



Review

Scales for measuring user engagement with social network sites: A systematic review of psychometric properties



Leif Sigerson, Cecilia Cheng*

Department of Psychology, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Article history:

Received 12 April 2017

Received in revised form

19 January 2018

Accepted 20 January 2018

Available online 1 February 2018

Keywords:

Social network site

Social networking site

Social media

Psychometrics

Validation

Facebook

ABSTRACT

In the past decade, various scales have been developed for measuring engagement with social network sites (SNS), but validity concerns have recently been raised about some of them. The present study thus provides a systematic review of the psychometric properties of these scales. This review included articles that aimed at either developing an SNS engagement scale or providing a systematic test of the psychometric properties of the scale. We conducted keyword-based searches of several broad multidisciplinary databases, along with reference list searches and article citation searches. These search strategies yielded a total of 14 reports, revealing validation evidence for 12 SNS engagement scales among 13,861 participants from 11 countries. There was mixed evidence for the various types of validity tests, with some scales having been validated more rigorously with multiple studies and samples while others having not yet been systematically validated. Sampling and acquiescence biases were also present for some scales. The present review provides recommendations for researchers intending to study SNS engagement. Although the literature search was multi-faceted, it may conceivably have missed studies that provided less rigorous validity evidence. Overall, this study contributes to evaluating and strengthening the methodological foundations of SNS research.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction	88
1.1. Social network sites	88
1.2. Measuring SNS engagement	88
1.3. Measurement issues addressed in this systematic review	89
1.3.1. Reliability	89
1.3.2. Validity	89
2. Methods	90
2.1. Selection criteria	90
2.2. Literature search and study selection	90
2.3. Coding process	90
2.4. Evaluations of psychometric properties of individual scales	91
2.4.1. Content validity	91
2.4.2. Reliability	92
2.4.3. Structural validity	92
2.4.4. Associations with other variables	94
2.4.5. Response processes	95
2.5. Evaluations of generalizability	95
2.6. Efforts to reduce publication bias	95
3. Results	95

* Corresponding author.

E-mail address: ceci-cheng@hku.hk (C. Cheng).

3.1.	Scales validated in at least two samples	95
3.1.1.	Facebook Intensity (FBI) scale	95
3.1.2.	Multi-dimensional Facebook Intensity Scale (MFIS)	96
3.1.3.	Media and Technology Usage and Attitudes Scale (MTUAS)	97
3.1.4.	Social Networking Time Use Scale (SONTUS)	97
3.1.5.	Social Networking Activity Intensity Scale (SNAIS)	97
3.1.6.	Facebook Psychological Involvement Scale (FPIS)	98
3.1.7.	Gravitating Towards Facebook (GoToFB) scale	98
3.1.8.	Social Media Use Integration Scale (SMUIS)	98
3.1.9.	Psycho-social Aspects of Facebook Use (PSAFU)	98
3.2.	Scales validated in one sample	99
3.2.1.	The Impact of Student's Social Network Use Scale (ISSNU) scale	99
3.2.2.	Facebook Usage Questionnaire (FBUQ)	99
3.2.3.	Facebook Questionnaire	99
4.	Discussion	100
4.1.	Implications for SNS engagement research	100
4.1.1.	Recommendations for scale selection	100
4.1.2.	Measurement of SNS intensity	101
4.2.	Implications for validation studies	101
4.2.1.	Additional tests of criterion validity	102
4.2.2.	Studies with diverse samples	102
4.2.3.	Studies with additional platforms	102
4.3.	Limitations	102
4.4.	Conclusions	102
	Conflicts of interest	102
	Acknowledgements	102
	References	103

1. Introduction

1.1. Social network sites

The rise of the Internet has spurred interest across multiple social sciences disciplines in how this new technology can transform social dynamics (e.g., Kraut et al., 2002; Wellman, 2002). The subsequent prevalence of social network sites (SNS) since the mid 2000's has only increased this academic interest. Boyd and Ellison (2007, see also Ellison & Boyd, 2013), define an SNS as “a networked communication platform in which participants 1) have uniquely identifiable profiles that consist of user-supplied content, content provided by other users, and/or system-provided data; 2) can publicly articulate connections that can be viewed and traversed by others; and 3) can consume, produce, and/or interact with streams of user-generated content provided by their connections on the site.” (Ellison & Boyd, 2013, p. 158). Although Facebook is the most widely-known SNS with the greatest number of users worldwide, many others, such as Twitter and LinkedIn, are also popular (Dreamgrow.com, 2017).

1.2. Measuring SNS engagement

Though SNS are becoming an integral part of many people's daily lives (Pew Research Center, 2011; Statista, 2018), multiple studies have shown that there are important individual differences in how SNS are used (C. Y. Lai & Yang, 2014; N. Park, Lee, & Kim, 2012). To capture these differences, researchers have used a variety of measurement techniques. Some techniques focus on objective behaviors (e.g., time spent on SNS, frequency of SNS behaviors), and data mining-based methods provide promising new opportunities to collect this type of data (Burke & Kraut, 2013; February). These methods may be particularly useful in situations where self-report methods are likely to be inaccurate, such as in estimating amount of time spent on Facebook or the Internet (Junco, 2013;

Scharkow, 2016).

These techniques, however, may be less adequate in measuring the psychological aspects of SNS use, such as emotional connection to the SNS (Jenkins-Guarnieri, Wright, & Johnson, 2013; Orosz, Tóth-Király, & Bóthe, 2016) or the motivations underlying SNS behaviors (Lomborg & Bechmann, 2014; Mahrt & Scharkow, 2013). Thus, misinterpretations of SNS use and behavior are more likely for studies that rely solely on objective measures (Ellison & Boyd, 2013). For example, an objective behavior such as re-posting someone else's status can indicate an affirmation of their relationship, or simply a superficial passing along of information (Boyd, Golder, & Lotan, 2010).

To capture psychological aspects of SNS use, several scales have been developed that assess the level of an individual's engagement with the SNS. Engagement refers to “a quality of user experience with technology that is characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect” (O'Brien & Toms, 2008, p. 23).

SNS engagement has been linked to a variety of important psychosocial outcomes, such as social capital (e.g., Steinfield, Ellison, & Lampe, 2008), self-esteem (e.g., Faraon & Kaipainen, 2014), and well-being (Ybarra, Résibois, Jonides, & Kross, 2017). McCay-Peet and Quan-Haase (2016) provide a comprehensive working definition of SNS engagement, stating that it is an umbrella construct that encompasses six major facets. These are (a) self-presentation, or the creation of a virtual identity on the SNS; (b) action and participation, comprising the different behaviors that can be performed on an SNS (e.g., retweeting, liking, friending); (c) uses and gratifications, which cover the user's motivations for using the SNS, such as for social or informational purposes; (d) positive experiences on the SNS, which help maintain the user's engagement; (e) usage and activity counts, which can represent either overall usage (e.g., time spent on the site), or data presented to the user about their behavior or the behavior of their SNS contacts (e.g.,

responses to their posts); and (f) social context, which shapes SNS behaviors, for example with norms and standards for a particular SNS community. This definition of engagement was developed for social media generally; but we consider that it applies equally well to SNS, a specific type of social media.

As SNS engagement reflects users' phenomenal experiences that are not observable, this construct is best assessed by self-reports. Among the various self-report measures of SNS engagement, the most widely adopted is the Facebook Intensity (FBI) scale, developed as part of a heavily cited study that examined the association between Facebook engagement and social capital (Ellison, Steinfield, & Lampe, 2007). Several scholars have noted a dearth of efforts to validate the FBI scale (e.g., Jenkins-Guarnieri et al., 2013). In addition, the study by Kuru and Pasek (2016) demonstrated that this popular scale is vulnerable to acquiescence bias, which artificially inflates its estimated reliability and distorts its relationships with other variables (e.g., social capital). Recently, researchers have constructed and validated a number of new self-report scales on SNS engagement. Given the availability of these other SNS engagement scales and the concerns about the methodological issues of the FBI scale, SNS researchers may wonder which of these scales will be most suitable for their future study. Thus, there is a need for a systematic guide to the utility of these various scales in terms of their psychometric properties.

1.3. Measurement issues addressed in this systematic review

The overarching aim of the present systematic review is to identify a list of scales developed for the measurement of SNS engagement and provide a comprehensive psychometric review of each of them. Scholars with an interest in improving measurement in SNS research can use this guide to identify important areas where more validation effort is needed, whereas other SNS researchers can use this guide to determine whether a particular scale is reliable and valid when planning for their future studies.

1.3.1. Reliability

The psychometric properties of a scale are comprised primarily of its reliability and validity. Reliability refers to "the proportion of variance attributable to the true score of the latent variable" (Devellis, 2003, p. 27), or the ability of a scale to accurately assess the latent variable that cannot be directly observed (e.g., depression, need for affiliation) rather than error. Thus, scales with higher reliability are generally deemed more desirable.

1.3.2. Validity

Furr and Bacharach (2013) posit that a valid scale must include four essential aspects: (a) content validity, (b) structural validity, (c) associations with other variables, and (d) response processes. The first aspect of scale validity, content validity, is "the extent to which a specific set of items reflects a content domain" (Devellis, 2003, p. 49). This means that the items of a scale should be a representative sample of the facets of the latent variable (Anastasi & Urbina, 1997, pp. 114–115).

The second aspect, structural validity, refers to "the way the parts of a test are related to each other" (Furr & Bacharach, 2013, p. 206). Each scale presumes certain associations between its items and the latent variables they measure, which is commonly referred to as a factor structure. This theoretical structure can be modeled with various statistical techniques (e.g., factor analysis) to see how well it actually fits the data derived from a sample. For a scale to demonstrate structural validity, this model must have a good fit to

the data, indicating that the proposed theoretical structure of the scale is consistent with the empirical source.

The third aspect of scale validity is associations with other variables. If a scale is truly measuring the latent variable, it should behave in a certain way in its relation to other variables (Devellis, 2003, p. 53). This broad definition includes criterion validity, convergent validity, discriminant validity, and incremental validity.

To evaluate a scale's criterion validity, researchers examine its relationships with outcomes of the variable being measured. For example, as depression and social anxiety are known consequences of problematic online gaming (e.g., Kuss & Griffiths, 2012), significant positive associations with these variables can support the criterion validity of a new scale measuring problematic online gaming (e.g., Sigerson, Li, Cheung, Luk, & Cheng, 2017).

To test convergent and discriminant validity, developers of a new scale usually review theories or past studies to check whether the variable they intend to measure is related to other target variables or not. For a scale to show convergent validity, it should have significant associations with conceptually similar variables. For example, depression is known to be positively correlated with anxiety because both are major indicators of emotional problems (e.g., Cheng, Cheung, Chio, & Chan, 2013) and thus any valid measure of depression should have a significant positive correlation with a validated measure of anxiety.

For a scale to have discriminant validity, it should be empirically distinguished from measures of seemingly related but conceptually dissimilar variables. Discriminant validity is commonly demonstrated by checking that associations between the scale and another target variable are not excessively strong, which would indicate that the scale can discriminate between the variable it is measuring and the target variable. For example, time playing online games is one of the multiple indicators of problematic online gaming. Although individuals with this problem and heavy online gamers both spend plenty of time playing online games, only the former group experience psychosocial dysfunction (Sigerson et al., 2017). Hence, a problematic online gaming scale showing discriminant validity should not have an excessively strong correlation with hours spent on online gaming, indicating the scale's efficacy in discriminating between problematic online gaming and gaming time.

Incremental validity is less commonly tested, but still a major indicator of scale validity because it demonstrates the unique contribution of a new scale over existing ones measuring the same construct (Hunsley & Meyer, 2003). If a scale has incremental validity, it can explain the variance of a known correlate of the latent variable, even when controlling for the effects of other relevant predictors. If a new scale can account for a significant portion of variance of a major correlate beyond that explained by existing scales, such a result justifies the need to construct a new one.

The fourth aspect of scale validity, response processes, means that scales should be free of response biases, which are "a measurement artifact which emerges from the context of a particular situation" (Nunnally & Berstein, 1994, p. 376). One common type is acquiescence bias where responses are influenced by respondents' desire to be agreeable (Knowles & Nathan, 1997). Another common type is social desirability bias where responses are influenced by respondents' tendency to give socially acceptable answers instead of what they truly think.

In summary, this review outlines a list of key criteria for psychometrically sound scales. We conducted a comprehensive literature search to identify scales constructed to measure SNS engagement, and then applied all these criteria to evaluate each reviewed scale.

2. Methods

2.1. Selection criteria

The overarching objective of the present review is to evaluate available reports that aim to develop a new scale or systematically validate an existing scale measuring engagement with either a particular SNS or SNS in general. We rely on [Ellison and Boyd's \(2013\)](#) well-established definition of SNS cited in the Introduction, but err on the side of inclusiveness, as evidenced by our numerous search terms (see section 2.2).

As noted at the outset, SNS engagement is a multifaceted construct that comprises six major facets: self-presentation, action and participation, uses and gratifications, positive experiences on the SNS, usage and activity counts, and social context. [McCay-Peet and Quan-Haase \(2016\)](#) maintained that no single facet can provide a satisfactory measure of this multifaceted construct. In this light, it is unrealistic to require potential scales to measure all these six facets, and thus we included scales that measure at least two of these facets to ensure a minimum level of content validity (see Section 1.3.2 for details). On the basis of this criterion, we excluded scales that only focus on a single dimension of SNS engagement, such as uses and gratifications of SNS (e.g., [Joinson, 2008 April](#)) or usage and activity counts ([Mäntymäki & Islam, 2016](#)).

By “systematically validate,” we mean any article whose primary aim is to evaluate several psychometric properties of a scale. This type of article is distinct from “substantive” articles, where scale validation is only a secondary aim, and which generally only provide partial evidence of a scale's psychometric properties (e.g., articles in which only the reliability of a scale is assessed). Though these substantive articles are not included in the formal review, we also surveyed them for additional validity evidence (see Section 2.2 for details).

For conciseness, our review aimed at evaluating scales that are likely to be useful to a broad audience. Hence, we excluded one scale that reported no empirical psychometric evidence ([Elliott & Polyakova, 2014](#)), as well as those that were developed for highly specific samples (e.g., medical students; A. T. [Wang, Sandhu, Wittich, Mandrekar, & Beckman, 2012](#)).

All these a priori inclusion and exclusion criteria were applied in the literature search. We did not limit the searches by academic discipline, time period (year), cultural region, publication status, and language in order to locate the maximum number of eligible reports.

2.2. Literature search and study selection

Literature searches were initially conducted in September 2016 and then updated in April 2017. The first author, who received doctoral-level training in advanced statistics including meta-analysis and systematic reviews, conducted the initial search and selection process, and refined this process after checks and recommendations from the senior author as well as an independent expert in SNS research.

Though there are no universal standards for article selection processes ([Shamseer et al., 2015](#)), we conducted the searches by consulting multiple sources ([Cooper, 2010](#)). The entire literature search and selection process is summarized in [Fig. 1](#). We sought to obtain as many relevant reports as possible using multiple search strategies, which involved five major steps.

In the first step, we adopted database search strategies. Specifically, we searched three broad multidisciplinary electronic databases (i.e., Scopus, EBSCOhost, and ProQuest) because SNS research emerges across various disciplines, not just social sciences. The databases cover social sciences (e.g., PsycINFO), health and medical

sciences (e.g., MEDLINE[®]), education (e.g., ERIC), and business studies (e.g., ABI/INFORM Global). As EBSCOhost and ProQuest are meta-databases, our search covered a total of 79 individual databases.

Keywords were used to identify potentially relevant reports. We paired a set of SNS-related keywords (i.e., “social network site,” “social networking site,” “online social network,” “social media,” “Facebook,” “Twitter,” “Renren,” “WeChat,” “Weibo,” “Whatsapp,” “Youtube,” “Instagram,” “Flickr,” “Pinterest,” “Digg,” and “Tumblr”) with another set of keywords related to scale development and assessment (i.e., “scale development,” “valid”,¹ “measur*,” “assess*,” “reliab*,” “internal consisten*,” “internally consisten*”).

In the second step, we used the same set of SNS-related keywords to search the PSYCtests database, where researchers can register scales that they develop. Using all of these database search strategies, we obtained 655 potentially relevant reports from Scopus, 677 from EBSCOhost, 526 from ProQuest, and 460 from PSYCtests, for a total of 2,318 results. After removing duplicate results using the Mendeley Desktop software (version 1.17.9), 1,907 abstracts of these reports were read to identify reports to review, narrowing down to 51 of such reports (see [Fig. 1](#) for details). Upon perusing these reports in greater detail, we omitted 39 reports that failed to meet the pre-determined selection criteria.

In the third step, the references section of each of the 12 reviewed reports was checked for additional scales, yielding 2 more reports. In the fourth step, we conducted reference list searches using Scopus. We downloaded abstracts for all 3,623 citations of the 14 included articles, and read their abstracts to look for additional validation studies or new scales. Scopus was chosen because it has been found to outperform other electronic databases in citation analyses ([Norris & Oppenheim, 2007](#)). There were no new reports, leaving the total number of eligible reports at 14. Of the 3,621 citations of the 14 included reports in the review (all citation counts are from Scopus conducted in April 2017), 3,561 were for only two articles: 2,975 citations for [Ellison et al. \(2007\)](#) and 586 citations for [Ross et al. \(2009\)](#). Though none of the reports that cited these articles conducted a systematic validation of these scales, some of these reports used the SNS engagement scales in an empirical study, providing partial evidence for the psychometric properties of these scales. This evidence is described separately for each scale in the Results section.

In the fifth and final step, we carried out article citation searches. In these searches, we checked the references of these additional articles, as well as any citations of them, but failed to identify any new relevant reports.

Overall, the present review included 14 eligible reports, with a total of 12 scales on SNS engagement. These reviewed reports included data from 23 independent samples from 11 countries, including 13,861 participants. More detailed demographic characteristics for these samples are available in [Table 1](#). To allow readers to make judgments about the psychometric properties of each scale, we extracted the relevant psychometric information from the reviewed reports in the Results section and summarized the information in [Tables 1 and 2](#).

2.3. Coding process

The coding process was identical to those adopted in our previous reviews (see [Cheng, Cheung, & Lo, 2016](#); [Cheng & Li, 2014](#) for details). Specifically, data extraction was conducted by two

¹ The wildcard “*” indicates that any ending of that word would be included in the search results. For example, “valid” would include reports containing the keyword of “valid,” “validities,” “validity,” “validation,” and “validations.”

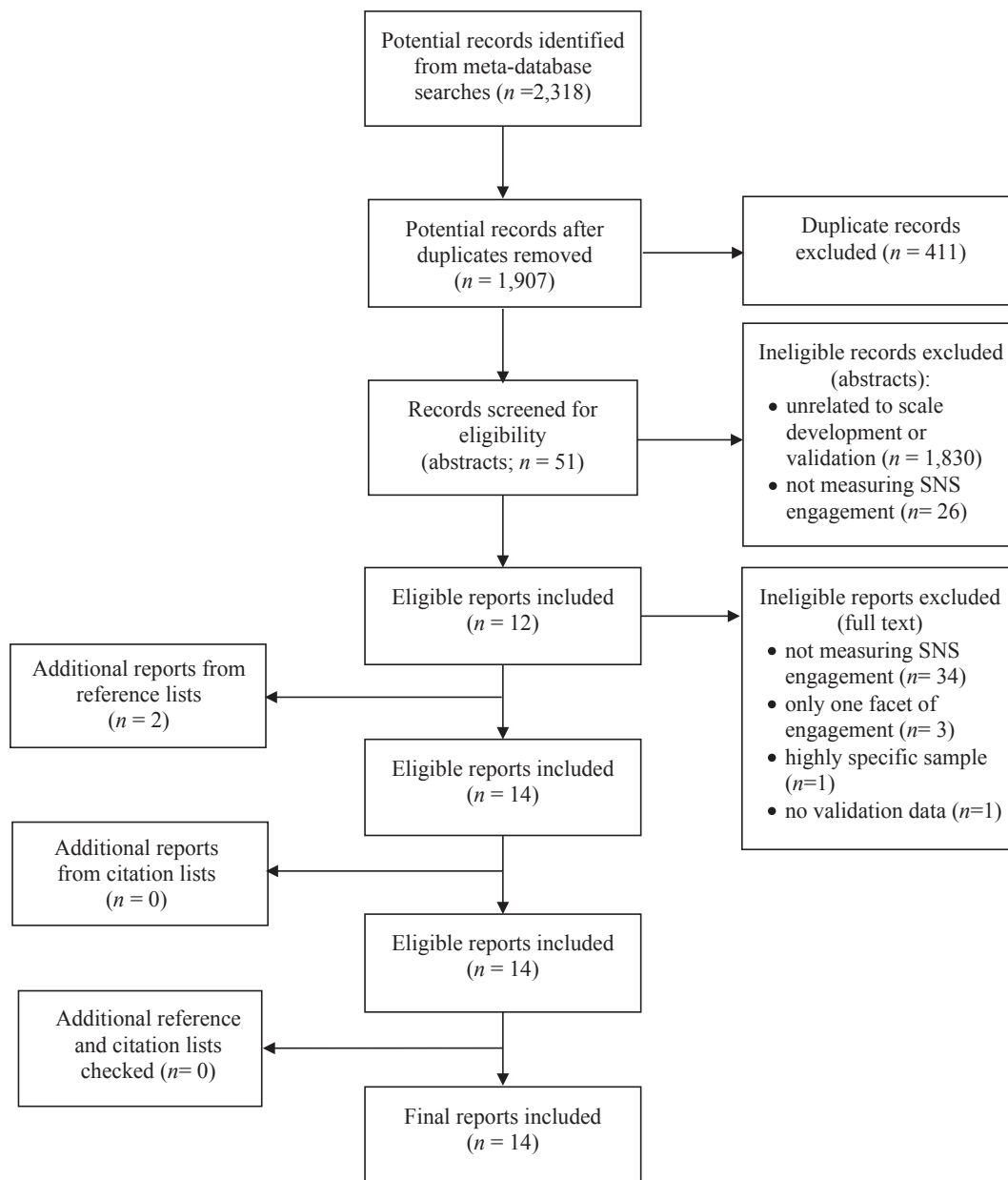


Fig. 1. A flow diagram summarizing the literature search process.

independent coders, both of whom have received prior training and practical experience in systematic and meta-analytic reviews. The procedures of triangulation were used to enhance reliability of the codings (Yin, 2003). Specifically, the coders initially carried out the task on their own. In the later stage, any discrepancies in data coding were discussed in ad hoc meetings and resolved by referencing the original report. The coder agreement was 95%.

2.4. Evaluations of psychometric properties of individual scales

In evaluating psychometric properties of the scales, we relied primarily on the five key criteria outlined in the Introduction: (a) content validity, (b) reliability, (c) structural validity, (d) associations with other variables (i.e., criterion, convergent, discriminant, and incremental validity), and (e) response processes (see Section 1.3).

Table 1 lists every article included in this review. Articles that

provide only partial evidence of validity are discussed separately for each scale.

2.4.1. Content validity

As the content validity of a scale is determined by its final list of items, performing good practices during the construction of this list is essential. To evaluate content validity, we summarize the item construction process of each scale. It is important to note that because the scales have different theoretical conceptions of SNS engagement, measuring more of these facets does not necessarily mean that the scale has “higher” content validity.

It is also important to note that the factor structures of the included scales rarely align neatly with the six facets of SNS engagement. For instance, the various items of the FBI scale measure three facets of SNS engagement (i.e., action and participation, usage and activity counts, and social context), but this scale contains only a single factor of “Facebook Intensity.” For researchers

Table 1
Publication information, characteristics, and sample demographics of each reviewed scale.

Scale	Publication information			Scale characteristics				Sample demographics				
	Author(s)	Year	Number of citations ^a	Number of items	Number of factors	SNS Platform	Facets of SNS engagement	Sample size (n)	Country	Students only	Gender (% Female)	Mean age (SD)
FBI Scale	Ellison et al.	2007	2,975	8	1	FB	AP, UAC, SC	286	U.S.	Yes	66	20 (1.6)
MFIS (Sample 1)	Orosz et al.	2016	1	13	4	FB	SP, AP	512	Hungary	No	63	22 (2.4)
MFIS (Sample 2)	Orosz et al.	2016	1	13	4	FB	SP, AP	566	Hungary	No	64	24 (8.1)
MFIS (Sample 3)	Orosz et al.	2016	1	13	4	FB	SP, AP	531	Hungary	No	74	24 (7.3)
MTUAS	Rosen, Whaling, Carrier, Cheever, & Rokkum	2013	31	60	15	SNS (general)	AP, UAC, SC	942	U.S.	No	62	30 (12.5)
MTUAS	Ozgur	2016	0	60	15	SNS (general)	AP, UAC, SC	913	Turkey	Yes	67	20 (N/A)
MTUAS (Sample 1)	Costa et al.	2016	0	56	14	SNS (general)	AP, UAC, SC	322	Portugal	Yes	59	15 (2.0)
MTUAS (Sample 2)	Costa et al.	2016	0	56	14	SNS (general)	AP, UAC, SC	479	Portugal	Yes	53	15 (2.0)
SONTUS (Sample 1)	Olufadi	2016	0	29	5	SNS (general)	UG, UAC	2,049	Nigeria	No	57	32 (7.7)
SONTUS (Sample 2)	Olufadi	2016	0	29	5	SNS (general)	UG, UAC	1,808	Nigeria	No	64	34 (3.9)
SNAIS (Sample 1)	Li et al.	2016	0	14	2	SNS (general)	SP, AP, UAC	455	Mainland China	Yes	41*	< 18
SNAIS (Sample 2)	Li et al.	2016	0	14	2	SNS (general)	SP, AP, UAC	455	Mainland China	Yes	41*	< 18
FPIS (Sample 1)	Yu	2015	0	27	3	FB	AP, UG, PE, SC	105	Taiwan	Yes	N/A	N/A
FPIS (Sample 2)	Yu	2015	0	27	3	FB	AP, UG, PE, SC	571	Taiwan	Yes	54	20 (1.3)
GoToFB Scale (Sample 1)	Aladwani	2014	18	34	8	FB	SP, AP, UG, SC	416	Kuwait	Yes	54	20 (1.4)
GoToFB Scale (Sample 2)	Aladwani	2014	18	34	8	FB	SP, AP, UG, SC	378	Kuwait	Yes	46	20 (1.4)
SMUIS (Sample 1)	Jenkins-Guarnieri et al.	2013	3	13	2	FB	AP, PE, SC	308	U.S.	Yes	70	18 (1.1)
SMUIS (Sample 2)	Jenkins-Guarnieri et al.	2013	3	10	2	FB	AP, PE, SC	308	U.S.	Yes	72	18 (0.9)
PSAFU Scale (Sample 1)	Bodroža & Jovanović	2016	5	43	5	FB	SP, AP, PE, UAC, SC	445	Serbia	No	79	27 (6.4)
PSAFU Scale (Sample 2)	Bodroža & Jovanović	2016	5	43	5	FB	SP, AP, PE, UAC, SC	359	Serbia	Yes	79	21 (3.0)
ISSNU Scale	Topaloglu et al.	2016	2	13	2	SNS (general)	SP, AP, PE, UAC, SC	1,005	Turkey	Yes	46	N/A
FBUQ	Blachnio et al.	2016	2	38	3	FB	SP, AP, PE, UAC, SC	551	Poland	No	71	20 (3.1)
FB Questionnaire	Ross et al.	2016	586	28	2	FB	AP, PE, UAC, SC	97	Canada	Yes	85	22 (5.4)

Note: * The numbers from this subsample are unavailable, so numbers from the combined sample are reported instead. SP = self-presentation, AP = action and participation, UG = uses and gratifications, PE = positive experiences on the SNS, UAC = usage and activity counts, SC = social context, FBI = Facebook Intensity, MFIS = Multi-Dimensional Facebook Intensity Scale, MTUAS = Media and Technology Usage Scale, SONTUS = Social Networking Time Use Scale, SNAIS = Social Networking Activity Intensity Scale, FPIS = Facebook Psychological Involvement Scale, GoToFB = Gravitating Towards Facebook, SMUIS = Social Media Use Integration Scale, PSAFU = Psycho-Social Aspects of Facebook Use, ISSNU = Impact of Student's Social Networking Use, FBUQ = Facebook Usage Questionnaire, FB = Facebook, SNS = social network sites, N/A = not available.

^a Scopus, April 2017.

who would like to check the content validity of a scale, they should examine the procedures involved in item construction and peruse the content of all of its items to make a judgment. To further assist scale selection, we report which of the six facets of SNS engagement are measured by its items.

2.4.2. Reliability

All of the scales in the review were reflective (rather than formative), and thus their reliability was assessed primarily with internal consistency (mainly with Cronbach's alpha/ α), and some also used test-retest reliability (with Pearson zero-order correlation coefficient r). For evaluating internal consistency, previous recommendations state that a Cronbach's alpha below .70 is less optimal (Nunnally & Bernstein, 1994, p. 265; Streiner, 2003). We point out any instances where the estimated Cronbach's alpha for the whole scale or any subscales is below .70. As test-retest reliability is dependent on multiple factors (e.g., length of time between tests), there is no cutoff and thus we simply report this value. If the reliability index for the full scale is not reported, we report the average reliability of the subscales.

2.4.3. Structural validity

To assess structural validity, we describe the findings from factor analyses that were conducted with the scale. We also specify whether principal components analysis (PCA) or exploratory factor analysis (EFA) has been used in exploring the factor structure of a scale. Several scholars maintain that both methods will yield essentially the same results (e.g., Devellis, 2003, pp. 128–129). However, a number of empirical studies have demonstrated that PCA can exert an adverse impact on the structural validity of the scale being developed (e.g., Costello & Osborne, 2005; Fabrigar, Wegener, MacCallum, & Strahan, 1999). For example, Costello and Osborne (2005) found that PCA would remove items from the final scale that would have been retained if EFA had been performed.

Given that an individual factor analysis may not yield stable, generalizable results (Brown, 2006, p. 30; Distefano, Zhu, & Mîndrilă, 2009), additional validation of the factor structure is essential for establishing structural validity. This is usually accomplished using a confirmatory factor analysis (CFA) with an independent sample, whose participants do not overlap with any of

Table 2
Descriptions of psychometric characteristics and assessments of each reviewed scale.

Scale	Author(s)	Year	Sample size (n)	Reliability		Structural validity		Associations with other variables
				Internal consistency (α)	Test-retest reliability (r)	Type of factor analysis	Fit indices	Type of validity and summary of results
FBI Scale	Ellison et al.	2007	286	0.83	N/A	N/A	N/A	Criterion validity- The whole scale had significant positive associations with bridging social capital and bonding social capital.
MFIS (Sample 1)	Orosz et al.	2016	512	0.79 †	0.82 †	Exploratory SEM	N/A	1. Incremental validity- Controlling for scores on the FBI scale, 3 of 4 factors had significant positive associations with self-reported FB liking and posting frequency. 2. Discriminant validity- None of these associations exceeded Brown's cutoff.
MFIS (Sample 2)	Orosz et al.	2016	566	0.77 †	N/A	CFA	RMSEA = .06 CFI = .95 TLI = .96	N/A
MFIS (Sample 3)	Orosz et al.	2016	531	0.88	N/A	N/A	N/A	1. Convergent validity- All 4 factors had significant positive associations with FB addiction. 2. Criterion validity- (a) All 4 factors had significant positive associations with online sociability, monetary value of FB, and obsessive FB passion. (b) 3 factors had significant positive associations with harmonious FB passion.
MTUAS	Rosen et al.	2013	942	0.85 †	N/A	EFA	N/A	1. Criterion validity- In a subsample (n = 545), 3 SNS-related factors (general social media usage, online friendships, and FB friendships) had significant positive associations with self-reported hours spent social networking. 2. Discriminant validity- None of these associations exceeded Brown's cutoff.
MTUAS	Özgür	2016	913	0.80 †	N/A	CFA	RMSEA = .04 CFI = .95 TLI = .95 SRMR = .046	N/A
MTUAS (Sample 1)	Costa et al.	2016	322	0.79 †	N/A	PCA	N/A	N/A
MTUAS (Sample 2)	Costa et al.	2016	479	N/A	N/A	CFA	RMSEA = .07 CFI = .88 TLI = .86	N/A
SONTUS (Sample 1)	Olufadi	2016	2,049	0.92	N/A	PCA	N/A	Convergent validity- In an unspecified sample, the whole scale as well as 4 of 5 factors had significant positive correlations with FB addiction and Internet addiction.
SONTUS (Sample 2)	Olufadi	2016	1,808	0.93	N/A	CFA	RMSEA = .04 CFI = .95 TLI = .94	N/A
SNAIS (Sample 1)	Li et al.	2016	455	0.89*	0.85*	PCA	N/A	Convergent validity- In a combined sample (n = 910), the whole scale as well as both factors had significant positive correlations with another measure of SNS engagement (the FBI scale), social networking addiction, and Internet addiction.
SNAIS (Sample 2)	Li et al.	2016	455	0.89*	0.85*	CFA	RMSEA < .08 CFI > .95 TLI > .95	N/A
FPIS (Sample 1)	Yu	2015	105	0.90	0.85	CFA	RMSEA = .08 CFI = .91 SRMR = .09	Convergent validity- The whole scale had a significant positive association with Internet addiction.
FPIS (Sample 2)	Yu	2015	571	0.92	N/A	CFA	RMSEA = .07 CFI = .96 SRMR = .07	N/A
GoToFB Scale (Sample 1)	Aladwani	2014	416	0.90	N/A	EFA	N/A	Criterion validity- The whole scale had a significant positive association with intentions to continue using FB
GoToFB Scale (Sample 2)	Aladwani	2014	378	0.90 †	N/A	CFA	RMSEA = .07 CFI = .89	N/A
SMUIS (Sample 1)	Jenkins-Guarnieri et al.	2013	308	0.92	0.80	EFA	N/A	Convergent validity- In a combined sample (n = 552), the whole scale as well as both factors had significant positive associations with another measure of SNS engagement (the FBI scale).

(continued on next page)

Table 2 (continued)

Scale	Author(s)	Year	Sample size (<i>n</i>)	Reliability		Structural validity		Associations with other variables
				Internal consistency (α)	Test-retest reliability (<i>r</i>)	Type of factor analysis	Fit indices	Type of validity and summary of results
SMUIS (Sample 2)	Jenkins-Guarnieri et al.	2013	308	0.86 †	N/A	CFA	RMSEA = .08 CFI = .96 TLI = .95	N/A
PSAFU Scale (Sample 1)	Bodroža & Jovanović	2016	445	0.87 †	N/A	PCA	N/A	N/A
PSAFU Scale (Sample 2)	Bodroža & Jovanović	2016	359	N/A	N/A	CFA	RMSEA = .05 CFI = .81 TLI = .80	N/A
ISSNU Scale	Topaloglu et al.	2016	1,005	0.88	N/A	PCA	N/A	N/A
FBUQ	Blachnio et al.	2016	551	0.81	N/A	PCA	N/A	N/A
FB Questionnaire	Ross et al.	2016	97	0.80 †	N/A	PCA	N/A	N/A

Note: † Reliability based on an average of subscales rather than full scale. * The numbers from this subsample are unavailable, so numbers from the combined sample are reported instead. FBI = Facebook Intensity, MFIS = Multi-Dimensional Facebook Intensity Scale, MTUAS = Media and Technology Usage Scale, SONTUS = Social Networking Time Use Scale, SNAIS = Social Networking Activity Intensity Scale, FPIS = Facebook Psychological Involvement Scale, GoToFB = Gravitating Towards Facebook, SMUIS = Social Media Use Integration Scale, PSAFU = Psycho-Social Aspects of Facebook Use, ISSNU = Impact of Student's Social Networking Use, FBUQ = Facebook Usage Questionnaire, FB = Facebook, SEM = structural equation modeling, N/A = not available, CFA = confirmatory factor analysis, EFA = exploratory factor analysis, PCA = principal components analysis, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker–Lewis index, SRMR = standardized root mean square residual.

those included in the PCA or EFA. For CFA's, we include any fit indices reported for the final structural model of the scale. To determine whether there is a good fit to the data, we adopt Hu and Bentler's (1999) widely-adopted criteria for fit indices: RMSEA < .06, CFI > .95, TLI > .95, and SRMR < .08. In addition, as post-hoc modifications of the model can threaten structural validity (Worthington & Whittaker, 2006), these results are also discussed in this review. Last, it is important to emphasize that reliability tests cannot be considered a substitute for factor analyses in evaluating structural validity (Devellis, 2003, p. 94; Furr, 2010, p. 11; Hinkin, 1998; Levine, 2005).

2.4.4. Associations with other variables

Studies generally rely on Pearson zero-order correlations or regression analysis to provide evidence for criterion, convergent, discriminant, and incremental validity. To evaluate this type of validity, studies have to rely on a previously established association between the variable being measured and a target variable grounded in theory or empirical work (Nunnally & Berstein, 1994, pp. 91–92). Hence, we only report associations with other variables that are relevant to the scales' validity while omitting associations with those without theoretical and empirical grounds.

To test the criterion validity of an SNS engagement scale, researchers should show that it is related to variables that are outcomes of SNS engagement. Below, we discuss two outcome variables that are particularly well-established: the amount that an individual uses the SNS, and social capital. Less widely-used criterion measures are discussed specifically for each scale in the Results section.

The criterion that an individual who is engaged with an SNS will use it more has been widely accepted to the point where this variable is often used as a proxy for SNS engagement by companies who run online platforms (McCay-Peet & Quan-Haase, 2016). Amount of SNS use can be measured in two major ways: average time (in minutes or hours) spent on the SNS, and the frequencies of exhibiting specific SNS behaviors (e.g., liking, retweeting). Empirical evidence has been obtained in a study by Turel and Serenko (2012), who found that SNS engagement was positively associated with both daily minutes spent on SNS's ($r = .33$) and the number of different SNS activities carried out by the user ($r = .32$). A significant positive association with any of these indicators would support the criterion validity of an SNS engagement scale. Ideally, amount of SNS use should not be assessed with self-report measures, particularly if asking about a respondent's average time spent on SNS because recall bias is likely to confound the findings (Junco,

2013; see Section 1.2 for details). Data mining methods would provide a helpful alternative to self-report measures in this case.

Many studies have focused on the relationship between SNS engagement and social capital, the latter of which refers to resources an individual can access from their social network. A recent meta-analysis by Liu, Ainsworth, and Baumeister (2016) found that SNS "intensity" (the most commonly used conceptualization of SNS engagement) had positive associations with both bridging and bonding social capital (r 's = .35 and .27, respectively). SNS engagement scales would have criterion validity if they had significant positive correlations with bridging social capital, bonding social capital, or both.

To establish convergent validity, researchers would need to show a significant correlation between an SNS engagement scale and a variable that is conceptually similar. Below, we describe the two most prominent variables for establishing convergent validity: SNS addiction and SNS engagement itself.

A significant positive association with a different well-validated measure of SNS engagement would support the convergent validity of an SNS engagement scale. As these two scales would be measuring the same latent variable, we would expect a significant positive relationship between the scales.

SNS addiction has also been frequently used to test convergent validity of SNS engagement scales (e.g., Li et al., 2016; Olufadi, 2016). This is because both engagement and addiction refer to a user's experience that can arise from interaction with SNS. Though engagement and addiction share some common characteristics such as euphoria and cognitive salience (Charlton & Danforth, 2007), a robust body of theoretical and empirical work has shown that they are distinct constructs, particularly their relationships with different indicators of psychological well-being (e.g., Lin, Hung, Fang, & Tu, 2015; Wan & Chiou, 2006). Factor analytic studies have shown that the two variables represent separate factors, though they are positively correlated (Charlton & Danforth, 2007; Charlton, 2002). Similar positive correlations have been found in the context of SNS as well (e.g., Turel & Serenko, 2012). Therefore, a significant positive association with a measure of SNS addiction would support the convergent validity of an SNS engagement scale.

To establish discriminant validity, an SNS engagement scale should not have an excessively strong association with scales that measure similar yet conceptually distinct constructs. We adopt Brown's (2006, p. 131) recommendation that a correlation between two factors above 0.80 indicates a lack of discriminant validity. We note that correlation attenuation, caused by the presence of

measurement error, would be present in any study where sum scores of scales are used in place of factor analysis (Bollen & Lennox, 1991), and would lower the estimated strength of the relationship. Thus, we adjust Brown's cutoff depending on the amount of measurement error present in the scales (indicated by scale reliabilities). For example, using the traditional formula described by Fan (2003), when both scales have an average reliability of $\alpha = .80$, Brown's cutoff for discriminant validity would be set at $r = .51$.

The most prominent variable for testing an SNS engagement scale's discriminant validity is amount of SNS use. As noted above, SNS engagement is conceptually distinct from amount of SNS use, primarily in its psychological components. If an SNS engagement scale had an association with amount of SNS use exceeding Brown's cutoff, this would indicate a lack of discriminant validity.

Last, on the basis of recommendations by Hunsley and Meyer (2003), the incremental validity of the new scale is supported by a significant association between a new SNS engagement scale and its related outcomes while statistically controlling for scores on another well-validated measure of SNS engagement.

2.4.5. Response processes

In measuring SNS engagement, there are two primary areas of concern for validity of response processes. First, Kuru and Pasek (2016) demonstrate that acquiescence bias has adverse effects on the psychometric properties of Facebook scales using "agree/disagree" response options, by artificially inflating their reliability estimates and distorting their estimated associations with other variables.

Second, all of the reviewed scales rely on self-reports, and this method is widely known for its susceptibility to social desirability bias (e.g., Gordon, 1987; Gravdal & Sandal, 2006).

2.5. Evaluations of generalizability

Just because a scale is valid with one sample or context, there is no guarantee that this validity will be generalizable to other samples or contexts (Messick, 1995). Thus, we describe the demographic characteristics of the samples with which the scales are validated. Future SNS researchers can judge whether these results are generalizable to their target sample.

We also report any tests of measurement invariance conducted with the scales. Measurement invariance tests are important to show that scores on the scale are not biased by demographics (Meredith & Millsap, 1992). Though measurement invariance is often assumed in practice, there are many factors that may violate this assumption, such as response biases that are specific to one cultural group (Wu, Li, & Zumbo, 2007). If an SNS engagement scale is intended for use in population surveys or epidemiological research on a demographically heterogeneous sample, this scale should have measurement invariance for various demographic variables (e.g., gender, age). If researchers would like to use an SNS engagement scale in their multinational research, they should test for measurement invariance across cultures before making cross-cultural comparisons (e.g., Cheng, Cheung, & Montasem, 2016).

2.6. Efforts to reduce publication bias

Publication bias is commonly recognized as a potential threat to the comprehensiveness of systematic reviews (see e.g., Bradley & Gupta, 1997; Rosenberg, 2005). Studies having large samples or significant findings are generally more likely to be published, and published work is more easily accessible in literature searches. Hence, studies that use small samples or yield insignificant findings are more likely to be precluded from publications. To reduce this

possible bias, we did not place any limits regarding publication status in our database searches, and also endeavored to locate more unpublished work by browsing through conference proceedings and thesis databases (e.g., COS Conference Papers Index; Proquest Dissertations & Theses A&I).

3. Results

In this section, we categorized the included scales according to how extensively they have been validated. It is important to note that these scales are derived from distinct, complementary theoretical approaches that may suit the specific aims of diverse studies, and thus there is no single "best" scale.

3.1. Scales validated in at least two samples

3.1.1. Facebook Intensity (FBI) scale

Created as part of a substantive study by Ellison et al. (2007), the FBI scale is the pioneer of SNS engagement scales, and is undoubtedly the most widely used. This scale was designed to assess a respondent's level of "Facebook intensity," which the authors define as active Facebook use, emotional connection to Facebook, and integration of Facebook into daily life. Unlike most other scales in this review, the authors did not describe a process of generating or refining the list of items (e.g., by consulting other researchers, previous literature, or SNS users). The scale includes six items measuring attitude towards Facebook and two items assessing objective Facebook behavior (i.e., total minutes spent on Facebook per day, and number of Facebook friends). The 8-item scale has one factor. According to the list of items, the FBI scale covers three of the six facets of SNS engagement: (a) action and participation, (b) usage and activity counts, and (c) social context.

In the researchers' original (2007) study, the FBI scale was not systematically validated. It was administered in a sample of 286 U.S. undergraduates, who were young (mean age = 20) and mostly female (66% women).

Reliability, estimated with internal consistency, was acceptable (Cronbach's $\alpha = .83$). In addition, significant positive relationships of the FBI scale with both bridging and bonding social capital provide support for its criterion validity. However, no systematic validation has been conducted in this study, at least explicitly.

Our extensive literature search (including all 2,975 articles that have cited the investigators' 2007 study) shows that no other study has attempted a systematic validation of the FBI scale. Despite this lack of systematic validation evidence, the FBI scale's psychometric properties may be inferred to some extent from the many substantive studies that have used this popular scale. For example, many of these studies have consistently shown that the FBI scale is reliable (i.e., Cronbach's $\alpha > .70$), including instances where it was adapted for use in diverse contexts such as SNS engagement in general (Salehan & Negahban, 2013), Snapchat (Piwek & Joinson, 2016), and cross-cultural Facebook interaction (Jiang & Bruijn, 2013). Moreover, Li et al. (2016) adapted the FBI scale for use with SNS generally and tested its test-retest reliability over a two-week period with Chinese adolescents (Pearson $r = .78$).

A number of studies have included the FBI scale in factor analyses, providing some support for its structural validity. The most direct evidence for this type of validity comes from a study by Beyens, Frison, and Eggermont (2016) with a sample of Belgian adolescents ($n = 402$, 57% women, mean age = 16). The original FBI scale was analyzed without any modifications, and the CFA of the scale had good fit to the data (i.e., RMSEA = .03, CFI = .97).

Additional support for the structural validity of the FBI scale comes from a recent study conducted with students in three Asian countries (Lee, Kim, Golden, Kim, & Park, 2016). The researchers

modified the wording of the FBI scale to measure general SNS engagement, and included six response options instead of the five options used in the original version. They conducted CFA's on this modified scale with university students from South Korea ($n = 113$, 52% women, mean age = 22), Malaysia ($n = 105$, 73% women, mean age = 23), and Mainland China ($n = 87$, 77% women, mean age = 21). Using the criteria proposed by [Hu and Bentler \(1999\)](#), we concluded that these models all had good fit to the data (all RMSEA's < .085, all CFI's > .095), providing further evidence for the structural validity of the FBI scale.

Less direct evidence of the FBI scale's structural validity can be found in a number of studies that either did not report crucial data for evaluation (e.g., model fit indices), or made major modifications to the scale's structure. [K. G. Park, Han, and Kaid \(2012\)](#) modified the FBI scale and used it with a sample of U.S. university students ($n = 339$, 49% women, no mean age reported). They conducted an EFA on it and found a one-factor solution, though no indices of model fit were reported. [Salehan and Negahban \(2013\)](#) used only five out of eight items of the FBI scale with U.S. university students ($n = 209$, 39% women, no mean age reported), and found that factor loadings were consistent with a good factor structure. [Li et al. \(2016\)](#) conducted a PCA on a modified version of the FBI scale with 910 Chinese students and found a one-factor solution with good factor loadings. In a sample of U.S. adult residents ($n = 193$, 53% women, mean age = 27), [Moqbel, Nevo, and Kock \(2013\)](#) modified the wording of the FBI scale for use with SNS generally. They conducted a PCA that included this modified scale as well as other variables (job performance, commitment, and satisfaction), and found factor loadings consistent with structural validity of the scale. [Alhabash, Chiang, and Huang \(2014\)](#) used six of the FBI scale's items with Facebook users in Taiwan ($n = 3,172$, 50% women, mean age = 33). They conducted an EFA on this revised FBI scale and found a one-factor solution, though no fit indices were reported.

Only one other study has conducted CFA ([C. Park, Jun, & Lee, 2015](#)). This study examined the factor structure of the FBI scale in two independent samples of smartphone-based SNS users, one from South Korea ($n = 487$, 49% women, no mean age reported) and the other from the U.S. ($n = 490$, 50% women, no mean age reported). They used only six of the items of the FBI scale. As the CFA included the modified FBI scale as well as measures of other constructs (i.e., innovativeness, propensity to share information, privacy concerns, and social capital), the fit indices of this model do not directly reflect the structural validity of the FBI scale. However, the factor loadings of the items in all of the scales do show that the items of the FBI scale are distinctive from those of all other measures, thus providing some converging evidence for the FBI scale's structural validity.

Criterion validity of the FBI scale could be inferred from the myriad of studies (including [Ellison et al.'s](#) original study) that have found this scale to have positive associations with social capital ([Liu et al., 2016](#)). The study by [Orosz et al. \(2016\)](#) revealed that the FBI scale had a significant association with the frequency of one Facebook behavior ("liking") but not another (posting), providing some mixed evidence for its validity in this particular criterion of interest. These associations did not exceed Brown's cutoff for discriminant validity, providing additional evidence of this type of validity for the scale.

Convergent validity of the FBI scale is supported by results from two other studies in this review that showed that the scale had significant positive correlations with other scales measuring SNS engagement, namely the Social Media Use Integration Scale ($r = .77$; [Jenkins-Guarnieri et al., 2013](#)) and the Social Networking Activity Intensity Scale ($r = .52$; [Li et al., 2016](#)).

There is a recent concern for the validity of the FBI scale's response processes. Specifically, the use of agree/disagree response

options makes the FBI scale susceptible to acquiescence bias. This problem has been empirically demonstrated in [Kuru and Pasek's \(2016\)](#) study, which indicates that the bias artificially inflates the estimated reliability of the FBI scale, with an increase in its Cronbach's alpha from .76 to .89. This issue deserves scrutiny because a good reliability is usually the only evidence given to support the validity of the FBI scale. In addition to inflating the estimated reliability, [Kuru and Pasek](#) further demonstrated that acquiescence bias distorted the FBI scale's correlations with other variables. For example, the investigators estimated that acquiescence bias artificially inflated the correlation (Pearson r) between the FBI scale and social capital from .49 to .55. Such a distortion may threaten the statistical validity of any future meta-analyses that include the FBI scale, and may increase the likelihood of Type I or Type II errors in individual studies.

In summary, the psychometric properties of the FBI scale are uncertain. Though evidence for its reliability and validity can be inferred from some of the many studies that have used it, the scale still lacks any formal kind of systematic validation. In addition, [Kuru and Pasek's \(2016\)](#) study shows that acquiescence bias may pose a threat to the scale's validity, and this problem of inflated results may weaken the credibility of findings that show high reliability of this scale and its consistent associations with other variables (e.g., social capital).

3.1.2. Multi-dimensional Facebook Intensity Scale (MFIS)

The MFIS ([Orosz et al., 2016](#)) is constructed on the basis of [Ellison et al.'s \(2007\)](#) conception of Facebook intensity. The investigators initially developed the scale's items with a focus group of university students and then utilized these results to refine the items. The final 13-item scale comprises four dimensions of Facebook engagement: persistent use, use to relieve boredom, overuse, and self-expression. The items of this scale cover two of the six facets of SNS engagement: (a) self-presentation and (b) action and participation.

The scale was then assessed with three large independent, community samples in Hungary (all n 's > 500). Each sample was composed of young adults (mean ages ranged from 22 to 24) and was majority women (ranging from 63% to 74%).

Reliability was estimated with internal consistency in all samples, with acceptable results (Cronbach's α ranged from .77 to .88). In addition, test-retest reliability was assessed with 93 participants over a four-week period (Pearson $r = .82$).

The structural validity of the MFIS was initially explored with an exploratory structural equation model. Then a CFA with a separate sample confirmed the existence of four first-order factors (i.e., all four MFIS subscales) as well as a second-order factor of Facebook Intensity (final model fit: RMSEA = .06, CFI = .95, TLI = .96). However, one post-hoc modification (one pair of correlated residuals) was necessary to achieve this fit, which may be a concern for the stability of the scale's structure.

Both criterion and convergent validity of the MFIS were assessed with concurrent regression models using its four subscales separately. In these regression models, the variables of interest were Facebook addiction, online sociability, passion for Facebook, and monetary value of Facebook (i.e., amount of money participants would have to hypothetically receive to give up using Facebook). In all these models, both criterion and convergent validity of the MFIS were supported by significant positive associations. In addition, tests of incremental validity showed that the MFIS was positively associated with self-reported Facebook posting and liking even after controlling for scores on the FBI scale. Last, none of these associations with Facebook behaviors exceeded Brown's cutoff, supporting the MFIS's discriminant validity.

There is one potential threat to the validity of the MFIS's

response processes. Similar to the FBI scale, the MFIS relies on agree/disagree response options and thus may be vulnerable to acquiescence bias, as proposed by Kuru and Pasek (2016).

To sum up, the MFIS has been tested extensively in three separate Hungarian samples, with support for its reliability, and various types of validity (i.e., structural, criterion, convergent, incremental, and discriminant). However, its validity may be threatened by acquiescence bias.

3.1.3. Media and Technology Usage and Attitudes Scale (MTUAS)

The MTUAS (Rosen et al., 2013) was created to measure frequency of technology use and attitudes, encompassing a broad range of technologies and platforms that include SNS. Items were developed and refined through previous theoretical frameworks and focus groups. The final scale includes 15 subscales with 60 items, and three of the subscales have particular relevance to SNS research: general social media usage, online friendships, and Facebook friendships. Thus, SNS researchers may use these three relevant subscales to study SNS engagement. The scale's items cover three facets of SNS engagement: (a) action and participation, (b) usage and activity counts, and (c) social context.

Psychometric properties of the MTUAS were evaluated in three different studies (Costa et al., 2016 July; Özgür, 2016; Rosen et al., 2013), including four samples from three countries: Portugal, Turkey, and the U.S. The samples were all adequate or large in size ($n = 322$ – 942) and included both adolescents and adults (mean ages ranged from 15 to 30). These samples varied in terms of gender balance (53%–67% women).

The SNS-related subscales had acceptable reliabilities in all four samples (Cronbach's α ranged from .79 to .85). Structural validity was originally explored with an EFA in a sample of U.S. community adults ($n = 942$). Then a separate study (Özgür, 2016) tested the factor structure with a CFA in a sample of Turkish undergraduates ($n = 913$), and had good model fit (RMSEA = .04, CFI = .95, TLI = .95, SRMR = .046).

Another study adapted the scale for Portuguese youth, and conducted a separate PCA and CFA to assess its structural validity (Costa et al., 2016 July). The final factor structure was slightly different from the original; the investigators removed one subscale and 4 items. Also, though they made 5 post-hoc modifications to their CFA model, the results still revealed inadequate model fit (RMSEA = .07, CFI = .88, TLI = .86).

Criterion validity of the MTUAS was supported in the U.S. sample by its significant positive correlations with self-reported hours spent social networking online (Rosen et al., 2013). All these associations did not exceed Brown's cutoff, providing evidence for the MTUAS's discriminant validity.

There are no obvious concerns for the acquiescence bias of the three SNS-related subscales of the MTUAS because they do not use agree/disagree response options.

In short, the MTUAS has been tested across relatively heterogeneous samples from distinct cultural regions, and has been found to be reliable and valid, though its convergent and incremental validity have not been tested yet. Although a few of its subscales may be susceptible to acquiescence bias, these subscales are less central to the measurement of SNS engagement.

3.1.4. Social Networking Time Use Scale (SONTUS)

The SONTUS (Olufadi, 2016) was developed to assess time spent on SNS. The items were constructed in consultation with multiple expert scholars as an attempt to enhance content validity. The scale comprises 29 items measuring five factors, with four factors representing different contexts where SNS are used (i.e., while relaxing, completing academic tasks, in public places, and in stressful encounters), and one factor representing motives for use. Thus, the

SONTUS covers two of the six facets of SNS engagement: uses and gratifications (the factor of motives for use), and usage and activity counts (the remaining four factors).

The scale's psychometric properties were evaluated with two large Nigerian community samples (both n 's > 1,800). The samples appeared to be balanced in terms of age, though some age data were not reported (mean ages = 32 and 34, age range = 20–58 in one sample but unknown in the other); but both samples comprised 57% and 64% women, respectively.

Reliability, estimated with internal consistency in both samples, was acceptable (Cronbach's $\alpha = .92$ and $.93$). Structural validity was evaluated with a PCA and CFA on the two samples respectively. The CFA confirmed the existence of a second-order factor in addition to the five first-order factors (i.e., 4 contextual and 1 motivational factors), and achieved good model fit (RMSEA = .04, CFI = .95, TLI = .94).

Convergent validity of the SONTUS was assessed with zero-order correlation and regression analyses. Most subscales had significant positive correlations with both Internet addiction and Facebook addiction. Last, as the SONTUS does not rely on agree/disagree response options, there are no obvious concerns for its susceptibility to acquiescence bias.

To sum up, the psychometric properties of the SONTUS have been thoroughly tested, and the results show that it demonstrates reliability as well as structural and convergent validity, with no apparent concerns for its response processes. However, its criterion, discriminant, and incremental validity are yet to be evaluated.

3.1.5. Social Networking Activity Intensity Scale (SNAIS)

The SNAIS (Li et al., 2016) was developed to measure "frequency of using multiple types of online social networking activities through multiple types of platforms" (Li et al., 2016, p. 3) among junior high school students in Mainland China. First, a list of items was generated by reviewing previous literature. Then, this list was refined by consulting various researchers, by pilot testing with a sample of 77 students, and by interviews with an additional 20 students. The final scale includes 14 items to measure two factors: social function use intensity and entertainment function use intensity. The items of the scale cover three of the facets of SNS engagement: (a) self-presentation, (b) action and participation, and (c) usage and activity counts.

The psychometric properties of the SNAIS were assessed with two separate, equally sized subsamples, from a single survey of 910 Chinese junior high school students. Demographics and reliability analyses were reported for the combined sample, rather than each subsample. This combined sample was comprised of 41% female participants. Though no age data were available, these student samples should include adolescents rather than adults.

For reliability, internal consistency of the whole scale was acceptable (Cronbach's $\alpha = .89$), but the subscale of entertainment function use intensity had an internal consistency below the standard cutoff of .70 (Cronbach's $\alpha = .60$). In addition, 114 of the participants from the original sample were surveyed again two weeks later to assess test-retest reliability (Pearson $r = .85$).

Structural validity of the SNAIS was first explored with a PCA (using varimax rotation) with one of the subsamples. This factor structure was then confirmed using a CFA with the other subsample. Using Hu and Bentler's (1999) criteria, this model's fit to the data is unclear: both the CFI and TLI were above .95, but the authors only stated that the RMSEA was below .08, without reporting the exact statistics.

Associations between the SNAIS and other scales support its convergent validity. Specifically, the whole scale as well as each of its two factors was found to be significantly positively correlated with the FBI scale, and also with two additional scales measuring

social networking addiction and Internet addiction.

As the items of the SNAIS do not have agree/disagree response options, they are not susceptible to acquiescence bias like some other scales in this review are.

Overall, the SNAIS has received multiple tests in two samples, though its criterion, discriminant, and incremental validity have not been established, and there are minor concerns about the reliability of one of its subscales and its structural validity.

3.1.6. Facebook Psychological Involvement Scale (FPIS)

The FPIS (Yu, 2015) incorporates multiple theories (i.e., personal involvement, Internet addiction, usage motivation) and focuses on three aspects of psychological involvement in Facebook: importance, emotional support, and amusement. The scale has 27 items to measure these three factors. The FPIS items assess four facets of SNS engagement: (a) action and participation, (b) uses and gratifications, (c) positive experiences, and (d) social context.

The FPIS was empirically assessed with two samples of university students in Taiwan. Demographic data were not available for the first sample, but the second sample ($n = 571$) was young (mean age = 20) and had gender balance (54% women). For reliability, internal consistency was acceptable in both samples (Cronbach's $\alpha = .90$ and $.92$), and the authors also assessed test-retest reliability over a four-week period (Pearson $r = .85$).

Structural validity of the FPIS was assessed with separate CFA models in the two samples. The model fit was marginal in the first sample (RMSEA = .08, CFI = .91, SRMR = .09), but only after making two minor post-hoc modifications. Then the refined model was tested in the second sample, and had acceptable model fit (RMSEA = .07, CFI = .96, SRMR = .07). These findings indicate the stability of the factor structure, though Yu suggested future researchers to further modify the items according to their specific needs.

The convergent validity of FPIS was supported by a significant positive correlation with Internet addiction. Though the investigators do not explicitly state it, the FPIS appears to rely on agree/disagree response options. If this is the case, the scale may be vulnerable to acquiescence bias.

Overall, the findings show that the FPIS is reliable and has convergent validity. Additional effort is essential to further examine its structural, criterion, discriminant, and incremental validity, and the validity of its response processes.

3.1.7. Gravitating Towards Facebook (GoToFB) scale

Aladwani (2014) developed a scale to measure "the extent to which one feels gravitated towards exploiting Facebook capabilities influencing one's social experiences" (p. 273). He constructed and refined the list of items by consulting other researchers and a focus group comprising Facebook users. The scale includes 34 items measuring eight factors, such as connecting, sharing, and relaxing. The scale items cover four facets of SNS engagement: (a) self-presentation, (b) action and participation, (c) uses and gratifications, and (d) social context.

The scale was validated with two independent Kuwaiti samples (both n 's > 350). The samples comprised entirely of university students, so both were quite young (both mean ages = 20). Yet, there was gender balance for these two samples (54% and 46% women).

Reliability for the GoToFB scale was estimated with a statistic called composite reliability. Similar to Cronbach's alpha, this statistical method uses the correlations among a scale's items to estimate its reliability. Using this method, the investigators found acceptable reliability for the GoToFB scale (composite reliability = .90 for both samples). Structural validity was evaluated with separate EFA and CFA's on the two samples. The CFA of the

final model had marginal fit to the data (RMSEA = .07, CFI = .89), leaving the structural validity inconclusive.

Criterion validity of the GoToFB scale was supported by a significant positive association with participants' intentions to continue using Facebook. Lastly, as with many other scales in this review, the GoToFB scale relies on agree/disagree response options, which may make it susceptible to acquiescence bias.

In summary, though the GoToFB scale has been shown to be reliable and demonstrate criterion validity, it lacks strong evidence for its structural validity. Moreover, its convergent, discriminant, and incremental validity remain untested, and this scale may be threatened by acquiescence bias.

3.1.8. Social Media Use Integration Scale (SMUIS)

The SMUIS (Jenkins-Guarnieri et al., 2013) was constructed to measure engaged use of social media, or "the degree to which social media is integrated into the social behavior and daily routines of users, and the importance of and emotional connection to this use" (p. 39). The investigators constructed the scale specifically for use among emerging adults, though its items may also be applicable to other age groups. Some scale items were adapted from the FBI scale (Ellison et al., 2007) and the Facebook Questionnaire (Ross et al., 2009) included in this review, and additional items were created in consultation with two psychologists. The 10-item scale contains two factors: integration into social routines, and social integration and emotional connection. These items cover three facets of SNS engagement: (a) action and participation, (b) positive experiences, and (c) social context.

The scale was psychometrically evaluated with two separate, equal sized subsamples from a single survey of 616 U.S. undergraduates. The subsamples were both young (both mean ages = 18) and predominantly female (70% and 72% women).

Internal consistency for the final model was acceptable (Cronbach's $\alpha = .92$ and $.86$). The authors also assessed the test-retest reliability over a three-week period, with a Pearson r of $.80$ for the entire scale.

Structural validity was assessed with an EFA (with 13 items) and a CFA in the two subsamples separately. The final CFA model had acceptable model fit (RMSEA = .08, CFI = .96, TLI = .95), but this fit was attained only after a post-hoc removal of three items. This type of model modification suggests that the factor structure is not stable or generalizable to other samples.

In support of its convergent validity, the SMUIS subscale scores and its combined score all had very strong positive correlations with the FBI scale (all r 's > .69). Given the substantial amount of shared variance between the SMUIS and the FBI scale, it remains unknown whether the development of a new scale is necessary if there is already an existing scale that measures a similar construct. Also, as the scale relies on agree/disagree response options, it may be vulnerable to acquiescence bias.

To sum up, the validity evidence indicates that the SMUIS has good reliability and convergent validity, though its discriminant, incremental, and criterion validity have not yet been established, and it may be prone to acquiescence bias. Moreover, as large modifications have been made to achieve CFA model fit, its structural validity is uncertain. Also, although this scale was originally developed for the study of general social media use, only a single platform of Facebook has been examined.

3.1.9. Psycho-social Aspects of Facebook Use (PSAFU)

The PSAFU scale (Bodroža & Jovanović, 2016) was constructed to measure psychological aspects of Facebook use as well as objective Facebook behavior. The researchers developed this scale by adapting items from the Facebook Questionnaire (Ross et al., 2009) included in this review, as well as items from previous measures of

SNS behavior (Bodroža, Popov, & Poljak, 2009) and Internet addiction (Young, 1998). In addition, the researchers generated some new items in consultation with Facebook users. The scale comprised 43 items measuring five different psychosocial aspects of Facebook use: compensation, self-presentation, socialization, addiction, and virtual self. The items of this scale cover five of the six facets of SNS engagement: (a) self-presentation, (b) action and participation, (c) positive experiences on the SNS, (d) usage and activity counts, and (e) social context.

The psychometric properties of the scale were evaluated with two independent Serbian samples, one of which was from the community, and the other that was composed of students (both n 's > 350). The samples were largely female (both 79% women) and young (mean ages = 27 and 21).

Reliability for the PSAFU scale, estimated with internal consistency in the first sample, was acceptable (average Cronbach's $\alpha = .87$). Structural validity was evaluated with a PCA and CFA in the two samples, respectively. The initial CFA model fit for the scale was not acceptable (RMSEA = .05, CFI = .81, TLI = .80), but the investigators were able to achieve better CFA model fit by removing 17 of the 43 items (RMSEA = .04, CFI = .93, TLI = .92). However, they chose to proceed with the unmodified version of the 43-item scale, arguing that it had better content validity.

The subscales of the PSAFU scale were regressed onto a number of other variables (e.g., personality traits, sensation seeking), but none were relevant to the evaluation of its validity as a measure of SNS engagement.

As the investigators did not report the format of the scale's response options, it is unknown whether acquiescence bias is a concern for the validity of the PSAFU scale.

In short, the PSAFU scale has multiple validity concerns and gaps in its validation evidence, most importantly the inadequate fit of the factor structure.

3.2. Scales validated in one sample

3.2.1. The Impact of Student's Social Network Use Scale (ISSNU) scale

The ISSNU scale (Topaloglu, Caldibi, & Oge, 2016) was developed to measure "social network use, the purpose of social networks use and the preferences of the university students between social life and social networks" (Topaloglu et al., 2016, p. 350). The investigators developed and refined the items in consultation with a sample of university students as well as expert scholars. The scale included 13 items, measuring two factors: aims of social network use, and social network communication preferences. The scale covers five of the six facets of SNS engagement: (a) self-presentation, (b) action and participation, (c) positive experiences on the SNS, (d) usage and activity counts, and (e) social context.

The psychometric properties of the ISSNU were evaluated with a sample of 1,005 Turkish university students, which was predominantly young (92% between ages 18–23), with gender balance (46% women). Reliability, assessed with internal consistency, was acceptable for both subscales as well as for the entire scale (Cronbach's $\alpha = .87, .86, \text{ and } .88$, respectively).

Structural validity of the ISSNU was explored with a PCA, using varimax rotation; and so no model fit indices were available. Moreover, associations between the ISSNU and other variables were not examined. As the ISSNU relies on agree/disagree response options, it may be susceptible to acquiescence bias.

In summary, the ISSNU is found to be reliable, but there is no conclusive support for any of its other psychometric properties.

3.2.2. Facebook Usage Questionnaire (FBUQ)

The FBUQ was developed as part of a substantive study by

Blachnio, Przepiorka, Boruch, and Balakier (2016). Items were created following a review of previous literature, with the aim of measuring different types of Facebook use and attachment to Facebook. The scale consists of 38 items that measure three factors: standard Facebook use, Facebook addiction, and Facebook entertainment. The FBUQ items cover five of the facets of SNS engagement: (a) self-presentation, (b) action and participation, (c) positive experiences on the SNS, (d) usage and activity counts, and (e) social context.

Psychometric properties of the FBUQ were tested in a large Polish community sample ($n = 551$), comprised primarily of young women (mean age = 20; 71% women). Average reliability (assessed with Cronbach's α) of the subscales was .81, and it is noteworthy that the subscale of Facebook entertainment had an estimated internal consistency barely below the standard cutoff of .70 for acceptable reliability (Cronbach's $\alpha = .69$).

Structural validity of the FBUQ was explored with a PCA, with a varimax rotation; and so no model fit indices were available. Associations between the FBUQ subscales and a number of other variables were examined, but none of these variables were relevant to testing the scale's validity.

As the FBUQ does not use agree/disagree response options, there are no clear concerns for its vulnerability to acquiescence bias.

In short, the FBUQ is currently limited primarily by a lack of evidence for structural validity (see Section 2.4.3 for details on the issues with PCA), and no systematic testing of its associations with other variables.

3.2.3. Facebook Questionnaire

The Facebook Questionnaire was developed by Ross et al. (2009) as part of a substantive study on personality and motivations on Facebook. The questionnaire was designed to measure Facebook use, attitudes, and privacy behavior, but the authors did not describe a process of generating and refining the item list. It includes all six items from the FBI Scale, as well as a number of other items added by the investigators. The scale includes a total of 28 items, but the investigators conducted validation tests on only 12 of the items. The final list of items covers four facets of SNS engagement: (a) action and participation, (b) positive experiences, (c) usage and activity counts, and (d) social context.

The scale was tested with a sample of 97 Canadian university students who were predominantly young (mean age = 22) and female (85% women). The 12 items that were included in the reliability tests had acceptable internal consistency (average Cronbach's $\alpha = .80$).

A single PCA was used to explore the scale's factor structure, and thus no model fit indices were available. Also, it is noteworthy that a sample size of less than 100 (97 for this study) would usually be too small to yield a stable factor solution (see Costello & Osborne, 2005 for a detailed discussion).

The investigators did not check the entire scale's association with other variables to test criterion, convergent, discriminant, or incremental validity. Moreover, as it includes items from the FBI scale, along with agree/disagree response options, the Facebook Questionnaire may be susceptible to acquiescence bias.

The study that developed the Facebook Questionnaire (Ross et al., 2009) has been cited 586 times. However, this may be due to its contributions to research on personality and SNS usage rather than its creation of the Facebook Questionnaire. By perusing the reports that cited this article, we ruled out the existence of any additional systematic validation studies, and could identify only one substantive study that used the Facebook Questionnaire (J.-L. Wang, Jackson, Gaskin, & Wang, 2014). The investigators of this study adapted this questionnaire for use with Qzone, the most popular SNS in Mainland China (Niu, Sun, Zhou, Kong, & Tian,

2016), but did not report any data on its psychometric properties.

In summary, the Facebook Questionnaire does not have much validity evidence of any type apart from acceptable reliability, and an exploration of its structural validity using only 12 out of the 28 total items. Moreover, this questionnaire's validation evidence is also limited by its sample, which was composed of young Canadians and was predominantly female.

4. Discussion

The present review provides psychometric evaluations of scales for measuring SNS engagement. Apart from the widely used FBI scale, we found 11 other scales that may provide useful alternatives for researchers who would like to study SNS engagement. The quantity of these scales and their distinctive theoretical frameworks show that multiple options are available for researchers who are interested in this topic. However, we also found some yet-to-be-resolved validity concerns for these SNS engagement scales, implying that researchers should check the psychometric properties of a scale before using it for data collection. These concerns also imply that there is a wealth of opportunities for validation studies to improve how SNS engagement is measured and studied. Both of these implications are discussed in detail below.

4.1. Implications for SNS engagement research

The array of validity concerns that we identified indicates that SNS engagement researchers should exercise caution when selecting a scale to use in their study. Only one scale in our review

has no major validity concerns: the MTUAS. However, the MTUAS may not be suitable for all studies, which may vary vastly in their research goal, theoretical framework, or sample demographics. Hence, SNS engagement researchers may use this review as a reference to choose an appropriate scale that best meets their needs. We hereby offer a set of recommendations for guiding researchers in their scale selection process. As SNS intensity is the most widely used operationalization of SNS engagement, and because there are concerns about acquiescence bias among the scales that measure it, we also offer a discussion of how to validly measure SNS intensity.

4.1.1. Recommendations for scale selection

Given that SNS is an emerging but rapidly growing area of study, SNS engagement scales have mushroomed in the past decade but not many of them have undergone thorough psychometric evaluations. Hence, researchers are unlikely to find a "perfect" SNS engagement scale free of all validity concerns. We provide below some general advice on how to select a psychometrically sound scale (summarized in Table 3), and how to mitigate possible threats to its validity.

A researcher's first consideration should be the content validity of the scale, and this is where their own theoretical expertise will be most important. Researchers should critically assess whether the reported item development process would be sufficient for content validity. They should also judge whether the scale measures all of the facets of SNS engagement that are necessary for their study.

Next, researchers should examine the demographics of the

Table 3
Steps in selecting an appropriate scale for measuring SNS engagement.

Step	Description
1. Content validity	<ul style="list-style-type: none"> Evaluate the item development process to judge whether the scale has followed typical good practices: <ol style="list-style-type: none"> construct operationalization and item construction through literature review item refinement through expert consultations, pilot test with samples from target populations, or preferably both Identify which facets of SNS engagement the scale covers Decide whether the facets meet the intended study's objectives
2. Context of previous validation research	<ul style="list-style-type: none"> Check whether the scale has been validated in: <ol style="list-style-type: none"> a sample with similar demographic characteristics a country with similar cultural values the same SNS platform (e.g., Facebook) Look for measurement invariance evidence that indicates whether the scale functions in the same way for diverse demographic groups and with different SNS platforms.
3. Reliability	<ul style="list-style-type: none"> Judge whether the whole scale and each subscale have acceptable internal consistency (Cronbach's $\alpha \geq .70$)
4. Structural validity	<ul style="list-style-type: none"> Determine whether stable factor structure has been shown with: <ol style="list-style-type: none"> exploratory factor analysis, and confirmatory factor analysis (CFA) tested in an independent sample Determine whether the CFA model has good data fit with few or no modifications, using Hu and Bentler's (1999) criteria (RMSEA < .06, CFI > .95, TLI > .95, and SRMR < .08)
5. Associations with other variables	<ul style="list-style-type: none"> Evaluate whether the scale has demonstrated any of the following types of validity: <ol style="list-style-type: none"> Criterion validity: significant associations with outcomes grounded in theory or empirical findings (e.g., amount of SNS use, social capital) Convergent validity: significant associations with conceptually similar variables (e.g., SNS addiction) or another validated SNS engagement scale Discriminant validity: no excessively strong associations with similar yet distinct variables (e.g., amount of SNS use) based on Brown's (2006) cutoff (factor correlation < .80) Incremental validity: significant associations with outcomes (described above for criterion validity) while statistically controlling for scores on another validated SNS engagement scale
6. Response processes	<ul style="list-style-type: none"> Check whether the scale uses "agree/disagree" response format, which should have been tested for acquiescence bias (e.g., evaluated with CFA models) If acquiescence bias exists, make effort to mitigate it by any of the following methods: <ol style="list-style-type: none"> modifications of item wordings and response options reverse wording of items structural equation modeling techniques Alert to social desirability as a possible bias by examining whether the scale has a significant association with a validated social desirability measure Make effort to minimize social desirability bias with certain study conditions (e.g., identity anonymity).

Note: More details, as well as instructions for dealing with scales that do not meet these criteria, can be found in Section 4.1.1.

samples that were recruited to test the psychometric properties of the scales. SNS behavior has been found to vary by demographic variables, such as age (e.g., Barker, 2012), gender (e.g., Ryan, Chester, Reece, & Xenos, 2014), and culture (Na, Kosinski, & Stillwell, 2014); so researchers should endeavor to select a scale that has been validated with a sample whose demographics are similar to those of their target sample. If a potentially useful scale does not meet this criterion, we recommend that researchers test the scale's psychometric properties to ensure that it is valid with their new target sample. It is noteworthy that none of the reviewed scales have been tested for measurement invariance, so researchers are cautioned that the scale may function differently amongst the demographic subgroups of their study (e.g., men vs. women, old vs. young), and they may perform separate analyses for these subgroups to check for this potential issue.

Similar to demographics, there is robust evidence documenting that SNS engagement varies depending on which specific online platform is used (e.g., Hughes, Rowe, Batey, & Lee, 2012; Panek, Nardis, & Konrath, 2013). Thus, researchers should endeavor to select a scale that has been validated on the platform they will be investigating (see Table 1). Again, if this is not possible, we recommend that researchers validate the scale in this new context and report the results.

Researchers should then check the estimated reliability of the potential scales. Unlike validity, reliability is not a major concern because most of the reviewed scales had sufficient reliability estimates.

To ensure good structural validity, researchers should find a scale with a clear factor structure that has been explored, and then confirmed (without modifications needed) in a separate sample. However, this criterion was rarely met in our review, and so researchers may have no choice but to use a scale whose factor structure is insufficiently validated. In this case, these researchers should perform CFA to check the scale's structural validity after data collection.

As with structural validity, associations with other variables have rarely been sufficiently established amongst the scales in our review. Again, if forced to use a scale which has not established this type of validity, we recommend that researchers conduct their own test of this psychometric property. Researchers may consider using the Bergen Facebook Addiction Scale (Andreassen, Torsheim, Brunborg, & Pallesen, 2012) as an ad-hoc test of convergent validity of an SNS engagement scale. This 6-item scale is suitable for studies that require a short protocol, relatively well-validated, and should have significant positive correlations with a valid scale of SNS engagement. In addition, a significant positive correlation with amount of SNS use (e.g., average time spent on SNS, frequency of SNS behaviors) would support the criterion validity of the SNS engagement scale, and researchers should check that this association does not exceed Brown's cutoff to ensure discriminant validity.

Acquiescence bias is a common issue among the reviewed scales. This bias has been empirically shown to influence the FBI scale, and is a potential concern for six others that use the agree/disagree response format (i.e., the MFIS, SMUIS, FPIS, GoToFB scale, ISSNU, and Facebook Questionnaire). Researchers can use the CFA techniques employed by Kuru and Pasek (2016) to empirically evaluate this issue and mitigate its effect on the results if the bias is present (see Section 4.1.2 for details).

Though social desirability bias is a common concern for self-report measures, none of the scales in our review have been tested for their possible vulnerability to this problem. An example of good practice can be found in the recent study by Mäntymäki and Islam (2016), whose scale was not included in this review because it covered only one facet of SNS engagement (usage and activity counts). The researchers tested for the influence of social

desirability by including a validated social desirability measure in a partial least squares model with their SNS engagement scale. Thus, we recommend researchers to adopt similar techniques to detect the possible influence of social desirability bias, and even design their study conditions to further minimize it, such as by ensuring that participants are alone and anonymous when completing the survey (Richman, Kiesler, Weisband, & Drasgow, 1999).

In summary, given that most if not all of the reviewed scales have at least one validity concern, additional work is essential to test one or more types of validity of a SNS engagement scale. The recommendations in this section may help researchers minimize these concerns and thereby enhance study quality.

4.1.2. Measurement of SNS intensity

As evidenced by the citation counts (see Table 1), the FBI scale is by far the most widely used measure of SNS engagement. It is the pioneer of the scales in our review, but this popularity is not in accordance with the actual evidence of its psychometric properties.

Our review shows that no systematic effort has been made to validate the FBI scale, and it has been empirically shown to be susceptible to acquiescence bias, which artificially inflates its reliability and distorts its relationships with other scales (Kuru & Pasek, 2016). Hence, the widespread use of this scale cannot be taken as sufficient evidence for its validity. In short, the common practice in SNS engagement research of using the FBI scale while only checking its reliability may threaten the findings due to some validity concerns.

For studies that aim to measure SNS intensity, there are a number of alternative methods to ensure valid measurement. For researchers who would like to stick with the original items of the FBI scale, they may consider using any of the three solutions proposed by Kuru and Pasek (2016) to mitigate the acquiescence bias that this scale is susceptible to. Briefly, these researchers may consider (a) altering both the item wordings and response options, (b) reversing the wording of some items to balance out the acquiescence bias, and (c) performing structural equation modeling techniques to statistically remove the potential effects of this bias. All of these proposed solutions have been tested and described in detail by Kuru and Pasek (2016).

As these solutions are designed to deal with a single concern of acquiescence bias, there is still no full guarantee of the validity of the FBI scale. We recommend that in addition to using one of the solutions described above, researchers using the FBI scale should test its reliability, structural validity, and associations with other variables.

Alternatively, researchers may consider using either the MFIS or SMUIS, both of which were developed on the basis of Ellison et al. (2007)'s concept of Facebook Intensity. These scales were constructed using a multi-stage item development process and were systematically validated in multiple samples. However, the MFIS and SMUIS both use the same response format as the FBI scale (i.e., agree/disagree), and thus they may also be susceptible to acquiescence bias, though this potential problem has not been empirically evaluated (see Sections 3.1.2 and 3.1.8 for the psychometric properties of both scales).

4.2. Implications for validation studies

Our review identified some validity concerns for most of the included scales, indicating ample opportunities for future validation studies to make a major impact on the assessment of SNS engagement, and to further advance the field. Validation researchers looking for areas where more work is needed can target specific issues raised for each scale (see the Results section). Requirements for a reliable, valid SNS engagement scale are

summarized in Table 3 (see also Sections 2.4.1 to 2.4.5). As none of the scales in our review have been tested for their vulnerability to social desirability bias, this could be an impactful test for future validation studies. We discuss below a few additional broader issues that emerge from the present review.

4.2.1. Additional tests of criterion validity

We found that evidence supporting the criterion validity of SNS engagement scales is often derived from respondents' self-report of their estimated time spent on the SNS or frequency of undertaking specific SNS behaviors. Recent concerns have arisen regarding the possible recall bias from these self-report data that may threaten the credibility of validation findings (Junco, 2013; Scharnow, 2016). To obtain more solid evidence for the criterion validity of SNS engagement scales, validation researchers can improve the field by adopting more sophisticated alternative methods such as objective measures (e.g., objective logs) and mixed methods (e.g., subjective reports and objective logs).

We recommend that one promising alternative would be to use an SNS platform's application programming interface (API) to collect publicly available objective records of user's activities on that platform. This method can be carried out together with any self-report survey. Survey participants can report their user ID on the SNS platform, and researchers can use this ID to collect participants' data from the API. This type of mixed-methods data collection has already been done with Twitter (Riedl, Köbler, Goswami, & Krcmar, 2013), though this study did not focus on SNS engagement. Researchers who are interested in adopting this novel method in studying SNS engagement may consult a useful guide for data collection from Twitter with the R programming language (Murphy, 2017).

4.2.2. Studies with diverse samples

In view of the known demographic variations in SNS behavior, and the importance of checking a scale's validity across contexts (Messick, 1995), scales of SNS engagement should demonstrate sound psychometric properties in samples with wide demographic ranges. These scales should show measurement invariance across different demographic groups as well.

Most of the scales in our review, however, were limited by their sample homogeneity. For example, the majority (> 80%) of the included scales were validated exclusively with young samples whose mean age was below 30, and half of these scales were validated only with samples containing more than 60% women. Moreover, slightly more than half of the reviewed scales were validated only with student samples. None of these studies have conducted any tests of measurement invariance. Thus, future studies may contribute to this research area by validating SNS engagement scales with community samples characterized by a good gender balance and broad age range, and more important, by testing measurement invariance across gender and age.

Despite these demographic concerns, it is encouraging to note the cultural diversity of the samples reported in the studies we reviewed. The surveys have been conducted in 11 different countries from diverse cultural regions around the world, namely Africa (i.e., Nigeria; Olufadi, 2016), Asia (e.g., Taiwan; Yu, 2015), Europe (e.g., Portugal; Costa et al., 2016), the Middle East (i.e., Kuwait; Aladwani, 2014), and North America (e.g., the U.S.; Jenkins-Guarnieri et al., 2013). However, because only one of these scales has been validated in multiple locations (the MTUAS), a fruitful research direction would be to adapt these scales for use in diverse cultural regions (see Wild et al., 2005 for guidelines). Cross-cultural tests of measurement invariance (e.g., Cheng et al., 2011; C. Lai et al., 2015), which have not been conducted for any of the scales, would contribute to fill this knowledge gap. This type of cross-cultural

comparison can lay the foundation for a better understanding of potential country differences in SNS engagement.

4.2.3. Studies with additional platforms

Multiple studies have shown that SNS engagement varies by platform. Facebook has been the most popular SNS for many years and has garnered a large amount of attention from scholars. In view of the growing popularity of a number of alternative online platforms such as WeChat (Business Insider, 2016) and Instagram (eMarketer, 2016), more studies should be conducted with these platforms that have so far received scant research attention. We recommend that SNS engagement scales need to have good psychometric properties across multiple online platforms.

The scales in our review were focused almost entirely on Facebook. More than 65% of the scales were assessed exclusively with Facebook users, and this narrow focus may impede knowledge advancement in research on engagement with SNS as a whole. For example, an exclusive focus on Facebook precludes SNS research in countries where Facebook is inaccessible (e.g., Mainland China) or not popular (e.g., Japan). We recommend that future validation studies should focus on general SNS engagement, as well as specific online platforms other than Facebook.

4.3. Limitations

As an attempt to include all articles that constructed or systematically validated SNS engagement scales, we relied on multiple literature search strategies: database searches, reference list searches, and lists of citing articles. The only articles that may have been missed are those that are not primarily focused on scale development or validation (and did not include the corresponding keywords), were not cited by any articles in our review, or did not cite any articles in our review. We consider that these types of study are less likely to provide rigorous, important contributions to the literature, and thus their absence should not pose a major threat to the comprehensiveness of this review.

4.4. Conclusions

In the present review we provide a psychometric guide to the scales constructed to measure SNS engagement. Given the recent flourishing of SNS research as social networking has become an essential part of people's lives, SNS engagement is a critical research area that urgently needs further development. Though there are a number of promising scales, each one still needs additional validation work. This review may provide an effective guide for researchers planning to conduct studies on SNS engagement, and may spur new psychometric research for further evaluations of SNS engagement scales. Continued psychometric research is essential for providing a foundation of valid measurement in this timely research topic.

Conflicts of interest

All authors declare that they have no conflicts of interest.

Acknowledgements

The authors thank Mike Cheung, King-wa Fu, and Timothy Levine for their constructive comments on earlier drafts of this manuscript. Thanks also go to Fanny Cheng and Stella Tam for clerical assistance. This study was funded by Hong Kong Research Grants Council's General Research Fund (17400714), and the University of Hong Kong's Seed Fund for Incubating Group-based Collaborative Research Projects (102009405) and Seed Fund for

Basic Research (201411159152). The funders had no role in the study design and administration, data analysis or interpretation, manuscript writing, or the decision to submit the paper for publication.

References

- Aladwani, A. M. (2014). Gravitating towards Facebook (GoToFB): What it is? and How can it be measured? *Computers in Human Behavior*, 33, 270–278. <https://doi.org/10.1016/j.chb.2014.01.005>.
- Alhabash, S., Chiang, Y. H., & Huang, K. (2014). MAM & U&G in Taiwan: Differences in the uses and gratifications of Facebook as a function of motivational reactivity. *Computers in Human Behavior*, 35, 423–430. <https://doi.org/10.1016/j.chb.2014.03.033>.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.
- Andreassen, C. S., Torsheim, T., Brunborg, G. S., & Pallesen, S. (2012). Development of a Facebook addiction scale. *Psychological Reports*, 110(2), 501–517. <https://doi.org/10.2466/02.09.18.PR0.110.2.501-517>.
- Barker, V. (2012). A generational comparison of social networking site use: The influence of age and social identity. *The International Journal of Aging and Human Development*, 74(2), 163–187. <https://doi.org/10.2190/AG.74.2.d>.
- Beyens, I., Frison, E., & Eggermont, S. (2016). "I don't want to miss a thing": Adolescents' fear of missing out and its relationship to adolescents' social needs, Facebook use, and Facebook related stress. *Computers in Human Behavior*, 64, 1–8. <https://doi.org/10.1016/j.chb.2016.05.083>.
- Blachnio, A., Przepiorka, A., Boruch, W., & Balakier, E. (2016). Self-presentation styles, privacy, and loneliness as predictors of Facebook use in young people. *Personality & Individual Differences*, 94, 26–31. <https://doi.org/10.1016/j.paid.2015.12.051>.
- Bodroža, B., & Jovanović, T. (2016). Validation of the new scale for measuring behaviors of Facebook users: Psycho-social aspects of Facebook use (PSAFU). *Computers in Human Behavior*, 54, 425–435. <https://doi.org/10.1016/j.chb.2015.07.032>.
- Bodroža, B., Popov, B., & Poljak, I. (2009). Procena virtuelnih pona sanja u dru stvenim mre zama [The assessment of virtual behaviours on social networks]. In M. Biro, S. Smederavac, & Z. Novovic (Eds.), *Procena psiholoskih i psihopatoloskih fenomena [The assessment of psychological and psychopathological phenomena]* (pp. 179–192). Beograd: Centar za primenjenu psihologiju.
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305–314. <https://doi.org/10.1037/0033-2909.110.2.305>.
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-mediated Communication*, 13(1), 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>.
- Boyd, D., Golder, S., & Lotan, G. (2010). *Tweet, tweet, retweet: Conversational aspects of retweeting on twitter*. Proceedings of the annual Hawaii international conference on system sciences (pp. 1–10). <https://doi.org/10.1109/HICSS.2010.412>.
- Bradley, M. T., & Gupta, R. D. (1997). Estimating the effect of the file drawer problem in meta-analysis. *Perceptual & Motor Skills*, 85(2), 719–722. <https://doi.org/10.2466/pms.1997.85.2.719>.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research* (1st ed.). New York, NY: Guilford Publications.
- Burke, M., & Kraut, R. (2013). *Using Facebook after losing a job: Differential benefits of strong and weak ties*. Proceedings of the 2013 conference on computer supported cooperative work <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84874824661&partnerID=40&md5=1c70ff4219d7c28f7717d556dc9507d1>.
- Business Insider. (2016). WeChat breaks 700 million monthly active users. Retrieved from: <http://www.businessinsider.com/wechat-breaks-700-million-monthly-active-users-2016-4>.
- Charlton, J. P. (2002). A factor-analytic investigation of computer 'addiction' and engagement. *British Journal of Psychology*, 93(3), 329–344. <https://doi.org/10.1348/000712602760146242>.
- Charlton, J. P., & Danforth, I. D. W. (2007). Distinguishing addiction and high engagement in the context of online game playing. *Computers in Human Behavior*, 23(3), 1531–1548. <https://doi.org/10.1016/j.chb.2005.07.002>.
- Cheng, C., Cheung, S., Chio, J. H., & Chan, M. S. (2013). Cultural meaning of perceived control: A meta-analysis of locus of control and psychological symptoms across 18 cultural regions. *Psychological Bulletin*, 139(1), 152–188. <https://doi.org/10.1037/a0028596>.
- Cheng, C., Cheung, M. W. L., & Lo, B. C. Y. (2016). Relationship of health locus of control with specific health behaviours and global health appraisal: A meta-analysis and effects of moderators. *Health Psychology Review*, 10(4), 460–477. <https://doi.org/10.1080/17437199.2016.1219672>.
- Cheng, C., Cheung, M. W. L., & Montasem, A. (2016). Explaining differences in subjective well-being across 33 nations using multilevel models: Universal personality, cultural relativity, and national income. *Journal of Personality*, 84(1), 46–58. <https://doi.org/10.1111/jopy.12136>.
- Cheng, C., Jose, P. E., Sheldon, K. M., Singelis, T. M., Cheung, M. W. L., Tiliouine, H., et al. (2011). Sociocultural differences in self-construal and subjective well-being: A test of four cultural models. *Journal of Cross-cultural Psychology*, 42(5), 832–855. <https://doi.org/10.1177/00220222110381117>.
- Cheng, C., & Li, A. Y. (2014). Internet addiction prevalence and quality of (real) life: A meta-analytic review of 31 nations across seven world regions. *Cyberpsychology, Behavior, and Social Networking*, 17(12), 755–760. <https://doi.org/10.1089/cyber.2014.0317>.
- Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Los Angeles, CA: Sage.
- Costa, J. J. M., Paula, A., Pinheiro, R., Salvador, C., Vale-dias, L., & Zenha-rela, M. (2016 July). *Evaluating use and attitudes towards social media and ICT for Portuguese youth: The MTUAS-PY scale*. Paper presented at the 2nd International Conference on Health and Health Psychology, Porto, Portugal.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor Analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 10(7), 1–9. <https://doi.org/10.1111/j.1442-9993.2009.02074.x>.
- Devellis, R. F. (2003). *Scale Development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
- Distefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research and Evaluation*, 14(20), 1–11. doi:10.1.1.460.8553.
- Dreamgrow.com. (2017). *Top 15 most popular social networking sites and apps*. Retrieved from <https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/>.
- Elliott, L. J., & Polyakova, V. (2014). Beyond Facebook: The generalization of social networking site measures. *Computers in Human Behavior*, 33, 163–170. <https://doi.org/10.1016/j.chb.2014.01.023>.
- Ellison, N. B., & Boyd, D. M. (2013). Sociality through social network sites. In W. H. Dutton (Ed.), *Oxford handbook of Internet studies* (pp. 151–172). Oxford, United Kingdom: Oxford University.
- Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-mediated Communication*, 12(4), 1143–1168. <https://doi.org/10.1111/j.1083-6101.2007.00367.x>.
- eMarketer. (2016). Instagram continues double-digit growth. Retrieved from: <https://www.emarketer.com/Article/Instagram-Continues-Double-Digit-Growth/1013612>.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299.
- Fan, X. (2003). Two approaches for correcting correlation attenuation caused by measurement error: Implications for research practice. *Educational & Psychological Measurement*, 63(6), 915–930.
- Faraon, M., & Kaipainen, M. (2014 August). *Much more to it: The relation between facebook usage and self-esteem*. Paper presented at the IEEE 15th international conference, Huddinge, Sweden.
- Furr, R. M. (2010). *Scale construction and psychometrics for social and personality psychology*. CA: Sage; Thousand Oaks.
- Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.
- Gordon, R. A. (1987). Social desirability bias: A demonstration and technique for its reduction. *Teaching of Psychology*, 14(1), 40–42. https://doi.org/10.1207/s15328023top1401_11.
- Gravdal, L., & Sandal, G. M. (2006). The two-factor model of social desirability: Relation to coping and defense, and implications for health. *Personality & Individual Differences*, 40(5), 1051–1061. <https://doi.org/10.1016/j.paid.2005.11.004>.
- Hinkin, T. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1, 104–121.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2), 561–569. <https://doi.org/10.1016/j.chb.2011.11.001>.
- Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*, 15(4), 446–455. <https://doi.org/10.1037/1040-3590.15.4.446>.
- Jenkins-Guarnieri, M. A., Wright, S. L., & Johnson, B. (2013). Development and validation of a social media use integration scale. *Psychology of Popular Media Culture*, 2(1), 38–50. <https://doi.org/10.1037/a0030277>.
- Jiang, Y., & Bruijn, O. D. (2013). Facebook helps: A case study of cross-cultural social networking and social capital. *Information, Communication & Society*, 17(6), 732–749. <https://doi.org/10.1080/1369118X.2013.830636>.
- Joinson, A. N. (2008 April). 'Looking at', 'looking up' or 'keeping up with' People? *Motives and uses of Facebook*. Paper presented at the 26th annual SIGCHI conference on human factors in computing systems, Florence, Italy.
- Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, 29(3), 626–631. <https://doi.org/10.1016/j.chb.2012.11.007>.
- Knowles, E. S., & Nathan, K. T. (1997). Acquiescent responding in self-reports: Cognitive style or social concern? *Journal of Research in Personality*, 31(2), 293–301. <https://doi.org/10.1006/jrpe.1997.2180>.
- Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues*, 58(1), 49–74. <https://doi.org/10.1111/1540-4560.00248>.
- Kuru, O., & Pasek, J. (2016). Improving social media measurement in surveys: Avoiding acquiescence bias in Facebook research. *Computers in Human Behavior*,

- 57, 82–92. <https://doi.org/10.1016/j.chb.2015.12.008>.
- Kuss, D. J., & Griffiths, M. D. (2012). Internet gaming addiction: A systematic review of empirical research. *International Journal of Mental Health and Addiction*, 10(2), 278–296. <https://doi.org/10.1007/s11469-011-9318-5>.
- Lai, C., Mak, K., Cheng, C., Watanabe, H., Nomachi, N., Bahar, N., et al. (2015). Measurement invariance of the Internet addiction test among Hong Kong, Japanese, and Malaysian adolescents. *Cyberpsychology, Behavior, and Social Networking*, 18(10), 609–617. <https://doi.org/10.1016/j.puhe.2015.07.031>.
- Lai, C. Y., & Yang, H. L. (2014). Determinants and consequences of Facebook feature use. *New Media & Society*, 18(7), 1310–1330. <https://doi.org/10.1177/1461444814555959>.
- Lee, S. L., Kim, J. A., Golden, K. J., Kim, J. H., & Park, S. A. (2016). A cross-cultural examination of SNS usage intensity and managing interpersonal relationships online: The role of culture and the autonomous-related self-construal. *Frontiers in Psychology*, 7(376), 1–12. <https://doi.org/10.3389/fpsyg.2016.00376>.
- Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports*, 22(4), 335–338. <https://doi.org/10.1080/00036810500317730>.
- Li, J., Lau, J. T. F., Mo, P. K. H., Su, X., Wu, A. M. S., Tang, J., et al. (2016). Validation of the social networking activity intensity scale among junior middle school students in China. *PLoS One*, 11(10). <https://doi.org/10.1371/journal.pone.0165695>.
- Lin, C., Hung, W., Fang, K., & Tu, C. (2015). Understanding players' achievement values from MMORPGs: An exploratory study. *Internet Research*, 25(5), 829–851. <https://doi.org/10.1108/IntR-12-2013-0268>.
- Liu, D., Ainsworth, S. E., & Baumeister, R. F. (2016). A meta-analysis of social networking online and social capital. *Review of General Psychology*, 20(4), 369–391. <https://doi.org/10.1037/gpr0000091>.
- Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, 30(4), 256–265. <https://doi.org/10.1080/01972243.2014.915276>.
- Mahrt, M., & Scharrow, M. (2013). The value of big data in digital media research. *Journal of Broadcasting & Electronic Media*, 57(1), 20–33. <https://doi.org/10.1080/08838151.2012.761700>.
- Mäntymäki, M., & Islam, A. K. M. N. (2016). The Janus face of Facebook: Positive and negative sides of social networking site use. *Computers in Human Behavior*, 61, 14–26. <https://doi.org/10.1016/j.chb.2016.02.078>.
- McCay-Peet, L., & Quan-Haase, A. (2016). A model of social media engagement: User profiles, gratifications, and experiences. In *Why engagement matters: Cross-disciplinary perspectives of user engagement in digital media* (pp. 199–217).
- Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, 57(2), 289–311.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>.
- Moqbel, M., Nevo, S., & Kock, N. (2013). Organizational members' use of social networking sites and job performance. *Information Technology and People*, 26(3), 240–264. <https://doi.org/10.1108/ITP-10-2012-0110>.
- Murphy, S. C. (2017). A hands-on guide to conducting psychological research on Twitter. *Social Psychological and Personality Science*, 8(4), 396–412. <https://doi.org/10.1177/1948550617697178>.
- Na, J., Kosinski, M., & Stillwell, D. J. (2014). When a new tool is introduced in different cultural contexts: Individualism-collectivism and social network on Facebook. *Journal of Cross-Cultural Psychology*, 46(3), 355–370. <https://doi.org/10.1177/0022022114563932>.
- Niu, G., Sun, X., Zhou, Z., Kong, F., & Tian, Y. (2016). The impact of social network site (Qzone) on adolescents' depression: The serial mediation of upward social comparison and self-esteem. *Acta Psychologica Sinica*, 48(10), 1282–1291. <https://doi.org/10.3724/SPJ.1041.2016.01282>.
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics*, 1(22), 161–169. <https://doi.org/10.1016/j.joi.2006.12.001>.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
- O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A Conceptual Framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938–955. <https://doi.org/10.1002/asi.20801.1>.
- Olufadi, Y. (2016). Social networking time use scale (SONTUS): A new instrument for measuring the time spent on the social networking sites. *Telematics & Informatics*, 33(2), 452–471. <https://doi.org/10.1016/j.tele.2015.11.002>.
- Orosz, G., Tóth-Király, I., & Bóthe, B. (2016). Four facets of Facebook intensity — the development of the multidimensional Facebook intensity scale. *Personality & Individual Differences*, 100, 95–104. <https://doi.org/10.1016/j.paid.2015.11.038>.
- Özgür, H. (2016). Adapting the media and technology usage and attitudes scale to Turkish. *Kuram ve Uygulamada Eğitim Bilimleri*, 16(5), 1711–1735. <https://doi.org/10.12738/estp.2016.5.0085>.
- Panek, E. T., Nardis, Y., & Konrath, S. (2013). Mirror or Megaphone?: How relationships between narcissism and social networking site use differ on Facebook and Twitter. *Computers in Human Behavior*, 29(5), 2004–2012. <https://doi.org/10.1016/j.chb.2013.04.012>.
- Park, K. G., Han, S., & Kaid, L. L. (2012). Does social networking service usage mediate the association between smartphone usage and social capital? *New Media & Society*, 15(7), 1077–1093. <https://doi.org/10.1177/1461444812465927>.
- Park, C., Jun, J., & Lee, T. (2015). Consumer characteristics and the use of social networking sites. *International Marketing Review*, 32(3), 414–437. <https://doi.org/10.1108/IMR-09-2013-0213>.
- Park, N., Lee, S., & Kim, J. H. (2012). Individuals' personal network characteristics and patterns of Facebook use: A social network approach. *Computers in Human Behavior*, 28(5), 1700–1707. <https://doi.org/10.1016/j.chb.2012.04.009>.
- Pew Research Center, Washington D.C. (2011). Social networking sites and our lives. Retrieved from: <http://www.pewinternet.org/2011/06/16/social-networking-sites-and-our-lives/>.
- Piwek, L., & Joinson, A. (2016). What do they snapchat about? Patterns of use in time-limited instant messaging service. *Computers in Human Behavior*, 54, 358–367. <https://doi.org/10.1016/j.chb.2015.08.026>.
- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, 84(5), 754–775. <https://doi.org/10.1037/0021-9010.84.5.754>.
- Riedl, C., Köbler, F., Goswami, S., & Krcmar, H. (2013). Tweeting to feel connected: A model for social connectedness in online social networks. *International Journal of Human-Computer Interaction*, 29(10), 670–687. <https://doi.org/10.1080/10447318.2013.768137>.
- Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, 59(2), 464–468. <https://doi.org/10.1111/j.0014-3820.2005.tb01004.x>.
- Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A., & Rökkum, J. (2013). The media and technology usage and attitudes Scale: An empirical investigation. *Computers in Human Behavior*, 29(6), 2501–2511. <https://doi.org/10.1016/j.chb.2013.06.006>.
- Ross, C., Orr, E. S., Sisis, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior*, 25, 578–586. <https://doi.org/10.1016/j.chb.2008.12.024>.
- Ryan, T., Chester, A., Reece, J., & Xenos, S. (2014). The uses and abuses of Facebook: A review of Facebook addiction. *Journal of Behavioral Addictions*, 3(3), 133–148. <https://doi.org/10.1556/JBA.3.2014.016>.
- Salehan, M., & Negahban, A. (2013). Social networking on smartphones: When mobile phones become addictive. *Computers in Human Behavior*, 29(6), 2632–2639. <https://doi.org/10.1016/j.chb.2013.07.003>.
- Scharrow, M. (2016). The accuracy of self-reported Internet use—a validation study using client log data. *Communication Methods and Measures*, 10(1), 13–27. <https://doi.org/10.1080/19312458.2015.1118446>.
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ*, 349, g7647.
- Sigerson, L., Li, A. Y., Cheung, M. W., Luk, J. W., & Cheng, C. (2017). Psychometric properties of the Chinese Internet gaming disorder scale. *Addictive Behaviors*, 74, 20–26. <https://doi.org/10.1016/j.addbeh.2017.05.031>.
- Statista. (2018). Most famous social network sites 2017, by active users. Retrieved from: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.
- Steinfeld, C., Ellison, N. B., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, 29(6), 434–445. <https://doi.org/10.1016/j.appdev.2008.07.002>.
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18.
- Topaloglu, M., Caldibi, E., & Oge, G. (2016). The scale for the individual and social impact of students' social network use: The validity and reliability studies. *Computers in Human Behavior*, 61, 350–356. <https://doi.org/10.1016/j.chb.2016.03.036>.
- Turel, O., & Serenko, A. (2012). The benefits and dangers of enjoyment with social networking websites. *European Journal of Information Systems*, 21(5), 512–528. <https://doi.org/10.1057/ejis.2012.1>.
- Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues & Policy Review*, 11(1), 274–302. <https://doi.org/10.1111/sipr.12033>.
- Wan, C.-S., & Chiou, W.-B. (2006). Psychological motives and online games addiction: A test of flow theory and humanistic needs theory for Taiwanese adolescents. *CyberPsychology & Behavior*, 9(3), 317–324.
- Wang, J.-L., Jackson, L. A., Gaskin, J., & Wang, H.-Z. (2014). The effects of Social Networking Site (SNS) use on college students' friendship and well-being. *Computers in Human Behavior*, 37, 229–236. <https://doi.org/10.1016/j.chb.2014.04.051>.
- Wang, A. T., Sandhu, N. P., Wittich, C. M., Mandrekar, J. N., & Beckman, T. J. (2012). Using social media to improve continuing medical education: A survey of course participants. *Mayo Clinic Proceedings*, 87(12), 1162–1170. <https://doi.org/10.1016/j.mayocp.2012.07.024>.
- Wellman, B. (2002). Little boxes, globalization, and networked individualism. In M. Tanabe, P. V. D. Besselaar, & T. Ishida (Eds.), *Digital cities II: Computational & sociological approaches* (pp. 10–25). Berlin, Germany: Springer.
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., et al. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health*, 8(2), 94–104.
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *Counseling Psychologist*, 34(6),

- 806–838. <https://doi.org/10.1177/0011000006288127>.
- Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation, 12*(3), 1–26.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Young, K. (1998). *Caught in the net: How to recognize the signs of internet addiction and a winning strategy for recovery*. New York, NY: John Wiley & Sons.
- Yu, S.-C. (2015). Happiness or Addiction: An example of taiwanese college students' use of Facebook. *International Journal of Technology & Human Interaction, 11*, 26–40. <https://doi.org/10.4018/IJTHI.2015100102>.