Itemschwierigkeiten, *thresholds*) verschieden sein können. Ungleiche Schwellenwerte bedeuten, dass das Ankreuzen desselben Itemwertes in der einen Gruppe eine stärkere Ausprägung auf dem latenten Konstrukt zur Folge hat als in der anderen Gruppe. DAS bedeutet beispielsweise auf den CIUS bezogen, wenn die eine Gruppe (z.B. Männer) auf dem Item xy eine 3 ankreuzt dies für einen stärkeren/schwächeren kompulsiver Internetgebrauch (dem Gesamttest des CIUS) hindeutet als wenn die andere Gruppe (im Beispiel Frauen) eine 3 auf demselben Item ankreuzt.

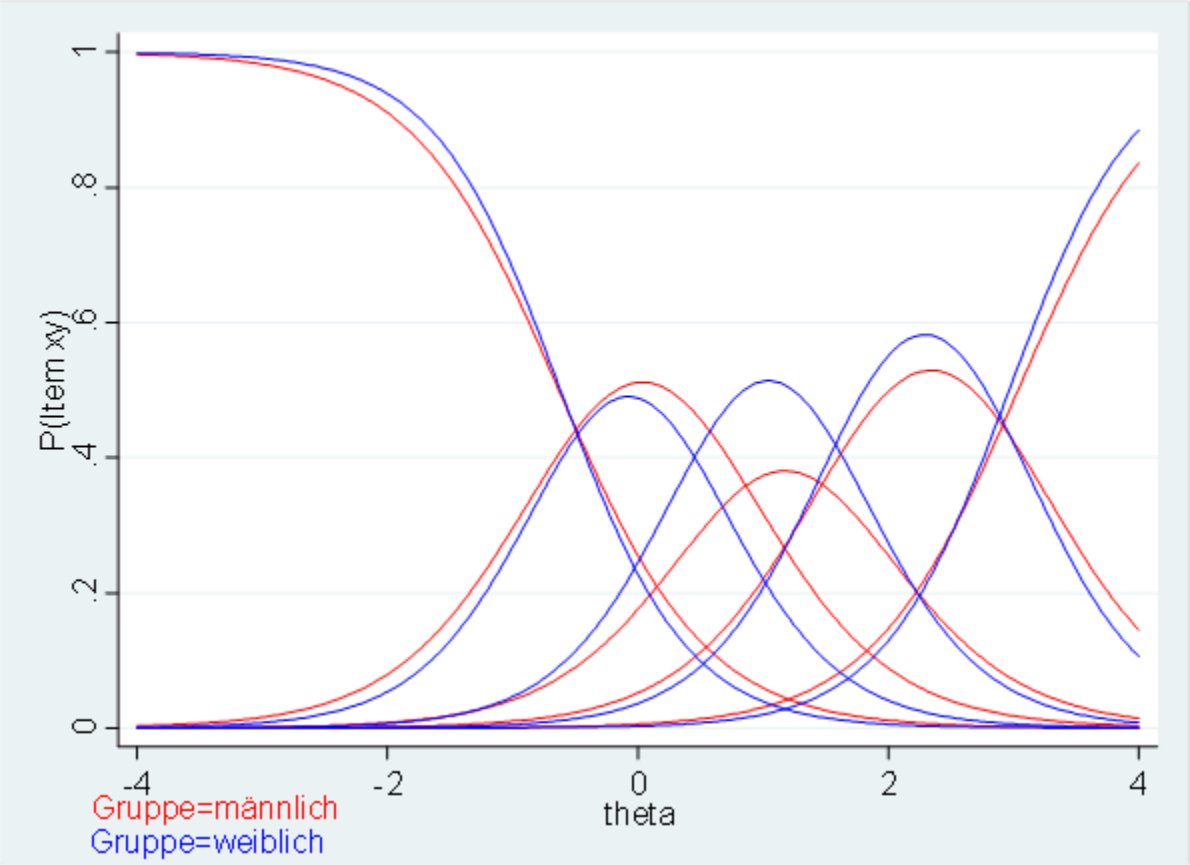
Ein Beispiel (vgl. Abbildung 2) aus der IRT-Analyse soll das verdeutlichen (vgl. Cohen et al., 1993). Auf der X-Achse (kompulsiver Gebrauch) findet man die Ausprägung des latenten (geschätzten) Konstruktes. Die b-Parameter (Schwierigkeit bzw. Schwellenwert) geben an, wo sich im Durchschnitt Personen auf dem Konstrukt "kompulsiver Internetgebrauch" befinden, wenn sie eine bestimmte Kategorie ankreuzen. Die y-Achse gibt die geschätzte Wahrscheinlichkeit, welche Ausprägung die Befragten auf dem Item wählen. Befindet man sich also ganz links im negativen Bereich, geben die Personen mit hoher Wahrscheinlichkeit die geringste Ausprägung des Items an, also die 1 (Range des Items fünfstufig von 1 bis 5). An den Schnittstellen der Kurven, wird jeweils die nächsthöhere Itemausprägung wahrscheinlicher. Dies bedeutet also im Beispiel, dass bei einem Wert von theta = -0.63 (Ausprägung auf dem latenten Konstrukts "kompulsiver Internetgebrauch) die Personen eher die Itemausprägung 2 als die Itemausprägung 1 ankreuzen (vgl. Tabelle 7 für die entsprechenden Werte, die der Abbildung zugrunde liegen). Bei einem Wert (theta) von etwa 1.6 auf dem latenten Konstrukt geben sie eher die Itemausprägung 4 als die Itemausprägung 3 an. Schliesslich kreuzen Personen bei einem Wert (theta) von etwa 3 auf dem latenten Konstrukt eher (d.h. wahrscheinlicher) den Skalenwert 5 des Items als den Skalenwert 4 an. Wie man erkennt kreuzen sich die Linien zwischen 1 und 2, 3 und 4 und 4 und 5 bei Männern und Frauen in etwa auf dem gleichen Niveau des latenten Konstruktes. Ein Unterschied ist jedoch zwischen der Ankreuzmöglichkeit 2 und 3 zu sehen. Bei Frauen liegt der Schnittpunkt weiter "links" als bei Männern (vgl. Tabelle 5: b2 bei Frauen = 0.4684; bei Männern = 0.7015). Dies bedeutet, dass Frauen bei diesem Beispielitem bei gleich starkem problematischem Internetgebrauch eher bereit sind schon eine 3 als eine 2 anzukreuzen als Männer. Das DIF ist umso grösser je mehr diese Schnittpunkte voneinander abweichen und insbesondere dann, wenn sie über alle Ausprägungen abweichen. Dies nennt man gleichförmige (uniform) DIF.

Ein höherer Trennschärfeparameter (Diskriminationsparameter) in der einen Gruppe im Vergleich zur anderen Gruppe bedeutet, dass das Item besser zwischen nicht-kompulsiven und kompulsiven Internetgebrauchenden unterscheidet. Dies drückt sich graphisch dadurch aus, dass die Kurven im Durchschnitt höher sind, also die Personen mit höherer Wahrscheinlichkeit Ausprägungen wählen. Im Beispiel liegen die Kurven der Frauen (Diskrimination = 1.97) höher als bei den Männern (1.69), insbesondere bei den hohen Itemausprägungen (rechte Seite der Abbildung 2). Man sieht insbesondere bei den Männern, dass die Itemausprägung 3 nur sehr schlecht zwischen den Ausprägungen 2 und 4 diskriminiert (also nur einen kleinen Bereich auf dem latenten Konstrukt abdeckt und dieses mit geringer Wahrscheinlichkeit).

Tabelle 7: IRT-Parameter für das Beispielitem (Abbildung 2)

Parameter	Männlich	Weiblich
Discrim.	1.6949	1.9749
b1	-0.6307	-0.6180
b2	0.7015	0.4684
b3	1.6457	1.6177
b4	3.0357	2.9641

Abbildung 2: Item information function eines Beispielitems mit differential item functioning nach Geschlecht



3.3.1 Tests für differential item functioning

Es gibt verschiedene Möglichkeiten DIF zu testen. Zumbo (1999) schlägt vor, drei Regressionsmodelle zu berechnen (eine entsprechende Software wurde z.B. von Choi et al., 2011; Crane et al., 2006, entwickelt):

(1) $Y = b_0 + b_1 * Tot$
(2) $Y = b_0 + b_{1\#} * Tot + b_2 * Gruppe$
(3) $Y = b_0 + b_1 * Tot + b_2 * Gruppe + b_3 * (Gruppe * Tot)$

Wobei "Tot" den Summenscore der Skala repräsentiert und "Gruppe" die Variable darstellt, welche die Gruppen bezeichnet (z.B. Männer und Frauen), für die die DIFs ermittelt werden sollen. Zumbo (1999) erweiterte den Ansatz von Zumbo und Thomas (1997) von dichotomen und kontinuierlichen Variablen auf ordinale Variablen, wobei *ordered logit* Regressionen berechnet werden. Die zweite Regression untersucht im Vergleich mit Model 1 uniforme DIFs, wogegen die dritte Regression im Vergleich mit Model 2 nicht-uniforme DIFs untersucht. Model 3 im Vergleich mit Model 1 untersucht gleichzeitig uniforme und nicht-uniforme DIFs. Auf statistische Signifikanz lassen sich die Modellunterschiede über *likelihood ratio* Tests prüfen. Diese *likelihoods* sind $\chi 2$ verteilt und deren *ratios* stellen $\chi 2$-Differenzen dar, die ebenfalls $\chi 2$-verteilt sind.

Da χ2-Tests (*likelihood ratio* Tests) jedoch wieder stark durch grosse Stichprobenumfänge beeinflusst werden, sind verschiedene zusätzliche Effektstärken vorgeschlagen worden. Eine Effektstärke für uniforme DIFs ist die prozentuale Änderung in den Regressionskoeffizienten b_1 zwischen Model 1 und Model 2, also $I(b_1 - b_{1\#})/ b_{1\#}I$, wobei eine Änderung um 10% für praktisch bedeutsam erachtet wird (Crane et al., 2006). Zumbo (1999) schlägt vor, den Unterschied zwischen den R^2-Werten, also ΔR^2 als Teststärke heranzuziehen, wobei der pseudo-R^2 nach McKelvey und Zavoina (1975) verwendet werden sollte, weil er am ehesten als Varianzaufklärung im Sinne des R^2 in multiplen linearen Regression zu interpretieren ist (Latila, 1993; Thomas et al., 2008).

Zumbo and Thomas (1997) schlagen folgende Werte als Effektstärken für DIF Items vor:

- Type A Items—vernachlässigbares DIF: $\Delta R^2 < 0.13$.
- Type B Items—moderates DIF: $0.13 \leq \Delta R2 \leq 0.26$.
- Type C Items—grosses DIF: $\Delta R2 > 0.26$.

Dabei sollten zusätzlich der likelihood ratio (LR) Tests bei den letzten beiden Typen B und C signifikant sein. Allerdings wurden diese von Zumbo and Thomas (1997) vorgeschlagenen Schwellenwerte als zu grosszügig angesehen (Hidalgo & López-Pina, 2004). Strengere Kriterien wurden von Jodoin and Gierl (2001) vorgeschlagen:

- Type A Items—vernachlässigbares DIF: $\Delta R2 < 0.035$
- Type B Items—moderates DIF: $0.035 \leq \Delta R2 \leq 0.070$,
- Type C Items—grosses DIF: $\Delta R2 > 0.070$,

wobei wiederum bei den letzten beiden Itemtypen der LR-Test signifikant sein sollte.

Methoden der (eindimensionalen) IRT berechnen DIF direkt basierend auf IRT Modellen. Dabei wird eine Gruppe als Fokusgruppe angesehen und eine Gruppe als Referenzgruppe. Der statistische Test wird über LR-Test durchgeführt (χ2-Differenz, die sich als gute Absicherung über Typ I Fehler gezeigt haben (Suh & Cho, 2014)). Für jedes Item wird ein χ2-Differenz berechnet, wobei zwei hierarchisch geschachtelte Modelle berechnet werden. Im ersten Modell werden die Parameter (Trennschärfe und Schwierigkeit) in beiden Gruppen gleichgesetzt (*compact model*) und gegen ein Modell getestet, in dem sie ungleich sind (*augmented model*, siehe Suh, 2016).

Im vorliegenden Bericht wurden die Testungen mit dem UIRT-Befehl (unidimensional IRT) in Stata durchgeführt. Dieses Programm liefert auch Effektstärken nach Wainer (1993), wobei Items als DIF betrachtet werden, wenn die Effektstärke grösser als 0.05 (Suh, 2016) oder grösser als 0.10 (Kim et al., 2007) ist. Erstere dürfte vermutlich zu schnell Items als DIF identifizieren, letztere etwas zu konservativ sein. Das Programm weist zwei Effektstärken aus, leicht variieren, je nachdem welche Gruppe als Fokus- bzw. Referenzgruppe angenommen wird. Dieses Prinzip der Invarianztestung wird als ideale Testung im Rahmen von IRT Modellen gesehen (Kim et al., 2007).

Als dritte Variante werden hier Tests der ordinalen konfirmatorischen Faktorenanalyse vorgeschlagen, die ähnlich wie im IRT-DIF-Test die geschätzten Parameter als in den Gruppen gleich fixieren und gegen ein Model mit freien Parametern testen. Diese Methode zur Testung einzelner Items, also nicht gesamthaft für alle Items gleichzeitig wie bei der Testung der starken und schwachen Invarianz, stellen Elousa (2011) und Stark, Chernyshenko, and Drasgow (2006) dar. Im "*free baseline approach*" werden zunächst die Parameter (Trennschärfe und Schwellenwerte) zwischen den Gruppen frei geschätzt, und dann für das entsprechende Item, für das DIF getestet werden soll, gleichgesetzt (*constrained*). Der Unterschied dieser beiden genesteten Modelle kann über LR-Tests (χ2-Differenzen) auf Signifikanz getestet werden, bzw. können ΔCFI und ΔRMSEA-Werte herangezogen werden, da die LR-Tests stark auf grosse Fallzahlen reagieren und zu schnell praktisch irrelevante DIFs identifizieren.

Eine weitere Methode ist die Testung der Gleichheit von Faktorladungen (Trennschärfen) in der konfirmatorischen Faktoranalyse (Meade & Bauer, 2007). Gleiches ist ebenfalls für die den Diskriminationsparameter in der IRT möglich. Prinzipiell wäre das auch für die Schwierigkeits- oder Schwellenwertparameter denkbar, es besteht aber Konsens, dass die Testung der Ladungen (Trennschärfe) die wichtigere Testung auf Messinvarianz ist (Meade & Kroustalis, 2006). Unter Annahme der Nullhypothese (Gleichheit der Ladungen) ist dieser Test verhältnismässig einfach zu berechnen:

$$CI_{\hat{\lambda}_2-\hat{\lambda}_2} = |\hat{\lambda}_2 - \hat{\lambda}_1| \pm z_{crit} \times SE_{(\hat{\lambda}_2-\hat{\lambda}_1)}$$

,

wobei

$$SE_{(\hat{\lambda}_2-\hat{\lambda}_1)} = \sqrt{SE(\hat{\lambda}_1)^2 + SE(\hat{\lambda}_2)^2}$$

,

und z_{crit} den kritischen Wert eine Standardnormalverteilung darstellt. Man würde also bei einem kritischen Wert von 1.96 man das 95%ige Konfidenzintervall erhalten.

3.4 Analytische Vorgehensweise zur Findung einer Kurzform

Im ersten Schritt wurde ein IRT-Modell mit allen 14 CIUS-Items gerechnet, um zu sehen, ob Trennschärfen einzelner Items zu gering sind. Im zweiten Schritt wurde mit allen Items eine CFA berechnet und entsprechende Modifikationsindizes herangezogen, die auf hohe Fehlerkorrelationen und damit Redundanzen von Items hinweisen. Letzteres stellte den Ansatz von Cartierre und Kollegen (2011) für die Itemreduktion in ihrer Kurzform dar, wobei die Autoren nicht angaben, warum sie welches Item des jeweiligen korrelierenden Itempaares eliminiert haben. Deshalb werden im dritten Schritt verschiedene DIF-Analysen herangezogen, um zu entscheiden, welches Item des jeweiligen hochkorrelierenden Itempaares ausgeschlossen werden sollte. Anschliessend werden die Schritte mit der reduzierten Skala wiederholt. Dieses iterative Verfahren nennt man "purification" (Hidalgo-Montesinos & Gómez-Benito, 2003; Zumbo, 1999).

Anschliessend wurde die gekürzte Skala nochmals insgesamt über CFA auf Invarianz getestet, wobei schwache Invarianz (metrische Invarianz) die Gleichheit der Ladungen/Trennschärfen annimmt und starke Invarianz zusätzlich Gleichheit der Schwellenwerte/Schwierigkeiten annimmt. Die schwache Invarianz wird gegen ein Modell mit freien Parametern zwischen den Gruppen (jedoch gleicher Faktorenzahl also konfiguraler Invarianz) getestet. Die starke Invarianz wird gegen das Modell mit schwacher Invarianz getestet. Im letzten Schritt wird die Kurzform der Skala in der Validierungsstichprobe überprüft. Da der CIUS in allen Untergruppen eindimensional ist, wird die konfigurale Invarianz nicht getestet sondern nur als Ausgangsmodel für die Testung der metrischen und skalaren Invarianz herangezogen.

3.5 Ergebnisse

3.5.1 Trennschärfen nach IRT

Die IRT-Analyse über 14 Items ergab hohe Trennschärfen (> 1.34) für alle Items, wobei die geringsten Trennschärfen für die Items 3, 4, 6, 7, und 14 festgestellt werden konnten (Tabelle 8).

Tabelle 8: Trennschärfen der item response theory mit graduated response model

Item	Trennschärfe
Item 1	1.938
Item 2	1.940
Item 3	1.476
Item 4	1.416
Item 5	1.640
Item 6	1.345
Item 7	1.371
Item 8	2.026
Item 9	2.009
Item 10	2.074
Item 11	1.771
Item 12	1.911
Item 13	1.820
Item 14	1.380

3.5.2 Modellanpassung der CFA

Die konfirmatorische Faktorenanalyse über alle 14 Items ergibt einen RMSEA von 0.107 und einen CFI von 0.922. Das sind zwar keine schlechten Werte jedoch für das CFI gerade noch akzeptabel (Werte zwischen 0.90 und 0.95 wären akzeptabel). Nicht mehr akzeptabel ist das Modell im Hinblick auf den RMSEA (akzeptabel wären Werte zwischen 0.08 und 0.05).

Die Modifikationsindizes der konfirmatorischen Faktorenanalyse weisen die höchsten Werte für Korrelationen zwischen den Items 12 und 13 (M.I. = 643.5), den Items 1 und 2 (M.I. = 242.3), den Items 8 und 9 (M.I. = 193.4) und den Items 6 und 7 (M.I. = 93.7) auf. Modifikationsindizes > 10 gelten als beachtenswert (es ist die Voreinstellung in Mplus). Lässt man die Korrelation zwischen diesen Items zu, so erhöht sich der CFI auf 0.988 und der RMSEA sinkt auf 0.044, was gute bis ausgezeichnete Modelanpassungen wären. Alle 4 Itempaare wurden auch in den ursprünglichen Analysen von (Meerkerk, 2007) erkannt und dienten deswegen auch als potenzielle Kürzungen für Cartierre und Kollegen (2011), wobei diese nicht angaben, warum sie welches Item des jeweiligen Paares ausschlossen. Die Paare mit den beiden höchsten M.I.-Werten wurden auch von Khazaal und Kollegen (2012) identifiziert.

3.5.3 DIF Analyse für vier Itempaare mit hoher Iteminterkorrelation

Um dasjenige Item eines Paares mit dem grösseren DIF auszuwählen, wurden Gruppenvergleiche über die oben beschriebenen Methoden für diese Itempaare durchgeführt. Die DIF-Analysen wurden für Geschlecht, Sprachregion und Alter berechnet (Tabelle 9). Da fast alle Verfahren nur jeweils 2 Gruppen vergleichen, wurden die französisch- und die Italienischsprachige Schweiz zusammengelegt und gegen die deutschsprachige Schweiz getestet. Die italienischsprachige Schweiz hat alleine betrachtet auch zu geringe Fallzahlen. Beim Alter wurden die 15- bis 19-Jährigen mit den 20- bis 34-Jährigen verglichen, da diese in etwa gleich grosse Fallzahlen für Personen mit problematischem Internetgebrauch nach der vollständigen CIUS Version mit 28 Punkten aufweisen (vgl. Marmet et al., 2013). Ab 35 Jahren liegt praktisch kaum noch problematischer (kompulsiver) Internetgebrauch vor.

Tabelle 9: Vergleich der verschiedenen DIF Kriterien für die Itempaare 1 und 2, 6 und 7, 8 und 9 und 12 und 13.

			Item 1	Item 2	Item 6	Item 7	Item 8	Item 9	Item 12	Item 13
Geschlecht	UIRT	LR-Test		*	*		**	***	*	***
		Ladungen/Trennschärfen		*			***	***		
	Zambo	ΔR^2 uniform								*
		ΔR^2 non-uniform								
		ΔR^2 total								*
		LR uniform	*		**			**	**	***
		LR non-uniform					***	**		
		LR total			*		***	***	**	***
		Δb (uniform)								*
	CFA	ΔRMSEA								
		ΔCFI								
		$\Delta\chi 2$	9.66	23.05	18.37	4.6	42.21	19.37#	11.85	19.34
		Ladungen/Trennschärfen		**			***	*		
Sprache	UIRT	LR-Test	*	***	*	*	**		***	
		Ladungen/Trennschärfen								
	Zambo	ΔR^2 uniform		*						
		ΔR^2 non-uniform								
		ΔR^2 total		*						
		LR uniform		***	*	**	**		***	
		LR non-uniform		**						
		LR total		***		*	**		***	
		Δb (uniform)		**	*	*				
	CFA	ΔRMSEA								
		ΔCFI								
		$\Delta\chi 2$	15.95	42.53	11.94	11.98	5.07	10.53#	6.05	9.38
		Ladungen/Trennschärfen								
Alter	UIRT	LR-Test			*		*			
		Ladungen/Trennschärfen			*	*				
	Zambo	ΔR^2 uniform								
		ΔR^2 non-uniform	**	*						
		ΔR^2 total	**	*						
		LR uniform			*					
		LR non-uniform			*	*				
		LR total			**	*				
		Δb (uniform)	*		***	*		*		
	CFA	ΔRMSEA								
		ΔCFI								
		$\Delta\chi 2$	13.52	35.73	11.76	13.65	5.36	6.85#	8.09	9.96

Kriterien: Für LR, Ladungen/Trennschärfen: * p < 0.05, ** p < 0.01, *** p < 0.001; für ΔCFI, ΔRMSEA: * > 0.02; ** > 0.01, für ΔR2: * > 0.01, ** > 0.035, *** > 0.07; für Δb * >1%, **> 2%; *** >3%; für $\Delta\chi 2$ werden Werte statt Signifikanzniveaus ausgewiesen, da fast alle hochsignifikant sind, jedoch Unterschiede in den Werten zusätzliche Hinweise liefern. Signifikanzniveaus bei 5 Freiheitsgraden wären 11.07 (p < 0.05), 15.09 (p<0.01) und 20.52 (p < 0.01). Bei Item 9 wurde der letzte Schwellenwert mit dem davorliegenden zusammengelegt, da die Ausprägung so selten war, dass sie nicht in allen Gruppen vorkam; die entsprechenden Signifikanzniveaus bei 4 Freiheitsgraden lägen bei 9.49; 13.28 und 14.47

Insgesamt ist festzuhalten, dass bei allen Items das DIF nicht sehr stark ausgeprägt ist. LR-Tests (χ2-Tests) gelten bei den Fallzahlen als zu progressiv. ΔRMSEA und ΔCFI weisen in keinem Fall auf DIF hin und auch bei den Effektstärken der Regressionsanalysen nach Zumbo (1999) haben wir noch schärfere Kriterien festgelegt als in der Literatur vorgeschlagen worden sind, so dass sich auch mit diesem Test kaum Belege für DIF finden lassen. Da aber wegen der Korrelationen zwischen den Items auf jeden Fall ein Item aus jedem Paar ausgeschlossen werden soll, haben wir strengere Kriterien angelegt, um eine Entscheidung herbeiführen zu können. Diese Entscheidung fällt für den Vergleich der Items 1 und 2 relativ klar aus: Item 2 zeigt etwas stärkeres DIF. Beim Itempaar 6 und 7 spricht das stärkere DIF zwischen Männern und Frauen für den Ausschluss des Items 6 und beim Itempaar 8 und 9 das stärkere DIF bei der Sprachregion für den Ausschluss des Items 8. Schwierig ist die Entscheidung zwischen Item 12 und Item 13. Item 13 zeigt ein stärkeres DIF beim Vergleich der beiden Geschlechter, Item 12 jedoch beim Vergleich der Sprachregionen, wobei dies im Wesentlichen für den Regressionsvergleich nach Zumbo (1999) gilt und sich UIRT und CFA widersprechen. Da das Item 12, jedoch nicht das Item 13, auch in den Kurzformen nach Cartierre et al. (2011) und dem deutschen PIEK (Bischof et al., 2016) enthalten ist, haben wir uns für den Ausschluss von Item 13 entschieden. Bei Cartierre et al. (2011) und bei Bischof et al. (2016) wurden ebenfalls die Items 2, 6 und 8 ausgeschlossen.

3.5.4 Testung der 10-Itemlösung mit CFA und IRT

Diese 10-Itemvariante wurde erneut einer CFA unterzogen und ergab einen RMSEA von 0.51 und CFI von 0.983 also einer guten Modellanpassung. Die Modifikationsindizes geben den höchsten Wert für eine Korrelation zwischen Item 3 und Item 9 an (M.I. = 39.19). Die Invarianztestungen in der CFA beziehungsweise für DIF im UIRT zeigen Folgendes (vgl. Tabelle 10):

- Geschlechtsunterschiede: Beim Ausschluss von Item 3 wird der RMSEA für die metrische Invarianz im Vergleich zum Model der konfiguralen Invarianz besser, der CFI bleibt gleich. Im Modell der starken Invarianz verbessert sich der RMSEA jedoch wird der CFI deutlich (um 0.017) schlechter. Die Effektstärke für das DIF im UIRT liegt bei >0.10.
  Beim Ausschluss von Item 9 verbessert sich der RMSEA und auch leicht der CFI beim Modell der schwachen Invarianz. Auch bei der starken Invarianz verbessert sich weiterhin der RMSEA und der CFI verschlechtert sich weniger. Die Effektstärken für das DIF liegen mit 0.15 und 0.14 höher als beim Item 3. Insgesamt gibt es leichte aber unbedeutende Vorteile für die Beibehaltung von Item 3 bzw. für den Ausschluss von Item 9.
- Regionale Unterschiede: Sowohl beim Ausschluss von Item 3 als auch beim Ausschluss von Item 9 behielte die Skala metrische Invarianz (sogar eine Verbesserung des CFI und RMSEA) jedoch besteht keine starke Invarianz, wobei ohne Item 3 (also mit Item 9) der CFI deutlich schlechter wird als umgekehrt. Die Effektstärken des DIF beim UIRT sind für beide Variablen in etwa gleich. Keine klare Entscheidung für den Ausschluss des einen oder anderen Items.
- Altersunterschiede: Bei Ausschluss von Item 3 verbessert sich der RMSEA um 0.01 sogar bei starker Invarianz, der CFI wird nur geringfügig schlechter. Bei Ausschluss von Item 9 verschlechterte sich sowohl der RMSEA als auch der CFI, wobei letzteres um mehr als 0.02 sinken würde. Item 3 zeigt im UIRT eine extreme Effektstärke von -0.40 bzw. 0.30, wobei das im UIRT Modell hoch nichtsignifikant ist (p = 0.92). Dies spricht eindeutig für den Ausschluss von Item 3 im Vergleich zu Item 9.

Zusammenfassend ist die Beibehaltung von Item 9 im Vergleich zu Item 3 zu bevorzugen. Item 9 zeigt auch die grössere Trennschärfe (siehe Tabelle 8).

Tabelle 10: UIRT DIF-Analyse und CFA-Testung auf Invarianz Entscheidung für Item 3 oder Item 9

		CFA						UIRT					
		ohne Item 3			ohne Item 9			Item 3			Item 9		
		RMSEA	CFI	χ2-Differenz / p-Wert*	RMSEA	CFI	χ2-Differenz /p-Wert*	p-Wert des χ2-Tests	Effektstärke (Ref: Gruppe 1)	Effektstärke (Ref. Gruppe 2)	p-Wert des χ2-Tests	Effektstärke (Ref Gruppe 1)	Effektstärke (Ref. Gruppe 2)
	total	0.048	0.987		0.041	0.990							
Geschlecht	konfigural	0.044	0.99		0.041	0.991		0.0173	-0.1092	0.1044	0.0002	0.1493	-0.1352
	metrisch	0.041	0.99	0.019	0.036	0.992	0.073						
	skalar	0.04	0.982	81.0	0.033	0.989	58.2						
Sprache	konfigural	0.048	0.987		0.044	0.988		0.2323	0.0027	-0.0024	0.0847	-0.017	0.0231
	metrisch	0.042	0.988	0.028	0.039	0.990	0.079						
	skalar	0.052	0.967	131.7	0.047	0.977	120.9						
Alter	konfigural	0.05	0.981		0.042	0.986		<0.0001	-0.3989	0.2965	0.9233	-0.0466	0.0336
	metrisch	0.047	0.981	0.0169	0.042	0.984	0.0199						
	skalar	0.04	0.979	48.9	0.05	0.963	106.6						

*Bei der skalaren (starken) Invarianz werden χ2 Werte angegeben, da alle Tests hochsignifikant ($p < 0.001$) sind, so dass die Unterschiede in den χ2 Werten mehr Information geben als die Signifikanzniveaus

3.6 9-Itemlösung einer CIUS-Kurzform

Bei einer 9-Item-Kurzform (Items 1, 4, 5, 7, 9, 10, 11, 12, 14) erhalten wir ein sehr gute CFA-Modellanpassung mit einem RMSEA von 0.48 und einem CFI von 0.987. Diese zeigen zwar signifikante schwache Invarianzen gemäss LR-Tests, jedoch nur auf dem 0.05-Niveau, wobei sich sowohl die CFI als auch die RMSEA für Geschlecht und Alter im schwachen Invarianzmodell verbessern. Man kann also durchaus von zumindest schwacher Invarianz ausgehen. Mit Ausnahme der Sprachregion verbessern sich sogar CFI und RMSEA bei Modellen starker Invarianz. Eine weitere UIRT Analyse ergab, dass DIF für die Region insbesondere für die Items 10 und 11 (und 12) besteht, wobei bei signifikanten χ2-Tests ($p < 0.001$) insbesondere die Effektstärken gegen Item 10 sprächen. Item 10 bietet also weiteres Kürzungsmaterial. Dieser Ausschluss ist auch in Betracht zu ziehen, da Item 10 auch in der Kurzform von Cartierre et al. (2011) bzw. der deutschen Kurzform PIEK (Bischof et al., 2016) ausgeschlossen worden ist. Dieser Möglichkeit wird im Abschnitt 3.9.4 weiter nachgegangen.

Eine solche Skala bestehend aus 9 Items (aber auch bei zusätzlicher Kürzung um das Item 10) behält auch die fünf ursprünglichen Unterdimensionen von Meerkerk (2007) mit zumindest jeweils einem Item bei:

- Konflikt: Items 10 und 11
- Coping: Item 12
- Kontrollverlust: Items 1, 5 und 9
- Vertieftsein (preoccupation): Items 4 und 7
- Entzugserscheinungen: Item 14

3.6.1 Vergleich der 9-Itemlösung mit dem Gesamttest: Fläche unter der Receiver Operating Characteristics (ROC)

ROC Analysen bieten die Möglichkeit, einen Test gegen einen sogenannten "goldenen Standard" zu evaluieren bzw. Schwellenwerte (*cut-offs*) zu bestimmen, die möglichst sowohl eine gute a) Sensitivität (wie gut werden "Kranke" erkannt) als auch b) Spezifität (wie gut werden "Gesunde" ausgeschlossen) haben. Je näher die Sensitivität und Spezifität an der "1" (=100% werden richtig klassifiziert) liegen, desto besser sind sie.

Hosmer and Lemeshow (2000) schlagen als Daumenregel vor, dass bei einer Fläche unter der ROC Werte von 0.70 bis 0.80 "akzeptabel" seien, Werte von 0.80 bis 0.90 wären "exzellent" und Werte von 0.9 oder grösser "herausragend". Ein Wert von 0.5 besagt, dass Personen nur zufällig als "gesund" oder "krank" klassifiziert werden, also der Test nur die Qualität eines Münzwurfes hat.

Leider liegt uns kein Aussenkriterium als goldener Standard vor. Wir können jedoch den Gesamttest mit 14 Items heranziehen. Dafür wurden Schwellenwerte (*cut-offs*) von 28 (problematische Internetnutzung, vgl. Abschnitt 3.9.2) bzw. 20 (symptomatische Internetnutzung) in der Literatur vorgeschlagen (Marmet et al., 2013). Aussagen darüber, wie gut unsere Kurzform ist, kann man nur unter der Annahme treffen, dass der CIUS gesamthaft ein guter Test ist. Tabelle 9 zeigt die Fläche unter der ROC mit den dazugehörigen Sensitivitäten, Spezifitäten sowie den jeweiligen Schwellenwerten für die Kurzform. Dabei wurden als Grenzbereiche Werte mit einer Sensitivität und Spezifität von jeweils über 0.9 angenommen (also jeweils mindestens 90% der "Gesunden" und "Kranken" werden korrekt klassifiziert).

Tabelle 11: Fläche unter der ROC, Sensitivität, Spezifizität für die gesamte Stichprobe und nach Alter, Geschlecht und Sprachregion

		CIUS >= 20				Cius >= 28			
		Area (S.E.)	**Sensi-tivität**	**Spezi-fität**	**Cut-off**	**Area (S.E.)**	**Sensi-tivität**	**Spezi-fität**	**Cut-off**
Total		0.990 (0.002)	.968-.923	.929-.961	11-12	.993 (.002)	1-.900	.928 - .982	14-17
Geschlecht	Männer	0.989 (0.003)	.980 - .910	.972-.959	11-12	.995 (.003)	1-.941	.924-.975	14-17
	Frauen	0.991 (0.002)	1-.936	.904-.963	10-12	.993 (.003)	1-.938	.904-.975	13-16
Region	deutsch	0.991 (0.002)	.991-.935	.933-.963	11-12	.995 (.002)	1-.966	.926-.979	13-16
	lateinisch	0.989 (0.003)	.980 - .911	.920-.957	11-12	.990 (.005)	1-.950	.924-.968	15-17
Alter	15-19	0.985 (0.004)	0.941	0.934	12	.985 (.005)	0.970	0.936	16
	20-34	0.989 (0.004)	.984-.934	.918-.954	11-12	.995 (.004)	1-.933	.916-980	13-17

Wie Tabelle 11 zeigt, sind die Kennwerte der ROC-Analysen herausragend. Unter der Annahme, dass der CIUS ein gutes Messinstrument ist, kann man davon ausgehen, dass es sicherlich auch die Kurzform ist. Nimmt man an, dass der CIUS ein Screeningverfahren ist, man also eher "Kranke" korrekt identifizieren möchte als "Gesunde" korrekt auszuschliessen, so sollte man eine höhere Sensitivität gegenüber einer höheren Spezifität als Kriterium bevorzugen. Aus diesem Grund schlagen wir für die CIUS-Kurzform einen Schwellenwert von 11 im Vergleich zu einem Schwellenwert von 20 beim gesamten CIUS bzw. einen Schwellenwert von 15 im Vergleich zum gesamten CIUS mit dem Schwellenwert von 28 vor. Forscher, die eher an einer höheren Spezifizität interessiert sind, sollten höhere Cut-Offs von 12 bzw. 16 wählen.

3.6.2 IRT: Test Characteritic Curve (TCC) und Test Information Function (TIF) als Vergleiche mit dem Gesamttest

Sowohl die TCC als auch die TIF für den Gesamttest und die Kurzform zeigen zweierlei (vgl. Abbildung 3 und 4). Erstens diskriminieren beide Skalen nicht im unteren Bereich (also den unteren 50% (entspricht Theta < 0) der Skalenwerte). Beide Skalen (Gesamttest und Kurzform) messen also das Konstrukt "kompulsiven Internetgebrauch" nur präzise im hohen Wertebereich. Zweitens erkennt man, dass die Kurzform praktisch identische Information im Vergleich zum Gesamttest liefert. Der Wertebereich der "Information" (y-Achse) ist im Gesamttest absolut gesehen grösser, was aber nur daran liegt, dass jedes Item zur Gesamtinformation beiträgt, also ist der Informationswert mit 14 Items höher liegen muss als mit 9 Items. Der Informationsgehalt verläuft aber bei der Kurzform nahezu identisch zum Gesamttest, der Informationsgehalt ist im entscheidenden Bereich (hohe Werte des Konstrukts sogar relativ gesehen besser (siehe Abbildung 5), was daran liegt, dass "schlechte" Items ausgeschlossen worden sind. Bei der *test characteristic curve* (Abbildung 3) haben wir für den Gesamttest die Schwellenwerte für 20 und 28 Punkte und die entsprechenden Werte auf dem latenten Konstrukt (kompulsiver Internetgebrauch) auf die Kurzform übertragen. Erreicht jemand 28 oder mehr Punkte, also die Hälfte der maximal möglichen Punkte (bzw. beantwortet alle Fragen mit durchschnittlich "manchmal"), gilt er oder sie nach den Empfehlungen von Meerkerk (2007) als problematisch Internetnutzender oder problematisch Internetnutzende. In Deutschland wurde ergänzend als Screening ein Grenzwert bei mehr als 20 Punkten gesetzt, was als symptomatische Internetnutzung bezeichnet werden kann. Die *test characteristic curve* (Abbildung 3) zeigt, dass in der Kurzform die entsprechenden Schwellenwerte von 20 und 28 der Langform bei etwa 12 bzw. 17 lägen. Dies entspräche den hohen Schwellenwerten basierend auf den ROC Analyse (vgl. Abschnitt 3.9.1). Diese Schwellenwerte lägen somit bei Werten, welche einer höheren Spezifität (also den besseren Ausschluss "Gesunder" (nicht kompulsiv Nutzender)) entsprächen. Anders herum könnte man auch sagen, dass vermutlich die Schwellenwerte von 20 und 28 auf der ursprünglichen

Originalskala des CIUS mit 14 Items zu hoch gesetzt worden sind, also nicht genügend Sensitivität hätten. Wie bereits gesagt, plädieren wir eher für eine höhere Sensitivität und damit für geringere Schwellenwerte. Dies entspräche auch dem Vorgehen der Kurzversion PIEK in Deutschland (Bischof et al., 2016). Wer an höherer Spezifität interessiert ist sollte Cut-offs von 12 bzw. 16 nehmen (obere Grenze in Tabelle 11, die auf alle Untergruppen zuträfe).

Abbildung 3: Test characteristic curves für den gesamten CIUS (links) und die Kurzform (rechts) und Ausprägungen des latenten Konstruktes bei Schwellenwerten von 20 und 28 Punkten auf dem Gesamttest

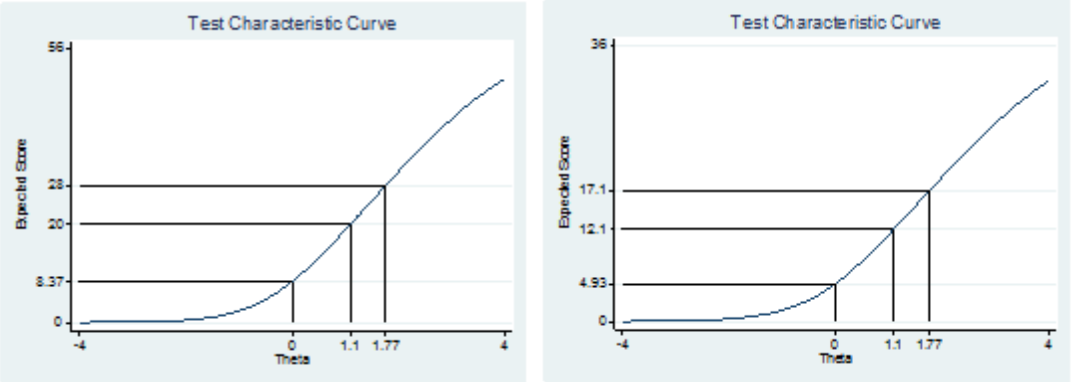


Abbildung 4: Test information functions für den gesamten CIUS (links) und die Kurzform (rechts)

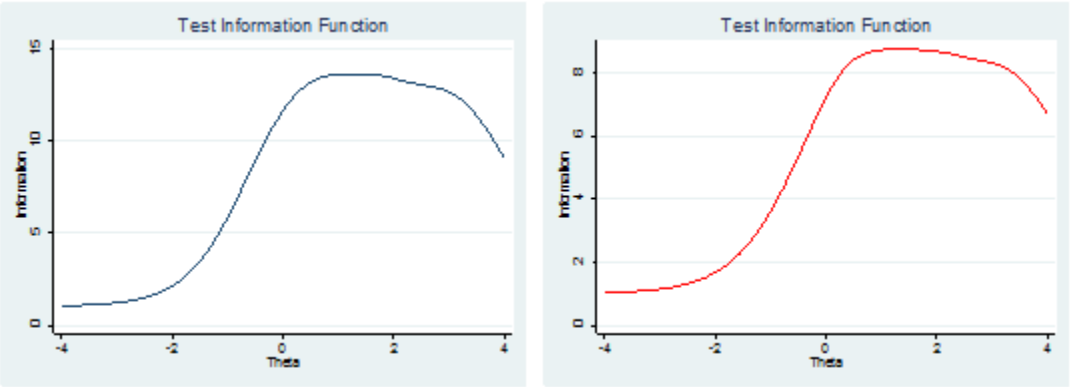
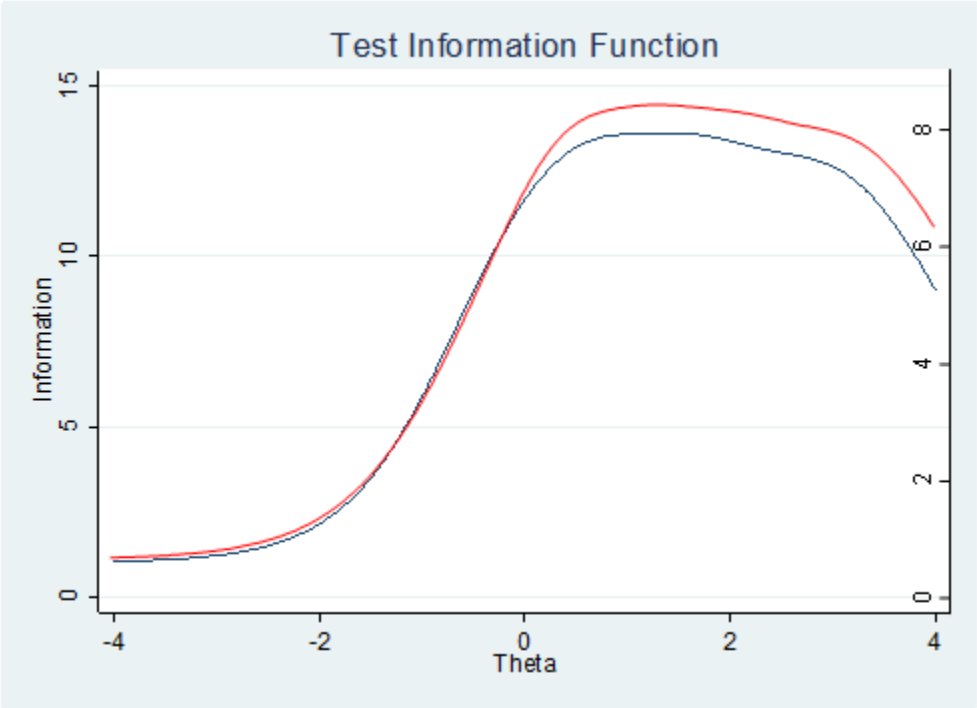


Abbildung 5: Direkter Vergleich der test information functions für den Gesamtest (linke Skala, schwarze Kurve) sowie der Kurzform (rechte Skala, rote Kurve)



3.6.3 Übereinstimmung mit Erkenntnissen der klassischer Testtheorie

Im Vergleich zur der 9-Itemlösung der klassischen Testtheorie (vgl. Kapitel 2) ist genau ein Item anders. Danach hätten wir eher Item 9 statt Item 8 ausgeschlossen.

- 9-Itemlösung gemäss klassischer Testtheorie:
  Item 1, Item 4, Item 5, Item 7, Item 8, Item 10, Item 11 Item 12, Item 14
- 9-Itemlösung nach probabilistischer Testtheorie und ordinaler CFA:
  Items 1, Item 4, Item 5, Item 7, Item 9, Item 10, Item 11, Item 12, Item 14

Wie jedoch im Abschnitt 3.8.4 angedeutet wurde, war die Beibehaltung von Item 9 (anstelle von Item 3) weniger eindeutig.

Die Lösung von Cartierre und Kollegen (2011) sah wie folgt aus:

- Item 1, Item 3, Item 4, Item 5, Item 7, Item 9, Item 11, Item 12, Item 14

Das heisst, dass die Lösung von Cartierre und Kollegen (2011) sich ebenfalls von unserer Lösung nur um ein Item unterscheidet. Bei der Skala von Cartierre et al. ist das Item 3 beibehalten worden, während bei unserem Vorschlag das Item 10 beibehalten worden ist. Wiederum war der Ausschluss von Item 3 (vgl. Abschnitt 3.8.4) weniger eindeutig als bei den anderen Items. Es hätte aber in jedem Fall eine Abweichung von einem Item im Vergleich zum Vorschlag von Cartierre und Kollegen gegeben, da die Entscheidung bei unserem Vorschlag zwischen Item 3 und Item 9 lag und beide Items bei der "Cartierre-Skala" beibehalten worden sind.

Generell weichen aber die verschiedenen Vorschläge kaum voneinander ab. Was auch die Korrelationen (Tabelle 12, vgl. auch Tabelle 6) untereinander und mit dem Gesamttest belegen.

Tabelle 12: Interkorrelationsmatrix der verschiedenen 9-Itemlösungen und des Gesamttests

		9-Item probabilistisch	9-Item klassisch	9-Item Cartierre	Gesamtskala CIUS
9-Item probabilistisch	Pearson Korrelation	1			
9-Item klassisch	Pearson Korrelation	0.988	1		
9-Item Cartierre	Pearson Korrelation	0.980	0.968	1	
Gesamtskala CIUS	Pearson Korrelation	0.972	0.975	0.975	1

3.6.4 Weiteres Kürzungspotential über Ausschluss von Item 10.

Wie im Abschnitt 3.9 angesprochen worden ist, böte sich der zusätzliche Ausschluss von Item 10 an. Dieser Ausschluss wurde als Möglichkeit angesehen, die starke Invarianz zwischen den Sprachregionen zu verbessern sowie eine stärkere Vergleichbarkeit mit der Cartierre Lösung herzustellen. Tabelle 13 informiert über den Vergleich mittels ordinaler konfirmatorischer Faktorenanalyse.

Tabelle 13: Vergleich der 9-Item- und 8-Itemlösung

		9-Itemlösung				8-Itemlösung			
		RMSEA	**CFI**	**Δχ2**	**p-Wert**	**RMSEA**	**CFI**	**Δχ2**	**p-Wert**
	Total	0.048	0.987			0.045	0.989		
Invarianz Geschlecht	konfigural	0.044	0.990			0.047	0.988		
	metrisch	0.041	0.990	18.2	0.019	0.043	0.988	16.5	0.021
	skalar	0.040	0.982	81.0	< .001	0.041	0.982	62.3	< .001
Invarianz Sprache	konfigural	0.048	0.987			0.043	0.990		
	metrisch	0.042	0.988	15.7	0.028	0.041	0.989	16.2	0.023
	skalar	0.052	0.967	131.7	< .001	0.047	0.976	97.9	< .001
Invarianz Alter	konfigural	0.050	0.981			0.048	0.983		
	metrisch	0.047	0.981	17.1	0.017	0.045	0.983	13.7	0.054
	skalar	0.040	0.979	48.9	< .001	0.036	0.981	38.4	0.139

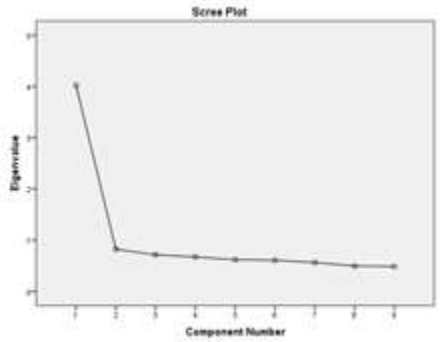
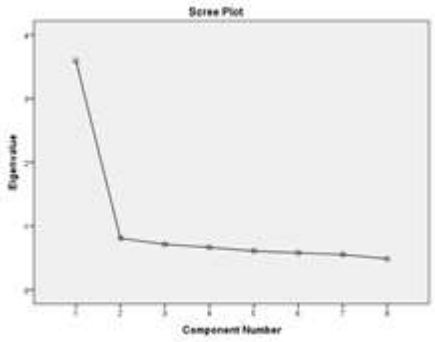
Insgesamt kann man sagen, dass der Ausschluss von Item 10 weiterhin zu hervorragender Modellanpassung führte. Zwar ergibt sich eine leichte Verbesserung hinsichtlich der starken Invarianz beim Vergleich der Sprachregionen, beispielsweise sinkt der ΔRMSEA von 0.021 auf 0.013, jedoch bliebe der Chi^2-Test auf sprachregionale starke Invarianz weiterhin hochsignifikant. Zusätzlich ergäbe sich jedoch auch eine Verbesserung in der starken Invarianz beim Vergleich der Altersgruppen und des Geschlechts. Bei letzterem wäre selbst der Chi^2-Test für schwache und starke Invarianz nicht mehr signifikant. Item 10 stellt also weiteres Kürzungspotential dar, ohne dass sich eine psychometrische Verschlechterung für die Skala ergäbe.

3.7 Überprüfung der Kurzformen in der Validierungsstichprobe

Im Jahr 2015 wurde der CIUS zum 2. Mal im Rahmen des Suchtmonitorings bei 1550 Personen erhoben (Marmet et al, 2015).

Tabelle 14 zeigt die Kennwerte der klassischen Testtheorie für die gesamte Population der Validierungsstichprobe (für den Gesamttest in der Lernstichprobe vgl. Kapitel 2).

Tabelle 14: Kennwerte der klassischen Testtheorie und der Fläche unter der ROC in der Validierungsstichprobe

	9-Itemlösung	8-Itemlösung
KMO	.915	.869
Bereich der MSA	.908 - .928	.886 - .908
Scree-Test		
Cronbachs Alpha	.841	.821
Trennschärfe	Nur Item 4 < 0.5 (.451)	Nur Item 4 und Item 14 < 0.5 (.452; .494)
Fläche unter ROC (cut-off = 20 im Gesamttest); cut-offs im Bereich Sens. und Spez. > .9	.992; Schwellenwerte: 10-12	.989; Schwellenwerte: 9-10
Fläche unter ROC (cut-off = 28 im Gesamttest); Cut-offs im Bereich Sens. und Spez. > .9	.995; Schwellenwerte 13 - 16	.993; Schwellenwerte: 12 - 15

Wie sich zeigt, ist auch der Datensatz der Validierungsstichprobe ausgezeichnet für Faktorenanalysen (KMO > .8 und MSA-Werte nahe an oder grösser als .9) geeignet. Es ist eindeutig, selbst nach dem strengen Kaiser-Guttman Kriterium, dass es sich um eine eindimensionale Faktorlösung handelt. Das Cronbachs Alpha ist gut (> .8). Die Werte für die Fläche unter der ROC sind herausragend (> .9) und die Schwellenwerte für die 9-Itemlösung bestätigen sich (bei hoher Sensitivität und Spezifität) im Vergleich zur Lernstichprobe mit 11 Punkten (im Vergleich zu einem Gesamttestwert des CIUS mit einem Schwellenwert von 20) bzw. 15 Punkten (Gesamttest CIUS mit einem Schwellenwert von 28).

Die Ergebnisse der ordinalen konfirmatorischen Faktorenanalyse finden sich in Tabelle 15.

Tabelle 15: Ergebnisse der ordinalen konfirmatorischen Faktorenanalyse in der Validierungsstichprobe

		9-Itemlösung				8-Itemlösung			
		RMSEA	CFI	$\Delta\chi 2$	p-Wert	RMSEA	CFI	$\Delta\chi 2$	p-Wert
	Total	0.034	0.995			0.036	0.995		
Invarianz Geschlecht	konfigural	0.036	0.994			0.029	0.996		
	metrisch	0.035	0.994	28.2	0.0195	0.028	0.995	10	0.1861
	skalar	0.040	0.987	101.12	< .001	0.041	0.982	58.4	0.0014
Invarianz Region	konfigural	0.033	0.995			0.043	0.990		
	metrisch	0.028	0.996	11.5	0.1763	0.041	0.989	16.3	0.0228
	skalar	0.043	0.985	137.6	< .001	0.044	0.983	81.5	< .001
Invarianz Alter	konfigural	0.041	0.990			0.047	0.987		
	metrisch	0.030	0.994	6	0.6519	0.035	0.992	5	0.6631
	skalar	0.029	0.991	46.7	0.072	0.029	0.991	34.7	0.254

Generell zeigen sich noch bessere Modelanpassungen (kleinerer RMSEA und CFI) in der Validierungsstichprobe im Vergleich zur Lernstichprobe. Nach den Kriterien des ΔRMSEA und ΔCFI kann man von starker Invarianz bezüglich des Geschlechts und Alters ausgehen. Die 9-Itemlösung scheitert knapp an dem engen Kriterium (ΔRMSA und ΔCFI < .01) für sprachregionale Invarianz. Diese wäre jedoch bei der 8-Itemlösung gegeben.

4. Fazit

Sowohl die Kurzform mit neun Items (Items 1, Item 4, Item 5, Item 7, Item 9, Item 10, Item 11, Item 12, Item 14) als auch die mit acht Items (zusätzlicher Ausschluss von Item 10) weisen hervorragende psychometrische Eigenschaften auf und messen mit vergleichbarer Präzision dasselbe wie der CIUS-Gesamttest, wobei alle fünf Dimensionen (Konflikt, Coping: Kontrollverlust, Vertieftsein (preoccupation) und Entzugserscheinungen) der Originalskala durch Items vertreten bleiben. Bei der Itemfindung für die Kurzform decken sich Ansätze der klassischen und probabilistischen Testtheorie.

Tabelle 16: Items der schweizerischen Kurzform mit 8 (9) Items sowie der Kurzform nach Cartierre et al und Bischof et al.

Item-nummer	Di-men-sion	Wortlaut der Frage	Ver-sion Schweiz mit 8 (9) Items	Kurz-form von Car-tierre et al.	Kurz-form von 5 (7) Items nach Bi-schof et al.
Item 1	KV	Wie häufig finden Sie es schwierig, mit dem Internetgebrauch aufzuhören, wenn Sie online sind?	X	X	X
Item 3	K	Wie häufig sagen Ihnen andere Menschen (z.B. Partner, Kinder, Eltern, Freunde), dass Sie das Internet weniger nutzen sollten?		X	X
Item 4	V	Wie häufig bevorzugen Sie das Internet, statt Zeit mit anderen zu verbringen (z.B. Partner, Kinder, Eltern, Freunde)?	X	X	
Item 5	KV	Wie häufig schlafen Sie zu wenig wegen des Internets?	X	X	X
Item 7	V	Wie oft freuen Sie sich bereits auf Ihre nächste Internetsitzung?	X	X	(X)
Item 9	KV	Wie häufig haben Sie erfolglos versucht, weniger Zeit im Internet zu verbringen?	X	X	(X)
Item 10	K	Wie häufig erledigen Sie Ihre Aufgaben (zuhause oder auf der Arbeit) hastig, damit Sie früher ins Internet können?	(X)		
Item 11	K	Wie häufig vernachlässigen Sie Ihre Alltagsverpflichtungen (Arbeit, Schule, Familienleben), weil Sie lieber ins Internet gehen?	X	X	X
Item 12	C	Wie häufig gehen Sie ins Internet, wenn Sie sich niedergeschlagen fühlen	X	X	X
Item 14	E	Wie häufig fühlen Sie sich unruhig, frustriert oder gereizt, wenn Sie das Internet nicht nutzen können?	X	X	

Bemerkung: K = Konflikt, C=Coping, KV= Kontrollverlust, V=Vertieftsein, E=Entzugserscheinung

Im Vergleich mit der zur Zeit verwendeten Kurzform nach Cartierre und Kollegen gibt es aber praktisch keine Unterschiede, so dass auch die Cartierre et al. Version verwendet werden kann. Kürzt man letztere um Item 3 wäre sie sogar mit unserer 8-Item Lösung identisch.

5. Referenzen

An, X. & Yung, Y.-F. (2014). Item response theory: what it is and how you can use the IRT procedure to apply it. SAS Institute Inc. SAS364-2014.

Baker, F. B. (2001). *The basics of item response theory. 2nd edition.*, Portsmouth NH: Heinemann.

Bischof, G., Bischof, A., Besser, B. & Rumpf, H. J. (2016). Problematische und pathologische Internetnutzung: Entwicklung eines Kurzscreenings (PIEK). Abschlussbericht an das Bundesministerium für Gesundheit. Lübeck: Universität zu Lübeck, Klinik für Psychiatrie und Psychotherapie.

Bischof, G., Bischof, A., Meyer, C., John, U. & Rumpf, H.-J. (2013). Prävalenz der Internetabhängigkeit–Diagnostik und Risikoprofile (PINTA-DIARI) - Kompaktbericht. Lübeck: Klinik für Psychiatrie und Psychotherapie.

Bortz, J. & Döring, N. (2005). *Forschungsmethoden und Evaluation.* Heidelberg: Springer-Verlag.

Cartierre, N., Coulon, N. & Demerval, R. (2011). Validation d'une version courte en langue française pour adolescents de la Compulsive Internet Use Scale. *Neuropsychiatrie de l'Enfance et de l'Adolescence 59*, 415-419.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling 14*, 464-504.

Choi, S. W., Gibbons, L. E. & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software 39*, 1.

Cohen, A. S., Kim, S.-H. & Baker, F. B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement 17*, 335-350.

Crane, P. K., Gibbons, L. E., Jolley, L. & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care 44*, S115-S123.

DeVellis, R. F. (2012). *Scale development: Theory and applications. 3rd edition.* Los Angeles: Sage publications.

Elosua, P. (2011). Assessing Measurement Equivalence in Ordered-Categorical Data. *Psicologica: International Journal of Methodology and Experimental Psychology 32*, 403-421.

Hidalgo-Montesinos, M. D. & Gómez-Benito, J. (2003). Test Purification and the evaluation of differential item functioning with multinomial logistic regression. *European Journal of Psychological Assessment 19*, 1.

Hidalgo, M. D. & López-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement 64*, 903-915.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika 30*, 179-185.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression, Second Edition.* New York: Wiley.

Jodoin, M. G. & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education 14*, 329-349.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika 39*, 31-36.

Khazaal, Y., Chatton, A., Horn, A., Achab, S., Thorens, G., Zullino, D. & Billieux, J. (2012). French validation of the compulsive internet use scale (CIUS). *Psychiatric Quarterly 83*, 397-405.

Kim, S. H., Cohen, A. S., Alagoz, C. & Kim, S. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement 44*, 93-116.

Latila, T. (1993). A pseudo-R 2 measure for limited and qualitative dependent variables. *Journal of Econometrics 56*, 341-356.

Marmet, S., Notari, L. & Gmel, G. (2013). Suchtmonitoring Schweiz - Themenheft Internetnutzung und problematische Internetnutzung in der Schweiz im Jahr 2013. Lausanne, Schweiz: Sucht Schweiz.

McKelvey, R. D. & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology 4*, 103-120.

Meade, A. W. & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling 14*, 611-635.

Meade, A. W. & Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods 9*, 369-403.

Meerkerk, G.-J. (2007). Pwned by the Internet. *Erasmus University Rotterdam*18-36.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education 15*, 625-32.

Stark, S., Chernyshenko, O. S. & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology 91*, 1292-306.

Suh, Y. (2016). Effect Size Measures for Differential Item Functioning in a Multidimensional IRT Model. *Journal of Educational Measurement 53*, 403-430.

Suh, Y. & Cho, S.-J. (2014). Chi-square difference tests for detecting differential functioning in a multidimensional IRT model: A Monte Carlo study. *Applied Psychological Measurement 38*, 359-375.

Thomas, D. R., Zhu, P., Zumbo, B. D. & Dutta, S. (2008). On measuring the relative importance of explanatory variables in a logistic regression. *Journal of Modern Applied Statistical Methods 7*, 4.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika 41*, 321-327.

Wainer, H. (1993). Model-based standardized measurement of an item's differential impact. In H. Wainer & P. W. Holland (Eds.), *Differential item functioning*, (pp. 123-135). Hillsdale, NJ: Lawrence Erlbaum.

Zumbo, B. & Thomas, D. (1997). A measure of effect size for a model-based approach for studying DIF. Prince George, Canada: Edgeworth Laboratory for Quantitative Behavioral Science, University of Northern British Columbia.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF).* Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules of determining the number of components to retain. *Psychological Bulletin 99*, 432-442.

6. Anhang Fragebogen

Deutsch

cius_ig1	**Wie häufig finden Sie es schwierig, mit dem Internetgebrauch aufzuhören, wenn Sie online sind?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig2	**Wie häufig setzen Sie Ihren Internetgebrauch fort, obwohl Sie eigentlich aufhören wollten?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig3	**Wie häufig sagen Ihnen andere Menschen (z.B. Partner, Kinder, Eltern, Freunde), dass Sie das Internet weniger nutzen sollten?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig4	**Wie häufig bevorzugen Sie das Internet, statt Zeit mit anderen zu verbringen (z.B. Partner, Kinder, Eltern, Freunde)?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig5	**Wie häufig schlafen Sie zu wenig wegen des Internets?**

Nie	1
Selten …	2
Manchmal	3
Häufig	4
Sehr häufig	5
weiss nicht	98
Keine Angabe/verweigert	99

cius_ig6	**Wie häufig denken Sie an das Internet, auch wenn Sie gerade nicht online sind?**

Nie	1
Selten …	2
Manchmal	3
Häufig	4
Sehr häufig	5
weiss nicht	98
Keine Angabe/verweigert	99

cius_ig7	**Wie oft freuen Sie sich bereits auf Ihre nächste Internetsitzung?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig8	**Wie häufig denken Sie darüber nach, dass Sie weniger Zeit im Internet verbringen sollten?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig9	**Wie häufig haben Sie erfolglos versucht, weniger Zeit im Internet zu verbringen?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig10	**Wie häufig erledigen Sie Ihre Aufgaben (zuhause oder auf der Arbeit) hastig, damit Sie früher ins Internet können?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig11	**Wie häufig vernachlässigen Sie Ihre Alltagsverpflichtungen (Arbeit, Schule, Familienleben), weil Sie lieber ins Internet gehen?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig12	**Wie häufig gehen Sie ins Internet, wenn Sie sich niedergeschlagen fühlen?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig13	**Wie häufig nutzen Sie das Internet, um Ihren Sorgen zu entkommen oder um sich von einer negativen Stimmung zu entlasten?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	weiss nicht	98
	Keine Angabe/verweigert	99

cius_ig14	**Wie häufig fühlen Sie sich unruhig, frustriert oder gereizt, wenn Sie das Internet nicht nutzen können?**	
	Nie	1
	Selten …	2
	Manchmal	3
	Häufig	4
	Sehr häufig	5
	Weiss nicht	98
	Keine Angabe/verweigert	99

Französisch

cius_ig1	**A quelle fréquence trouvez-vous difficile d'arrêter d'utiliser internet une fois que vous êtes en ligne ?**	
	Jamais	1
	Rarement	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig2	**A quelle fréquence continuez-vous d'utiliser internet malgré votre intention d'arrêter ?**	
	Jamais	1
	Rarement	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig3	**A quelle fréquence d'autres personnes (ex : partenaire, enfants, parents) vous disent-elles que vous devriez moins utiliser internet ?**	
	Jamais	1
	Rarement …	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig4	**A quelle fréquence préférez-vous aller sur internet au lieu de passer du temps avec d'autres personnes (partenaires, enfants, parents) ?**	
	Jamais	1
	Rarement …	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig5	**A quelle fréquence vous arrive-t-il de manquer de sommeil à cause d'internet ?**	
	Jamais	1
	Rarement	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig6	**A quelle fréquence pensez-vous à internet même quand vous n'êtes pas en ligne ?**	
	Jamais	1
	Rarement	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig7	**A quelle fréquence vous réjouissez-vous de votre prochaine connexion à internet ?**	
	Jamais	1
	Rarement …	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig8	**A quelle fréquence pensez-vous que vous devriez utiliser Internet moins souvent ?**	
	Jamais	1
	Rarement …	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig9	**A quelle fréquence avez-vous essayé sans succès de passer moins de temps sur internet ?**	
	Jamais	1
	Rarement	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig10	**A quelle fréquence vous arrive-t-il de vous dépêcher de faire quelque chose (travail, ménages…) pour aller sur internet ?**	
	Jamais	1
	Rarement	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig11	**A quelle fréquence négligez-vous vos obligations quotidiennes (travail, école, ou vie familiale) parce que vous préférez aller sur internet ?**

Jamais	1
Rarement …	2
Parfois	3
Souvent	4
Très souvent	5
ne sait pas	98
pas de réponse / refus	99

cius_ig12	**A quelle fréquence allez-vous sur internet quand vous n'avez pas le moral ?**

Jamais	1
Rarement …	2
Parfois	3
Souvent	4
Très souvent	5
ne sait pas	98
pas de réponse / refus	99

cius_ig13	**A quelle fréquence utilisez-vous internet pour fuir vos soucis ou vous soulager d'un sentiment négatif ?**	
	Jamais	1
	Rarement	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

cius_ig14	**A quelle fréquence vous sentez-vous agité, frustré ou irrité lorsque vous ne pouvez pas utiliser internet ?**	
	Jamais	1
	Rarement	2
	Parfois	3
	Souvent	4
	Très souvent	5
	ne sait pas	98
	pas de réponse / refus	99

Italienisch:

cius_ig1	**Con quale frequenza trova difficile smettere di usare internet quando è online?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig2	**Con quale frequenza continua ad usare internet nonostante stia cercando di smettere?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig3	**Con quale frequenza altre persone (es. partner, genitori, amici) le dicono che dovrebbe ridurre l'uso di internet?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig4	**Con quale frequenza preferisce utilizzare Internet piuttosto che trascorrere tempo con altre persone (es. partner, genitori, amici)?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig5	**Con quale frequenza dorme poco a causa di internet?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig6	**Con quale frequenza pensa ad Internet anche quando non è online?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig7	**Con quale frequenza le capita di aspettare con impazienza di intraprendere una nuova sessione?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig8	**Con quale frequenza pensa che dovrebbe usare internet meno spesso?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig9	**Con quale frequenza ha provato a trascorrere meno tempo su internet senza riuscirci?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig10	**Con quale frequenza le capita di sbrigare velocemente i suoi impegni a casa/lavoro per poter utilizzare internet?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig11	**Con quale frequenza le capita di trascurare i suoi obblighi quotidiani (lavoro, scuola, famiglia) perché preferisce utilizzare internet?**

Mai	1
Raramente	2
Qualche volta	3
Spesso	4
Molto spesso	5
non lo sa	98
nessuna risposta / rifiuto	99

cius_ig12	**Con quale frequenza utilizza internet quando è triste?**

Mai	1
Raramente	2
Qualche volta	3
Spesso	4
Molto spesso	5
non lo sa	98
nessuna risposta / rifiuto	99

cius_ig13	**Con quale frequenza usa internet per sfuggire ai suoi dispiaceri o trarre sollievo dai suoi sentimenti negativi?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99

cius_ig14	**Con quale frequenza si sente inquieto, frustrato, o irritato quando non può usare Internet?**	
	Mai	1
	Raramente	2
	Qualche volta	3
	Spesso	4
	Molto spesso	5
	non lo sa	98
	nessuna risposta / rifiuto	99