ORIGINAL ARTICLE

Development of a short form of the compulsive internet use scale in Switzerland

Gerhard Gmel^{1,2,3,4} | Yasser Khazaal^{5,6,7} | Joseph Studer¹ | Stéphanie Baggio^{8,9,10,11} | Simon Marmet¹

¹Alcohol Treatment Centre, Lausanne University Hospital, Lausanne, Switzerland

²Research Department, Addiction Switzerland, Lausanne, Switzerland

³Centre for Addiction and Mental Health, Institute for Mental Health Policy Research, Toronto, Ontario, Canada

⁴University of the West of England, Faculty of Health and Social Science, Frenchay Campus, Bristol, UK

⁵Department of Mental Health and Psychiatry, Geneva University Hospitals, Geneva, Switzerland

⁶Department of Psychiatry, Geneva University, Geneva, Switzerland

⁷Research Centre, University Institute of Mental Health at Montreal, Montreal, Quebec, Canada

⁸Division of Prison Health, Geneva University Hospitals and University of Geneva, Geneva, Switzerland

⁹Life Course and Social Inequality Research Centre, University of Lausanne, Lausanne, Switzerland

¹⁰Centre for Excessive Gambling, Lausanne University Hospital, Lausanne, Switzerland

¹¹Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia

Correspondence

Gerhard Gmel, Alcohol Treatment Centre, Lausanne University Hospital CHUV, Av. Beaumont 21 bis, Pavillon 2, Lausanne CH-1011, Switzerland.
Email: gerhard.gmel@chuv.ch

Funding information

Federal Office of Public Health, Grant/Award Number: FOPH 16.924108/204.0001-1511/1

Abstract

Objectives: The study aims to develop a short form of the compulsive internet use scale (CIUS), which can be used in multitopic and general population health surveys and is invariant across different sexes, linguistic regions, and ages.

Methods: Two general population surveys from 2013 and 2015 were used as learning ($n = 1,371$) and validation samples ($n = 1,550$), respectively. Reducing items from the original CIUS was based on the following: (a) correlated errors between items, (b) differential item functioning, and (c) measurement invariance. Methods used item response theory and latent confirmatory factor analysis for ordinal variables.

Results: The eight-item short form maintained the five dimensions of the original scale and was metric and mostly scale invariant for sex, region, and age. It fell marginally short of scale invariance ($\Delta CFI < 0.01$) for regions in the learning sample and for sexes in the validation sample (both $\Delta CFI = 0.013$, $p < 0.01$). Root mean square error of approximation was 0.045 and 0.036, and comparative fit index was 0.989 and 0.995, in the learning and validation samples, respectively, showing excellent fit of the model to data. Correlations with the full scale were $r = 0.966$ (learning) and $r = 0.969$ (validation).

Conclusion: If the full 14-item CIUS is a valid, reliable screening instrument, then the short eight-item form is too, and can be used in multitopic, general population health surveys.

KEYWORDS

compulsive internet use, differential item functioning, French, German, Italian, short form

1 | INTRODUCTION

Given the wide spread of the internet today, it comes as little surprise that some individuals may use it in problematic ways. Clinically, relevant cases of internet use have existed for at least 20 years, when the term *internet addiction* first emerged (Griffiths, 1999; Young, 1996). Since then, research on internet addiction has dramatically increased, with the emergence of a multitude of terms such as

compulsive, problematic, excessive, or pathological internet use (Breslau, Aharoni, Pedersen, & Miller, 2015; De Cock et al., 2014; Kuss, Griffiths, Karila, & Billieux, 2014). However, there is no internationally agreed definition of these different terms (Tokunaga & Rains, 2016), nor any agreement about whether there is an independent pathology justifying terms such as *internet addiction* or *internet use disorder* (Breslau et al., 2015; Weinstein & Lejoyeux, 2010). Gambling disorder is the only currently accepted behavioral addiction (Bélanger-Lejars,

2015) in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5®; American Psychiatric Association [APA], 2013). Internet gaming disorder was added to the DSM-5 under the category of potential disorders needing further research and clinical experience before inclusion as an official disorder. Gaming disorder will be included in the revision of the *International Classification of Diseases, 11th Edition*.

Addicted internet users may experience similar symptoms to those traditionally felt in substance use disorders (Kuss & Griffiths, 2012; Kuss, Shorter, van Rooij, Griffiths, & Schoenmakers, 2014; Widyanto & Griffiths, 2006), which was strengthened by neuroscience (Brand, Young, & Laier, 2014; Kuss & Griffiths, 2012; Park, Han, & Roh, 2016). There is also increasing criticism of the use of concepts like withdrawal and tolerance for behavioral addictions (Kardefelt-Winther et al., 2017; Starcevic & Aboujaoude, 2017). Internet is a medium and the excessive behavior (internet addiction) is the response to specific content or activities (Davis, 2001; Davis, Flett, & Besser, 2002; Griffiths, 2000; Kardefelt-Winther, 2017; Weinstein & Lejoyeux, 2010; Yellowlees & Marks, 2007). Nevertheless, even critics of an internet disorder do not deny the internet's *addictive* features, which may lead to negative consequences. It therefore makes sense to screen for *problematic internet use* (PIU).

Due to the lack of consensus about any theoretical criteria and definitions of PIU, there is a large number of instruments measuring the behavior. Laconi, Rodgers, and Chabrol (2014) counted 45 tools for assessing internet addiction. Many were specialized PIU tools (e.g., gaming) or not psychometrically tested. The compulsive internet use scale (CIUS) is a general internet problem scale and one of the best-validated instruments (Breslau et al., 2015; Kuss, Griffiths, et al., 2014; Laconi et al., 2014; Lortie & Guitton, 2013). The CIUS has been shown to be unidimensional with high internal consistency (Dhir, Chen, & Nieminen, 2015; Dhir, Chen, & Nieminen, 2016). A few studies have found more than one dimension (Alavi, Jannatifard, Eslami, & Rezapour, 2011; Yong, Inoue, & Kawakami, 2017), probably due to reliance on the Kaiser criterion (Kaiser, 1970), which overestimates the number of dimensions.

As with other PIU scales—at 14 items—the CIUS is relatively long. There is a need for a shorter instrument for inclusion in general health surveys, like the Swiss Health Survey, where many health topics are assessed simultaneously. The present study was mandated by Switzerland's Federal Office of Public Health to derive a valid short form consisting of fewer than 10 items for future inclusion in Swiss Health Surveys.

Short forms have the advantage of reducing participants' fatigue, frustration, and boredom (Robins, Hendin, & Trzesniewski, 2001). Short instruments or scales can improve assessment by reducing response time and effort, minimizing burden, decreasing fatigue effects, and thus increasing response rates and representativeness (Dodeen & Al-Darmaki, 2016). Short, high-quality instruments are ideal for large surveys (Green & Frantom, 2002) and can often measure constructs well when they retain acceptable psychometric properties (Thalmayer, Saucier, & Eigenhuis, 2011). Methods like item response theory (IRT), which allows the selection of items providing the maximum information within a specific trait range along the underlying latent dimension, can be used to produce precise and relatively brief

instruments (Edelen & Reeve, 2007), which can be even more reliable than corresponding longer tests (Embretson & Reise, 2000).

To the best of our knowledge, there are two short forms of the CIUS. One was developed in French by Cartierre, Coulon, and Demerval (2011), but only used a small sample ($n = 289$) of adolescents (mean age = 13.8). Hence, no information on adults is available for that short form. Moreover, they used nine items based on the error correlations between items found by Meerkerk (2007). There was no psychometric development of a short form, but only an empirical testing of a preassumed short form. In Germany, Besser et al. (2017) validated five-item and seven-item short forms against a clinical interview using DSM-5 criteria with the Munich-Composite International Diagnostic Interview (Wittchen et al., 1995). They used two high-risk group samples, namely, (a) job-seekers from an employment agency and four unemployment offices and (b) positively screened young people from two vocational schools. Thus, this sample was not representative of the general population either.

The CIUS has been shown to be relatively sex and age invariant (Guertler et al., 2014; Meerkerk, Van Den Eijnden, Vermulst, & Garretsen, 2009; Yong et al., 2017). Furthermore, Switzerland requires an instrument, which is invariant concerning its culturally distinct linguistic regions (there are three official *main* languages as follows: German, French, and Italian). As indicated by Van de Vijver and Poortinga (1997), even if an instrument shows excellent properties in separate assessments in different societies/cultures, this is not a sufficient condition for valid cross-cultural comparison: tests for invariance must be done jointly in one sample. The present study sought to develop a short form of the CIUS using two general population samples (learning and validation samples), which is invariant as regards age, gender, and linguistic region.

2 | METHOD

2.1 | Participants and procedure

Data came from two samples from the Addiction Monitoring System in Switzerland with core questions on substance use addressed to about 11,000 respondents annually (since 2011). In order to include many topics (e.g., e-cigarette use, health literacy, and acceptance of national policies) without overloading participants, a modular system was added to the core questions using a random split-half with question modules changing every 6 months. The CIUS was distributed twice between January and June 2013 and 2015. Response rates (proportion of contacted individuals minus incapable target persons [hearing difficulties and language problems]) were 55% and 48%, respectively. In 2013, Addiction Monitoring System in Switzerland used a two-stage random sampling (households and target person in household) of noninstitutionalized people aged 15 and above. In 2015, sampling was based on the Federal Statistical Office's private individualized telephone register, which had not previously been available. This enabled simple random sampling based directly on individuals, where possible, and avoiding household sampling as the primary sampling unit. In 2013 (2015) 2,459 (2,613) individuals participated, whereby 344 (367) mostly older people did not use the

internet at all. Only people using the internet for at least 1 hr on at least 1 day during the week for private reasons (not professional) were asked to respond to the CIUS items. Final sample sizes were 1,371 people in 2013 and 1,550 in 2015.

2.1.1 | Learning sample and validation sample

The short form was developed using the 2013 sample as a learning sample and was then independently tested using the 2015 sample as the validation sample.

2.2 | Instrument

The CIUS consists of 14 items (Table 1) with a 5-point Likert-type scale from 0 (*never*) to 4 (*very often*). According to Meerkerk (2007), the CIUS represents five dimensions of compulsive internet use: conflict (Items 3, 8, 10, and 11), coping (Items 12 and 13), loss of control (Items 1, 2, 5, and 9), preoccupation (Items 4, 6, and 7), and withdrawal symptoms (Item 14). Part of the strategy to develop a short form involved retaining at least one item per dimension, which logically resulted in retaining Item 14. We used the French version developed by Khazaal et al. (2012) together with the German version developed by Guertler et al. (2014) and slightly adapted them to make them more comparable with the English version. The Italian version was based on these French and German versions and was translated by two bilingual native Italian speakers.

TABLE 1 Items of the compulsive internet use scale, full form

- Item 1: How often do you find it difficult to stop using the Internet when you are online? (Loss of control)
- Item 2: How often do you continue to use the Internet despite your intention to stop? (Loss of control)
- Item 3: How often do others (e.g., partner, children, and parents) say you should use the Internet less? (Conflict)
- Item 4: How often do you prefer to use the Internet instead of spending time with others (e.g., partner, children, and parents)? (Preoccupation)
- Item 5: How often are you short of sleep because of the Internet? (Loss of control)
- Item 6: How often do you think about the Internet, even when not online? (Preoccupation)
- Item 7: How often do you look forward to your next Internet session? (Preoccupation)
- Item 8: How often do you think you should use the Internet less often? (Conflict)
- Item 9: How often have you unsuccessfully tried to spend less time on the Internet? (Loss of control)
- Item 10: How often do you rush through your (home) work in order to go on the Internet? (Conflict)
- Item 11: How often do you neglect your daily obligations (work, school, or family life) because you prefer to go on the Internet? (Conflict)
- Item 12: How often do you go on the Internet when you are feeling down? (Coping)
- Item 13: How often do you use the Internet to escape from your sorrows or get relief from negative feelings? (Coping)
- Item 14: How often do you feel restless, frustrated, or irritated when you cannot use the Internet? (Withdrawal symptoms)

Note. Shaded items were retained for the short form.

It was back translated by bilingual native French and German speakers. Minor discrepancies between translation and back translation were resolved by discussions among the group of three translators.

2.3 | Statistical analysis strategy

The statistical analysis strategy is summarized in Figure 1.

2.3.1 | Unidimensionality

Unidimensionality is a prerequisite for constructing a single scale, particularly in IRT. We tested unidimensionality using classic test theory methods assuming an interval scale level of measurement and using probabilistic test theory assuming an ordinal level of measurement. For classic test theory, we estimated each item's measure sampling adequacy (MSA) and the corresponding overall Kaiser–Meyer–Olkin criterion for the total test battery to determine the adequacy of items for exploratory factor analysis. According to Kaiser (Kaiser, 1974), values from 0.6 to <0.8 are mediocre or middling, whereas values from 0.8 to <0.9 are meritorious and 0.9 and above are marvelous. As the

Learning sample

Validation sample

FIGURE 1 Flowchart of statistical analysis strategy. CIUS: Compulsive Internet Use Scale; UIRT: unidimensional item response theory (add-on in Stata); IRT: item response theory; CFA: confirmatory factor analysis; RMSEA: root mean square error of approximation; CFI: comparative fit index; Zumbo: regression approach for differential item functioning analysis developed by Zumbo (1999); DIF: differential item functioning

common Kaiser–Guttman criterion overestimates the number of factors, we applied Cattell's scree plot, parallel analysis, following Horn (1965), and the minimum average partial test (Velicer, 1976) to determine unidimensionality. We then used latent confirmatory factor analysis (CFA) for ordinal variables using the weighted least squares means and variance adjusted estimator for testing the model fit of a one-dimensional solution. The rule of thumb, also used in the present study, for model fit to the data: (a) comparative fit index (CFI) values close to 0.95 or higher indicate good fit (Hu & Bentler, 1999), whereas values between 0.90 and 0.95 are considered acceptable (Brown, 2006); (b) for the root mean square error of approximation (RMSEA) values close to 0.06 or below are indicative of good fit (Hu & Bentler, 1999), and values between 0.06 and 0.08 can be considered acceptable (Browne & Cudeck, 1993).

2.3.2 | Item reduction

For item reduction, we used models based on ordinal item information, namely, CFA and graduated response models for IRT. Both models of the ordered polytomous item responses estimate a discrimination parameter and $k-1$ (k = number of response options) difficulty (threshold) parameters per item. A high discrimination parameter suggests an item that has a high ability to differentiate subjects. When discrimination is high, the item provides more information on the latent trait, and the information is concentrated around item difficulties. For IRT, Baker (2001) proposed different ranges of discrimination parameters: no discriminative power (0), very low (0.01–0.34), low (0.35–0.64), moderate (0.65–1.34), high (1.35–1.69), and very high (≥ 1.70). The item difficulties in graduated response model refer to the average level of the latent variable, where individuals have the maximum probability of choosing a response option.

The reduction of items was based on the following strategy. First, items should have a high discriminative power. Second, there should be a low error correlation between items. Error correlation should be avoided (Hermida, 2015), as it indicates that variable covariance is not only explained by the latent construct but may have some other exogenous common cause or, more often, be due to very similar wording, indicating redundancy (Brown, 2006). Eliminating items with correlated errors also increases local independence—an assumption in IRT (Revicki, Chen, & Tucker, 2015). Meerkerk (2007) could only get satisfactory fit of a unidimensional latent construct by letting errors of pairs of items correlate, namely, CUIS Items 1 and 2, 6 and 7, 8 and 9, 10 and 11, and 12 and 13. Other studies have also found high correlations between these item errors (Dhir et al., 2015; Dhir et al., 2016; Guertler et al., 2014). Deleting one item from each of these five pairs was the strategy adopted by Cartierre et al. (2011) in a sample of students around 13 years old in France. However, they neither tested whether Meerkerk's redundancy Meerkerk (2007) applied to their sample nor gave any criterion for which item in each pair should be deleted.

In the present study, we tested for correlated errors by applying modification indices (MI) to the CFA using Mplus Version 8 (Muthén & Muthén, 1998). As a default, Mplus identifies $MI \geq 10$ as relevant. Second, differential item functioning (DIF) analysis was used to delete the item in each pair with the greatest resulting DIF with regard to sex,

linguistic region, and age. DIF refers to the instrument's bias to specific individual items, indicating when persons from different groups, but with the same instrument score, do not have the same expected score on that item (Shepard, Camilli, & Averill, 1981). DIF analysis identifies items related to measurement invariance. For the measurement of ordinal items, invariance can be tested using multigroup CFA (MG-CFA) and IRT. Both MG-CFA and IRT showed reasonable power in the identification of DIF (Kim & Yoon, 2011) and resulted in similar detection of DIF (Stark, Chernyshenko, & Drasgow, 2006).

2.3.3 | Test for DIF

There are several tests for DIF or item invariance.

A first approach using IRT defined one group as the focus group and the second as the reference group. Two nested models were compared via likelihood ratio (χ^2 tests). The first model (compact model) constrained discrimination and difficulty parameters to make them equal in both groups; the second model (augmented model) had unconstrained parameters. IRT models were estimated with Stata, and DIF tests were estimated using the unidimensional IRT (UIRT) add-on. UIRT provides effect size estimates following Wainer (1993), whereby effect sizes >0.05 (Suh, 2016) or >0.10 (Kim, Cohen, Alagoz, & Kim, 2007) are indicative of DIF. Effect sizes may slightly vary according to which group is chosen as the focus or reference group. We used effect sizes of 0.05 to <0.10 (small), 0.10 to <0.15 (medium), and ≥ 0.15 (large).

Second, Zumbo (1999) suggested using regression models, whereby the first model estimates the regression of the total test score on each single item, and the second model adds the grouping variable and its interaction with the total score. The original approach for interval scaled variables (Zumbo & Thomas, 1997) was expanded to ordinal variables using ordered logit regressions. Likelihoods of model fits can be tested via χ^2 tests; χ^2 tests are highly sensitive to large sample size, and thus, trivial discrepancy may lead to rejection of models (Chen, 2007). Therefore, Zumbo (1999) suggested that when χ^2 tests were significant, differences in R^2 (i.e., ΔR^2) should be used as a measure of effect size. The pseudo- R^2 developed by McKelvey and Zavoina (1975) was used, as the best approximation of the concept of explained variance in multiple linear regression models (Laitila, 1993; Thomas, Zhu, Zumbo, & Dutta, 2008). The effects sizes originally suggested by Zumbo and Thomas (1997) were seen as too generous (Hidalgo & López-Pina, 2004), potentially overlooking DIF. More conservative criteria were suggested by Jodoin and Gierl (2001) and used in the present study, with $\Delta R^2 < 0.035$ being negligible, $0.035 \leq \Delta R^2 \leq 0.070$ flagging moderate DIF, and $\Delta R^2 > 0.070$ suggesting large DIF. Tests were run in SPSS.

The third approach used, similar to IRT, was MG-CFA: separate models for each item must be run. Mplus was used to test for DIF. Following Elosua (2011) and Stark et al. (2006), we first tested a model with unconstrained parameters (free baseline approach) and then determined DIF, separately for each item, by setting difficulties and discrimination parameters to be equal across groups. Again, significance testing using χ^2 tests of loglikelihoods can be performed but is sensitive to sample size. ΔCFI and $\Delta RMSEA$ smaller than 0.01 or 0.015 (Chen, 2007) can be used as effect size measures pointing to

negligible DIFs. Both cutoffs were used to indicate weak (>0.01 to <0.015) and strong (≥ 0.015) indication of DIF.

2.3.4 | Measurement invariance

The short form CIUS was tested for invariance in three steps using hierarchically nested models. First, the model for configural invariance (same number of factors) was estimated and used as the comparison model for metric (weak) invariance (same discrimination parameters across groups). The model for metric invariance was then used as the counterfactual model for scalar (strong) invariance, that is, for the same item difficulties across groups (Brown, 2006). Differences between models were tested using χ^2 tests. To avoid high sensitivity to large sample sizes, when χ^2 tests were significant, we used differences between RMSEA (Δ RMSEA) and CFI (Δ CFI) as effect size measures. Significant χ^2 differences are assumed to be negligible if Δ RMSEA and Δ CFI < 0.01 (Elosua, 2011). According to Chen (2007), a $|\Delta$ CFI| value ≥ 0.01 supplemented by a $|\Delta$ RMSEA| value ≥ 0.015 indicates the metric and scalar noninvariance between groups. CFA invariance models were tested with Mplus. We distinguished weak (>0.01 to <0.015) and strong (≥ 0.015) indication of noninvariance. The methods used for invariance testing allow only two group comparisons. As the sample size for the Italian-speaking sample was too small, French- and Italian-speaking (Latin languages) samples were combined. Similarly, compulsive internet use is very rare among older people. Therefore, invariance testing excluded the age group of those 35 years and older. Steps were repeated after each item's elimination—an iterative process called *purification* (Hidalgo-Montesinos & Gómez-Benito, 2003; Zumbo, 1999).

2.3.5 | Area under the receiver operating characteristic (ROC) curve and cutoffs

ROC curves to determine the cutoffs for the short form with the long form cutoffs as comparison were used. According to Hosmer and Lemeshow (2000), an area under the ROC curve of >0.9 shows

outstanding properties. Generally, sensitivities and specificities >0.9 are considered good and were used in the present study. There are little validation studies of a cutoff for CIUS. Recommendations for cutoffs go as low as 18–21 (e.g., Besser et al., 2017; Guertler et al., 2014), which may be useful for case finding for further in-depth investigation or early interventions (low cutoff and high sensitivity). In general population studies, the aim commonly is on prevalence which may be overestimated with low cutoffs and are aimed at high specificity (Maraz, Kiraly, & Demetrovics, 2015). We compared the short and full form using Meerkert's (2007) suggested cutoff of 28 points and above, which was most often used internationally. Studies using latent class analysis (Rumpf et al., 2014; Van Rooij, Schoenmakers, Vermulst, Van den Eijnden, & Van de Mheen, 2011) found groups demarcating potential addiction and high-risk use from high engagement corresponding to approximately CIUS summary scores of 28. If there were more than one cutoff with sensitivity and specificity >0.9 , Youden's index (Youden, 1950), which is basically the sum of sensitivity and specificity, was used to recommend a cutoff.

3 | RESULTS

The sample was a good representation of Switzerland's population, with more than two thirds being German-speaking (Table 2). PIU showed a rapid decrease as respondents got older.

3.1 | Unidimensionality

The Kaiser–Meyer–Olkin value for the total learning sample was a *marvelous* 0.918, varying in subgroups between 0.869 (age 35+) and 0.925 (women). MSA for items in the total sample varied between a *meritorious* 0.861 and a *marvelous* 0.959. MSA values for items were >0.8 in all subgroups, and mostly >0.9 , demonstrating the adequacy of items for exploratory factor analysis. Despite eigenvalues for a second factor being >1 (the biggest second eigenvalue was 1.34 among those age 35+), Catell's scree plots clearly showed a single factor for

TABLE 2 Sample characteristics of training and learning samples

	Learning sample (n = 1,371)		Validation sample (n = 1,550)	
	Percentage	Percentage with full CIUS screened positive (cutoff = 28)	Percentage	Percentage with full CIUS screened positive (cutoff = 28)
Sex				
Men	50.3	2.5	51.4	2.5
Women	49.7	4.7	48.6	2.8
Region				
German	67.0	3.2	71.2	2.0
French/Italian	33.0	4.4	28.8	4.3
Age (years)				
15–19	34.2	7.0	22.5	7.5
20–34	37.5	2.9	27.4	2.4
35+	28.3	0.3	50.1	0.6

Note. CIUS = Compulsive Internet Use Scale.

the total sample and in all subgroups by sex, linguistic region, and age. Neither the parallel analysis (Horn, 1965) nor Velicer's minimum average partial test indicated any second factors, either in the total sample or in subgroups. The total sample's Cronbach's alpha was 0.889 and varied between 0.842 and 0.903 in subsamples, showing good internal consistency (DeVellis, 2012). CFA yielded an acceptable CFI of 0.922 but a mediocre RMSEA of 0.107. Discriminatory parameters in IRT varied between 1.38 (Item 14) and 2.07 (Item 10), considered high to very high, according to Baker (2001).

3.2 | First item selection based on MI and DIF

The highest MI in CFA suggested to free error correlations between Items 12 and 13 ($MI = 643.5$), 1 and 2 ($MI = 242.3$), 8 and 9 ($MI = 193.4$), and 6 and 7 ($MI = 93.7$). These were four of the five item pairs identified in Meerkerk's (2007) original study and used for item reduction by Cartierre et al. (2011).

Table 3 shows flags for the DIF tests (test values available in Table S1) for the four item pairs. According to the DIF tests, Item 2 should be excluded in favor of Item 1. Item 6 showed more DIF as regards sex and was comparable with Item 7 as regards linguistic

region and age: thus, Item 6 was excluded. Item 8 showed similar DIF to Item 9 for sex but more DIF as regards linguistic region and age: thus, Item 8 was excluded. Deciding between Items 12 and 13 was difficult. Item 13 demonstrated slightly worse DIF for sex and age, but more DIF than Item 12 as regards linguistic region. Furthermore, findings were contradictory as regards UIRT and CFA for region. Item 12 was eventually retained, also because it was kept in the German-language short form (Besser et al., 2017) and the Cartierre et al. (2011) short form. Table 1 shows the retained items.

3.3 | Purification steps for item selection

The 10-item solution resulted in a good RMSEA of 0.051 and an excellent CFI of 0.983 in CFA, and its largest MI (39.19) was found for the error correlation between Items 3 and 9. We therefore tested for invariance using CFA, alternately excluding Items 3 and 9, and tested for DIF using IRT (Table 4). Invariance testing with CFA showed that excluding Item 3 or 9 yielded similar results as regards sex (decreased RMSEA, decreased CFI, and significant χ^2 tests for DIF with IRT). The items also performed similarly for linguistic region (with increased RMSEA and decreased CFI for scalar invariance, and nonsignificant

TABLE 3 Comparison of differential item functioning criteria for item pairs 1 and 2, 6 and 7, 8 and 9, and 12 and 13

			Item pairs							
			Item 1	Item 2	Item 6	Item 7	Item 8	Item 9	Item 12	Item 13
Sex										
UIRT	LR test		*		*		**	***	*	***
	Effect size				*		*	**	***	***
Zumbo	ΔR^2									*
	$\Delta \chi^2$				*		***	***	**	***
CFA	$\Delta RMSEA$									
	ΔCFI									
	$\Delta \chi^2$				***	**	***	***	*	**
Region										
UIRT	LR test		*	***	*	*	**		***	
	Effect size			***		*	***		***	
Zumbo	ΔR^2			*					*	
	$\Delta \chi^2$			***		*	**		***	
CFA	$\Delta RMSEA$									
	ΔCFI									
	$\Delta \chi^2$			**	***	*	*		*	
Age										
UIRT	LR test				*		*			
	Effect size				**	**				
Zumbo	ΔR^2		*		*	*	*	**	*	**
	$\Delta \chi^2$		***	**	**	***	***	**	***	***
CFA	$\Delta RMSEA$									
	ΔCFI									
	$\Delta \chi^2$		*	***	*	*				

Note. UIRT: unidimensional item response theory (add-on in Stata); CFA: confirmatory factor analysis; LR: likelihood ratio; Zumbo: regression approach for differential item functioning analysis developed by Zumbo (1999); RMSEA: root mean square error of approximation; CFI: comparative fit index; Δ : difference; χ^2 : chi-squared. Significance levels for LR test in UIRT, for $\Delta \chi^2$ in the Zumbo approach and in the CFA: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Levels for ΔCFI , $\Delta RMSEA$: * >0.01 , ** >0.015 . Levels for ΔR^2 : * >0.01 , ** >0.035 , *** >0.07 . Effect size after Wainer: * >0.05 , ** >0.10 , *** >0.15 ; effect sizes are only shown for significant LR.

TABLE 4 Unidimensional item response theory analysis of DIF and CFA invariance testing for decision to exclude Item 3 or 9

	Type of invariance	CFA						DIF testing with IRT					
		Without Item 3			Without Item 9			Item 3			Item 9		
		RMSEA	CFI	χ^2 p value	RMSEA	CFI	χ^2 p value	χ^2 p value	Effect size (ref. Group 1)	Effect size (ref. Group 2)	χ^2 p value	Effect size (ref. Group 1)	Effect size (ref. Group 2)
Total		0.048	0.987		0.041	0.990							
Sex	Configural	0.044	0.990		0.041	0.991		0.017	-0.109	0.104	< 0.001	0.149	-0.135
	Metric	0.041	0.990	0.019	0.036	0.992	0.073						
	Scalar	0.040	0.982	<0.001	0.033	0.989	<0.001						
Region	Configural	0.048	0.987		0.044	0.988		0.232	0.003	-0.002	0.085	-0.017	0.023
	Metric	0.042	0.988	0.028	0.039	0.990	0.079						
	Scalar	0.052	0.967	<0.001	0.047	0.977	<0.001						
Age	Configural	0.050	0.981		0.042	0.986		<0.001	-0.399	0.296	0.923	-0.047	0.034
	Metric	0.047	0.981	0.017	0.042	0.984	0.020						
	Scalar	0.040	0.979	<0.001	0.050	0.963	<0.001						

Note. UIRT: unidimensional item response theory (add-on in Stata); RMSEA: root mean square error of approximation; CFI: comparative fit index; χ^2 : chi-squared; DIF: differential item functioning; CFA: confirmatory factor analysis. χ^2 p value in CFA as a test against the preceding model, that is, test for metric invariance against configural invariance, and for scalar invariance against metric invariance. χ^2 p value in IRT models are the tests for DIF of the corresponding item. Ref group: reference group; Group 1 (Group 2) for sex: male (female); for region: German (French/Italian), for age: 15–19 years (20–34 years).

DIF tests in IRT models), with a small advantage when excluding Item 9 (excluding Item 3 would lead to Δ CFI = 0.02; RMSEA > 0.05 in the scalar invariance model, and the DIF for Item 3 was smaller than for Item 9, i.e., χ^2 p = 0.2323 for Item 3 and p = 0.0847 for Item 9). Item 3 performed worse with regard to invariance and DIF for differences in age groups. Excluding Item 3 decreased RMSEA in the scalar invariance model and decreased CFI by only 0.002. RMSEA with Item 9 excluded increased RMSEA and decreased CFI by 0.023. DIF testing in IRT indicated highly significant DIF (LR = 47.97; p < 0.0001) and a very large effect size for Item 3, whereas DIF testing was highly non-significant (LR = 0.909, p = 0.9233) for Item 9. Furthermore, the discrimination parameter in IRT was lower for Item 3 (a = 1.48) than for Item 9 (a = 2.01). Thus, Item 3 was excluded.

As a final purification step, we rechecked for DIF in IRT and found that Item 10 had the poorest overall DIF of all the remaining items, that is, the second largest LR value for sex (behind Item 9) and region (behind Item 11), and the largest LR value for age. After excluding Item 10 (Table 5), RMSEA for the total sample further decreased (from 0.048 to 0.045) and CFI increased (from 0.987 to 0.989). The short form scale was metric invariant for sex, region, and age as regards Δ RMSEA (criterion < 0.015) and Δ CFI (criterion < 0.01). Scalar invariance was also found, but not for linguistic

region as regards Δ CFI (=0.013), which fell short of the strong criterion (Δ CFI < 0.01).

The eight-item short form's Cronbach's alpha was a good 0.829 (DeVellis, 2012) and correlated with the full CIUS form at r = 0.966.

3.4 | Area under the ROC curve

Compared with the full CIUS (Cutoff 28), cutoffs for the short form were generally between 12 and 14 (except for age, where a cutoff from 12 to 16 was found for the older age group), with both specificities and sensitivities for the total sample and subsamples being >0.9 (Table 6). Youden's index indicated a better fit with slightly lower sensitivity. Hence, we suggest a cutoff of 13.

3.5 | Validation sample

The eight-item solution revealed RMSEA of 0.036 and CFI of 0.995 (Table 7) in the validation sample. The scale was metric invariant for sex, region, and age as regards Δ RMSEA (criterion < 0.01) and Δ CFI (criterion < 0.01). Scalar invariance was found for region and age, and just fell short for sex (Δ RMSEA = 0.013; Δ CFI = 0.013), however,

TABLE 5 Comparison of the eight-item solution with the nine-item solution

	Type of invariance	Nine-item solution				Eight-item solution			
		RMSEA	CFI	$\Delta\chi^2$	p Wert	RMSEA	CFI	$\Delta\chi^2$	p Wert
Total		0.048	0.987			0.045	0.989		
Sex	Configural	0.044	0.990			0.047	0.988		
	Metric	0.041	0.990	18.2	0.019	0.043	0.988	16.5	0.021
	Scalar	0.040	0.982	81.0	<0.001	0.041	0.982	62.3	<0.001
Region	Configural	0.048	0.987			0.043	0.990		
	Metric	0.042	0.988	15.7	0.028	0.041	0.989	16.2	0.023
	Scalar	0.052	0.967	131.7	<0.001	0.047	0.976	97.9	<0.001
Age	Configural	0.050	0.981			0.048	0.983		
	Metric	0.047	0.981	17.1	0.017	0.045	0.983	13.7	0.054
	Scalar	0.040	0.979	48.9	<0.001	0.036	0.981	38.4	0.139

Note. RMSEA: root mean square error of approximation; CFI: comparative fit index; $\Delta\chi^2$: difference chi-squared. The excluded item in the eight-item solution compared with the nine-item solution is Item 10 (cf., Table 1).

TABLE 6 Area under the ROC curve, sensitivities, and specificities for the total sample and for sex, region, and age for the eight-item solution, with both sensitivity and specificity >0.9

	CIUS ≥ 28			
	Area (SE)	Sensitivity	Specificity	Cutoff
Total	0.989 (0.003)	0.98–0.94	0.91–0.96	12–14
Sex				
Men	0.990 (0.005)	1.00–0.94	0.93–0.95	13–14
Women	0.990 (0.004)	0.97–0.94	0.92–0.97	12–14
Region				
German	0.991 (0.004)	0.97–0.93	0.93–0.97	12–14
French/Italian	0.990 (0.005)	1.00–0.95	0.90–0.94	13–14
Age				
15–19	0.985 (0.005)	0.94	0.93	14
20–34	0.991 (0.007)	0.93–0.92	0.92–0.99	12–16

Note. SE: standard error; CIUS: compulsive internet use scale; ROC: receiver operating characteristic.

TABLE 7 Invariance testing in with confirmatory factor analysis in the validation sample

	Type of invariance	Eight-item solution			
		RMSEA	CFI	$\Delta\chi^2$	p value
Total		0.036	0.995		
Sex	Configural	0.029	0.996		
	Metric	0.028	0.995	10.0	0.186
	Scalar	0.041	0.982	58.4	0.001
Region	Configural	0.043	0.990		
	Metric	0.041	0.989	16.3	0.023
	Scalar	0.044	0.983	81.5	<0.001
Age	Configural	0.047	0.987		
	Metric	0.035	0.992	5.0	0.663
	Scalar	0.029	0.991	34.7	0.254

Note. RMSEA: root mean square error of approximation; CFI: comparative fit index; $\Delta\chi^2$: difference chi-squared.

even the scalar invariant model for sex had an excellent fit (RMSEA = 0.041, CFI = 0.982). The area under the ROC curve again indicated outstanding accuracy (Hosmer & Lemeshow, 2000), and sensitivities and specificities indicated a cutoff similar to the learning sample (area = 0.989, sensitivity = 0.976, specificity = 0.950 for a cutoff of 13). Cronbach's alpha was .821 (good) and correlated with the full CIUS at $r = 0.969$.

4 | DISCUSSION

The present study suggests that a short form of CIUS could be used with no loss of accuracy in multitopic, general population surveys. The eight-item form was constructed with a view to (a) avoiding redundant items (correlated errors) and (b) being invariant with regard to sex, linguistic region, and age. The scale can be assumed to be scalar invariant (borderline for sex in the validation sample; Δ RMSEA = 0.013; Δ CFI = 0.013). The unidimensional short form showed good internal consistency (Cronbach's $\alpha > 0.8$ in both the learning and validation samples). The psychometric properties were in line with validations for the full scale in different countries (Dhir et al., 2015; Dhir et al.,

2016) and with the seven-item and five-item short forms in German (Besser et al., 2017).

To the best of our knowledge, there have been only two previous suggestions for a short form CIUS: one in France (Cartierre et al., 2011) and one in Germany (Besser et al., 2017). The eight items in our short form are all in Cartierre et al.'s nine-item short form Cartierre et al. (2011), confirming their findings. It extends their work, as their short form was developed on a sample of school children aged around 13, whereas the present scale was developed using a general population survey and is invariant across different age groups. Our short form also contains six of the seven items in the seven-item German short form (Besser et al., 2017). The advantage of Besser et al.'s study is its validation against structured clinical interviews. However, it did not use a general population sample. The present study could not use clinical diagnosis, but the congruence between the items in the Swiss, German, and French short form versions shows that virtually the same items have proven valid and clinically relevant. Our approach was also different from Besser et al.'s (2017) as we aimed to keep at least one item from each of the underlying dimensions suggested by Meerkerk (2007): loss of control, conflict, preoccupation, coping, and withdrawal. Besser et al.'s seven-item short form could not cover withdrawal (Item 14), and their even shorter, five-item form could not cover preoccupation.

One limitation of the present study is that the short form could not be validated against a gold standard diagnosis, but only against the full 14-item form. Nevertheless, the CIUS is widely used in many different countries and has proved useful in many studies (see, e.g., overviews in Dhir et al., 2016; Guertler et al., 2014). Correlation with the full CIUS was >0.96 in both the learning and validation samples. Hence, at the very least, we can conclude that if the full CIUS is a valid form, then the short form presented here is too. The high correlation with the full form is probably mainly due to the elimination of redundant items. The item-pair error correlations in the present study have been shown in many others (see, e.g., overviews in Dhir et al., 2015; Dhir et al., 2016), indicating that they are a general problem of the CIUS. In conclusion, the CIUS short form is a highly reliable instrument, which is invariant across sex, linguistic region, and age. It provides a reasonable alternative to the full CIUS form when

questionnaire space is limited. A next steps would be to compare this scale with similar short scales like the Internet Gaming Disorder Scale-SF9 (Pontes & Griffiths, 2015), which is on gaming only, or the eight-item of the Internet Addiction Diagnostic Questionnaire (Young, 1998), which was based on gambling disorder. Such comparisons would also be helpful to analyze the appropriateness of some items in general population studies among subgroups. For example, items referring to work or family may not be appropriate for someone living alone or being unemployed. This concerns also our choice of keeping at least one item of all six dimensions suggested by Meerkerk (2007), which may not be necessary for a screening tool. For example, there is still debate on withdrawal as a criterion of behavioral addictions. Withdrawal has not been established using physiological measures comparable with those used in patients dependent on substances (Van Rooij & Prause, 2014) and may be either metaphorical or describe behavioral criteria, for example, the patient's complaints of feeling irritable or anxious (Pies, 2009). The lack of chemically operating tolerance is one of the major arguments of some researchers not to accept PIU as a "real" disorder, because it lacks criteria related to physical dependence (see also Griffiths et al., 2016).

ACKNOWLEDGEMENT

This study was funded by the Federal Office of Public Health (FOPH 16.924108/204.0001-1511/1).

CONFLICTS OF INTEREST

None of the authors has any conflict of interest.

ORCID

Gerhard Gmel <https://orcid.org/0000-0001-9403-9405>

Stéphanie Baggio <https://orcid.org/0000-0002-5347-5937>

REFERENCES

- Alavi, S. S., Jannatifard, F., Eslami, M., & Rezapour, H. (2011). Validity, reliability and factor analysis of compulsive internet use scale in students of Isfahan's universities. *Health Information Management*, 7, 715–724.
- American Psychiatric Association (APA) (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596>
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). Portsmouth NH: Heinemann.
- Bélanger-Lejars, V. O. (2015). Internet gaming disorder and gambling disorder: A comparison of individual psychological factors. *Journal of Psychology*, 3, 122–140. <https://doi.org/10.15640/jpbs.v3n2a12>
- Besser, B., Rumpf, H. J., Bischof, A., Meerkerk, G. J., Higuchi, S., & Bischof, G. (2017). Internet-related disorders: Development of the short compulsive internet use scale. *Cyberpsychology, Behavior and Social Networking*, 20, 709–717. <https://doi.org/10.1089/cyber.2017.0260>
- Brand, M., Young, K. S., & Laier, C. (2014). Prefrontal control and internet addiction: A theoretical model and review of neuropsychological and neuroimaging findings. *Frontiers in Human Neuroscience*, 8, 375. <https://doi.org/10.3389/fnhum.2014.00375>
- Breslau, J., Aharoni, E., Pedersen, E. R., & Miller, L. L. (2015). *A review of research on problematic internet use and well-being: With recommendations for the US Air Force*. Santa Monica, Calif: RAND Corporation.
- Brown, T. A. (2006). *Confirmatory factor analyses for applied research*. New York, NY: Guilford Press.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen, & J. S. Long (Eds.), *Testing structural models* (pp. 136–162). Newbury Park, CA: Sage.
- 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. <https://doi.org/10.1016/j.neurenf.2011.06.003>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504. <https://doi.org/10.1080/10705510701301834>
- Davis, R. A. (2001). A cognitive-behavioral model of pathological Internet use. *Computers in Human Behavior*, 17, 187–195. [https://doi.org/10.1016/S0747-5632\(00\)00041-8](https://doi.org/10.1016/S0747-5632(00)00041-8)
- Davis, R. A., Flett, G. L., & Besser, A. (2002). Validation of a new scale for measuring problematic Internet use: Implications for pre-employment screening. *Cyberpsychology & Behavior*, 5, 331–345. <https://doi.org/10.1089/109493102760275581>
- De Cock, R., Vangeel, J., Klein, A., Minotte, P., Rosas, O., & Meerkerk, G. J. (2014). Compulsive use of social networking sites in Belgium: Prevalence, profile, and the role of attitude toward work and school. *Cyberpsychology, Behavior and Social Networking*, 17, 166–171. <https://doi.org/10.1089/cyber.2013.0029>
- DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Los Angeles: Sage publications.
- Dhir, A., Chen, S., & Nieminen, M. (2015). Psychometric validation of the Chinese compulsive internet use scale (CIUS) with Taiwanese high school adolescents. *Psychiatric Quarterly*, 86, 581–596. <https://doi.org/10.1007/s11126-015-9351-9>
- Dhir, A., Chen, S., & Nieminen, M. (2016). Psychometric validation of the compulsive internet use scale: Relationship with adolescents' demographics, ICT accessibility, and problematic ICT Use. *Social Science Computer Review*, 34, 197–214. <https://doi.org/10.1177/0894439315572575>
- Dodeen, H., & Al-Darmaki, F. (2016). The application of item response theory in developing and validating a shortened version of the Emirate Marital Satisfaction Scale. *Psychological Assessment*, 28, 1625–1633. <https://doi.org/10.1037/pas0000296>
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16(Suppl 1), 5–18. <https://doi.org/10.1007/s11136-007-9198-0>
- Elosua, P. (2011). Assessing measurement equivalence in ordered-categorical data. *Psicologica: International Journal of Methodology and Experimental Psychology*, 32, 403–421. doi: v32 n2 p403-421 2011
- Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Green, K. E., & Frantom, C. G. (2002). *Survey development and validation with the Rasch model*. Paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing, Charleston, SC.
- Griffiths, M. (1999). Internet addiction: Fact or fiction? *The Psychologist*, 12, 246–250.
- Griffiths, M. (2000). Does Internet and computer "addiction" exist? Some case study evidence. *Cyber Psychology and Behavior*, 3, 211–218. <https://doi.org/10.1089/109493100316067>
- Griffiths, M., Van Rooij, A., Kardefelt-Winther, D., Starcevic, V., Király, O., Pallesen, S., ... King, D. L. (2016). Working towards an international consensus on criteria for assessing internet gaming disorder: A critical commentary on Petry et al. (2014). *Addiction*, 111, 167–175. <https://doi.org/10.1111/add.13057>
- Guertler, D., Broda, A., Bischof, A., Kastirke, N., Meerkerk, G. J., John, U., ... Rumpf, H. J. (2014). Factor structure of the compulsive internet use scale. *Cyberpsychology, Behavior and Social Networking*, 17, 46–51. <https://doi.org/10.1089/cyber.2013.0076>
- Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: Concerns and considerations. *Computational Methods in Social Sciences*, 3, 5.

- 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. <https://doi.org/10.1177/0013164403261769>
- 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). <https://doi.org/10.1027//1015-5759.19.1.1>, 1–11.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185. <https://doi.org/10.1007/BF02289447>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471722146>
- Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>
- 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. https://doi.org/10.1207/S15324818AME1404_2
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35, 401–415. <https://doi.org/10.1007/BF02291817>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31–36. <https://doi.org/10.1007/BF02291575>
- Kardefelt-Winther, D. (2017). Conceptualizing internet use disorders: Addiction or coping process? *Psychiatry and Clinical Neurosciences*, 71, 459–466. <https://doi.org/10.1111/pcn.12413>
- Kardefelt-Winther, D., Heeren, A., Schimmenti, A., van Rooij, A., Maurage, P., Carras, M., ... Billieux, J. (2017). How can we conceptualize behavioural addiction without pathologizing common behaviours? *Addiction*, 112, 1709–1715. <https://doi.org/10.1111/add.13763>
- 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. <https://doi.org/10.1007/s1126-012-9210-x>
- Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling: A Multidisciplinary Journal*, 18, 212–228. <https://doi.org/10.1080/10705511.2011.557337>
- 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. <https://doi.org/10.1111/j.1745-3984.2007.00029.x>
- Kuss, D., Griffiths, M., Karila, L., & Billieux, J. (2014). Internet addiction: A systematic review of epidemiological research for the last decade. *Current Pharmaceutical Design*, 20, 4026–4052. <https://doi.org/10.2174/13816128113199990617>
- Kuss, D. J., & Griffiths, M. D. (2012). Internet and gaming addiction: A systematic literature review of neuroimaging studies. *Brain Sciences*, 2, 347–374. <https://doi.org/10.3390/brainsci2030347>
- Kuss, D. J., Shorter, G. W., van Rooij, A. J., Griffiths, M. D., & Schoenmakers, T. M. (2014). Assessing internet addiction using the parsimonious internet addiction components model—A preliminary study. *International Journal of Mental Health and Addiction*, 12, 351–366. <https://doi.org/10.1007/s11469-013-9459-9>
- Laconi, S., Rodgers, R. F., & Chabrol, H. (2014). The measurement of Internet addiction: A critical review of existing scales and their psychometric properties. *Computers in Human Behavior*, 41, 190–202. <https://doi.org/10.1016/j.chb.2014.09.026>
- Laitila, T. (1993). A pseudo- R^2 measure for limited and qualitative dependent variables. *Journal of Econometrics*, 56, 341–356. [https://doi.org/10.1016/0304-4076\(93\)90125-O](https://doi.org/10.1016/0304-4076(93)90125-O)
- Lortie, C. L., & Guitton, M. J. (2013). Internet addiction assessment tools: Dimensional structure and methodological status. *Addiction*, 108, 1207–1216. <https://doi.org/10.1111/add.12202>
- Maraz, A., Kiraly, O., & Demetrovics, Z. (2015). Commentary on: Are we overpathologizing everyday life? A tenable blueprint for behavioral addiction research. The diagnostic pitfalls of surveys: If you score positive on a test of addiction, you still have a good chance not to be addicted. *Journal of Behavioral Addictions*, 4, 151–154. <https://doi.org/10.1556/2006.4.2015.026>
- McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4, 103–120. <https://doi.org/10.1080/0022250X.1975.9989847>
- Meerkerk, G.-J. (2007). Pwned by the Internet. *Erasmus University Rotterdam*. doi:hdl.handle.net/1765/10511
- Meerkerk, G. J., Van Den Eijnden, R. J., Vermulst, A. A., & Garretsen, H. F. (2009). The compulsive internet use scale (CIUS): Some psychometric properties. *Cyberpsychology & Behavior*, 12, 1–6. <https://doi.org/10.1089/cpb.2008.0181>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (Eighth ed.). Los Angeles, CA: Muthén & Muthén.
- Park, B., Han, D. H., & Roh, S. (2016). Neurobiological findings related to Internet use disorders. *Psychiatry and Clinical Neurosciences*, 71, 467–478. <https://doi.org/10.1111/pcn.12422>
- Pies, R. (2009). Should DSM-V designate “Internet addiction” a mental disorder? *Psychiatry (Edmont)*, 6, 31–37.
- Pontes, H. M., & Griffiths, M. D. (2015). Measuring DSM-5 Internet gaming disorder: Development and validation of a short psychometric scale. *Computers in Human Behavior*, 45, 137–143. <https://doi.org/10.1016/j.chb.2014.12.006>
- Revicki, D. A., Chen, W. H., & Tucker, C. (2015). Developing item banks for patient-reported health outcomes. In S. P. Reise, & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 334–363). New York, NY: Routledge.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 27, 151–161. <https://doi.org/10.1177/0146167201272002>
- Rumpf, H. J., Vermulst, A. A., Bischof, A., Kastirke, N., Gurtler, D., Bischof, G., ... Meyer, C. (2014). Occurrence of internet addiction in a general population sample: A latent class analysis. *European Addiction Research*, 20, 159–166. <https://doi.org/10.1159/000354321>
- Shepard, L., Camilli, G., & Averill, M. (1981). Comparison of procedures for detecting test-item bias with both internal and external ability criteria. *Journal of Educational Statistics*, 6, 317–375. <https://doi.org/10.3102/10769986006004317>
- Starcevic, V., & Aboujaoude, E. (2017). Internet addiction: Reappraisal of an increasingly inadequate concept. *CNS Spectrums*, 22, 7–13. <https://doi.org/10.1017/s1092852915000863>
- 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, 1292–1306. <https://doi.org/10.1037/0021-9010.91.6.1292>
- Suh, Y. (2016). Effect size measures for differential item functioning in a multidimensional IRT model. *Journal of Educational Measurement*, 53, 403–430. <https://doi.org/10.1111/jedm.12123>
- Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief to medium-length Big Five and Big Six Personality Questionnaires. *Psychological Assessment*, 23, 995–1009. <https://doi.org/10.1037/a0024165>
- 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). <https://doi.org/10.22237/jmasm/1209614580>, 21–38.
- Tokunaga, R. S., & Rains, S. A. (2016). A review and meta-analysis examining conceptual and operational definitions of problematic internet use. *Human Communication Research*, 42, 165–199. <https://doi.org/10.1111/hcre.12075>

- van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 29–37. <https://doi.org/10.1027/1015-5759.13.1.29>
- Van Rooij, A., & Prause, N. (2014). A critical review of "Internet addiction" criteria with suggestions for the future. *Journal of Behavioral Addictions*, 3, 203–213. <https://doi.org/10.1556/JBA.3.2014.4.1>
- Van Rooij, A. J., Schoenmakers, T. M., Vermulst, A. A., Van den Eijnden, R. J., & Van de Mheen, D. (2011). Online video game addiction: Identification of addicted adolescent gamers. *Addiction*, 106, 205–212. <https://doi.org/10.1111/j.1360-0443.2010.03104.x>
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321–327. <https://doi.org/10.1007/BF02293557>
- 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.
- Weinstein, A., & Lejoyeux, M. (2010). Internet addiction or excessive internet use. *Journal of Drug and Alcohol Abuse*, 36, 277–283. <https://doi.org/10.3109/00952990.2010.491880>
- Widyanto, L., & Griffiths, M. (2006). 'Internet addiction': A critical review. *International Journal of Mental Health and Addiction*, 4, 31–51. <https://doi.org/10.1007/s11469-006-9009-9>
- Wittchen, H., Beloch, E., Garczynski, E., Holly, A., Lachner, G., Perkonig, A., ... Ziegler, S. (1995). *Münchener Composite International Diagnostic Interview (M-CIDI), Version 2.2*. München: Max-Planck-Institut für Psychiatrie.
- Yellowlees, P. M., & Marks, S. (2007). Problematic Internet use or Internet addiction? *Computers in Human Behavior*, 23, 1447–1453. <https://doi.org/10.1016/j.chb.2005.05.004>
- Yong, R. K. F., Inoue, A., & Kawakami, N. (2017). The validity and psychometric properties of the Japanese version of the compulsive internet use scale (CIUS). *BMC Psychiatry*, 17, 201. <https://doi.org/10.1186/s12888-017-1364-5>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3, 32–35. [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)
- Young, K. S. (1996). Psychology of computer use: XL. Addictive use of the Internet: A case that breaks the stereotype. *Psychological Reports*, 79, 899–902. <https://doi.org/10.2466/pr0.1996.79.3.899>
- Young, K. S. (1998). Internet addiction: The emergence of a new clinical disorder. *Cyberpsychology & Behavior*, 1, 237–244. <https://doi.org/10.1089/cpb.1998.1.237>
- 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.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Gmel G, Khazaal Y, Studer J, Baggio S, Marmet S. Development of a short form of the compulsive internet use scale in Switzerland. *Int J Methods Psychiatr Res*. 2019;28:e1765. <https://doi.org/10.1002/mpr.1765>