

Using Item Response Theory to Develop Measures of Acquisitive and Protective Self-Monitoring From the Original Self-Monitoring Scale

Assessment
1–15
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1073191115615213
asm.sagepub.com


Michael P. Wilmot¹, Jack W. Kostal¹, David Stillwell², and Michal Kosinski³

Abstract

For the past 40 years, the conventional univariate model of self-monitoring has reigned as the dominant interpretative paradigm in the literature. However, recent findings associated with an alternative bivariate model challenge the conventional paradigm. In this study, item response theory is used to develop measures of the bivariate model of acquisitive and protective self-monitoring using original Self-Monitoring Scale (SMS) items, and data from two large, nonstudent samples ($N_s = 13,563$ and 709). Results indicate that the new acquisitive (six-item) and protective (seven-item) self-monitoring scales are reliable, unbiased in terms of gender and age, and demonstrate theoretically consistent relations to measures of personality traits and cognitive ability. Additionally, by virtue of using original SMS items, previously collected responses can be reanalyzed in accordance with the alternative bivariate model. Recommendations for the reanalysis of archival SMS data, as well as directions for future research, are provided.

Keywords

self-monitoring, acquisitive and protective self-monitoring, item response theory, Five-Factor Model, metatraits, plasticity and stability

Self-monitoring of expressive behavior (Snyder, 1974), or *self-monitoring*, is a major construct in the personality and social psychological literature. Self-monitoring concerns the extent to which people can and do engage in the expressive control of their public self-presentations and is typically measured using full-scale scores from the Self-Monitoring Scale (SMS; Snyder, 1987). Traditionally, self-monitoring has been interpreted as a univariate and categorical construct (i.e., high and low self-monitor classes; Gangestad & Snyder, 1985), which is largely distinct from other personality traits (Barrick, Parks, & Mount, 2005). Results from quantitative reviews provide support for the construct validity of self-monitoring and its *conventional univariate model* (Day, Schleicher, Unckless, & Hiller, 2002; Gangestad & Snyder, 2000), which has prompted a surge of research interest in the construct.

Despite its empirical success, emerging evidence challenges key assumptions and measurement practices of the conventional self-monitoring model. First, factor-analytic evidence indicates that the SMS is multidimensional in both its original 25-item version (Briggs, Cheek, & Buss, 1980) and its revised 18-item version (SMS-R; Briggs & Cheek, 1988; Snyder & Gangestad, 1986). Second, recent findings indicate that self-monitoring is not a single, categorical

variable (Wilmot, 2015), but rather it is composed of two uncorrelated dimensions: *acquisitive* and *protective* self-monitoring (Lennox, 1988). Third, when examined separately, these two dimensions can largely be integrated into the contemporary taxonomy of personality traits (Wilmot, DeYoung, Stillwell, & Kosinski, 2015). Taken together, new findings revive an earlier argument that, because the acquisitive and protective dimensions are independent and have divergent networks of relations to other psychological and external variables, they ought to be conceptualized and assessed separately (John, Cheek, & Klohnen, 1996). We refer to this competing model, in which self-monitoring is conceptualized as comprising the two independent dimensions of acquisitive and protective self-monitoring, as the *alternative bivariate model*.

Given the growing evidence challenging the conventional univariate model and supporting the alternative

¹University of Minnesota, Minneapolis, MN, USA

²University of Cambridge, Cambridge, UK

³Stanford University, Stanford, CA, USA

Corresponding Author:

Michael P. Wilmot, Psychology Department, University of Minnesota, Twin Cities, 75 East River Road, Minneapolis, MN 55455, USA.
Email: wilmo040@umn.edu

bivariate conceptualization, we use the modern psychometric approach of item response theory (IRT; de Ayala, 2009) to develop concise measures of acquisitive and protective self-monitoring.¹ We use the original SMS as the source of items for the new scales, as it enables the reanalysis of previously collected responses using the bivariate model. To construct and validate the scales, we run a series of analyses using data from two large, nonstudent samples. We use factor-analytic, classical psychometric, and IRT methods to select items for the new scales. Because dichotomous and polytomous response formats have seen equal use in prior self-monitoring research, we examine both formats. Next, we use tests of differential item functioning (DIF) to ascertain whether the newly developed measures are free from age and gender biases (see Day et al., 2002). Finally, to test the construct validity of the two scales, we examine convergent and divergent relations to personality factors and facets, and to cognitive ability. The interested reader can take advantage of this IRT-based scoring approach by applying our item parameters to their raw data or by visiting <http://www.mpwilmot.com/> to score their data online.

Method

Participants

Sample 1. Participants were 13,563 Facebook users (8,198 females, 5,365 males) who completed the 25-item SMS and provided gender and age information. Ages ranged from 18 to 78 years (median = 23, interquartile range = 8). Data were collected via the myPersonality application (www.mypersonality.org) between the years of 2007 and 2012 (Kosinski, Stillwell, & Graepel, 2013). Users of myPersonality elected to provide research data in exchange for the opportunity to complete assessments and display results on their Facebook profiles. Among users who provided their location (91.4%), the United States (67.7%) and Great Britain (18.61%) were most common. All myPersonality measures were administered in English.

Sample 2. Participants were 709 members of the Eugene Springfield Community Sample (303 males and 406 females), ranging from 18 to 85 years (median = 49, interquartile range = 19). Participants were homeowners recruited by mail to complete questionnaires, for payment, between 1994 and 2007. For both samples, to qualify for inclusion, complete self-report data were required for the SMS, gender, and age.

Measures

SMS. The SMS (Snyder, 1974) contains 25 items. Sample 1 completed the SMS using a true–false response format; a 5-point Likert-type format was used in Sample 2.

Five-Factor Model Factors and Facets. A subset of Study 1 participants ($N = 1,197$) completed the IPIP-NEO (International Personality Item Pool Neuroticism–Extraversion–Openness), which is a proxy of the NEO-PI-R (NEO Personality Inventory–Revised) based on the public domain IPIP (Goldberg et al., 2006). The IPIP-NEO contains 336 items that are scored using a 5-point Likert-type response format. The inventory breaks each of the Five-Factor Model (FFM) trait domains into six 10-item facet scales. There is high degree of correspondence between NEOPI-R and IPIP-NEO scores (observed $r_{\text{mean}} = .73$; corrected for attenuation, $r_{\text{mean}} = .90$). Internal consistency estimates for the facet scales ranged from $\alpha = .71$ to $.90$, and from $\alpha = .74$ to $.85$, for factor scales.

A subset of Study 2 ($N = 644$) completed the NEO-PI-R (Costa & McCrae, 1992), which contains 240 items scored using a 5-point Likert-type response format. The inventory breaks each of the FFM traits into six facets, each of which are assessed by eight items. Internal consistency estimates for the facet scales ranged from $\alpha = .61$ to $.85$, and from $\alpha = .70$ to $.87$ for factor scales.

Cognitive Ability. A subset of Sample 1 ($N = 1,651$) completed the IRT-based proxy for the Raven Standard Progressive Matrices test (Raven, 2000) developed for the myPersonality application (Stillwell & Kosinski, 2015). The test was delivered as a computerized adaptive test and had a length of 20 items. IRT parameters of the items were calibrated against Raven Standard Progressive Matrices using three items from the original test that were administered to all participants. Measurement properties of the new scale were equal or better than those of the original (see Stillwell & Kosinski, 2015).

A subset of Sample 2 ($N = 594$) completed Cattell's 16 Personality Factors Questionnaire (16PF; Conn & Rieke, 1994). The 16PF includes a 15-item cognitive ability test (i.e., Factor B) that includes knowledge and reasoning problems with multiple-choice answers ($\alpha = .73$). The scale has been used as a cognitive ability proxy in prior research (e.g., DeYoung, Grazioplene, & Peterson, 2012).

Results

Factor Analyses

To determine the number of factors present in the 25-item SMS in both data sets, we used Velicer's (1976) minimum average partial (MAP) test and parallel analysis (O'Connor, 2000).² In addition, we examined the eigenvalues extracted, and two indices of fit, as expressed in the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR).³ For Sample 1, MAP test results indicated two factors, whereas parallel analysis indicated three. For Sample 2, MAP test results and parallel analysis alike indicated the presence of three factors. The

relatively convergent findings were consistent with prior analyses of the SMS (e.g., Briggs et al., 1980) and provided support for our multidimensional approach. Since tests indicated three rather than two factors, we explored the validity of a three-factor solution first.

Three-Factor EFAs. We conducted exploratory factor analyses (EFAs) using principal axis factoring, and three factors were extracted. Across samples, eigenvalues were indicative of two robust factors, but a comparatively weak third factor (Sample 1 $\lambda_s = 2.37, 1.84,$ and $.64$; Sample 2 $\lambda_s = 3.36, 2.46,$ and 1.05 , respectively). Using the conventional loading criteria of $\geq .30$ in absolute value, results indicated that the third factor was defined by only one item in Sample 1, and four items in Sample 2; maximum factor loadings were relatively weak ($r_{S1} = .29, r_{S2} = .42$), and three of the five total items defining factor 3 had substantial cross-loadings. Fit indices for the three-factor model showed excellent fit (RMSEAs = $.036$ and $.050$; SRMRs = $.027$ and $.041$, for Samples 1 and 2, respectively). Fit notwithstanding, seeing as our purpose was not to fit a measurement model to the SMS, but rather to develop robust and reliable subscales from it, we decided in favor of the more parsimonious two-factor model; in the last analysis, the comparatively small third eigenvalue, and fewer and weaker item indicators, provided persuasive evidence against the three-factor model.

Two-Factor EFAs. Next, we conducted a second set of EFAs using principal axis factoring, and two factors were extracted and rotated; we used a direct oblimin rotation, which allows for correlated factors, to rotate factors toward simple structure. Nevertheless, even on rotation, results indicated that the two factors were virtually orthogonal (factor $r_s = .08$ and $.07$, respectively), a finding which is consistent with that of Briggs and Cheek (1988). Fit indices for the two-factor model remained in the range of good fit (RMSEAs = $.044$ and $.065$; SRMRs = $.037$ and $.057$, for Samples 1 and 2, respectively). With a few minor exceptions (e.g., Items 6 and 23 for the protective factor), pattern matrix loadings mirrored loadings of the unrotated matrix. Consequently, as potential items for confirmatory factor analysis (CFA), we selected all items that met the conventional factor-loading criterion of $\geq .30$ in absolute value.⁴ Nine items met this criterion for the acquisitive factor (Items 1, 5, 8, 12, 14, 18, 20, 21, and 22), eight items met the criterion for the protective factor (Items 2, 7, 13, 15, 16, 17, 19, and 25), and two items (Items 6 and 23) met criteria for both factors; both cross-loading items were included as indicators of both factors.⁵ EFA results are included in the supplementary online materials (all supplementary materials available online at <http://asm.sagepub.com/content/by/supplemental-data>).

Two-Factor CFAs. Having found that a two-factor model provided a good fit to the data, and having selected item

Table 1. Confirmatory Factor-Analytic Results of the Self-Monitoring Scale (SMS): Two-Factor Model.

SMS item	Sample 1		Sample 2	
	Acquisitive	Protective	Acquisitive	Protective
1	.40	—	.30	—
2	—	.36	—	.40
3	—	—	—	—
4	—	—	—	—
5	.40	—	.52	—
6	.33	.37	.50	.33
7	—	.36	—	.31
8	.54	—	.65	—
9	—	—	—	—
10	—	—	—	—
11	—	—	—	—
12	.52	—	.61	—
13	—	.54	—	.58
14	.38	—	.38	—
15	—	.35	—	.41
16	—	.49	—	.63
17	—	.27	—	.26
18	.49	—	.59	—
19	—	.50	—	.61
20	.50	—	.59	—
21	.30	—	.36	—
22	.40	—	.60	—
23	.45	$-.27$.56	$-.29$
24	—	—	—	—
25	—	.38	—	.48

Note. Ns for Samples 1 and 2 = 13,563 and 709, respectively. Maximum likelihood estimation of the full covariance matrix. Factor loadings $\geq .30$ in absolute value are bolded. Factor intercorrelations (i.e., ϕ) = $.03$ for both samples. For Sample 1, RMSEA = $.055$, SRMR = $.064$; for Sample 2, RMSEA = $.076$, SRMR = $.091$. All item numbers refer to the order of items as presented in the original 25-item SMS (Snyder, 1974).

indicators of the acquisitive and protective self-monitoring factors, we tested the validity of our two sets of items using CFA. The nine acquisitive items were used to define the latent acquisitive factor, the eight protective items were used to define the latent protective factor, paths to Items 6 and 23 were freed to load on both factors, and the correlation between factors was freely estimated. Table 1 presents CFA results. In both samples, the two factors were uncorrelated (factor $r_s = .03$). Fit indices for the two-factor CFA models were poorer compared with those of the two-factor EFA models (RMSEAs = $.055$ and $.076$; SRMRs = $.064$ and $.091$, for Samples 1 and 2, respectively), but remained in the range of adequate fit. Finally, factor loadings of all items (with the exceptions of Item 17 in both samples, and Item 23 in Sample 1 for the protective factor) remained above the $|\geq .30|$ criterion. Based on findings, these 11 acquisitive items and 10 protective items (both sets of which included

the two cross-loading items) were included in the two respective item pools for subsequent IRT analyses.

Taken together, the above findings suggest that a two-factor model of acquisitive and protective self-monitoring had a reasonably good fit to the data in both samples. Items selected for the two items pools are also highly consistent with those included in subscales developed in prior studies using factor analysis (i.e., Briggs & Cheek, 1988; Briggs et al., 1980; Snyder & Gangestad, 1986) or other quantitative methods (i.e., Gangestad & Simpson, 1993; Gangestad & Snyder, 1985). In short, the two respective item pools comprise all of the best SMS item indicators of the two self-monitoring factors.

Descriptive Statistics and Classical Psychometric Item Analyses

Table 2 presents the descriptive statistics, internal consistencies, and correlations between the SMS, SMS-R, and the respective acquisitive and protective item pools identified above. The SMS and SMS-R were strongly related to the acquisitive item pool (Sample 1: $r_{\text{SMS}} = .78$, $r_{\text{SMS-R}} = .89$; Sample 2: $r_{\text{SMS}} = .78$, $r_{\text{SMS-R}} = .90$), and moderately related to the protective pool (Sample 1: $r_{\text{SMS}} = .56$, $r_{\text{SMS-R}} = .34$; Sample 2: $r_{\text{SMS}} = .57$, $r_{\text{SMS-R}} = .37$). The two item pools, by contrast, were uncorrelated ($r_{S1} = .03$ and $r_{S2} = .02$). For the dichotomous items in Sample 1, internal consistencies were adequate for the acquisitive pool ($\alpha = .71$), but low for the protective pool ($\alpha = .63$). Similar patterns emerged for polytomous sets in Sample 2; the acquisitive coefficient was desirable ($\alpha = .80$), while the estimate for the protective set was somewhat low ($\alpha = .68$).

Parameter Estimates, Information Functions, and Standard Errors of Measurement

Having conducted classical psychometric analyses, we performed IRT analyses using the Two-parameter logistic (2PL) model and Samejima's (1969) graded response model. In both cases, item parameters are estimated using marginal maximum likelihood (Bock & Aitken, 1981). IRT models were fit to the acquisitive and protective items separately within each sample. Across samples, fit indices supported the use of our chosen IRT models. For Sample 1, χ^2 -based RMSEAs = .01 and .01 for the acquisitive and protective scales, respectively; for Sample 2, M_2 -based RMSEAs = .04 for both scales. All RMSEAs are rounded to two decimal places.⁶

Parameter estimates for all *candidate* items, which are described below, are included in the supplementary online materials; estimates for *final* items are presented in Table 3. In IRT, the probability of an examinee responding correctly to an item is modeled as a function of examinee trait standing and item characteristics. In the 2PL and graded response

model, items are characterized by *discrimination* and *difficulty* parameters. The discrimination parameter corresponds to the maximum value of the function relating examinee trait levels to item response probabilities. It is analogous to a factor loading (cf. McDonald, 1999), and determines the extent to which an item's inclusion in a test will decrease the standard error of estimated examinee trait standing (de Ayala, 2009). Concerning the discrimination parameter, most (7 out of 11) of the acquisitive candidate items showed discrimination estimates of $a \approx 1.0$, or greater, in both samples. In contrast, only 4 out of 10 protective items evidenced similar discriminatory qualities across samples. These findings indicate that items in the acquisitive candidate item pool tend to function better as single indicators of examinee trait standing (Fraley, Waller, & Brennan, 2000) than do items in the protective pool. Correspondingly, standard errors of trait estimates obtained from the acquisitive candidate item pool would be expected to be lower than those observed for the protective item pool.

In IRT, the difficulty parameter corresponds to the trait level at which the item response function obtains its steepest slope; items provide the most information regarding trait standing for examinees near this value. In our samples, item difficulties tended to range rather widely across the trait continuum. For instance, although median b values for items with discrimination parameters $\geq .85$ in Sample 1 were both near 0 ($b = -0.01$ and $b = -0.20$, for acquisitive and protective scales, respectively), item difficulties nonetheless tended to range close to 1 standard deviation above and below these values (from -0.84 to 0.75 , and from -0.81 to 1.43 , for the respective scales). Likewise, items in Sample 2 tended to possess difficulty parameters spanning a wide trait range. Nonetheless, item difficulties in each sample tended to cluster near the mid-to-upper regions of the trait scale (i.e., near or above 0). What this means is that, although the range of individual item difficulties was somewhat wide in our samples, when considered as a whole, both sets of items provided the most information at the mid-to-upper regions of the measured traits. Taken together, these findings mean that item properties provide more information for accurate assessment near scale midpoints (i.e., from -1 SD to $+1$ SD) or higher, in the case of the polytomous response format, than they do at lower ends of the traits (see Figure 1). To aid in the selection of final items, we used item information functions (IIF) as visual aids; IIFs are also included in the supplementary materials.

Final Item Selection

Acquisitive Self-Monitoring Scale. Across samples, Items 8, 12, 18, and 20 were characterized by relatively high discrimination ($\text{Mean } a_{S1} = 1.41$, $\text{Mean } a_{S2} = 1.67$) and information. In contrast, Item 21 was excluded due to low discrimination ($as = 0.72$ and 0.76) and little information

Table 2. Scale Descriptive Statistics, Part–Whole Correlations, and Relations to Demographic Correlates.

Measure	No. of items							Demographic correlates		
		1	2	3	4	5	6	Gender <i>d</i>	Age group <i>d</i>	Age <i>r</i>
1. Self-Monitoring Scale	25	<i>.68/.74</i>	.93	.78	.56	.65	.58	.20/.28	.38/.12	-.15/-.10
2. Self-Monitoring Scale–Revised	18	.94	<i>.69/.77</i>	.89	.34	.78	.35	.20/.29	.25/.20	-.10/-.14
3. Acquisitive candidate items	11	.78	.90	<i>.71/.80</i>	.03	.89	.04	.13/.23	.11/.28	-.04/-.15
4. Protective candidate items	10	.57	.37	.02	<i>.63/.68</i>	-.04	.94	.10/.15	.37/-.19	-.16/.03
5. Acquisitive final items	6	.69	.82	.93	-.01	<i>.65/.77</i>	-.03	.13/.32	.08/.20	-.03/-.10
6. Protective final items	7	.57	.36	.02	.95	-.01	<i>.61/.69</i>	.07/.11	.35/-.20	-.16/.03
Sample 1	<i>M</i>	12.13	9.38	5.69	3.54	2.84	3.38	.40	.08	25.63
	<i>SD</i>	4.08	3.50	2.72	2.28	1.77	1.80	.49	.27	8.36
Sample 2	<i>M</i>	13.16	9.64	5.97	3.80	2.98	3.54	.42	.83	51.41
	<i>SD</i>	2.31	2.05	1.60	1.20	1.07	0.96	.50	.37	12.61

Note. *N*s for Samples 1 and 2 = 13,563 and 709, respectively. For Sample 1, a dichotomous (true–false) response format was used; for Sample 2, a polytomous (5-point Likert-type) response format was used. Sample 1 correlations are reported above, and Sample 2 correlations are reported below, the diagonal. Internal consistency estimates (i.e., Cronbach's α) reported along the diagonal in italics (i.e., Sample 1/Sample 2).

Among demographic correlates, positive *d* values for gender correspond to higher scores for males; positive *d* values for age group correspond to higher scores for younger individuals (i.e., younger than 40 years vs. older than 40 years). Age *r* values correspond to zero-order correlations with age. Based on evidence of factor cross-loadings, Items 6 and 23 were included in both acquisitive and protective candidate item pools.

Acquisitive candidate item pool = 1, 5, 6, 8, 12, 14, 18, 20, 21, 22, and 23; final items = 5, 8, 12, 18, 20, and 22.

Protective candidate item pool = 2, 6, 7, 13, 15, 16, 17, 19, 23, and 25; final items = 2, 7, 13, 15, 16, 19, and 25.

across samples. Having selected the strongest and excluded the weakest items, we turned our attention to the remaining candidate items. Results indicated that Items 5 and 22 showed relatively high discrimination in the polytomously scored Sample 2 ($as = 1.25$ and 1.61 , respectively), but somewhat lower estimates using the dichotomous format ($as = 0.96$ and 0.98 , respectively). On the other hand, Items 1 and 14 evidenced adequate discrimination in Sample 1 ($as = 0.98$ and 0.95), but not in Sample 2 ($as = 0.62$ and 0.82). Inspecting IIFs, visual evidence showed that all four items had virtually identical IIFs in Sample 1; by contrast, in Sample 2, Items 5 and 22 evidenced moderate information, but Items 1 and 14 showed very little information. In the end, we selected six items for inclusion in the final acquisitive scale: 5, 8, 12, 18, 20, and 22 (see Table 3).

Protective Self-Monitoring Scale. Items 13, 16, and 19 showed high discrimination ($Mean a_{s1} = 1.56$, $Mean a_{s2} = 1.57$) and high information across samples. Item 25 also had relatively high discrimination ($a = 1.19$) and information in the polytomously scored Sample 2, but somewhat lower discrimination using the dichotomous format in Sample 1 ($a = 0.91$). By comparison, Item 17 showed the lowest information and poorest discriminations across samples and was therefore excluded. This left us with three remaining items, 2, 7, and 15, of which Item 2 showed the highest discrimination ($as = 1.03$ and 0.85 , respectively) and most information. Nevertheless, in view of the comparatively weaker properties of the protective candidate item set vis-à-vis the acquisitive pool, we made the decision to include both 7 and 15 to enhance the reliability of the new scale. Thus, seven

items were selected for the final protective scale: 2, 7, 13, 15, 16, 19, and 25 (see Table 3). The cross-loading items, 6 and 23, were ultimately excluded for non-IRT reasons.⁷

Following item selection, we reran IRT analyses using only selected items to compute the final discrimination and difficulty parameter estimates reported in Table 3. In addition, we computed test information functions for candidate item pools and final scales alike, which are depicted in adjacent panes in Figure 1. Impressively, test information functions for items selected using IRT achieved comparative shapes and information using only 65% of the items. Both scales showed strong part–whole correlations with their respective candidate pools (Sample 1: $r_{ACQ} = .89$, $r_{PRO} = .94$; Sample 2: $r_{ACQ} = .93$, $r_{PRO} = .95$); internal consistencies were acceptable, if low, for the dichotomous items (Sample 1: $\alpha_{ACQ} = .65$, $\alpha_{PRO} = .61$; Sample 2: $\alpha_{ACQ} = .77$, $\alpha_{PRO} = .69$). Scale correlations ($r_{s1} = -.03$, $r_{s2} = -.01$) reflected the independence of their respective factors.

Having selected items to compose the acquisitive and protective SMSs, next, we turned our attention toward examining questions of measurement bias, which might be associated with items in the two new scales.

Differential Item Functioning

Meta-analytic evidence indicates that SMS scores differ across males and females and are negatively correlated with age (Day et al., 2002). To determine if these findings reflect biased measurement, we conducted tests to detect the presence of DIF.⁸ DIF occurs when the relation between a latent trait and item responses differs across subgroups. Although

Table 3. IRT-Based Parameter Estimates of Items Included in the Final Acquisitive and Protective Self-Monitoring Scales.

SMS item	Text	Sample 1		Sample 2				
		<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i> ₁	<i>b</i> ₂	<i>b</i> ₃	<i>b</i> ₄
<i>Acquisitive self-monitoring</i>								
5	I can make impromptu speeches even on topics about which I have almost no information.	0.99 (0.03)	0.21 (0.02)	1.12 (0.11)	-0.21 (0.08)	0.79 (0.10)	1.23 (0.12)	2.80 (0.25)
8	I would probably make a good actor.	1.94 (0.06)	-0.14 (0.01)	2.66 (0.26)	-0.44 (0.06)	0.20 (0.05)	0.63 (0.06)	1.55 (0.10)
12	In a group of people, I am rarely the center of attention. ^a	1.13 (0.04)	0.37 (0.02)	1.20 (0.11)	-1.41 (0.13)	0.20 (0.08)	0.99 (0.10)	2.78 (0.23)
18	I have considered being an entertainer.	1.49 (0.04)	0.14 (0.02)	2.29 (0.21)	0.24 (0.06)	0.76 (0.07)	1.00 (0.08)	1.85 (0.13)
20	I have never been good at games like charades or improvisational acting. ^a	1.52 (0.04)	-0.44 (0.02)	1.59 (0.13)	-1.39 (0.11)	-0.26 (0.07)	0.25 (0.07)	1.44 (0.11)
22	At a party, I let others keep the jokes and stories going. ^a	0.85 (0.05)	0.84 (0.03)	1.22 (0.11)	-1.56 (0.14)	0.25 (0.08)	0.88 (0.10)	2.70 (0.23)
<i>Protective self-monitoring</i>								
2	My behavior is usually an expression of my true inner feelings, attitudes, and beliefs. ^a	1.08 (0.04)	1.38 (0.04)	0.83 (0.10)	-0.73 (0.13)	2.40 (0.28)	3.01 (0.35)	5.81 (0.77)
7	When I am uncertain how to act in a social situation, I look to the behavior of others for cues.	0.84 (0.03)	-0.83 (0.03)	0.65 (0.09)	-3.32 (0.46)	-1.69 (0.25)	-0.42 (0.13)	3.44 (0.47)
13	In different situations and with different people, I often act like very different persons.	1.56 (0.05)	-0.05 (0.02)	1.40 (0.13)	-0.75 (0.09)	0.38 (0.07)	0.89 (0.09)	2.45 (0.20)
15	Even if I am not enjoying myself, I often pretend to be having a good time.	0.81 (0.03)	-0.03 (0.02)	0.96 (0.11)	-2.54 (0.26)	-0.47 (0.10)	0.40 (0.09)	3.67 (0.38)
16	I'm not always the person I appear to be.	1.54 (0.05)	-0.70 (0.02)	1.70 (0.16)	-0.78 (0.08)	0.28 (0.06)	0.75 (0.07)	2.28 (0.17)
19	In order to get along and be liked, I tend to be what people expect me to be rather than anything else.	1.42 (0.05)	1.10 (0.03)	1.48 (0.14)	-0.46 (0.07)	0.99 (0.09)	1.66 (0.13)	3.33 (0.29)
25	I may deceive people by being friendly when I really dislike them.	0.93 (0.03)	-0.33 (0.02)	1.23 (0.12)	-1.16 (0.12)	0.18 (0.07)	0.72 (0.09)	3.29 (0.30)

Note. IRT = item response theory; SMS = Self-Monitoring Scale. *N*s for Samples 1 and 2 = 13,563 and 709, respectively. *a* = item discrimination parameter; *b* = item difficulty parameter; standard errors of parameter estimates are in parentheses. For more detailed explanation of the interpretation of IRT item parameters, see notes to Table S6 in the online supplement. Estimation of all parameters was conducted using IRTPRO (Cai, Thissen, & du Toit, 2011). Fit indices for final acquisitive (RMSEA = .04 and .04) and protective scales (RMSEA = .02 and .04). All item numbers refer to the order of items as presented in the original 25-item SMS (Snyder, 1974).

^aItem reverse-scored.

age is a continuous variable, we chose to artificially dichotomize it (i.e., <40 and ≥40 years), primarily to comport with U.S. federal age discrimination legislation, which is relevant for personnel selection in applied settings.⁹

We adopted a multistage approach to test for DIF. First, we conducted statistical significance tests against the null hypothesis of no DIF. Because different DIF tests need not lead to the same conclusions, we followed Hambleton's (2006) recommendation of performing multiple tests, using significant results to serve as flags for further review, and looking for areas of convergence. To this end, we utilized three statistical tests: (a) IRT-based likelihood ratio (IRT-LR) tests (Reise, Widaman, & Pugh, 1993), (b) ordinal logistic regression (ORL)-based approaches, and (c) the Mantel-Haenszel test (Penfield & Camilli, 2006).

In IRT-LR, a series of nested models are tested against one another. A baseline model is first created wherein item parameters, with the exception of one linking indicator item, are free to vary across groups. Next, additional models are

created by constraining parameters for individual items to be equal across groups. Chi-square likelihood ratio tests are then used to detect decreases in model fit and DIF.

The logic underlying the OLR approach is similar to that involved in IRT-LR testing. The key difference lies in OLR's use of observed test scores, which frees the test from parametric constraints. First, a dichotomous logistic model is fitted regressing the probability of answering an item in the keyed direction on total scale scores. Additional models include *group*, and *group* × *scale score* variables. Likelihood ratio tests are used to determine whether these additional variables result in improved model fit (i.e., ΔNagelkerke *R*²). Generalizations to the polytomous case follow the same logic and are described by Penfield and Camilli (2006). Finally, the Mantel-Haenszel odds ratio-based test was applied to dichotomously scored items (see Teresi & Fleishman, 2007).

In addition to tests of statistical significance, effect size indices were computed to determine the magnitude of DIF

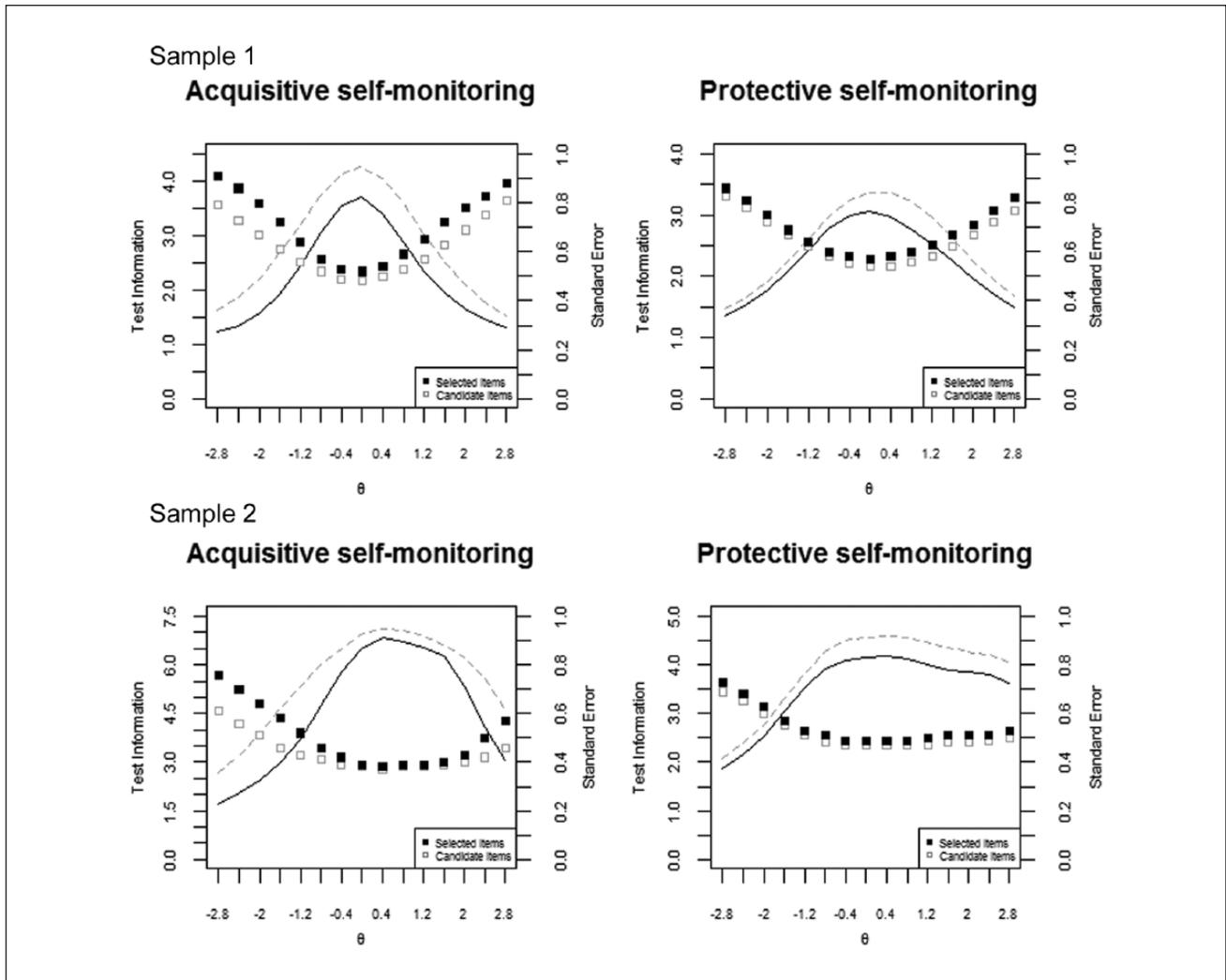


Figure 1. Test information functions and standard error curves for candidate items (dashed lines) and final items (solid lines) for acquisitive and protective self-monitoring scales.

Note. Sample 1 items scored using a dichotomous (true–false) response format (top panes); Sample 2 items scored using polytomous (5-point Likert-type) response format (bottom panes).

(e.g., Δ Nagelkerke R^2 for the OLR approach; cf. Penfield & Camilli, 2006). Effect sizes were classified as “negligible,” “moderate,” or “large,” using recommendations by prior authors (i.e., Dorans & Holland, 1993; Jodoin & Gierl, 2001). Sensitivity analyses were also conducted to determine the consequences of ignoring DIF. Specifically, item parameter estimates and person scores were first calculated using models wherein DIF-flagged item parameters were estimated freely within each subgroup. Estimates were then compared with those obtained when all item parameters were constrained to be equal across groups. What these results show are the practical consequences of ignoring DIF during item estimation and person scoring. All analyses were performed using the DIF options in IRTPRO, as well as functions provided in difR (Magis, Beland, & Raiche, 2013) and lordif (Choi, Gibbons, & Crane, 2011). As a final

check, we computed standardized between-group difference scores (i.e., Cohen’s d) across acquisitive and protective scale scores for males and females, and for both age groups.

DIF Analyses Across Gender and Age Groups. Results of DIF analyses are presented in Tables 4 and 5. A number of Sample 1 items were flagged for DIF in two or more tests. However, examining effect sizes revealed that most effects were negligible. Exceptions were Item 5, for gender, and Item 22, for age group, both of which showed moderate effects. Similar results were obtained in Sample 2, with universally negligible effect sizes across DIF-flagged items.

Sensitivity Tests. The evidence described above indicates that DIF does not appear to present a major hurdle to use of the

Table 4. DIF Analyses Across Males and Females.

SMS item	Sample 1							Sample 2			
	OLR-LR χ^2	IRT-LR χ^2	M-H χ^2	ΔR^2	λ_{MH}	JG	ETS	OLR-LR χ^2	IRT-LR χ^2	ΔR^2	JG
<i>Acquisitive self-monitoring</i>											
5	184.65 (2)*	111.29 (2)*	187.61 (1)*	<.01	1.34	A	B	19.80 (2)*	30.51 (5)*	.03	A
8	1.97 (2)	Referent	2.04 (1)	<.01	0.16	A	A	0.68 (2)	Referent	<.01	A
12	69.82 (2)*	40.79 (2)*	67.18 (1)*	<.01	-0.86	A	A	2.38 (2)	6.79 (5)	<.01	A
18	28.91 (2)*	10.95 (2)*	29.53 (1)*	<.01	0.56	A	A	2.64 (2)	7.18 (5)	<.01	A
20	0.44 (2)	0.35 (2)	0.24 (1)	<.01	0.06	A	A	7.22 (2)*	4.71 (5)	<.01	A
22	58.35 (2)*	36.25 (2)*	51.85 (1)*	<.01	-0.74	A	A	4.56 (2)	4.31 (5)	<.01	A
<i>Protective self-monitoring</i>											
2	19.90 (2)*	7.66 (2)*	18.49 (1)*	<.01	0.49	A	A	8.95 (2)*	8.62 (5)	.01	A
7	11.40 (2)*	9.05 (2)*	11.08 (1)*	<.01	-0.34	A	A	2.19 (2)	5.27 (5)	<.01	A
13	33.61 (2)*	6.97 (2)*	33.07 (1)*	<.01	0.60	A	A	4.27 (2)	8.95 (5)	<.01	A
15	39.49 (2)*	23.27 (2)*	38.07 (1)*	<.01	-0.60	A	A	15.20 (2)*	14.28 (5)*	.02	A
16	11.57 (2)*	0.72 (2)	9.43 (1)*	<.01	0.34	A	A	0.11 (2)	10.19 (5)	<.01	A
19	4.26 (2)	Referent	0.36 (1)	<.01	0.08	A	A	3.77 (2)	Referent	<.01	A
25	29.80 (2)*	18.01 (2)*	27.06 (1)*	<.01	-0.51	A	A	7.32 (2)*	3.62 (5)	<.01	A

Note. DIF = differential item functioning; IRT = item response theory. For Sample 1, male $N = 5,365$, female $N = 8,198$; for Sample 2, male $N = 303$, female $N = 406$. Parameter estimates followed by degrees of freedom in parentheses; for all significance tests, $\alpha = .05$. OLR-LR χ^2 = ordinal logistic regression likelihood ratio chi-square; IRT-LR χ^2 = IRT-based likelihood ratio chi-square, using the Stark, Chernyshenko, and Drasgow (2006) method; M-H χ^2 = Mantel-Haenszel chi-square; ΔR^2 = change in Nagelkerke R^2 above baseline IRT-LR model with inclusion of group and group \times sum-score interaction terms; λ_{MH} = common log-odds ratio effect size; JG = Jodoin and Gierl (2001) effect size scale for OLR-LR ΔR^2 ("A" = negligible; "B" = moderate; "C" = large); ETS = Educational Testing Service (ETS) effect size scale for λ_{MH} ("A" = negligible; "B" = moderate; "C" = large).

*Statistically significant.

Table 5. DIF Analyses Across Age Groups (Younger Than 40 Years and 40 Years and Older).

SMS item	Sample 1							Sample 2			
	OLR-LR χ^2	IRT-LR χ^2	M-H χ^2	ΔR^2	λ_{MH}	JG	ETS	OLR-LR χ^2	IRT-LR χ^2	ΔR^2	JG
<i>Acquisitive self-monitoring</i>											
5	7.29 (2)*	2.31 (2)	5.49 (1)*	<.01	0.41	A	A	4.16 (2)	3.69 (5)	<.01	A
8	3.24 (2)	2.44 (2)	1.35 (1)	<.01	0.23	A	A	3.18 (2)	4.26 (5)	<.01	A
12	28.71 (2)*	17.59 (2)*	25.93 (1)*	<.01	-0.93	A	A	1.99 (2)	4.06 (5)	<.01	A
18	23.53 (2)*	11.11 (2)*	19.70 (1)*	<.01	0.86	A	A	0.66 (2)	2.09 (5)	<.01	A
20	0.47 (2)	Referent	0.40 (1)	<.01	0.13	A	A	0.24 (2)	Referent	<.01	A
22	44.10 (2)*	26.30 (2)*	44.41 (1)*	<.01	-1.19	A	B	0.79 (2)	10.25 (5)	<.01	A
<i>Protective self-monitoring</i>											
2	4.01 (2)	1.72 (2)	2.14 (1)	<.01	0.34	A	A	1.93 (2)	4.40 (5)	<.01	A
7	0.84 (2)	1.17 (2)	0.33 (1)	<.01	-0.11	A	A	0.32 (2)	5.47 (5)	<.01	A
13	0.27 (2)	Referent	0.15 (1)	<.01	0.08	A	A	0.02 (2)	Referent	<.01	A
15	4.11 (2)	0.24 (2)	3.84 (1)*	<.01	-0.35	A	A	5.34 (2)	4.62 (5)	<.01	A
16	18.74 (2)*	5.10 (2)	17.45 (1)*	<.01	0.79	A	A	1.63 (2)	9.08 (5)	<.01	A
19	4.00 (2)	0.96 (2)	2.99 (1)	<.01	-0.42	A	A	3.34 (2)	4.34 (4)	<.01	A
25	0.24 (2)	1.22 (2)	0.09 (1)	<.01	0.06	A	A	5.28 (2)	6.53 (5)	<.01	A

Note. SMS = Self-Monitoring Scale; DIF = differential item functioning. For Sample 1, N_s for younger than 40 years and older than 40 years = 12,456 and 1,107, respectively. For Sample 2, N_s for younger than 40 years and older than 40 years = 119 and 590, respectively. Parameter estimates are followed by degrees of freedom in parentheses for all significance tests; $\alpha = .05$. OLR-LR χ^2 = ordinal logistic regression likelihood ratio chi-square; IRT-LR χ^2 = IRT likelihood ratio chi-square, using the Stark et al. (2006) method; M-H χ^2 = Mantel-Haenszel chi-square; ΔR^2 = change in Nagelkerke R^2 above baseline IRT-LR model with inclusion of group and group \times sum-score interaction terms; λ_{MH} = common log-odds ratio effect size; JG = Jodoin and Gierl (2001) effect size scale for OLR-LR ΔR^2 ("A" = negligible; "B" = moderate; "C" = large); ETS = ETS effect size scale for λ_{MH} ("A" = negligible; "B" = moderate; "C" = large).

*Statistically significant.

two new scales. Nevertheless, to determine the consequences of ignoring DIF, we estimated parameters for DIF-flagged items that were free to vary within each group. In contrast, non-DIF items were constrained to be equal across subgroups and fixed at the values obtained during prior item-fitting runs. Results produced a consistent pattern of small, between-group differences in item discrimination parameters (Sample 1: $M = 0.12$, range = 0.01-0.32; Sample 2: $M = 0.32$, range = 0.01-0.93) across samples and groups, and moderate differences in difficulty parameters (Sample 1: $M = 0.22$, range = 0.00-0.46; Sample 2: $M = 0.40$, range = 0.01-3.26). These findings mean that item slopes were relatively invariant across groups, but item intercepts differed slightly. Next, person parameter estimates obtained when constraining DIF-flagged item parameters to be equal were compared with those obtained by estimating DIF-flagged item parameters freely within subgroups. In all cases, results were indicative of high convergence in person parameter estimates across constrained and free models (mean $r = 1.0$ after rounding in each test and sample; mean absolute difference $\approx .02$ in Sample 1 across tests and grouping variables; mean absolute difference $\approx .04$ in Sample 2), indicating that in terms of estimating person scores, the consequences of ignoring DIF are trivial.

Finally, we computed standardized between-group differences in acquisitive and protective SMS scores for both males and females, and across age groups (see Table 2). In general, results mirrored meta-analytic findings showing that males and younger individuals tend to score higher than females and older individuals (Day et al., 2002). Gender differences were small for acquisitive and protective scale scores ($d_s \approx .20$ and $.10$, respectively) but variable across age groups. Variability may be partially attributable to sampling error; in Sample 2, the standard error for $d_{\text{age}} = .10$. Last, in examining age as a continuous variable, zero-order correlations of acquisitive and protective self-monitoring indicated slight, often negative relations for both scales (see Table 2). Results were similar when analyses were computed using IRT-based person scores obtained from models with DIF-flagged item parameters estimated freely across subgroups.

Taken together, results of DIF analyses and sensitivity tests showed little evidence of bias across gender and age groups for the two new scales. When DIF was observed, associated effect sizes were in the negligible range and did not distort the estimation of person scores. There was some evidence of small, between-group differences in intercepts of item difficulty parameters, and in full-scale acquisitive and protective self-monitoring scores. What these results mean is that there appear to be real, albeit small, differences in trait scores across gender and age groups, which are not an artifact of biased measurement.

Having determined that the new acquisitive and protective SMSs were not significantly biased in terms of gender

or age groups, finally, we turned our attention toward testing the construct validity of the new measures.

Convergent and Discriminant Validity

In a meta-analysis reported by Barrick et al. (2005), self-monitoring was found to be poorly represented by the Five-Factor Model (FFM) of personality traits; although it showed a moderate relation to Extraversion, self-monitoring had negligible relations to the other FFM dimensions. However, when separated into acquisitive and protective self-monitoring, Wilmot et al. (2015) found that both dimensions were, in fact, well represented in the FFM, but at the metatrait level (DeYoung, 2006; Digman, 1997). The first metatrait, labeled Alpha or Stability, is the higher-order personality trait composed of the shared variance of the Emotional Stability (Neuroticism reversed), Agreeableness, and Conscientiousness factors; the second metatrait, named Beta or Plasticity, is composed of the covariance of Extraversion and Openness/Intellect. Authors found that acquisitive self-monitoring and Plasticity are *equivalent* constructs, whereas protective self-monitoring is moderately negatively related to Stability (Wilmot et al., 2015).

Metatrait findings provide further evidence that hierarchy is an intrinsic and pervasive feature of personality trait structure (Markon, Krueger, & Watson, 2005). With hierarchy in mind, in validating our new measures, we sought to examine convergent and discriminant relations across two levels of personality traits. To do so, we correlated them with the FFM factor domains, and their lower level facets, as measured by the NEO inventory (Costa & McCrae, 1992). As a measure of Plasticity, we hypothesized that acquisitive self-monitoring would have moderate-to-strong relations to the Extraversion and Openness/Intellect factors and demonstrate relatively uniform relations across their constituent facets. We hypothesized that protective self-monitoring would have a moderate, positive relation to Neuroticism, and small-to-moderate negative relations to Agreeableness and Conscientiousness, but differential relations would be observed across their respective facets (cf. Wolf, Spinath, Riemann, & Angleitner, 2009). In contrast, extant evidence indicates nil/negligible relations between the SMS and cognitive ability measures (e.g., Côté, Lopes, Salovey, & Miners, 2010). To our knowledge, no research has examined relations between acquisitive and protective self-monitoring and cognitive ability, and we have no theoretical reason to believe these constructs should be related. Thus, we hypothesized that both scales would demonstrate discriminant validity to cognitive ability. In sum, personality trait and cognitive ability relations provide a test of construct validity (Cronbach & Meehl, 1955) while also expanding knowledge of the two self-monitoring factors.

Relations to FFM Factors, Facets, and Cognitive Ability. Correlations between acquisitive and protective self-monitoring and personality trait factors and facets are presented in Table 5. For purpose of comparison, relations to the original 25-item SMS are also reported. Across samples, acquisitive self-monitoring was strongly related to Extraversion ($r_{S1} = .51, r_{S2} = .47$) and moderately related to Openness/Intellect ($r_{S1} = .25, r_{S2} = .38$); facet correlations were relatively uniform across factors, with Assertiveness ($r_{S1} = .53, r_{S2} = .49$) and Ideas ($r_{S1} = .27, r_{S2} = .33$) having the strongest relations within their respective domains. By comparison, protective self-monitoring had moderate relations to Neuroticism ($r_{S1} = .24, r_{S2} = .34$) and Conscientiousness ($r_{S1} = -.19, r_{S2} = -.25$) but weaker relations to Agreeableness ($r = -.12$). Facet relations were differentiated within domains. Among Neuroticism facets, protective self-monitoring had moderate relations ($rs \approx .30$) to Self-Consciousness, Anxiety, Depression, and Vulnerability. Relations to Conscientiousness facets were modest ($rs \approx -.15$ to $-.20$), with Competence, Self-Discipline, and Dutifulness having the strongest relations. Similar to factor-level relations, relations to Agreeableness facets were generally weak ($rs \approx .10$). Taken together, patterns of relations support hypotheses, indicating that the scales measure their intended constructs and evidence divergent relations. However, there was also evidence of noteworthy facet-level relations across self-monitoring dimensions. Despite their orthogonality, both constructs had moderate negative relations in the same direction to Straightforwardness (acquisitive: $r_{S1} = -.30, r_{S2} = -.29$; protective: $r_{S1} = -.18, r_{S2} = -.20$); this nonredundant variance was reflected in stronger overall relations for the 25-item SMS ($r_{S1} = -.33, r_{S2} = -.37$). By comparison, acquisitive and protective constructs had relations of similar magnitudes, but in opposite directions, to Self-Consciousness (acquisitive: $r_{S1} = -.45, r_{S2} = -.29$; protective: $r_{S1} = .30, r_{S2} = .38$), which were largely obscured in SMS correlations ($r_{S1} = -.21, r_{S2} = .00$). Importantly, these results would not be detectable by examining factor-level relations alone and speak to the importance of validating new measures across multiple levels of the personality trait hierarchy.

Finally, concerning cognitive ability, despite evidence of a weak acquisitive relation in Sample 2 ($r = .16$), the general pattern of results support our hypothesis that both self-monitoring factors should demonstrate discriminant validity from measures of cognitive ability (Table 6).

Discussion

The purpose of this study was to use IRT to develop measures of acquisitive and protective self-monitoring using items present in the original 25-item SMS (Snyder, 1974). We used two large, nonstudent samples ($Ns = 13,563$ and 709) varying in methods of data collection (online vs.

paper-and-pencil) and response formats to the SMS (dichotomous vs. polytomous). First, we used EFA and CFA to select two pools of candidate items for the acquisitive and protective self-monitoring factors, which were then subjected to independent IRT analyses. Inspection of item parameter estimates and IIFs provided evidence that six acquisitive and seven protective items were characterized by item properties that were desirable for inclusion in final scales. Next, we further subjected items constituting the final scales to multiple DIF and sensitivity tests. Across samples and measures, results indicated that items displayed little evidence of bias across gender and age. Finally, we tested the convergent and discriminant validity of the scales using two different hierarchical measures of personality traits (IPIP-NEO vs. NEO-PI-R) and two cognitive ability tests (myIQ vs. 16PF Factor B). Results provided evidence that acquisitive self-monitoring has strong positive relations to the Extraversion and Openness/Intellect factors, and their facets, whereas protective self-monitoring has small-to-moderate negative relations to the Emotional Stability, Agreeableness, and Conscientiousness factors but more differentiated facet-level relations. Neither scale was meaningfully related to cognitive ability.

The conventional univariate model dominates self-monitoring research and the interpretation of its empirical findings (Day et al., 2002; Gangestad & Snyder, 2000); as a result, the call for bidimensional assessment (John et al., 1996) has been largely ignored in most self-monitoring scholarship. Nevertheless, the explosion of research interest in self-monitoring has resulted in an increased demand for a scale of more economical length. Although various authors have proposed abbreviations to the SMS (e.g., Allen, Weeks, & Moffitt, 2005; Chapman, Uggerslev, & Webster, 2003; Dunn, Biesanz, Human, & Finn, 2007; Glomb & Liao, 2003; Mullen & Noe, 1999; Scott, Barnes, & Wagner, 2012), all of these attempts have been based on the conventional univariate model, and all suffer, to varying degrees, from idiosyncratic item selection. By contrast, the two scales developed here use IRT to operationalize the empirically promising, alternative bivariate model of self-monitoring. Our scales are 65% shorter than their respective item pools, offering brevity while maintaining measurement precision.

Although the new scales have a number of advantages, they do have some limitations. For the acquisitive and protective dimensions alike, the limited size of the original item pool (i.e., the original SMS) prevented the development of measures that covered both the full range of the latent traits *and* did so with high precision. For both scales, item discrimination parameters mostly capture the moderate trait ranges with good accuracy (i.e., from $-1 SD$ to $1 SD$). Furthermore, despite limited accuracy at the low ends of the traits, assessment at the high ends is considerably more precise when using the polytomous (vs. the

Table 6. Correlations to NEO Factors and Facets, and Cognitive Ability.

Measure	Internal consistency		Original SMS		Acquisitive SMS		Protective SMS	
	S1	S2	S1	S2	S1	S2	S1	S2
Neuroticism	.87	.85	-.02	.11	-.12	-.15	.34	.24
Anxiety	.87	.83	-.09	.10	-.09	-.22	.26	.24
Hostility	.92	.80	-.02	.10	.04	-.06	.15	.15
Depression	.91	.85	-.05	.08	-.12	-.20	.29	.29
Self-consciousness	.85	.74	-.21	.00	-.29	-.45	.38	.30
Impulsiveness	.78	.72	.11	.17	.07	.03	.19	.18
Vulnerability	.88	.79	-.10	.06	-.17	-.22	.29	.23
Agreeableness	.81	.75	-.08	-.25	-.24	-.07	-.12	-.12
Trust	.90	.84	.01	-.01	.04	.01	-.12	-.13
Straightforwardness	.79	.74	-.33	-.37	-.29	-.30	-.20	-.18
Altruism	.85	.72	.07	-.09	-.03	.09	-.17	-.10
Compliance	.71	.73	-.13	-.15	-.18	-.21	.00	-.03
Modesty	.79	.75	-.30	-.31	-.37	-.44	-.04	.11
Tender-mindedness	.81	.61	-.07	-.04	-.09	-.03	.03	-.07
Conscientiousness	.83	.83	-.08	-.15	.01	.03	-.25	-.19
Competence	.84	.70	.08	-.03	.12	.23	-.23	-.24
Order	.84	.74	-.09	-.15	-.08	-.09	-.14	-.07
Dutifulness	.77	.67	-.23	-.20	-.08	-.20	-.18	-.15
Achievement-striving	.84	.67	.04	.00	.16	.12	-.17	-.12
Self-discipline	.89	.80	-.03	-.11	.04	.06	-.25	-.18
Deliberation	.83	.70	-.21	-.20	-.11	-.23	-.13	-.06
Extraversion	.83	.74	.33	.36	.47	.51	-.15	-.19
Warmth	.89	.80	.29	.18	.26	.37	-.17	-.18
Gregariousness	.89	.80	.31	.29	.30	.39	-.07	-.12
Assertiveness	.86	.80	.30	.28	.49	.53	-.18	-.20
Activity	.73	.72	.10	.18	.26	.22	-.11	-.14
Excitement-seeking	.84	.64	.32	.28	.26	.41	.06	-.03
Positive emotions	.86	.81	.25	.19	.29	.33	-.14	-.09
Openness/Intellect	.70	.77	.13	.23	.38	.25	-.06	-.05
Fantasy	.84	.82	.15	.28	.33	.20	.07	.11
Aesthetics	.82	.84	.11	.12	.23	.23	-.07	-.05
Feelings	.77	.75	.04	.13	.28	.09	-.12	.01
Actions	.83	.64	.16	.09	.24	.35	-.13	-.15
Ideas	.83	.82	.14	.18	.33	.27	-.05	-.07
Values	.80	.78	.02	.17	.13	.10	.03	-.01
Cognitive ability	—	.73	.04	.14	.16	-.02	.05	.04

Note. NEO = Neuroticism–Extraversion–Openness; SMS = Self-Monitoring Scale; S = sample. For personality trait variables, *N*s for Samples 1 and 2 = 1,197 and 644, respectively; for cognitive ability, *N*s = 1,651 and 594, respectively. For all *N*s, values >.09 in absolute value have 95% confidence intervals excluding 0. Bolded values indicate hypothesized relations; italicized values are internal consistency estimates.

dichotomous) response format. A second limitation is the relatively low internal consistency of the new protective scale. Although the seven-item scale is approximately as reliable as the 10 items comprising its candidate item pool, its associated coefficients fall short of the conventional baseline of $\alpha > .70$. To address these shortcomings, future researchers may consider including relevant items from other published measures of self-monitoring (i.e., Lennox & Wolfe, 1984) or by writing new items that better tap the lower ends of the two dimensions. Adding new items and/or replacing less informative ones (e.g., 7 and 15) would also

help boost the reliability of the protective SMS. Nevertheless, we believe the limitations are far outweighed by the considerable advantages associated with reanalyzing archival SMS data.

Recommendations for Data Analysis of Archival SMS Responses

Although the new acquisitive and protective SMSs hold promise for future research, they are particularly valuable for reanalyzing previously collected responses. To aid in

reanalysis efforts, we urge scholars to attend to three potential issues that may have been involved in the original data collection process. First, in some cases, archival data may constitute interitem correlation matrices but lack raw data. Although our IRT parameter estimates would be of little use in such a situation, linear composites of acquisitive and protective self-monitoring can be formed from correlations using formulae reported in the literature (see Ghiselli, Campbell, & Zedeck, 1981). A second issue concerns the matter of range enhancement. Due to conventional univariate model assumptions, early self-monitoring research was not infrequently characterized by the use of sampling methods that involved selecting individuals with high and low SMS scores but discarding moderate scorers. For data sets with evidence of this type of range enhancement, reanalysis efforts would probably best be bypassed. At most, results from such data sets should be interpreted with caution. A final issue concerns data that used the 18-item SMS-R instead of the 25-item SMS. Happily, all six items selected for the acquisitive scale are present in the SMS-R. In contrast, only three of the seven items in the protective scale (i.e., Items 13, 16, and 25) remain in the SMS-R. Researchers might consider using these three protective items, in addition to Item 17, which, although too weak to be selected for inclusion in our final protective scale, was nevertheless retained in the SMS-R. Although the reduced quality of this four-item scale is certainly *not* recommended for future research, it would provide a viable, if less reliable, means for reanalyzing SMS-R data for protective self-monitoring effects.

Concluding Thoughts

“Jangle fallacies” (i.e., calling the same construct by two different names) are a perennial problem in psychological research. To avoid yet another jangle fallacy, we believe that it is worth restating that acquisitive self-monitoring and the higher-order personality trait, composed of the shared variance of the Extraversion and Openness/Intellect factors, which has alternatively been labeled Beta (Digman, 1997) or Plasticity (DeYoung, 2006), are the *same construct*. As such, our six-item acquisitive self-monitoring scale represents a direct measure of metatrait Plasticity. Furthermore, in view of the comparatively stronger correlation of the acquisitive factor to both the SMS and the SMS-R (see Table 2), we are persuaded that the majority of past self-monitoring findings and theoretical advances (for a review, see Fuglestad & Snyder, 2010) are attributable to the acquisitive self-monitoring/Plasticity dimension. Nevertheless, due to the predominant use of full-scale SMS scores in self-monitoring research, prior findings remain contaminated, to an unknown degree, by protective self-monitoring. Thankfully, effects of both self-monitoring factors can be distinguished using the two new scales. Without the attenuating influence of the protective

dimension, prior acquisitive effects may turn out to be stronger (e.g., leadership emergence; cf. Ellis, 1988). In contrast, other findings will be found to be mostly attributable to the protective factor (e.g., attention and responsiveness to the behaviors of others; cf. Miell & le Voi, 1985). Still in other cases, both self-monitoring factors, despite being uncorrelated, may nevertheless contribute unique variance in the same direction to predict a common criterion (e.g., unrestricted orientation to sexuality; cf. Snyder, Simpson, & Gangestad, 1986). Regardless, in every case, abandoning the conventional univariate model and its associated full-scale scoring approach in favor of the new bivariate model measures will enable a clearer and more thorough explication of both self-monitoring dimensions. Not only are these bivariate scales more concise and interpretable than full-scale SMS scores but they also multiply researchers’ data-analytic possibilities to include examinations of (a) univariate effects, (b) joint effects in common or opposite directions, and (c) interactive effects. In view of these numerous advantages, the use of idiosyncratic abbreviations to the SMS should be abandoned.

Acknowledgments

We thank Mark Davison, for his helpful comments on an earlier version of the article, and Lew Goldberg, for making available data from the Eugene-Springfield Community Sample.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Despite a 40-year research history, the first published IRT analysis of the SMS occurred only recently. Using responses from a large, online sample ($N = 39,218$), from which our Sample 1 is a subset, He, Glas, Kosinski, Stillwell, and Veldkamp (2014) examined if textual features of individuals’ posts on the popular social-networking site, Facebook, could predict self-monitoring classification (i.e., high vs. low self-monitors). Authors fitted a unidimensional two-parameter logistic (2PL) model to the data, and investigated item fit using Lagrange multiplier tests (Glas, 1998) across three subgroups of respondents, which, based on their total SMS scores, were classified as low (0-9), intermediate (10-14), and high self-monitors (15-25). Researchers concluded that the IRT model provided adequate fit to the data, and that textual features of Facebook posts could accurately discriminate between high and low self-monitor classes. Intermediate self-monitors, which comprised the middle 50% of the distribution, were removed prior to using logistic regression to predict textual content.

2. All analyses were performed using the open-source statistics software R (R Development Core Team, 2015), using the “psych” package (Revelle, 2015). The sole exceptions were the confirmatory factor analyses (CFAs), which were performed using Amos, Version 22.0.0 (Arbuckle, 2012) with maximum likelihood estimation of full covariance matrices.
 3. For the RMSEA and SRMR, values $<.08$ indicate acceptable fit, and values with 95% confidence intervals overlapping $.05$ indicate excellent fit (Kline, 2005).
 4. The sole exception was Item 17 for the protective factor, the average loadings of which was $.27$ across samples. Item 17 was retained because it is one of the handful of protective self-monitoring items that remained following the revision to the SMS (Snyder & Gangestad, 1986).
 5. Throughout the article, all item numbers refer to the order of items as presented in the original 25-item SMS (Snyder, 1974).
 6. Due to the increased number parameters estimated and the smaller sample size in Sample 2, we were unable compute general multinomial goodness of fit statistics (i.e., χ^2 -based RMSEAs). Instead, we report M_2 -based RMSEAs, which were based on full marginal tables. For complete reporting, M_2 -based RMSEAs for Sample 1 were $.08$ and $.05$, for the two respective scales.
 7. Although results indicated that the cross-loading items, 6 and 23, demonstrated item properties in the somewhat desirable range for both candidate item sets, the decision was made to exclude them for technical weaknesses in item writing. Item 6 is double-barreled (“I guess I put on a show to impress or entertain people”), and Item 23 is both double-barreled and somewhat ambiguous (“I feel a bit awkward in company and do not show up quite so well as I should”). Unclear or grammatically problematic items can lead to comprehension problems among respondents, which can ultimately result in less reliable and valid measurement (Simms, 2008).
 8. As these definitions suggest, the concept of DIF is analogous to measurement invariance (cf. Stark et al., 2006).
 9. Results of DIF analyses for three alternative age categories (i.e., 18-30 years, 31-60 years, and 61+ years) are also reported in the supplemental online materials (all supplementary materials available online at <http://asm.sagepub.com/content/by/supplemental-data>).
- References**
- Allen, D. G., Weeks, K. P., & Moffitt, K. R. (2005). Turnover intentions and voluntary turnover: The moderating roles of self-monitoring, locus of control, proactive personality, and risk aversion. *Journal of Applied Psychology, 90*, 980-990. doi:10.1037/0021-9010.90.5.980
- Arbuckle, J. L. (2012). Amos (Version 22.0.0) [Computer software]. Chicago, IL: SPSS.
- Barrick, M. R., Parks, L., & Mount, M. K. (2005). Self-monitoring as a moderator of the relationships between personality traits and performance. *Personnel Psychology, 58*, 745-767. doi:10.1111/j.1744-6570.2005.00716.x
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459. doi:10.1007/BF02293801
- Briggs, S. R., & Cheek, J. M. (1988). On the nature of self-monitoring: Problems with assessment, problems with validity. *Journal of Personality and Social Psychology, 54*, 663-678. doi:10.1037/0022-3514.54.4.663
- Briggs, S. R., Cheek, J. M., & Buss, A. H. (1980). An analysis of the self-monitoring scale. *Journal of Personality and Social Psychology, 38*, 679-686. doi:10.1037/0022-3514.38.4.679
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO 2.1 for Windows [Computer software]. Chicago, IL: Scientific Software International.
- Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews: A field investigation. *Journal of Applied Psychology, 88*, 944-953. doi:10.1037/0021-9010.88.5.944
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*, 1-30.
- Conn, S. R., & Rieke, M. L. (1994). *The 16PF* (Technical manual; 5th ed.). Champaign, IL: Institute for Personality and Ability Testing.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO PI-R* (Professional manual). Odessa, FL: Psychological Assessment Resources.
- Côté, S., Lopes, P. N., Salovey, P., & Miners, C. T. H. (2010). Emotional intelligence and leadership emergence in small groups. *Leadership Quarterly, 21*, 496-508. doi:10.1016/j.leaqua.2010.03.012
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302. doi:10.1037/h0040957
- Day, D. V., Schleicher, D. J., Unckless, A. L., & Hiller, N. J. (2002). Self-monitoring personality at work: A meta-analytic investigation of construct validity. *Journal of Applied Psychology, 87*, 390-401. doi:10.1037/0021-9010.87.2.390
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91*, 1138-1151. doi:10.1037/0022-3514.91.6.1138
- DeYoung, C. G., Grazioplene, R. G., & Peterson, J. B. (2012). From madness to genius: The Openness/Intellect trait domain as a paradoxical simplex. *Journal of Research in Personality, 46*, 63-78. doi:10.1016/j.jrp.2011.12.003
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*, 1246-1256. doi:10.1037/0022-3514.73.6.1246
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.
- Dunn, E. W., Biesanz, J. C., Human, L. J., & Finn, S. (2007). Misunderstanding the affective consequences of everyday social interactions: The hidden benefits of putting one's best face forward. *Journal of Personality and Social Psychology, 92*, 990-1005. doi:10.1037/0022-3514.92.6.990
- Ellis, R. J. (1988). Self-monitoring and leadership emergence in groups. *Personality and Social Psychology Bulletin, 14*, 681-693. doi:10.1177/0146167288144004
- Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult

- attachment. *Journal of Personality and Social Psychology*, 78, 350-365. doi:10.1037//0022-3514.78.2.350
- Fuglestad, P. T., & Snyder, M. (2010). Status and the motivational foundations of self-monitoring. *Social & Personality Psychology Compass*, 4, 1031-1041. doi:10.1111/j.1751-9004.2010.00311.x
- Gangestad, S. W., & Simpson, J. A. (1993). Development of a scale measuring genetic variation related to expressive control. *Journal of Personality*, 61, 133-158.
- Gangestad, S. W., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete class variables in personality. *Psychological Review*, 92, 317-349. doi:10.1037/0033-295X.92.3.317
- Gangestad, S. W., & Snyder, M. (1991). Taxometric analysis redux: Some statistical considerations for testing a latent class model. *Journal of Personality and Social Psychology*, 61, 141-146. doi:10.1037/0022-3514.61.1.141
- Gangestad, S. W., & Snyder, M. (2000). Self-monitoring: Appraisal and reappraisal. *Psychological Bulletin*, 126, 530-555. doi:10.1037/0033-2909.126.4.530
- Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco, CA: W. H. Freeman.
- Glas, C. A. W. (1998). Detection of differential item functioning using Lagrange multiplier tests. *Statistica Sinica*, 8, 647-667.
- Glomb, T. M., & Liao, H. (2003). Interpersonal aggression in work groups: Social influence, reciprocal, and individual effects. *Academy of Management Journal*, 46, 486-496.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96. doi:10.1016/j.jrp.2005.08.007
- Hambleton, R. K. (2006). Good practices for identifying differential item functioning. *Medical Care*, 44(Suppl. 3), S182-S188. doi:10.1097/01.mlr.0000245443.86671.c4
- He, Q., Glas, C. A. W., Kosinski, M., Stillwell, D. J., & Veldkamp, B. P. (2014). Predicting self-monitoring skills using textual posts on Facebook. *Computers in Human Behavior*, 33, 69-78. doi:10.1016/j.chb.2013.12.026
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type 1 error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14, 329-349. doi:10.1207/S15324818AME1404_2
- John, O. P., Cheek, J. M., & Klohnen, E. C. (1996). On the nature of self-monitoring: Construct explication with Q-sort ratings. *Journal of Personality and Social Psychology*, 71, 763-776. doi:10.1037/0022-3514.71.4.763
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110, 5802-5805. doi:10.1073/pnas.1218772110
- Lennox, R. (1988). The problem with self-monitoring: A two-sided scale and a one-sided theory. *Journal of Personality Assessment*, 52, 58-73. doi:10.1207/s15327752jpa5201_5
- Lennox, R. D., & Wolfe, R. N. (1984). Revision of the self-monitoring scale. *Journal of Personality and Social Psychology*, 46, 1349-1364.
- Lynam, D. R., Sherman, E. D., Samuel, D., Miller, J. D., Few, L. R., & Widiger, T. A. (2013). Development of a short form of the elemental psychopathy. *Assessment*, 20, 659-669. doi:10.1177/1073191113502072
- Magis, D., Beland, S., & Raiche, G. (2013). *difR: Collection of methods to detect dichotomous differential item functioning (DIF) in psychometrics* (R package version 4.5). Retrieved from <http://artax.karlin.mff.cuni.cz/r-help/library/difR/html/00Index.html>
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, 88, 139-157. doi:10.1037/0022-3514.88.1.139
- McDonald, R. P. (1999). *Test theory: A unified approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Miell, D., & le Voi, M. (1985). Self-monitoring and control in dyadic interactions. *Journal of Personality and Social Psychology*, 49, 1652-1661. doi:10.1037/0022-3514.49.6.1652
- Mullen, E. J., & Noe, R. A. (1999). The mentoring information exchange: When do mentors seek information from their protégés? *Journal of Organizational Behavior*, 20, 233-242.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32, 396-402. doi:10.3758/BF03200807
- Penfield, R. D., & Camilli, G. (2006). Differential item functioning and item bias. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics* (pp. 125-167). Amsterdam, Netherlands: Elsevier.
- R Development Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Raven, J. C. (2000) The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41, 1-48.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566. doi:10.1037/0033-2909.114.3.552
- Revelle, W. (2015). *Procedures for psychological, psychometric, and personality research* (User's manual). Retrieved from <http://www.personality-project.org/r>
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Retrieved from <https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf>
- Scherbaum, C. A., Cohen-Charash, Y., & Kern, M. J. (2006). Measuring general self-efficacy: A comparison of three measures using item response theory. *Educational and Psychological Measurement*, 66, 1047-1063. doi:10.1177/0013164406288171
- Scott, B. A., Barnes, C. M., & Wagner, D. T. (2012). Chameleonic or consistent? A multilevel investigation of emotional labor variability and self-monitoring. *Academy of Management Journal*, 55, 905-926.
- Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social & Personality Psychology Compass*, 2, 414-433. doi:10.1111/j.1751-9004.2007.00044.x
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30, 526-537. doi:10.1037/h0037039

- Snyder, M. (1987). *Public appearances/private realities: The psychology of self-monitoring*. New York, NY: W. H. Freeman.
- Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology, 51*, 125-139. doi:10.1037/0022-3514.51.1.125
- Snyder, M., Simpson, J. A., & Gangestad, S. (1986). Personality and sexual relations. *Journal of Personality and Social Psychology, 51*, 181-190. doi:10.1037/0022-3514.51.1.181
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*, 1292-1306. doi:10.1037/0021-9010.91.6.1292
- Stillwell, D. J., & Kosinski, M. (2015). *myPersonality Project website*. Retrieved from <http://mypersonality.org>
- Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research, 16*, 33-42. doi:10.1007/s11136-007-9184-6
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.
- Wilmot, M. P. (2015). A contemporary taxometric analysis of the latent structure of self-monitoring. *Psychological Assessment, 27*, 353-364. doi:10.1037/pas0000030
- Wilmot, M. P., DeYoung, C. G., Stillwell, D., & Kosinski, M. (2015). Self-monitoring and the metatraits. *Journal of Personality*. Advance online publication. doi:10.1111/jopy.12162
- Wolf, H., Spinath, F. M., Riemann, R., & Angleitner, A. (2009). Self-monitoring and personality: A behavioural-genetic study. *Personality and Individual Differences, 47*, 25-29. doi:10.1016/j.paid.2009.01.040