Assessment

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

SCHOLARONE™
Manuscripts

IRT-BASED SUBSCALES OF THE SMS

Abstract

For the last 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 items (SMS; Snyder, 1974) and data from two large, non-student samples (Ns = 13,563 and 709). Results indicate that the new acquisitive (6-item) and protective (7-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 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

Using Item Response Theory to Develop Measures of Acquisitive and Protective

Self-Monitoring from the Original Self-Monitoring Scale

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 trait constructs (Barrick, Parks, & Mount, 2005).

Results from quantitative review provided 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. In fact, nearly

50% of the self-monitoring literature has been published since 2000 (Web of Science, July 2015).

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 comprised 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 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 scale 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, non-student 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 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/sm) 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, *IQR* = 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 was required for the

SMS, gender, and age.

Measures

  SMS. The Self-Monitoring Scale (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.

  FFM factors and facets. A subset of Study 1 participants (*N* = 1,197) completed the

IPIP-NEO, which is a proxy of the NEO-PI-R based on the public domain Internal Personality

Item Pool (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 FFM trait domains into six

10-item facet scales. There is high degree of correspondence between NEO-PI-R and IPIP-NEO

scores (observed r_{Mean} = .73; corrected for attenuation, r_{Mean} = .90). Internal consistency estimates

for the facet scales ranged from α = .71 to .90, and from α = .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 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 α = .61 to .85, and from α = .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 (RSPM; 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 RSPM 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 Factor Questionnaire

(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 (α = .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 minimum average partial (MAP) test (Velicer, 1975) 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 2 factors, whereas parallel analysis

indicated 3. For Sample 2, MAP test results and parallel analysis alike indicated the presence of

3 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 3 rather than 2 factors, we explored the validity of a 3-factor solution first.

3-Factor EFAs. We conducted exploratory factor analyses (EFA) using principal axis

factoring, and 3 factors were extracted. Across samples, eigenvalues were indicative of 2 robust

factors, but a comparatively weak third factor (Sample 1 λs = 2.37, 1.84, and .64; Sample 2 λs =

3.36, 2.46, and 1.05). Using the conventional loading criteria of ≥ .30 in absolute value, results

indicated that the third factor was defined by only 1 item in Sample 1, and 4 items in Sample 2;

maximum factor loadings were relatively weak (r_{S1} = .29, r_{S2} = .42), and 3 of the 5 total items

defining factor 3 had substantial cross-loadings. Fit indices for the 3-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 2-factor model; in the last analysis, the comparatively small third eigenvalue,

and fewer and weaker item indicators, provided persuasive evidence against the 3-factor model.

2-Factor EFAs. Next, we conducted a second set of EFAs using principal axis factoring,

and 2 factors were extracted and rotated; we used a direct oblimin rotation, which allows for

correlated factors, to rotate factors toward simple structure. Nevertheless, even upon rotation,

results indicated that the two factors were virtually orthogonal (factor rs = .08 and .07,

respectively), a finding which is consistent with that of Briggs and Cheek (1988, p. 669). Fit

indices for the 2-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, we selected

all items that met the conventional factor-loading criterion of $\geq .30$.[4] Nine items met this

criterion for the acquisitive factor (items 1, 5, 8, 12, 14, 18, 20, 21, and 22), 8 items met the

criterion for the protective factor (items 2, 6, 7, 13, 15, 16, 17, 19, and 25), and 2 items (items 6

and 23) met criteria for both factors; both cross-loading items were included as indictors of both

factors.[5] All EFA results are included in the supplementary online materials.

   2-Factor CFAs. Having found that a 2-factor model provided a good fit to the data, and

having selected item indicators of the acquisitive and protective self-monitoring factors, we

tested the validity of our two sets of items using confirmatory factor analysis (CFA). The 9

acquisitive items were used to define the latent acquisitive factor, the 8 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 2 factors were uncorrelated (factor rs = .03). Fit indices for the 2-factor CFA

models were poorer compared to those of the 2-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 the 10 protective items (both sets of which included the 2

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

<div align="center">[Insert Table 1 about here]</div>

   Taken together, the above findings suggest that a 2-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 et al., 1980; Briggs and Cheek, 1988; Snyder &

IRT-BASED SUBSCALES OF THE SMS   8

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_{SMS} = .78, r_{SMS-R}

= .89; Sample 2: r_{SMS} = .78, r_{SMS-R} = .90), and moderately related to the protective pool (Sample

1: r_{SMS} = .56, r_{SMS-R} = .34; Sample 2: r_{SMS} = .57, r_{SMS-R} = .37). The two item pools, by contrast,

were weakly correlated (r_{S1} = .03 and r_{S2} = .02). For the dichotomous items in Sample 1, internal

consistencies were adequate for the acquisitive pool (α = .71), but low for the protective pool (α

= .63). Similar patterns emerged for polytomous sets in Sample 2; the acquisitive coefficient was

desirable (α = .80), while the estimate for the protective set was somewhat low (α = .68).

[Insert Table 2 about here]

Parameter Estimates, Information Functions, and Standard Errors of Measurement

Having conducted classical psychometric analyses, we performed IRT analyses using the

2-Parameter Logistic model (2PL) and Samejima's (1969) Graded Response Model (GRM). In

both cases, item parameters are estimated using marginal maximum likelihood (MML; Bock &

Aiken, 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 2 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 GRM, 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 et al. 2000) than do items in the protective pool. Correspondingly, standard errors of trait estimates obtained from the acquisitive candidate item pool would 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 ≥ .85 in Sample 1 were both near zero (*b* = -.01 and *b* = -.20 for acquisitive and protective scales, respectively), item difficulties nonetheless tended to range close to one standard deviation above and below these values (from -.84 to .75, and from -.81 to 1.43 for the respective scales). Likewise, items in Sample 2 tended to possess difficulty parameters spanning

IRT-BASED SUBSCALES OF THE SMS   10

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. -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.

[Insert Table 3 about here]

Final Item Selection

Acquisitive self-monitoring scale. Across samples, Items 8, 12, 18, and 20 were characterized by relatively high discrimination ($M\ a_{S1}$ = 1.41, $M\ a_{S2}$ = 1.67) and information. In contrast, item 21 was excluded due to low discrimination (*a*'s = .72 and .76) and little information across samples. Having selected the strongest and excluded the weakest items, respectively, 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 (*a*'s = 1.25 and 1.61, respectively), but somewhat lower estimates using the dichotomous format (*a*'s = .96 and .98, respectively). On the other hand, items 1 and 14 evidenced adequate discrimination in Sample 1 (*a*'s = .98 and .95), but not in Sample 2 (*a*'s = .62 and .82). Inspecting IIFs, visual evidence showed that all 4 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 6 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 ($M\ a_{S1}$

$= 1.56$, $M\ a_{S2} = 1.57$) 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 ($a = .91$). 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 (a's $= 1.03$, and $.85$, 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, 7 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 (TIF) for candidate item pools and final scales alike, which

are depicted in adjacent panes in Figure 1. Impressively, TIFs 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 self-monitoring scales, next, we turned our attention toward examining questions of measurement bias, which might be associated with items in the two new scales.

[Insert Figure 1 about here]

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 differential item functioning (DIF).[8] DIF occurs when the relation between a latent trait and item responses differs across subgroups. Although age is a continuous variable, we chose to artificially dichotomize it (i.e. < 40 and ≥ 40 years) primarily to comport with United States federal age discrimination legislation, which is relevant for personnel selection in applied settings.[9]

We adopted a multi-stage 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: 1) IRT-based likelihood-ratio tests (IRT-LR; Reise, Widaman, & Pugh, 1993), 2) ordinal-logistic regression-based approaches (ORL), and 3) 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-x-scale score* variables. Likelihood-ratio tests are used to determine whether these additional variables result in improved model fit (i.e. Δ Nagelkerke 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 (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 to those obtained when all item parameters were constrained to be equal across groups. What these results show is 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, 2015). 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 3 and 4. 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.

[Insert Table 4 about here]

Sensitivity Tests. The evidence described above indicates that DIF does not appear to present a major hurdle to use of the 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 = .12, Range = .01 to .32; Sample 2: M = .32, Range = .01 to .93) across samples and groups, and moderate differences in difficulty parameters (Sample 1: M = .22, Range = .00 to .46; Sample 2: M = .40, Range = .01 to 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 to 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 ≈ .02 in Sample 1

IRT-BASED SUBSCALES OF THE SMS   15

across tests and grouping variables; mean absolute difference ≈ .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 self-monitoring scale scores for both males and females, and age groups (see Table 1).

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 ≈ .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_{age} = .10. Lastly, 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 1). 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 self-monitoring scales were

not significantly biased in terms of gender or age groups, finally, we turned our attention toward

to testing the construct validity of the new measures.

[Insert Table 5 about here]

Convergent and Discriminant Validity

In a meta-analyses 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 (Digman, 1997; DeYoung, 2006). 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 et al., 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-lever 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 et al., 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 ($rs = -.12$). 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 non-redundant 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 5).

[Insert Table 5 about here]

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 Self-Monitoring Scale

(SMS; Snyder, 1974). We used two large, non-student 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 exploratory and confirmatory factor analysis 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 6 acquisitive and 7 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 bi-dimensional 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, & Moffit, 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. -1.00 SD to 1.00 SD). Further, despite limited accuracy at the low ends of the traits, assessment at the high ends is considerably more precise when using the polytomous (versus the dichotomous) response format. A second limitation is the relatively low internal consistency of the new protective scale. Although the 7-item scale is approximately as reliable as the 10 items composing 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 to boost the reliability of the protective self-monitoring scale. 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 self-monitoring scales hold promise for future research, they are particular 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 inter-item 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

IRT-BASED SUBSCALES OF THE SMS   21

can be formed from correlations using formulae reported in the literature (Ghiselli, Campbell, &

Zedeck, 1980, Chapter 7). 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 6 items selected for the acquisitive scale

are present in the SMS-R. In contrast, only 3 of the 7 items in the protective scale (i.e., items 13,

16, and 25) remain in the SMS-R. Researchers might consider using these 3 protective items, in

addition to item 17, which, although it was 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 4-

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 6-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). In still 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.

Footnotes

[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 et al. (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 2PL logistic model to the data, and investigated item fit using Lagrange Multiplier tests (LM; Glas, 1999) 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 CFA analyses, 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 those 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 is one of the handful of protective self-monitoring items that remained following the revision to the Self-Monitoring Scale (Snyder & Gangestad, 1986).

[5] Throughout the manuscript, all item numbers refer to the order of items as presented in the original 25-item Self-Monitoring Scale (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, Chernyshenko, & Drasgow, 2006).

[9] Results of DIF analyses for three alternative age categories (i.e. 18 to 30, 31 to 60, and 61+ years) are also reported in the supplemental online materials.

IRT-BASED SUBSCALES OF THE SMS   25

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 program]. Chicago: SPSS.

Arkin, R. M. (1981). Self-presentation styles In J. T. Tedeschi (Ed.), *Impression management*

theory and social psychological research (pp. 311-333). New York: Academic Press.

Baker, F. B., & Kim, S.-H. (2004). *Item Response Theory: Parameter estimation techniques* (2nd

ed.). Monticello, NY: Marcel Dekker Inc.

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., & Buss (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

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

Burkhardt, M. E. (1994). Social interaction effects following a technological change: A

longitudinal investigation. *Academy of Management Journal, 37,* 869-898.

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO 2.1 for Windows. Chicago, IL:

Scientific Software International.

Chang, H.-H., & Ying, Z. (1996). A global information approach to computerized adaptive

testing. *Applied Psychological Measurement*, *20*, 213-229.

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.

http://www.jstatsof.org/v39/i08

Conn, S. R., & Rieke, M. L. (1994). *The 16PF fifth edition technical manual.* 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, Inc.

Côté, S., Lopes, P. N., Salovey, P., & Miners, C. T. H. (2010). Emotional intelligence and

leadership emergence in small groups. *The Leadership Quarterly*, *21*, 496–508.

http://doi.org/10.1016/j.leaqua.2010.03.012

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Mason, OH:

Cengage Learning.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological

Bulletin*, *52*, 281-302. http://dx.doi.org.ezp1.lib.umn.edu/10.1037/h0040957

Day, D. V., & Schleicher, D. J., (2006). Self-monitoring at work: A motive-based perspective.

Journal of Personality, 74, 685-713. doi:10.1111/j.1467-6494.2006.00389.x

IRT-BASED SUBSCALES OF THE SMS   27

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

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

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample.

Journal of Personality and Social Psychology, *91*(6), 1138–1151.

http://doi.org/10.1037/0022-3514.91.6.1138

DeYoung, C. G. (2013). The neuromodulator of exploration: A unifying theory of the role of

dopamine in personality. *Frontiers in Human Neuroscience, 7,* article 762.

doi:10.3389/fnhum.2013.00762

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. http://doi.org/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. http://dx.doi.org.ezp3.lib.umn.edu/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, England: Lawrence Erlbaum Associates, Inc.

Edwards, M. C. (2009). An introduction to item response theory using the Need for Cognition

Scale. *Social and Personality Psychology Compass*, *3*, 507-529. doi:10.1111/j.1751-

9004.2009.00194.x

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for psychologists (Multivariate

applications series)*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ellis, R. J. (1988). Self-monitoring and leadership emergence in groups. *Personality and Social

Psychology Bulletin, 14*, 681−693. http://doi.org/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

IRT-BASED SUBSCALES OF THE SMS   29

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral*

    sciences. San Francisco: W. H. Freeman and Company

Glas, C. A. W. (1998). Detection of differential item functioning using lagrange multiplier tests.

    Staistica 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:0.1016/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

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response*

    Theory. Newbury Park, CA: SAGE Publications.

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

Hirsh, J. B., DeYoung, C. G., & Peterson, J. B. (2009). Metatraits of the Big Five differentially

    predict engagement and restraint of behavior. Journal of Personality, 77, 1085-1102.

    doi:10.1111/j.1467-6494.2009.00575.x

Hoyle, R. H., & Lennox, R. D. (1991). Latent structure of self-monitoring. *Multivariate*

    Behavioral Research, 26, 511-540.

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

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment, 21*, 28-41. doi:10.1177/1073191113514105

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* New York: 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 (PNAS), 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*.

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

IRT-BASED SUBSCALES OF THE SMS   31

Miell, D., & le Voi, M. (1985). Self-monitoring and control in dyadic interactions. *Journal of*

   Personality and Social Psychology, 49, 1652-1661.

   http://dx.doi.org.ezp2.lib.umn.edu/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. (2007). 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, ISBN 3-900051-07-0, URL

   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.

Revelle, W. (2015). Procedures for psychological, psychometric, and personality research:

   User's manual [computer software and manual]. Retrieved February 1, 2015, from

   http://www.personality-project.org/r

Reise, S. P., & Waller, N. G. (2009). Item Response Theory and clinical measurement. *Annual*

   Review of Clinical Psychology, 5, 27-48.

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

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34* (4, Pt. 2)

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. http://dx,doi,org/10,5465/ainj.2010,105D

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

Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass 2/1,* 414-433, 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: 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*, 181-90. http://dx.doi.org.ezp3.lib.umn.edu/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.* Advance online publication.

http://dx.doi.org/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.

doi10.1016/j.paid.2009.01.040

IRT-BASED SUBSCALES OF THE SMS

Tables

Table 1
Confirmatory Factor Analytic Results of the Self-Monitoring Scale (SMS): 2-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. N*s for Samples 1 and 2 = 13,563 and 709, respectively. Maximum likelihood estimation of the full covariance matrix. Factor loadings ≥ .30 in absolute value are bolded.
Factor intercorrelations (i.e. phi) = .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 Self-Monitoring Scale (SMS; Snyder, 1974).

IRT-BASED SUBSCALES OF THE SMS

Table 2

Scale Descriptive Statistics, Part-Whole Correlations, and Relations to Demographic Correlates

Measure	# items	1	2	3	4	5	6	Demographic Correlates		
								Gender *d*	Age Group *d*	Age *r*
1. Self-Monitoring Scale	25	*.68/.74*	.93	.78	.56	.65	.58	.20/.28	.38/.12	-.15/-.10
2. Self-Monitoring Scale-Revised	18	.94	*.69/.77*	.89	.34	.78	.35	.20/.29	.25/.20	-.10/-.14
3. Acquisitive candidate item pool	11	.78	.90	*.71/.80*	.03	.89	.04	.13/.23	.11/.28	-.04/-.15
4. Protective candidate item pool	10	.57	.37	.02	*.63/.68*	-.04	.94	.10/.15	.37/-.19	-.16/.03
5. Acquisitive final items	6	.69	.82	.93	-.01	*.65/.77*	-.03	.13/.32	.08/.20	-.03/-.10
6. Protective final items	7	.57	.36	.02	.95	-.01	*.61/.69*	.07/.11	.35/-.20	-.16/.03
Sample 1	*M*	12.13	9.38	5.69	3.54	2.84	3.38	.40	.08	25.63
	SD	4.08	3.50	2.72	2.28	1.77	1.80	.49	.27	8.36
Sample 2	*M*	13.16	9.64	5.97	3.80	2.98	3.54	.42	.83	51.41
	SD	2.31	2.05	1.60	1.20	1.07	.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. Chronbach's α) reported along the diagonal in italics (i.e. S1/S2).

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. under 40 vs. 40 years and above). 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, 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.

IRT-BASED SUBSCALES OF THE SMS

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				
		a	b	a	b_1	b_2	b_3	b_4
Acquisitive self-monitoring								
5	I can make impromptu speeches even on topics about which I have almost no information.	.99(.03)	.21(.02)	1.12(.11)	-.21(.08)	.79(.10)	1.23(.12)	2.80(.25)
8	I would probably make a good actor.	1.94(.06)	-.14(.01)	2.66(.26)	-.44(.06)	.20(.05)	.63(.06)	1.55(.10)
12	In a group of people I am rarely the center of attention.[R]	1.13(.04)	.37(.02)	1.20(.11)	-1.41(.13)	.20(.08)	.99(.10)	2.78(.23)
18	I have considered being an entertainer.	1.49(.04)	.14(.02)	2.29(.21)	.24(.06)	.76(.07)	1.00(.08)	1.85(.13)
20	I have never been good at games like charades or improvisational acting.[R]	1.52(.04)	-.44(.02)	1.59(.13)	-1.39(.11)	-.26(.07)	.25(.07)	1.44(.11)
22	At a party, I let others keep the jokes and stories going.[R]	.85(.05)	.84(.03)	1.22(.11)	-1.56(.14)	.25(.08)	.88(.10)	2.70(.23)
Protective self-monitoring								
2	My behavior is usually an expression of my true inner feelings, attitudes, and beliefs.[R]	1.08(.04)	1.38(.04)	.83(.10)	-.73(.13)	2.40(.28)	3.01(.35)	5.81(.77)
7	When I am uncertain how to act in a social situation, I look to the behavior of others for cues.	.84(.03)	-.83(.03)	.65(.09)	-3.32(.46)	-1.69(.25)	-.42(.13)	3.44(.47)
13	In different situations and with different people, I often act like very different persons.	1.56(.05)	-.05(.02)	1.40(.13)	-.75(.09)	.38(.07)	.89(.09)	2.45(.20)
15	Even if I am not enjoying myself, I often pretend to be having a good time.	.81(.03)	-.03(.02)	.96(.11)	-2.54(.26)	-.47(.10)	.40(.09)	3.67(.38)
16	I'm not always the person I appear to be.	1.54(.05)	-.70(.02)	1.70(.16)	-.78(.08)	.28(.06)	.75(.07)	2.28(.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(.05)	1.10(.03)	1.48(.14)	-.46(.07)	.99(.09)	1.66(.13)	3.33(.29)
25	I may deceive people by being friendly when I really dislike them.	.93(.03)	-.33(.02)	1.23(.12)	-1.16(.12)	.18(.07)	.72(.09)	3.29(.30)

Note. Ns 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 conducting using IRTPRO (Cai, Thissen, & du Toit, 2011). Fit indices for final acquisitive (RMSEAs = .04 and .04) and protective scales (RMSEAs = .02 and .04).

All item numbers refer to the order of items as presented in the original 25-item Self-Monitoring Scale (SMS; Snyder, 1974).

[R] = Item reverse-scored.

IRT-BASED SUBSCALES OF THE SMS

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	λMH	JG	ETS	OLR-LR χ^2	IRT-LR χ^2	ΔR^2	JG
Acquisitive self-monitoring											
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	.16	A	A	.68(2)	Referent	<.01	A
12	69.82(2)*	40.79(2)*	67.18(1)*	<.01	-.86	A	A	2.38(2)	6.79(5)	<.01	A
18	28.91(2)*	10.95(2)*	29.53(1)*	<.01	.56	A	A	2.64(2)	7.18(5)	<.01	A
20	.44(2)	.35(2)	.24(1)	<.01	.06	A	A	7.22(2)*	4.71(5)	<.01	A
22	58.35(2)*	36.25(2)*	51.85(1)*	<.01	-.74	A	A	4.56(2)	4.31(5)	<.01	A
Protective self-monitoring											
2	19.90(2)*	7.66(2)*	18.49(1)*	<.01	.49	A	A	8.95(2)*	8.62(5)	.01	A
7	11.40(2)*	9.05(2)*	11.08(1)*	<.01	-.34	A	A	2.19(2)	5.27(5)	<.01	A
13	33.61(2)*	6.97(2)*	33.07(1)*	<.01	.60	A	A	4.27(2)	8.95(5)	<.01	A
15	39.49(2)*	23.27(2)*	38.07(1)*	<.01	-.60	A	A	15.20(2)*	14.28(5)*	.02	A
16	11.57(2)*	.72(2)	9.43(1)*	<.01	.34	A	A	.11(2)	10.19(5)	<.01	A
19	4.26(2)	Referent	.36(1)	<.01	.08	A	A	3.77(2)	Referent	<.01	A
25	29.80(2)*	18.01(2)*	27.06(1)*	<.01	-.51	A	A	7.32(2)*	3.62(5)	<.01	A

Note. For Sample 1, Male $N = 5,365$, Female $N = 8,198$; for Sample 2, Male $N = 303$, Female $N = 406$.

Parameter estimates followed by *df* 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 x sum-score interaction terms; λMH = 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 λMH ('A' = negligible; 'B' = moderate; 'C' = large).

* = Statistically significant.

IRT-BASED SUBSCALES OF THE SMS

Table 5

DIF Analyses Across Age Groups (Under 40 and 40 Years and Above)

SMS item	Sample 1							Sample 2			
	OLR-LR χ^2	IRT-LR χ^2	M-H χ^2	$\varDelta R^2$	λMH	JG	ETS	OLR-LR χ^2	IRT-LR χ^2	$\varDelta R^2$	JG
Acquisitive self-monitoring											
5	7.29(2)*	2.31(2)	5.49(1)*	<.01	.41	A	A	4.16(2)	3.69(5)	<.01	A
8	3.24(2)	2.44(2)	1.35(1)	<.01	.23	A	A	3.18(2)	4.26(5)	<.01	A
12	28.71(2)*	17.59(2)*	25.93(1)*	<.01	-.93	A	A	1.99(2)	4.06(5)	<.01	A
18	23.53(2)*	11.11(2)*	19.70(1)*	<.01	.86	A	A	.66(2)	2.09(5)	<.01	A
20	.47(2)	Referent	.40(1)	<.01	.13	A	A	.24(2)	Referent	<.01	A
22	44.10(2)*	26.30(2)*	44.41(1)*	<.01	-1.19	A	B	.79(2)	10.25(5)	<.01	A
Protective self-monitoring											
2	4.01(2)	1.72(2)	2.14(1)	<.01	.34	A	A	1.93(2)	4.40(5)	<.01	A
7	.84(2)	1.17(2)	.33(1)	<.01	-.11	A	A	.32(2)	5.47(5)	<.01	A
13	.27(2)	Referent	.15(1)	<.01	.08	A	A	.02(2)	Referent	<.01	A
15	4.11(2)	.24(2)	3.84(1)*	<.01	-.35	A	A	5.34(2)	4.62(5)	<.01	A
16	18.74(2)*	5.10(2)	17.45(1)*	<.01	.79	A	A	1.63(2)	9.08(5)	<.01	A
19	4.00(2)	.96(2)	2.99(1)	<.01	-.42	A	A	3.34(2)	4.34(4)	<.01	A
25	.24(2)	1.22(2)	.09(1)	<.01	.06	A	A	5.28(2)	6.53(5)	<.01	A

Note. For Sample 1, *N*s for under 40 and 40 years and above = 12,456, and 1,107, respectively. For Sample 2, *N*s for under 40 and 40 years and above = 119 and 590, respectively. Parameter estimates followed by *df* 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 x sum-score interaction terms; λMH = 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 λMH ('A' = negligible; 'B' = moderate; 'C' = large).

* = Statistically significant.

IRT-BASED SUBSCALES OF THE SMS

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

Measure	Internal Consistency		Original SMS		Acquisitive SM		Protective SM	
	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. 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 zero. Bolded values indicated hypothesized relations.

IRT-BASED SUBSCALES OF THE SMS

Figures

Sample 1
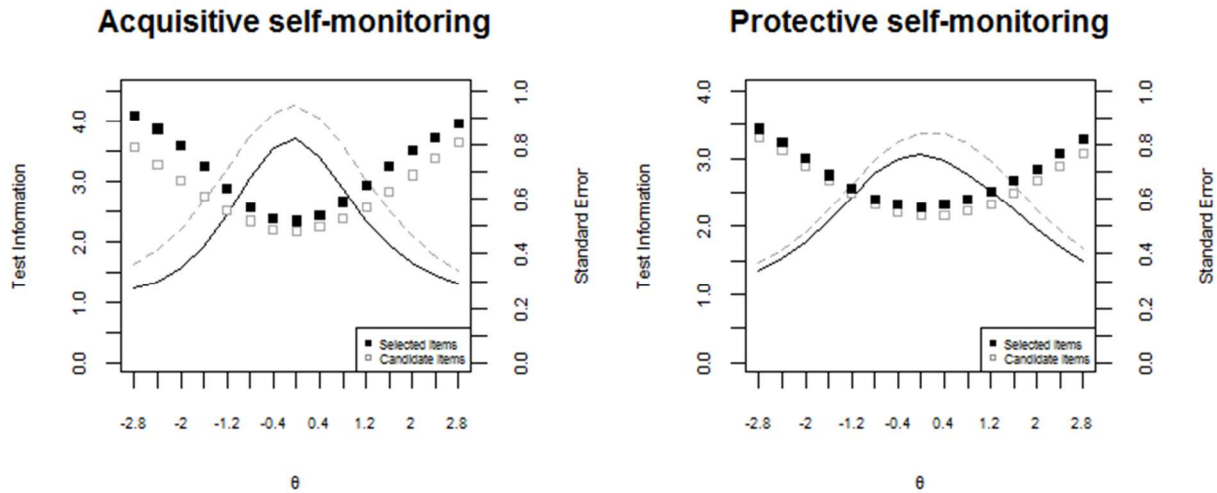
Sample 2

Figure 1. Test information functions (TIF) and standard error curves for candidate items (dashed lines) and final items (solid lines) for acquisitive and protective self-monitoring scales. Sample 1 items scored using a dichotomous response format (top panes); Sample 2 items scored using a polytomous response format (bottom panes).

IRT-BASED SUBSCALES OF THE SMS

Supplementary Tables

Table S1

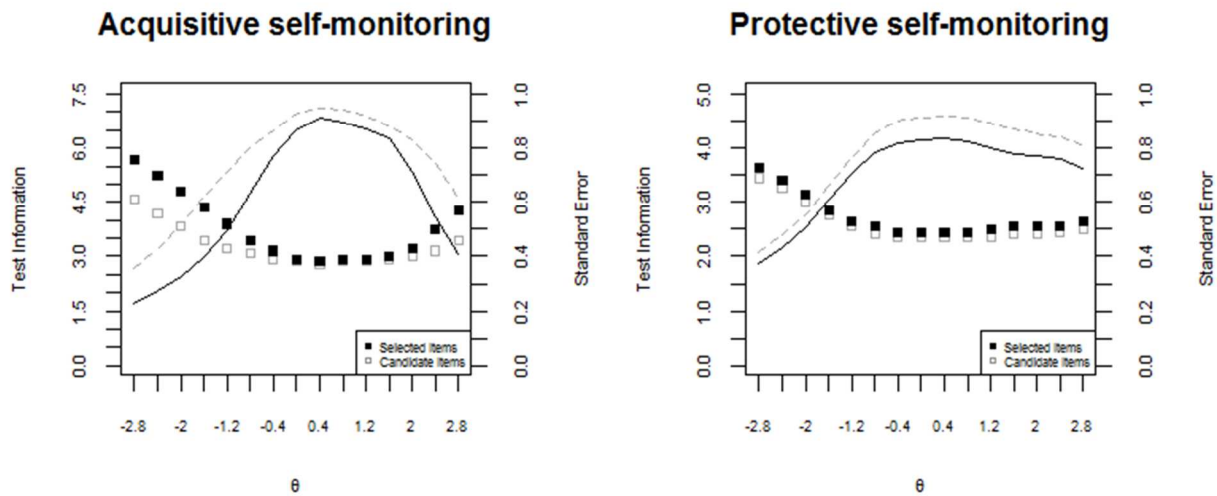*Items in the Self-Monitoring Scale and its Respective Empirically Developed Subscales*

Text	SMS	ACT	OD	EXT	SMSR	PP	AOD	TAX	GEC	ACQ	PRO
I find it hard to imitate the behavior of other people.[R]	1				1	1			1		
My behavior is usually an expression of my true inner feelings, attitudes, and beliefs.[R]	2		1								1
At parties and social gatherings, I do not attempt to say things that others will like.[R]	3		2		2						
I can only argue for ideas I already believe.[R]	4				3			1			
I can make impromptu speeches even on topics about which I have almost no information.	5	1			4	2			2	1	
I guess I put on a show to impress or entertain people.	6		3		5	3	1	2	3		
When I am uncertain how to act in a social situation, I look to the behavior of others for cues.	7		4								2
I would probably make a good actor.	8	2			6	4		3	4	2	
I rarely need the advice of my friends to choose movies, books, or music.[R]	9										
I sometimes appear to others to be experiencing deeper emotions than I actually am.	10										
I laugh more when I watch a comedy with others than when alone.	11										
In a group of people I am rarely the center of attention.[R]	12			1	7	5		4	5	3	
In different situations and with different people, I often act like very different persons.	13		5		8		2	5			3
I am not particularly good at making other people like me.[R]	14			2	9						
Even if I am not enjoying myself, I often pretend to be having a good time.	15		6								4
I'm not always the person I appear to be.	16		7		10		3				5
I would not change my opinions (or the way I do things) in order to please someone else or win their favor.[R]	17		8		11						
I have considered being an entertainer.	18	3			12	6		6	6	4	
In order to get along and be liked, I tend to be what people expect me to be rather than anything else.	19		9								6
I have never been good at games like charades or improvisational acting.[R]	20	4		3	13	7			7	5	
I have trouble changing my behavior to suit different people and different situations.[R]	21			4	14			7	8		
At a party, I let others keep the jokes and stories going.[R]	22			5	15	8		8	9	6	
I feel a bit awkward in company and do not show up quite so well as I should.[R]	23		10*	6	16	9	4*		10		
I can look anyone in the eye and tell a lie with a straight face (if for a right end).	24	5			17				11		
I may deceive people by being friendly when I really dislike them.	25		11		18		5				7

IRT-BASED SUBSCALES OF THE SMS

Note. ACT = Acting; OD = Other-Directedness; EXT = Extraversion. Subscales factor analytically derived (Briggs, Cheek, & Buss, 1980) from original 25-item Self-Monitoring Scale (SMS; Snyder, 1974). PP = Public Performing; AOD = Abbreviated Other-Directedness. Subscales factor analytically derived (Briggs & Cheek, 1988) from 18-item Self-Monitoring Scale-Revised (SMSR; Snyder & Gangestad, 1986). TAX = 8 items used in self-monitoring taxometric analysis (Gangestad & Snyder, 1985). GEC = Genic expressive control scale (Gangestad & Simpson, 1993), which corresponds to the first unrotated factor of the SMSR. ACQ = Acquisitive self-monitoring scale; PRO = Protective self-monitoring scale. Subscales developed using IRT (Wilmot et al., 2015).
[R] = Item reverse scored.
[*] = Cross-loading item included on multiple scales, but scored in the negative direction.

IRT-BASED SUBSCALES OF THE SMS

Table S2

Exploratory Factor Analytic Results of the Self-Monitoring Scale (SMS): 3-Factor Model

SMS item	Sample 1			Sample 2		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
1	**.45**	.02	.02	**.33**	.06	.10
2	.10	**.35**	-.06	.07	**.41**	.07
3	**.30**	.07	.29	.27	.12	**.42**
4	.23	.00	-.10	.26	.10	.03
5	**.40**	-.12	-.19	**.52**	-.08	-.14
6	**.43**	.21	.02	**.55**	.20	-.08
7	.10	**.37**	.14	-.01	**.35**	.10
8	**.54**	-.08	-.26	**.65**	-.06	**-.35**
9	.01	.09	.18	-.02	.01	.27
10	.09	.14	-.08	.04	**.34**	-.11
11	.13	.17	.13	.02	.22	.04
12	**.43**	**-.33**	.11	**.59**	-.22	.10
13	.23	**.48**	-.06	.21	**.54**	-.16
14	**.35**	-.23	.25	**.36**	-.29	**.36**
15	.14	**.34**	.12	.03	**.42**	.16
16	.17	**.46**	-.19	.19	**.57**	-.14
17	.15	.24	.21	.15	.25	**.40**
18	**.45**	-.13	-.13	**.57**	-.05	-.25
19	.14	**.48**	.17	.10	**.60**	.15
20	**.46**	-.18	-.14	**.55**	-.22	-.13
21	**.37**	.08	.16	**.42**	.13	.21
22	**.31**	**-.30**	.06	**.57**	-.19	.04
23	**.33**	**-.43**	.23	**.50**	**-.42**	.26
24	.29	.10	-.22	**.31**	.26	-.14
25	.20	**.33**	-.02	.17	**.47**	.02

*Note. N*s for Samples 1 and 2 = 13,563 and 709, respectively. Principal-axis factoring, unrotated factor matrices.
Factor loadings ≥ .30 in absolute value are bolded.
For Sample 1: eigenvalues = 2.37, 1.84, .64; RMSEA = .036, SRMR = .027.
For Sample 2: eigenvalues = 3.36, 2.46, 1.05; RMSEA = .050, SRMR = .041.
All item numbers refer to the order of items as presented in the original 25-item Self-Monitoring Scale (SMS; Snyder, 1974).

IRT-BASED SUBSCALES OF THE SMS

Table S3

Exploratory Factor Analytic Results of the Self-Monitoring Scale (SMS): 2-Factor Model

SMS item	Sample 1		Sample 2	
	Factor 1	Factor 2	Factor 1	Factor 2
1	**.40**	.20	**.30**	.14
2	-.06	**.36**	-.05	**.41**
3	.23	.18	.21	.18
4	.21	.09	.22	.17
5	**.41**	.05	**.52**	.06
6	**.30**	**.37**	**.47**	**.34**
7	-.06	**.38**	-.11	**.34**
8	**.52**	.14	**.62**	.11
9	-.02	.09	-.02	.00
10	.02	.16	-.06	**.34**
11	.05	.21	-.05	.22
12	**.53**	-.13	**.63**	-.06
13	.01	**.53**	.05	**.58**
14	**.40**	-.07	**.41**	-.18
15	-.01	**.36**	-.10	**.41**
16	-.04	**.48**	.01	**.60**
17	.03	.27	.07	.26
18	**.46**	.06	**.56**	.10
19	-.07	**.49**	-.08	**.60**
20	**.49**	.01	**.59**	-.07
21	**.30**	.22	**.37**	.23
22	**.42**	-.15	**.61**	-.03
23	**.47**	-.25	**.59**	-.27
24	.22	.20	.22	**.34**
25	.04	**.39**	.03	**.51**

*Note. N*s for Samples 1 and 2 = 13,563 and 709, respectively. Principal-axis factoring, pattern factor matrices.

Factor loadings ≥ .30 in absolute value are bolded.

Factor intercorrelations (i.e. phi) = .03 and .03, respectively.

For Sample 1: RMSEA = .044, SRMR = .037; for Sample 2: RMSEA = .065, SRMR = .057.

All item numbers refer to the order of items as presented in the original 25-item Self-Monitoring Scale (SMS; Snyder, 1974).

IRT-BASED SUBSCALES OF THE SMS

Table S4

Classical Item Statistics and IRT-based Parameter Estimates for Candidate Item Pools

SMS item	Sample 1					Sample 2								Final
	M	*SD*	CITC	*a*	*b*	*M*	*SD*	CITC	*a*	b_1	b_2	b_3	b_4	
Acquisitive self-monitoring														
1	.61	.49	.41	.98(.03)	-.57(.02)	2.72	1.21	.32	.62(.08)	-2.55(.35)	-.37(.14)	1.73(.25)	4.08(.54)	
5	.46	.50	.40	.96(.03)	.22(.02)	2.19	1.35	.51	1.25(.11)	-.20(.08)	.74(.09)	1.15(.11)	2.60(.22)	+
6	.40	.49	.34	.74(.03)	.59(.03)	2.12	1.21	.50	1.21(.11)	-.33(.08)	.83(.09)	1.35(.12)	3.80(.35)	
8	.54	.50	.53	1.53(.04)	-.15(.02)	2.48	1.39	.66	1.83(.15)	-.51(.07)	.21(.06)	.71(.07)	1.79(.12)	+
12	.42	.49	.52	1.42(.04)	.32(.02)	2.59	1.21	.61	1.59(.13)	-1.19(.10)	.17(.06)	.84(.08)	2.34(.16)	+
14	.66	.47	.40	.95(.03)	-.84(.03)	3.68	1.17	.40	.82(.09)	-4.22(.48)	-1.88(.22)	-.79(.13)	1.25(.16)	
18	.46	.50	.47	1.29(.04)	.15(.02)	1.98	1.35	.58	1.73(.16)	.26(.06)	.85(.08)	1.13(.09)	2.11(.15)	+
20	.61	.49	.48	1.39(.04)	-.46(.02)	3.01	1.37	.58	1.52(.12)	-1.41(.11)	-.26(.07)	.25(.07)	1.47(.11)	+
21	.67	.47	.33	.72(.03)	-1.07(.04)	3.09	1.22	.39	.76(.09)	-3.25(.37)	-.69(.13)	.42(.11)	2.72(.30)	
22	.35	.48	.39	.98(.03)	.75(.03)	2.62	1.20	.59	1.61(.13)	-1.33(.11)	.21(.06)	.74(.07)	2.26(.16)	+
23	.50	.50	.44	1.03(.03)	-.01(.02)	3.34	1.24	.57	1.26(.11)	-2.60(.21)	-.75(.09)	-.10(.07)	1.39(.12)	
Protective self-monitoring														
2	.23	.42	.37	1.03(.04)	1.43(.04)	1.87	.92	.40	.85(.10)	-.71(.12)	2.35(.27)	2.95(.33)	5.68(.73)	+
6	.40	.49	.36	.84(.03)	.54(.03)	2.12	1.21	.33	.73(.10)	-.46(.12)	1.22(.17)	1.97(.25)	5.67(.76)	
7	.65	.48	.36	.87(.03)	-.81(.03)	3.30	1.20	.34	.66(.09)	-3.26(.43)	-1.65(.24)	-.41(.13)	3.38(.45)	+
13	.51	.50	.51	1.56(.05)	-.05(.02)	2.46	1.31	.56	1.39(.13)	-.74(.09)	.39(.07)	.89(.09)	2.46(.20)	+
15	.51	.50	.36	.80(.03)	-.03(.02)	2.95	1.13	.44	.94(.10)	-2.58(.27)	-.48(.10)	.42(.10)	3.74(.39)	+
16	.68	.47	.48	1.48(.05)	-.72(.02)	2.50	1.28	.59	1.64(.15)	-.79(.08)	.29(.06)	.76(.08)	2.32(.17)	+
17	.26	.44	.28	.69(.03)	1.65(.06)	2.53	1.24	.28	.62(.09)	-1.92(.29)	.60(.15)	1.47(.23)	5.17(.75)	
19	.24	.43	.51	1.64(.05)	1.02(.02)	2.03	1.07	.63	1.68(.16)	-.42(.07)	.93(.08)	1.55(.12)	3.09(.25)	+
23*	.50	.50	.26	.48(.02)	.02(.04)	3.34	1.24	.28	.47(.08)	-3.07(.56)	.25(.17)	1.67(.33)	5.88(1.06)	
25	.57	.50	.39	.91(.03)	-.34(.02)	2.59	1.23	.49	1.19(.11)	-1.18(.12)	.19(.07)	.74(.09)	3.36(.30)	+

Note. Ns for Samples 1 and 2 = 13,563 and 709, respectively. CITC = corrected item-total correlation; a = discrimination parameter estimates; b = difficulty parameter estimates; standard errors are in parentheses. For Sample 1, χ^2-based RMSEAs = .01 and .01 for acquisitive and protective scales, respectively. For Sample 2, M_2-based RMSEAs = .04 for each scale. All RMSEAs are rounded to 2 decimal places.

Self-Monitoring Scale (SMS) item numbers refer to the original order as presented in Snyder (1974).

* Item scored in the negative direction.

Estimation of all parameters was conducting using IRTPRO (Cai, Thissen, & du Toit, 2011).

In the IRT framework, item response probabilities are modelled as a joint function of trait standing (θ) and item characteristics (e.g., discrimination and difficulty). In the 2-parameter logistic model:

IRT-BASED SUBSCALES OF THE SMS

$$p(x_{ij} = 1 | \theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - d_i)}}{1 + e^{a_i(\theta_j - d_i)}}.$$

Where x_{ij} indicates examinee j's score on item i, a refers to item discrimination, and b refers to item difficulty. Item discrimination is in turn proportional to the maximum slope of the item response curve. It is analogous to factor loadings and item-total correlations, in that it represents the degree to which an item indicates the measured trait, θ. Item difficulty corresponds to the point on the trait continuum at which the item characteristic curve achieves its maximum slope. In the two-parameter model, this corresponds to the trait level at which examinees have a .50 probability of responding to the item in the keyed direction. See Embretson & Reise (2000) for further information on the Item Response Theory approach to measurement. Samejima (1969) discusses the extension of IRT to polytomous item responses. Baker (2004) provides an excellent overview of parameter estimation within IRT models.

IRT-BASED SUBSCALES OF THE SMS

Table S5

Test Information Functions (TIF)

Θ	-2.80	-2.40	-2.00	-1.60	-1.20	-.80	-.40	.00	.40	.80	1.20	1.60	2.00	2.40	2.80
Sample 1															
Acquisitive candidate pool	1.62	1.87	2.22	2.67	3.20	3.74	4.14	4.26	4.05	3.58	3.03	2.51	2.09	1.77	1.53
Expected SE	.79	.73	.67	.61	.56	.52	.49	.48	.50	.53	.57	.63	.69	.75	.81
Acquisitive final items	1.22	1.35	1.58	1.93	2.44	3.05	3.56	3.70	3.39	2.86	2.35	1.94	1.65	1.44	1.30
Expected SE	.91	.86	.80	.72	.64	.57	.53	.52	.54	.59	.65	.72	.78	.83	.88
Protective candidate pool	1.46	1.65	1.91	2.24	2.61	2.98	3.25	3.38	3.37	3.24	2.98	2.62	2.25	1.93	1.68
Expected SE	.83	.78	.72	.67	.62	.58	.55	.54	.54	.56	.58	.62	.67	.72	.77
Protective final items	1.36	1.53	1.77	2.08	2.44	2.78	3.00	3.06	2.96	2.77	2.53	2.25	1.96	1.70	1.49
Expected SE	.86	.81	.75	.69	.64	.60	.58	.57	.58	.60	.63	.67	.71	.77	.82
Sample 2															
Acquisitive candidate pool	2.66	3.20	3.88	4.64	5.36	5.99	6.51	6.92	7.12	7.08	6.90	6.64	6.25	5.57	4.64
Expected SE	.61	.56	.51	.46	.43	.41	.39	.38	.37	.38	.38	.39	.40	.42	.46
Acquisitive final items	1.71	2.05	2.47	2.98	3.71	4.75	5.78	6.48	6.83	6.73	6.52	6.26	5.34	4.07	3.08
Expected SE	.76	.70	.64	.58	.52	.46	.42	.39	.38	.39	.39	.40	.43	.50	.57
Protective candidate pool	2.08	2.39	2.79	3.30	3.83	4.26	4.50	4.57	4.59	4.56	4.47	4.35	4.27	4.21	4.06
Expected SE	.69	.65	.60	.55	.51	.48	.47	.47	.47	.47	.47	.48	.48	.49	.50
Protective final items	1.87	2.16	2.55	3.03	3.53	3.91	4.09	4.16	4.18	4.12	4.01	3.90	3.86	3.79	3.62
Expected SE	.73	.68	.63	.57	.53	.51	.49	.49	.49	.49	.50	.51	.51	.51	.53

Note. *N*s for Samples 1 and 2 = 13,563 and 709, respectively.

IRT-BASED SUBSCALES OF THE SMS

Sensitivity Analyses

Sample 1

Table S6
Constrained Item Parameters and DIF-Flagged Item Parameters Estimated Freely Across Males and Females

SMS item	Constrained			Freely Estimated (Males)			Freely Estimated (Females)		
	a	*c*	*b*	*a*	*c*	*b*	*a*	*c*	*b*
Acquisitive self-monitoring									
5	1.00(.03)	-.11(.02)	.11(.02)	.98(.05)	.18(.03)	-.18(.03)	1.04(.05)	-.29(.03)	.28(.03)
8	1.93(.07)	.47(.04)	-.24(.02)	1.93(--)	.47(--)	-.24(--)	1.93(--)	.47(--)	-.24(--)
12	1.11(.04)	-.30(.03)	.27(.02)	1.15(.06)	-.49(.03)	.43(.03)	1.22(.06)	-.17(.03)	.14(.03)
18	1.49(.05)	-.05(.03)	.03(.02)	1.48(.07)	.05(.04)	-.03(.02)	1.55(.07)	-.11(.04)	.07(.02)
20	1.51(.05)	.82(.03)	-.54(.02)	1.56(.07)	.81(.04)	-.52(.03)	1.55(.07)	.83(.04)	-.54(.02)
22	.84(.03)	-.62(.02)	.74(.04)	.80(.05)	.77(.03)	.96(.06)	.97(.05)	-.52(.03)	.53(.04)
Protective self-monitoring									
2	1.07(.04)	-1.42(.03)	1.33(.05)	1.02(.06)	-1.31(.04)	1.28(.06)	1.05(.06)	-1.48(.04)	1.41(.07)
7	.82(.03)	.76(.02)	-.92(.04)	.83(.05)	.68(.03)	-.82(.05)	.81(.05)	.83(.03)	-1.03(.07)
13	1.54(.05)	.19(.03)	-.12(.02)	1.57(.08)	.31(.04)	-.19(.03)	1.44(.07)	.14(.04)	-.10(.03)
15	.79(.03)	.08(.02)	-.11(.03)	.81(.05)	-.04(.03)	.06(.04)	.78(.04)	.19(.03)	-.24(.04)
16	1.52(.06)	1.20(.04)	-.79(.03)	1.49(.08)	1.23(.05)	-.82(.04)	1.46(.08)	1.19(.05)	-.82(.05)
19	1.40(.05)	-1.46(.04)	1.05(.04)	1.40(--)	-1.46(--)	1.04(--)	1.40(--)	-1.46(--)	1.05(--)
25	.91(.03)	.38(.02)	-.41(.03)	.91(.05)	.26(.03)	-.28(.04)	.90(.05)	.48(.03)	-.53(.05)

Note. *N* for Sample 1 = 13,563; Male *N* = 5,365, Female *N* = 8,198. All items flagged for DIF in any one of the significance tests described in the main text are freely estimated across groups above.

IRT-BASED SUBSCALES OF THE SMS

Sample 2

Table S7

Constrained Item Parameters and DIF-Flagged Item Parameters Estimated Freely Across Males and Females

SMS item	Constrained					Freely Estimated (Males)					Freely Estimated (Females)				
	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4
Acquisitive self-monitoring															
5	1.03(.11)	-.45(.11)	.65(.12)	1.12(.15)	2.84(.31)	1.21(.17)	-.75(.14)	.24(.10)	.59(.13)	2.33(.30)	.87(.13)	-.13(.11)	1.17(.22)	1.86(.31)	3.53(.55)
8	2.39(.25)	-.70(.09)	.00(.07)	.48(.08)	1.49(.13)	2.39(--)	-.71(--)	.00(--)	.48(--)	1.49(--)	2.39(--)	-.71(--)	.00(--)	.48(--)	1.49(--)
12	1.07(.11)	-1.79(.19)	.01(.10)	.88(.13)	2.88(.31)	1.32(.18)	-1.48(.18)	.03(.10)	.91(.14)	2.53(.31)	.97(.14)	-1.94(.25)	.01(.13)	.82(.17)	3.07(.44)
18	2.09(.22)	.05(.08)	.63(.09)	.89(.10)	1.81(.16)	1.95(.26)	-.02(.08)	.66(.09)	.93(.11)	1.72(.17)	2.11(.28)	.11(.08)	.62(.10)	.88(.11)	2.00(.22)
20	1.44(.14)	-1.76(.16)	-.51(.09)	.05(.09)	1.37(.14)	1.44(.18)	-1.85(.21)	-.43(.10)	.10(.10)	1.53(.18)	1.45(.17)	-1.70(.18)	-.56(.11)	.02(.10)	1.24(.16)
22	1.11(.12)	-1.95(.20)	.06(.10)	.76(.12)	2.76(.28)	1.14(.16)	-2.09(.27)	.04(.11)	.81(.15)	2.87(.39)	1.08(.14)	-1.87(.23)	.09(.12)	.72(.15)	2.69(.37)
Protective self-monitoring															
2	.80(.11)	-0.85(.15)	2.40(.32)	3.04(.40)	5.94(.86)	.98(.17)	-.95(.20)	1.81(.30)	2.42(.39)	4.64(.83)	.64(.14)	-.81(.24)	3.23(.67)	3.90(.81)	7.90(1.82)
7	.62(.09)	-3.56(.52)	-1.86(.29)	-.54(.16)	3.51(.52)	.63(.15)	-3.62(.81)	-1.66(.40)	-.32(.21)	3.75(.84)	.58(.12)	-3.79(.80)	-2.17(.47)	-.78(.25)	3.51(.76)
13	1.34(.15)	-0.87(.12)	.31(.09)	.84(.11)	2.47(.25)	1.71(.24)	-.71(.12)	.22(.09)	.66(.11)	2.15(.24)	1.06(.19)	-1.09(.22)	.41(.14)	1.06(.20)	2.92(.48)
15	.91(.11)	-2.77(.32)	-.59(.13)	.33(.11)	3.78(.45)	.94(.17)	-2.49(.40)	-.24(.15)	.69(.18)	4.09(.72)	.89(.15)	-3.06(.51)	-.93(.22)	.02(.15)	3.6(.60)
16	1.62(.18)	-.91(.11)	.21(.08)	.70(.10)	2.30(.21)	2.12(.31)	-.78(.11)	.21(.08)	.60(.09)	2.14(.22)	1.29(.21)	-1.09(.21)	.20(.12)	.81(.16)	2.55(.38)
19	1.43(.15)	-.57(.10)	.94(.11)	1.64(.16)	3.37(.33)	1.43(--)	-.57(--)	.94(--)	1.63(--)	3.36(--)	1.43(--)	-0.57(--)	.94(--)	1.63(--)	3.36(--)
25	1.19(.13)	-1.3(.15)	.10(.09)	.66(.11)	3.33(.34)	1.04(.17)	-1.66(.26)	.11(.14)	.83(.17)	3.93(.63)	1.20(.19)	-1.24(.23)	.05(.13)	.55(.14)	3.20(.48)

Note. N for Sample 2 = 709; Males N = 303, Females N = 406. All items flagged for DIF in any one of the significance tests described in the main text are freely estimated across groups above.

IRT-BASED SUBSCALES OF THE SMS

Sample 1

Table S8

Constrained Item Parameters and DIF-Flagged Item Parameters Estimated Freely Across Age Groups (Below 40 and 40 and Above)

SMS item	Constrained			Freely Estimated (< 40)			Freely Estimated (≥ 40)		
	a	*c*	*b*	*a*	*c*	*b*	*a*	*c*	*b*
Acquisitive self-monitoring									
5	.98(.03)	-.20(.02)	.21(.02)	.99(.03)	-.19(.02)	.19(.02)	.80(.12)	-.33(.07)	.41(.12)
8	1.94(.06)	.29(.03)	-.15(.01)	1.94(--)	.29(--)	-.15(--)	1.94(--)	.29(--)	-.15(--)
12	1.12(.04)	-.41(.02)	.36(.02)	1.15(.04)	-.44(.02)	.38(.02)	.89(.14)	-.11(.07)	.12(.09)
18	1.49(.04)	-.19(.02)	.13(.02)	1.47(.04)	-.16(.02)	.11(.02)	1.71(.26)	-.48(.10)	.28(.07)
20	1.51(.05)	.68(.03)	-.45(.02)	1.51(--)	.68(--)	-.45(--)	1.51(--)	.68(--)	-.45(--)
22	.84(.03)	-.70(.02)	.83(.03)	.86(.03)	-.74(.02)	.86(.03)	.79(.13)	-.32(.07)	.41(.12)
Protective self-monitoring									
2	1.08(.04)	-1.46(.03)	1.36(.04)	1.08(--)	-1.46(--)	1.35(--)	1.08(--)	-1.46(--)	1.35(--)
7	.83(.03)	.73(.02)	-.88(.03)	.83(--)	.73(--)	-.88(--)	.83(--)	.73(--)	-.88(--)
13	1.53(.05)	.14(.03)	-.09(.02)	1.53(--)	.14(--)	-.09(--)	1.53(--)	.14(--)	-.09(--)
15	.80(.03)	.06(.02)	-.07(.02)	.80(.03)	.05(.02)	-.07(.02)	.87(.11)	.15(.08)	-.18(.09)
16	1.54(.05)	1.15(.03)	-.75(.02)	1.52(.05)	1.16(.03)	-.76(.02)	1.60(.20)	1.00(.13)	-.62(.06)
19	1.40(.05)	-1.51(.03)	1.08(.03)	1.40(.05)	-1.52(.03)	1.08(.03)	1.72(.23)	-1.32(.11)	.77(.11)
25	.92(.03)	.35(.02)	-.37(.02)	.92(--)	.35(--)	-.38(--)	.92(--)	.35(--)	-.38(--)

Note. *N* for Sample 1 = 13,563; *N*s for under 40 and 40 and above = 12,456, and 1,107, respectively. All items flagged for DIF in any one of the significance tests described in the main text are freely estimated across groups above.

IRT-BASED SUBSCALES OF THE SMS

Sample 2

Table S9

Constrained Item Parameters and DIF-Flagged Item Parameters Estimated Freely Across Age Groups (Below 40 and 40 and Above)

SMS item	Constrained					Freely Estimated (< 40)					Freely Estimated (≥ 40)				
	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4
Acquisitive self-monitoring															
5	1.14(.15)	-.39(.14)	.60(.19)	1.03(.23)	2.58(.40)	.91(.24)	-.26(.22)	.92(.30)	1.52(.42)	3.09(.80)	1.20(.13)	-.41(.08)	.54(.10)	.94(.12)	2.47(.26)
8	2.69(.33)	-.61(.13)	.02(.14)	.44(.16)	1.35(.23)	2.69(--)	-.61(--)	.02(--)	.44(--)	1.35(--)	2.69(--)	-.61(--)	.02(--)	.44(--)	1.35(--)
12	1.22(.15)	-1.56(.18)	.02(.15)	.79(.20)	2.55(.38)	1.63(.30)	-1.44(.25)	.09(.14)	.73(.17)	2.06(.35)	1.13(.12)	-1.60(.16)	.02(.09)	.83(.12)	2.73(.29)
18	2.34(.29)	.06(.15)	.57(.18)	.81(.19)	1.64(.26)	2.24(.47)	-.13(.12)	.44(.12)	.70(.14)	1.74(.27)	2.33(.25)	.10(.05)	.61(.07)	.85(.08)	1.63(.13)
20	1.62(.21)	-1.54(.16)	-.43(.13)	.07(.15)	1.24(.25)	1.62(--)	-1.54(--)	-.43(--)	.07(--)	1.24(--)	1.62(--)	-1.54(--)	-.43(--)	.07(--)	1.24(--)
22	1.25(.17)	-1.71(.18)	.07(.16)	.69(.21)	2.47(.40)	1.76(.32)	-1.68(.26)	.19(.13)	.61(.15)	1.71(.28)	1.15(.12)	-1.75(.16)	.05(.09)	.74(.11)	2.75(.29)
Protective self-monitoring															
2	.77(.11)	-.58(.18)	2.77(.39)	3.43(.48)	6.44(.98)	.77(--)	-.58(--)	2.79(--)	3.45(--)	6.47(--)	.77(--)	-.58(--)	2.79(--)	3.45(--)	6.47(--)
7	.60(.10)	-3.38(.57)	-1.62(.32)	-.26(.18)	3.91(.61)	.60(--)	-3.40(--)	-1.63(--)	-.25(--)	3.93(--)	.60(--)	-3.40(--)	-1.63(--)	-.25(--)	3.93(--)
13	1.29(.16)	-.61(.15)	.62(.14)	1.16(.17)	2.86(.33)	1.29(--)	-.61(--)	.62(--)	1.16(--)	2.86(--)	1.29(--)	-.61(--)	.62(--)	1.16(--)	2.86(--)
15	.90(.12)	-2.53(.36)	-.31(.15)	.64(.15)	4.15(.54)	.78(.24)	-2.46(.75)	.23(.26)	1.14(.38)	4.74(1.42)	.90(.11)	-2.63(.33)	-.43(.13)	.55(.11)	4.13(.45)
16	1.57(.18)	-.64(.14)	.51(.13)	1.02(.16)	2.67(.30)	1.61(.36)	-.62(.19)	.74(.18)	.98(.20)	2.99(.59)	1.53(.17)	-.68(.10)	.45(.08)	1.03(.10)	2.67(.22)
19	1.37(.17)	-.29(.13)	1.27(.18)	1.99(.24)		2.19(.56)	-.04(.13)	1.13(.19)	1.47(.23)		1.26(.14)	-.39(.10)	1.28(.12)	2.11(.18)	
25	1.13(.14)	-1.07(.18)	.40(.14)	.99(.17)	3.79(.47)	1.13(--)	-1.06(--)	.40(--)	.98(--)	3.78(--)	1.13(--)	-1.06(--)	.40(--)	.98(--)	3.78(--)

Note. N for Sample 2 = 709; <40 N = 119, ≥40 N = 590. Due to a lack of responses on the highest response option for SMS 19 in the <40 group, scores were collapsed into the next-lowest response option within both groups. All items flagged for DIF in any one of the significance tests described in the main text are freely estimated across groups above.

IRT-BASED SUBSCALES OF THE SMS

Table S10

Scale Descriptive Statistics for Adulthood, Adulthood, and Later Life Age Groups

Measure		Emerging Adulthood	Adulthood	Later Life	Standardized Between-Group Differences		
					$d_{(1-2)}$	$d_{(1-3)}$	$d_{(2-3)}$
Acquisitive self-monitoring							
Sample 1	*M*	2.87	2.74	2.90	.07	-.01	-.09
	SD	1.76	1.81	1.65			
	N	10,643	2,843	77			
Sample 2	*M*	3.59	3.00	2.85	.54	.73	.14
	SD	1.08	1.09	1.00			
	N	20	491	198			
Protective self-monitoring							
Sample 1	*M*	3.52	2.88	2.87	.36	.37	.01
	SD	1.77	1.81	1.77			
	N	10,643	2,843	77			
Sample 2	*M*	3.21	3.55	3.56	-.34	-.40	-.02
	SD	.71	1.00	.90			
	N	20	491	198			

Note. Emerging Adulthood (i.e., 18 to 30 years), Adulthood (31 to 60 years), and Later Life (61+ years).

IRT-BASED SUBSCALES OF THE SMS

Table S11

DIF Analyses Across Emerging Adulthood, Adulthood, and Later Life Age Groups

SMS item	Sample 1							Sample 2			
	OLR-LR χ^2	IRT-LR χ^2	M-H χ^2	ΔR^2	λMH	JG	ETS	OLR-LR χ^2	IRT-LR χ^2	ΔR^2	JG
Acquisitive self-monitoring											
5	24.53(4)*	18.76(4)*	22.26(2)*	<.01	--	A	--	.81(2)	5.27(5)	<.01	A
8	15.23(4)*	7.32(4)	12.87(2)*	<.01	--	A	--	11.98(2)*	1.29(5)	<.01	A
12	42.59(4)*	21.15(4)*	36.93(2)*	<.01	--	A	--	2.02(2)	10.82(5)	<.01	A
18	24.10(4)*	14.46(4)*	24.48(2)*	<.01	--	A	--	4.27(2)	5.07(5)	<.01	A
20	1.50(4)	Referent	.37(2)	<.01	--	A	--	5.36(2)	Referent	<.01	A
22	51.33(4)*	26.83(4)*	48.76(2)*	<.01	--	A	--	.96(2)	6.62(5)	<.01	A
Protective self-monitoring											
2	16.11(4)*	9.65(4)*	2.88(2)	<.01	--	A	--	3.91(2)	2.90(5)	<.01	A
7	.49(4)	2.73(4)	.10(2)	<.01	--	A	--	.76(2)	8.16(5)	<.01	A
13	3.09(4)	Referent	.48(2)	<.01	--	A	--	.72(2)	Referent	<.01	A
15	8.90(4)	11.99(4)*	7.65(2)*	<.01	--	A	--	17.03(2)*	15.68(5)*	.02	A
16	18.00(4)*	8.67(4)	14.07(2)*	<.01	--	A	--	.37(2)	13.98(5)*	<.01	A
19	5.39(4)	2.48(4)	3.96(2)	<.01	--	A	--	.94(2)	3.66(5)	<.01	A
25	4.66(4)	1.70(4)	2.75(2)	<.01	--	A	--	8.80(2)*	6.11(5)	<.01	A

Note. For Sample 1, *N*s for under Emerging Adulthood (i.e., 18 to 30 years), Adulthood (i.e., 31 to 60 years), and Later Life (i.e., 61+ years) age groups = 10,643, 2,843, and 77 respectively. For Sample 2, respective *N*s = 20, 491, and 198. Due to its small sample size, the emerging adult subgroup was merged with the adult subgroup for DIF and sensitivity analyses in Sample 2.

Parameter estimates followed by *df* in parentheses for all significance tests; α = .05. OLR-LR χ^2 = Generalized 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 = Generalized Mantel-Haenszel chi-square; ΔR^2 = change in Nagelkerke R^2 above baseline IRT-LR model with inclusion of group and group x sum-score interaction terms; λMH = common log-odds ratio effect size (not included, as effect size indicators for the generalized Mantel-Haenszel are still in need of development); 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 λMH ('A' = negligible; 'B' = moderate; 'C' = large).

* = Statistically significant.

IRT-BASED SUBSCALES OF THE SMS

Sample 1

Table S12

Constrained Item Parameters and DIF-Flagged Item Parameters Estimated Freely Across Age Groups (18-30, 31-60, and 61+)

SMS item	Constrained		Freely Estimated (18-30)		Freely Estimated (31-60)		Freely Estimated (61+)	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*
Acquisitive self-monitoring								
5	1.04(.04)	.29(.03)	1.03(.04)	.25(.03)	1.09(.07)	.43(.05)	.11(.36)	4.06(13.39)
8	2.06(.08)	-.04(.02)	2.00(.08)	-.06(.02)	2.18(.15)	.02(.03)	3.24(1.89)	-.09(.19)
12	1.18(.04)	.44(.03)	1.25(.05)	.47(.03)	1.06(.07)	.30(.04)	.78(.49)	-.01(.35)
18	1.58(.05)	.22(.03)	1.56(.06)	.19(.02)	1.63(.10)	.32(.03)	2.96(1.61)	-.03(.19)
20	1.60(.06)	-.33(.03)	1.60(--)	-.33(--)	1.60(--)	-.33(--)	1.60(--)	-.33(--)
22	.89(.04)	.88(.04)	.93(.04)	.92(.04)	.85(.06)	.66(.06)	.62(.45)	.54(.56)
Protective self-monitoring								
2	1.16(.05)	1.62(.05)	1.24(.06)	1.55(.04)	.91(.08)	1.98(.14)	1.22(.60)	2.04(.66)
7	.90(.04)	-.44(.04)	.90(--)	-.44(--)	.90(--)	-.44(--)	.90(--)	-.44(--)
13	1.65(.06)	.29(.03)	1.65(--)	.28(--)	1.65(--)	.28(--)	1.65(--)	.28(--)
15	.88(.04)	.30(.03)	.86(.04)	.27(.03)	.87(.07)	.44(.05)	.72(.34)	-.05(.37)
16	1.64(.06)	-.32(.03)	1.70(.08)	-.32(.03)	1.51(.10)	-.31(.04)	.63(.32)	-.15(.42)
19	1.52(.07)	1.36(.04)	1.52(--)	1.37(--)	1.52(--)	1.37(--)	1.52(--)	1.37(--)
25	1.00(.04)	.02(.03)	.99(.04)	.02(.03)	1.06(.07)	.01(04)	.62(.31)	.59(.46)

*Note. N*s for under Emerging Adulthood (i.e., 18 to 30 years), Adulthood (i.e., 31 to 60 years), and Later Life (i.e., 61+ years) age groups = 10,643, 2,843, and 77, respectively. Given the small sample size for the Later Life group, we encourage caution when interpreting results from IRT models freely estimated within this subgroup.

IRT-BASED SUBSCALES OF THE SMS

Sample 2

Table S13

Constrained Item Parameters and DIF-Flagged Item Parameters Estimated Freely Across Age Groups (18-30, 31-60, and 61+)

SMS items	Constrained					Freely Estimated (18-30)					Freely Estimated (31-60)					Freely Estimated (61+)				
	a	b_1	b_2	b_2	b_4	a	b_1	b_2	b_2	b_4	a	b_1	b_2	b_2	b_4	a	b_1	b_2	b_2	b_4
Acquisitive self-monitoring																				
5	1.14	-0.24	0.75	1.18	2.73	See Note					1.15	-0.31	0.69	1.07	2.64	1.01	-0.04	1.05	1.67	3.32
8	2.66	-0.47	0.16	0.59	1.50						2.57	-0.47	0.20	0.63	1.55	3.32	-0.45	0.10	0.49	1.39
12	1.22	-1.42	0.17	0.94	2.71						1.37	-1.39	0.08	0.80	2.51	0.79	-1.64	0.59	1.68	3.82
18	2.30	0.20	0.73	0.96	1.80						2.52	0.19	0.73	0.96	1.78	1.91	0.28	0.74	1.00	1.88
20	1.61	-1.40	-0.29	0.21	1.39						1.61	-1.40	-0.29	0.21	1.39	1.61	-1.40	-0.29	0.21	1.39
22	1.24	-1.57	0.22	0.84	2.63						1.30	-1.65	0.16	0.77	2.46	1.01	-1.46	0.46	1.18	3.48
Protective self-monitoring																				
2	0.86	-0.70	2.32	2.91	5.61	See Note					0.90	-0.78	2.14	2.74	5.38	0.74	-0.48	2.98	3.55	6.53
7	0.67	-3.21	-1.63	-0.41	3.33						0.67	-3.21	-1.63	-0.40	3.34	0.67	-3.21	-1.63	-0.40	3.34
13	1.45	-0.72	0.37	0.86	2.37						1.45	-0.72	0.37	0.86	2.37	1.45	-0.72	0.37	0.86	2.37
15	1.01	-2.43	-0.45	0.39	3.52						1.09	-2.09	-0.25	0.47	3.41	0.77	-4.17	-1.19	0.15	4.13
16	1.77	-0.75	0.28	0.73	2.20						1.87	-0.76	0.31	0.67	2.13	1.42	-0.76	0.20	0.97	2.60
19	1.53	-0.44	0.96	1.61	3.22						1.53	-0.44	0.96	1.61	3.22	1.53	-0.44	0.96	1.61	3.22
25	1.27	-1.12	0.18	0.71	3.20						1.22	-1.27	0.10	0.62	3.39	1.54	-0.75	0.35	0.89	2.61

Note. N for Sample 2 = 709; Emerging Adulthood and Adulthood N = 511, Later life N = 198. Standard errors of parameter estimates are available from the authors upon request. Due to its small sample size, the Emerging Adult subgroup was merged with the Adult subgroup for DIF and sensitivity analyses in Sample 2. Results for this combined group are reported under the 31-60 category above.

IRT-BASED SUBSCALES OF THE SMS

Supplementary Figures

Sample 1
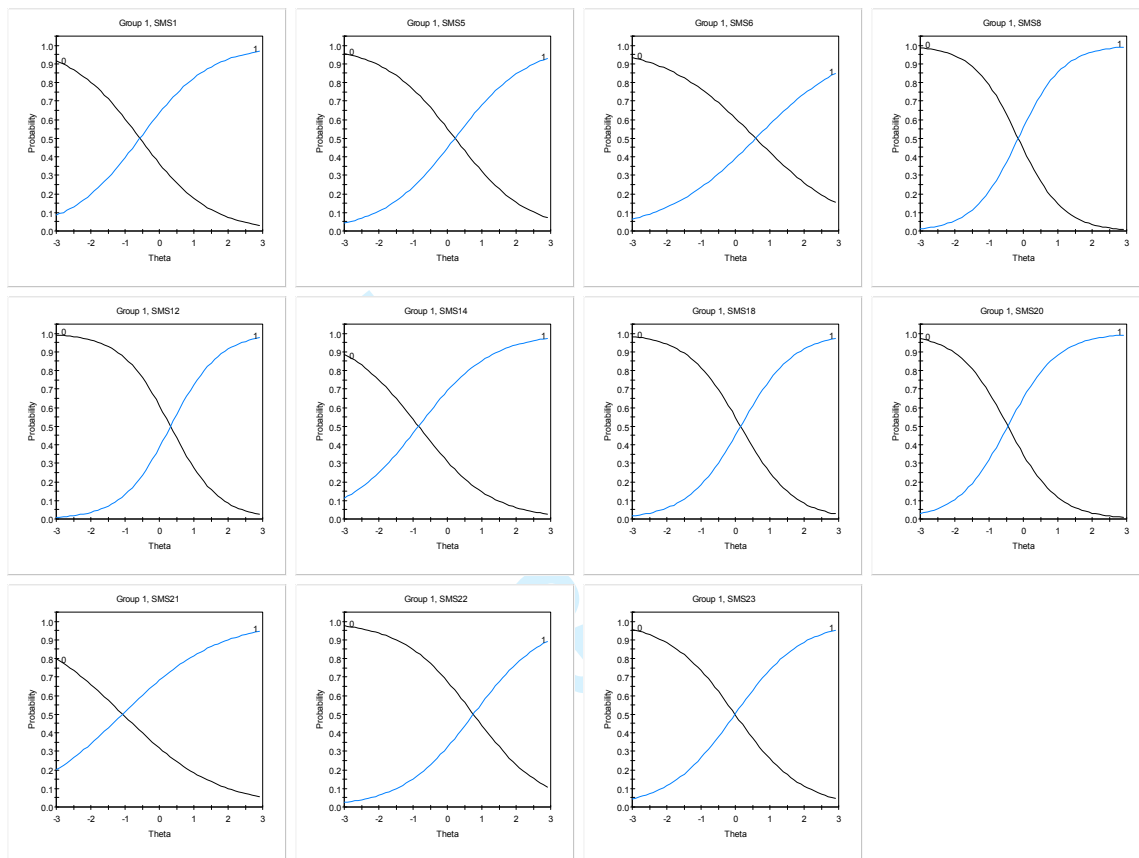
Figure S1. Item characteristic curves (ICC) of the 11 candidate items for the acquisitive self-monitoring scale. Sample 1 (*N* = 13,563) items scored using a dichotomous response format.
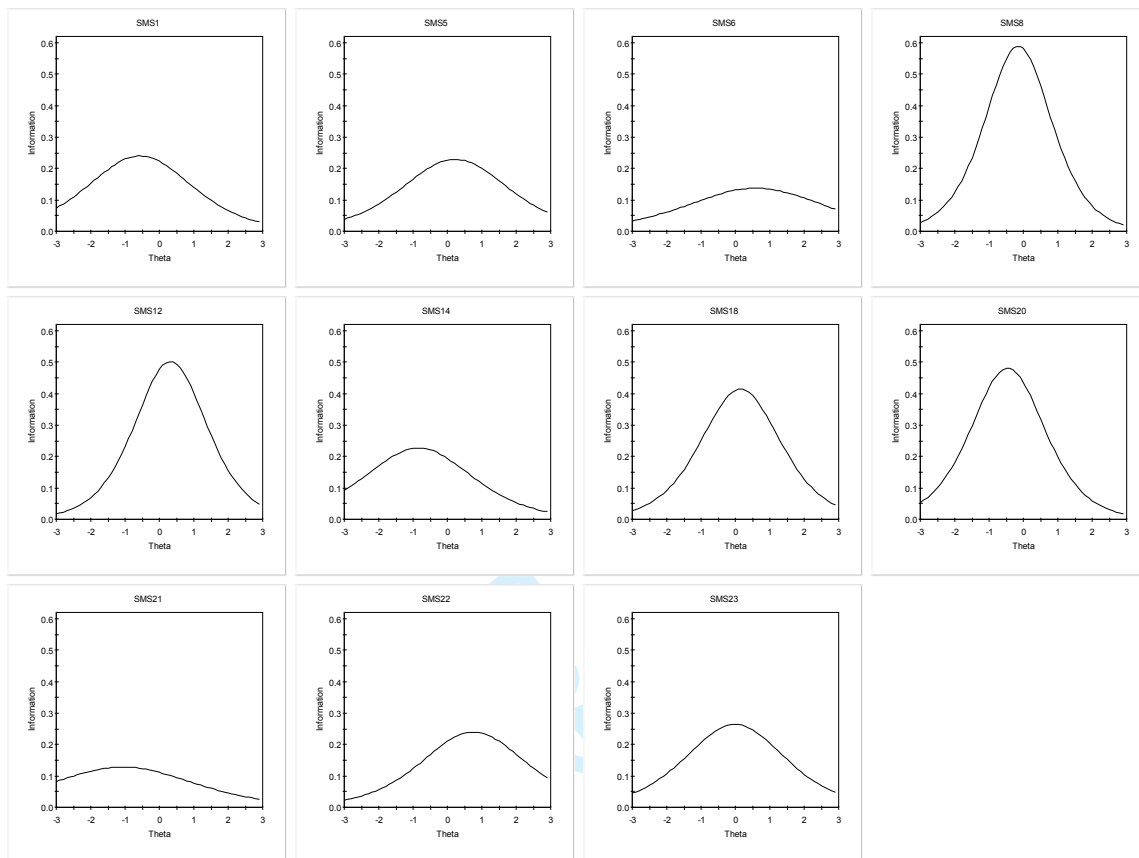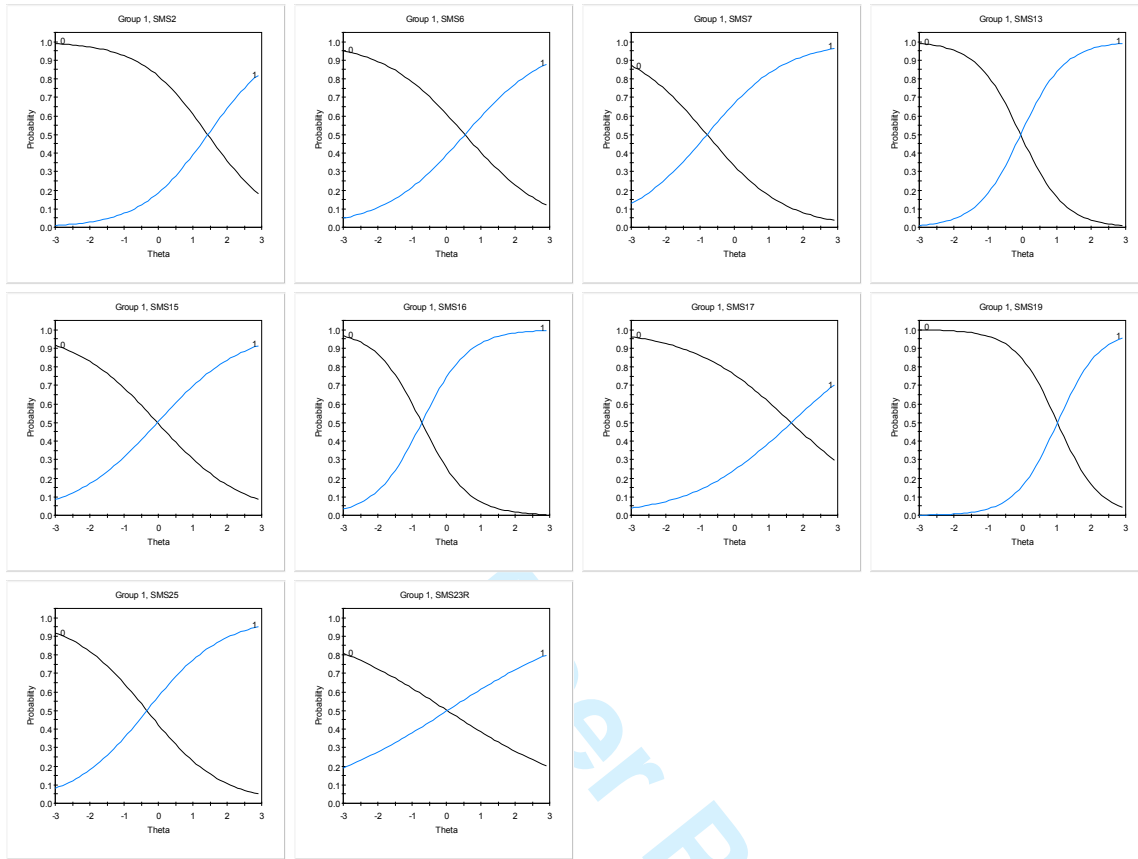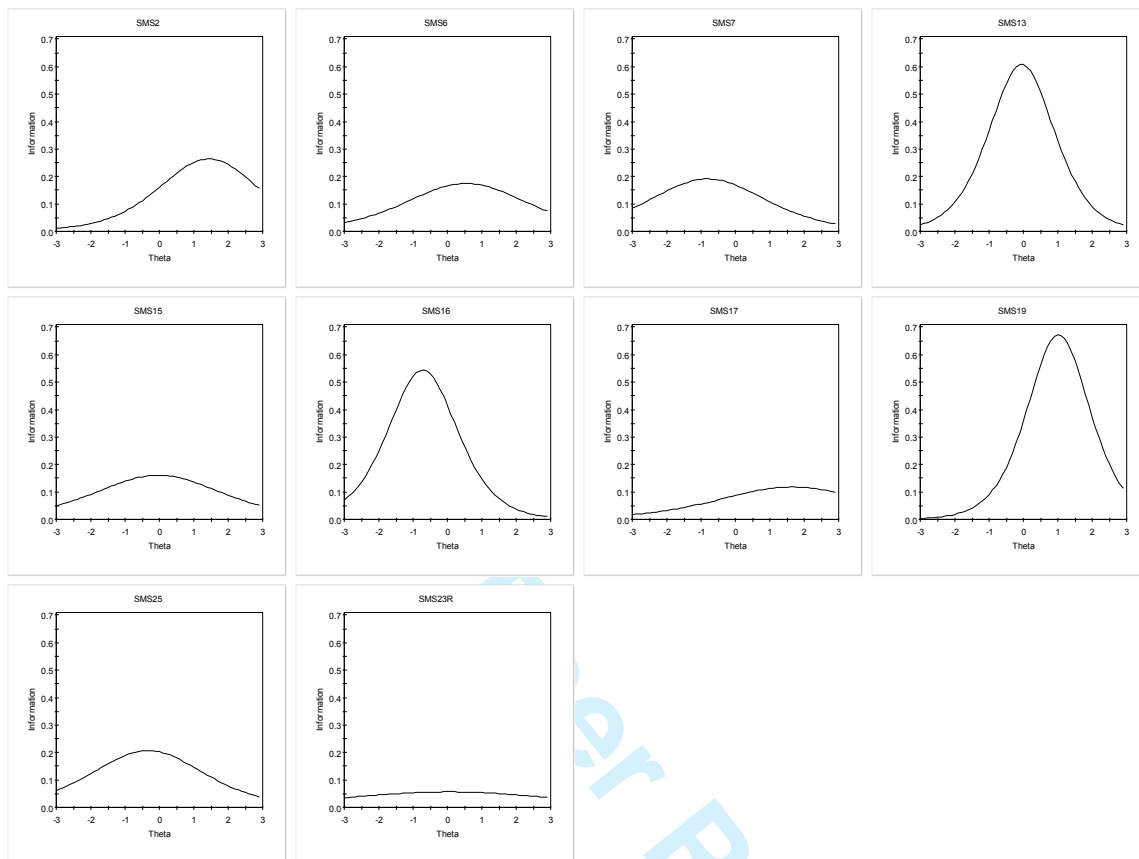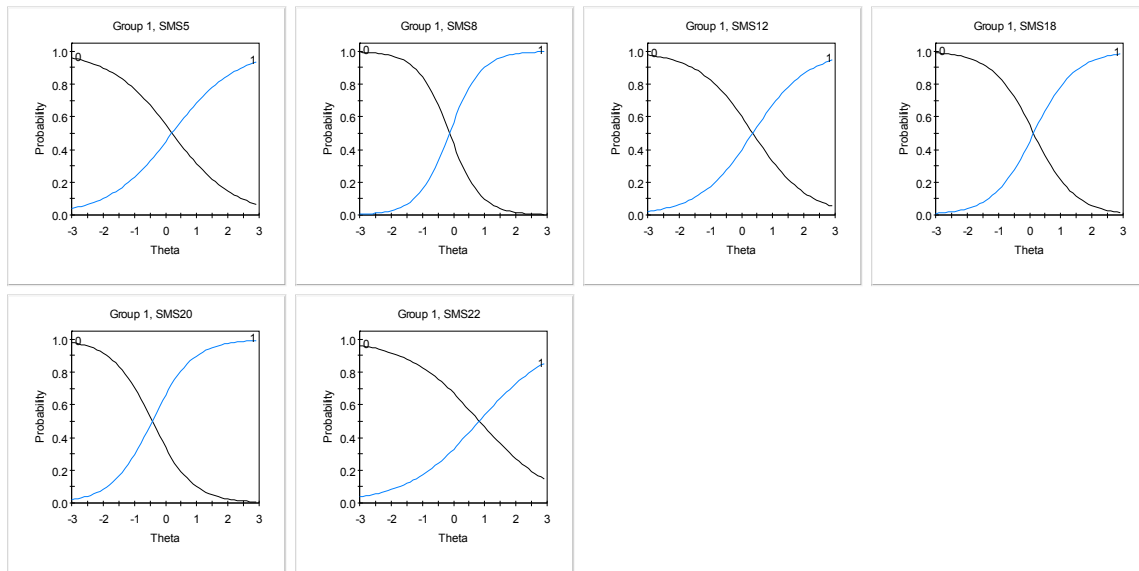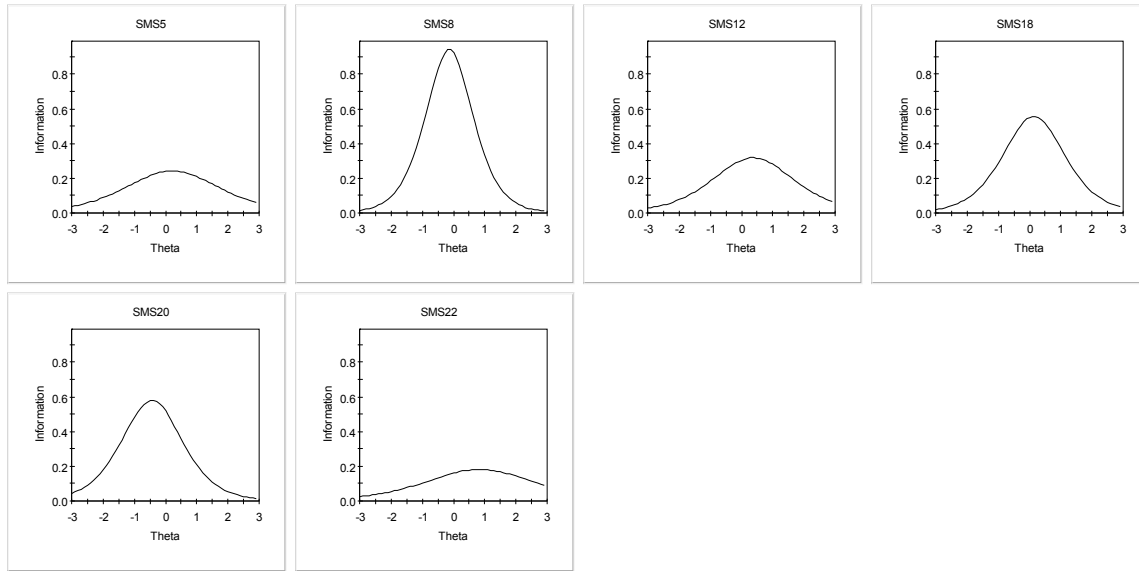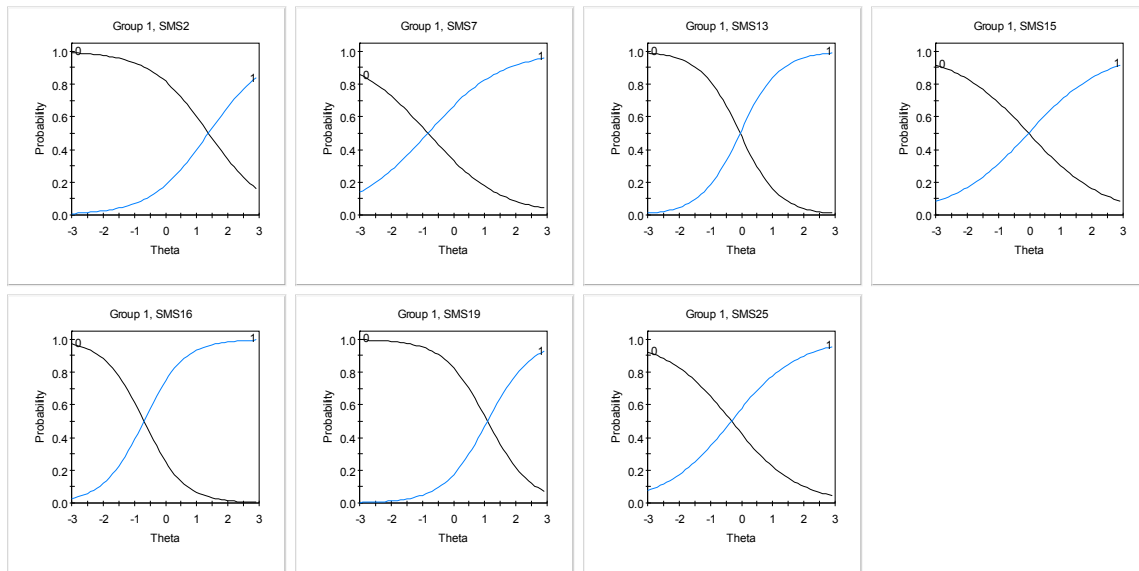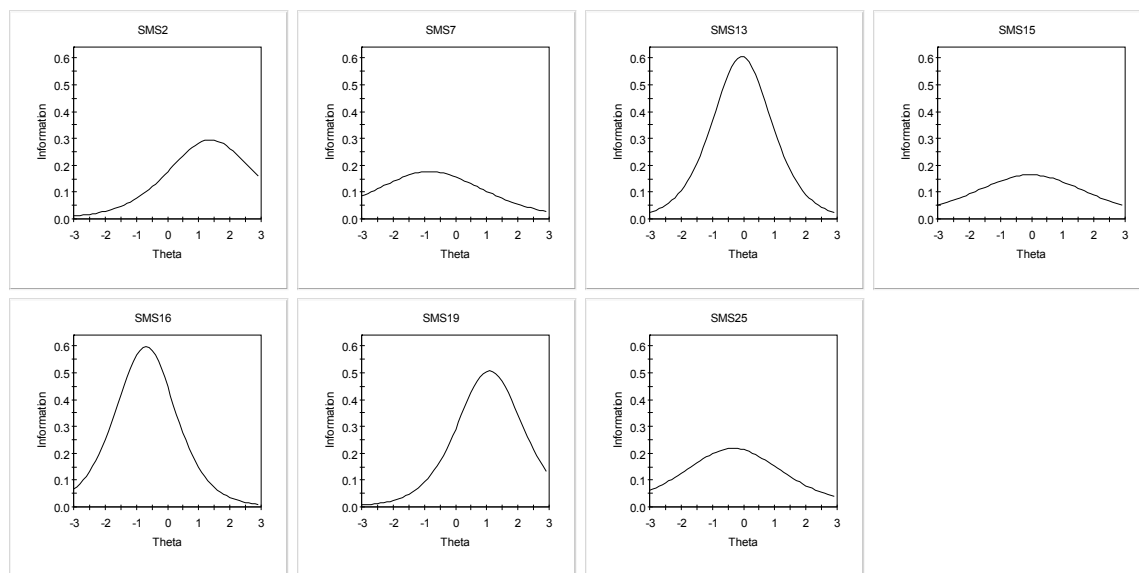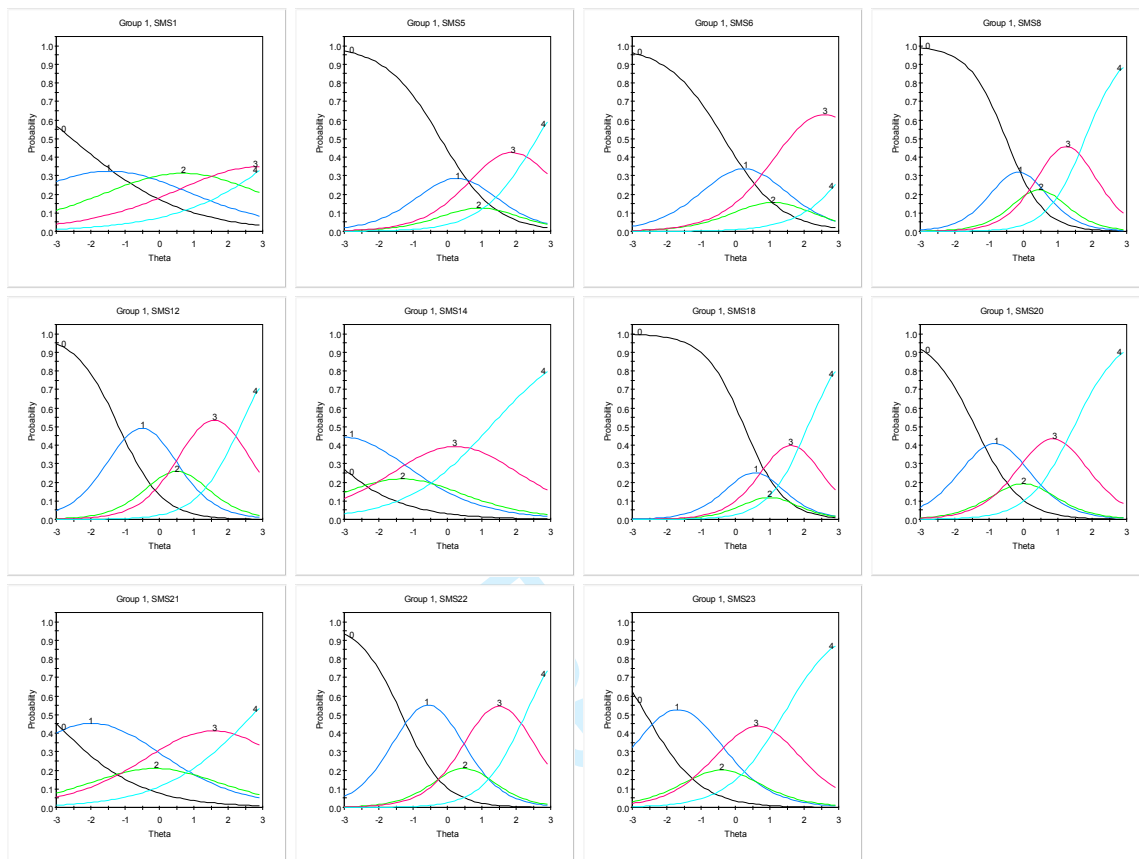
Sample 1

Figure S2. Item information functions (IIF) of the 11 candidate items for the acquisitive self-monitoring scale. Sample 1 (*N* = 13,563) items scored using a dichotomous response format.

Sample 1

Figure S3. Item characteristic curves (ICC) of the 10 candidate items for the protective self-monitoring scale. Sample 1 (*N* = 13,563) items scored using a dichotomous response format.
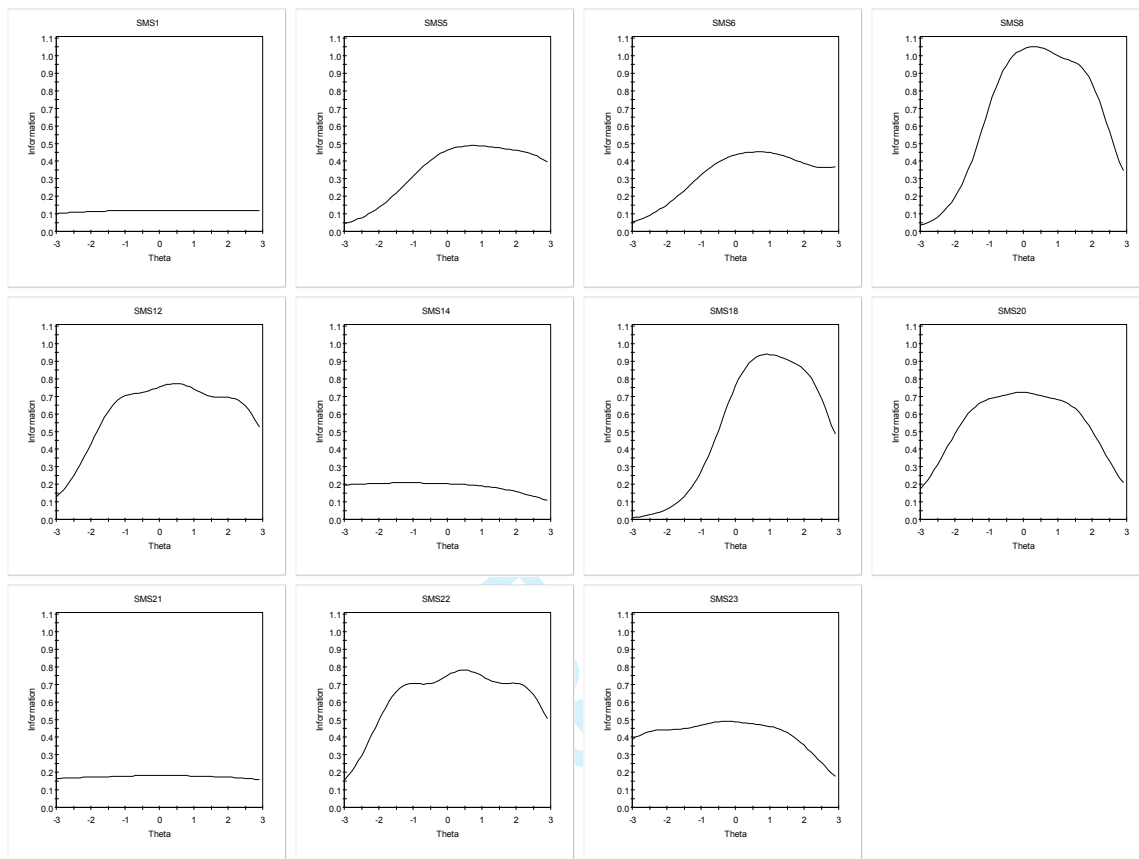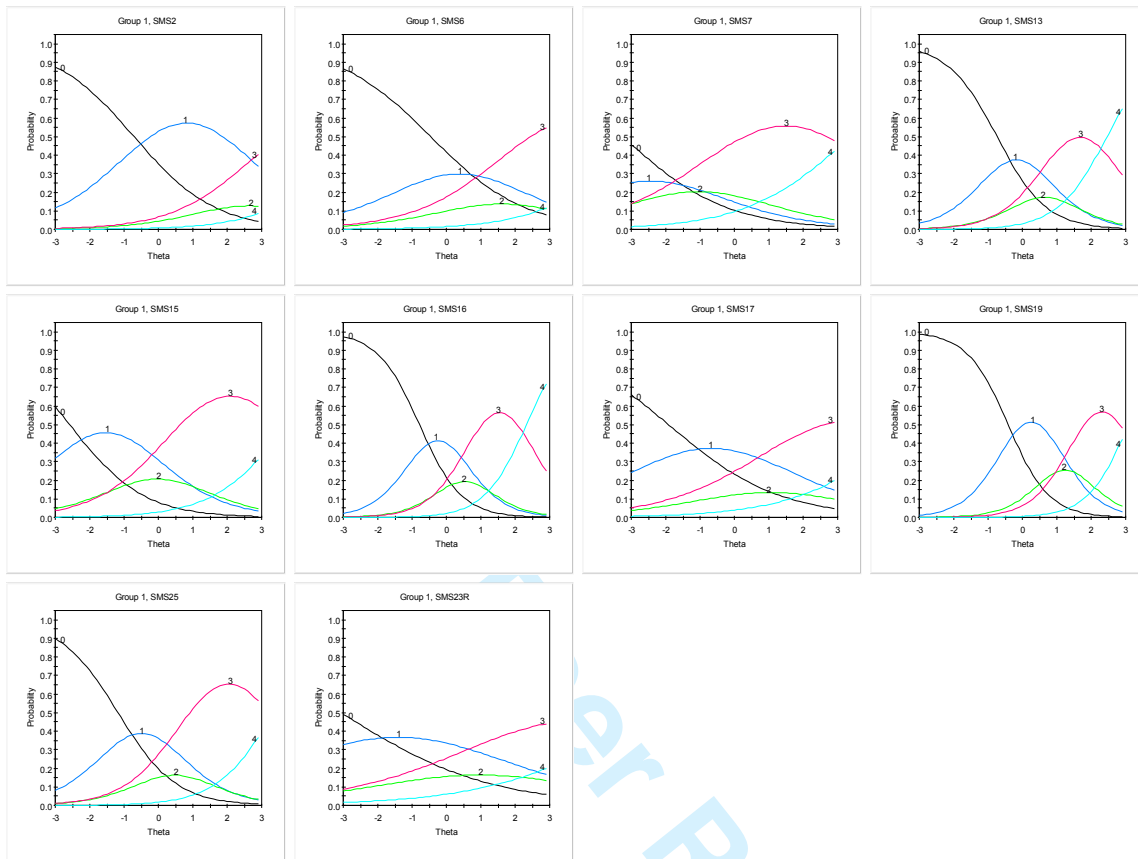
Sample 1

Figure S4. Item information functions (IIF) of the 10 candidate items for the protective self-monitoring scale. Sample 1 ($N = 13,563$) items scored using a dichotomous response format.

Sample 1

Figure S5. Item characteristic curves (ICC) of the 6 final items included in the acquisitive self-monitoring scale. Sample 1 (N = 13,563) items scored using a dichotomous response format.

Sample 1

Figure S6. Item information functions (IIF) of the 6 final items included in the acquisitive self-monitoring scale. Sample 1 ($N = 13,563$) items scored using a dichotomous response format.

Sample 1

Figure S7. Item characteristic curves (ICC) of the 7 final items included in the protective self-monitoring scale. Sample 1 (*N* = 13,563) items scored using a dichotomous response format.

Sample 1

Figure S8. Item information functions (IIF) of the 7 final items included in the protective self-monitoring scale. Sample 1 (*N* = 13,563) items scored using a dichotomous response format.

Sample 2

Figure S9. Item characteristic curves (ICC) of the 11 candidate items for the acquisitive self-monitoring scale. Sample 2 (*N* = 709) items scored using a polytomous response format.
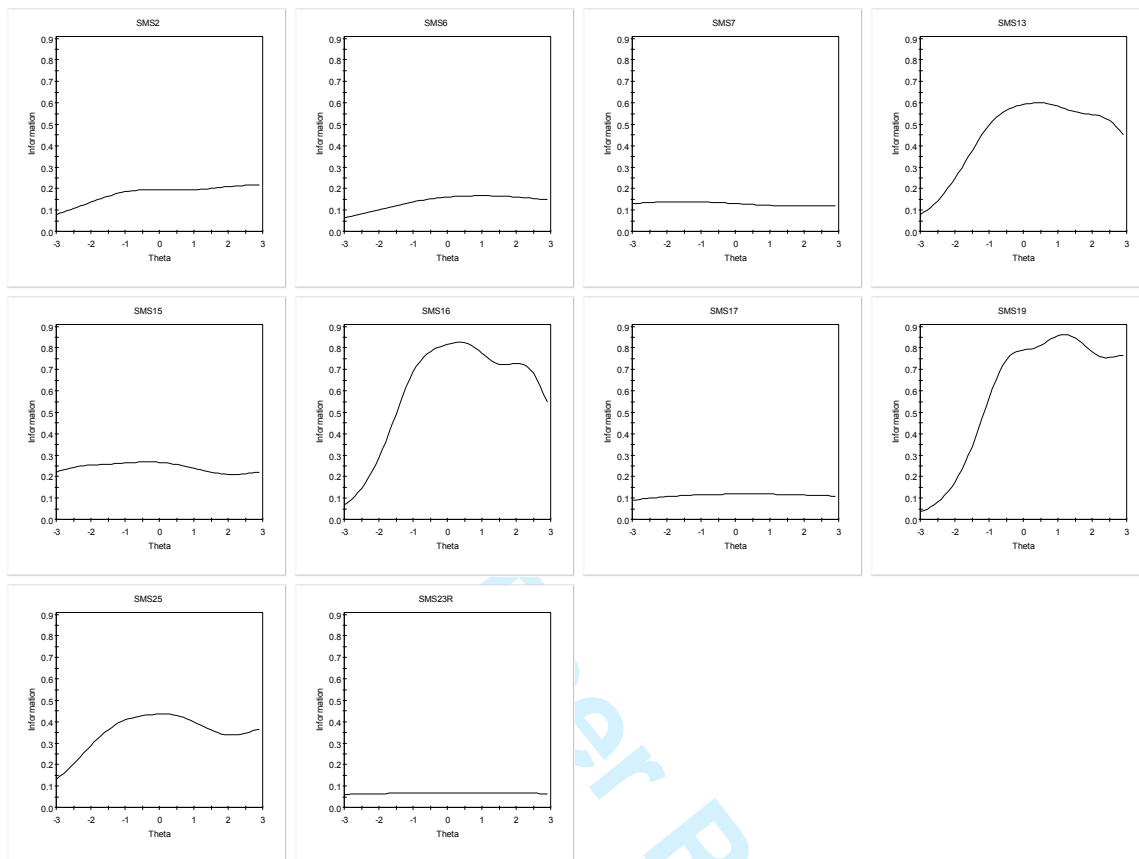
Sample 2

Figure S10. Item information functions (IIF) of the 11 candidate items for the acquisitive self-monitoring scale. Sample 2 (*N* = 709) items scored using a polytomous response format.
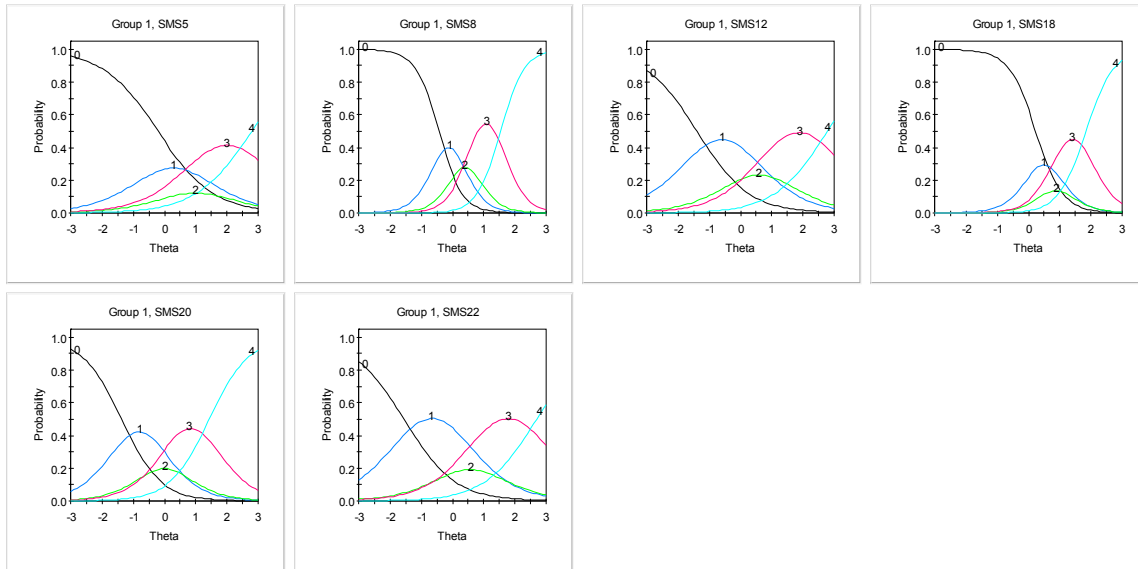
Sample 2

Figure S11. Item characteristic curves (ICC) of the 10 candidate items for the protective self-monitoring scale. Sample 2 (*N* = 709) items scored using a polytomous response format.

Sample 2

Figure S12. Item information functions (IIF) of the 10 candidate items for the protective self-monitoring scale. Sample 2 (*N* = 709) items scored using a polytomous response format.

Sample 2

Figure S13. Item characteristic curves (ICC) of the 6 final items included in the acquisitive self-monitoring scale. Sample 2 (*N* = 709) items scored using a polytomous response format.
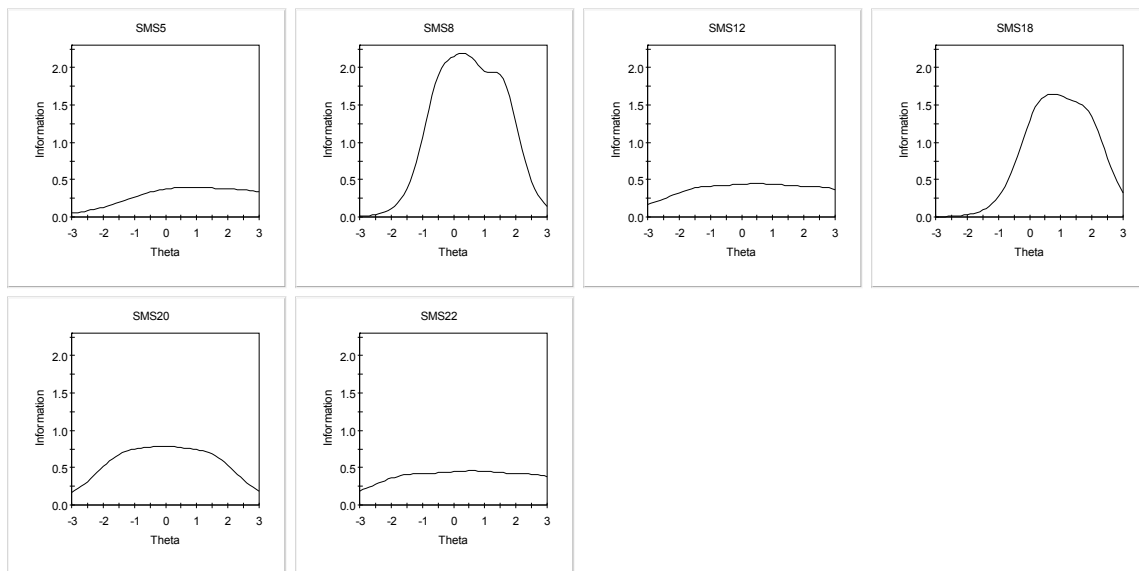
Sample 2

Figure S14. Item information functions (IIF) of the 6 final items included in the acquisitive self-monitoring scale. Sample 2 ($N = 709$) items scored using a polytomous response format.
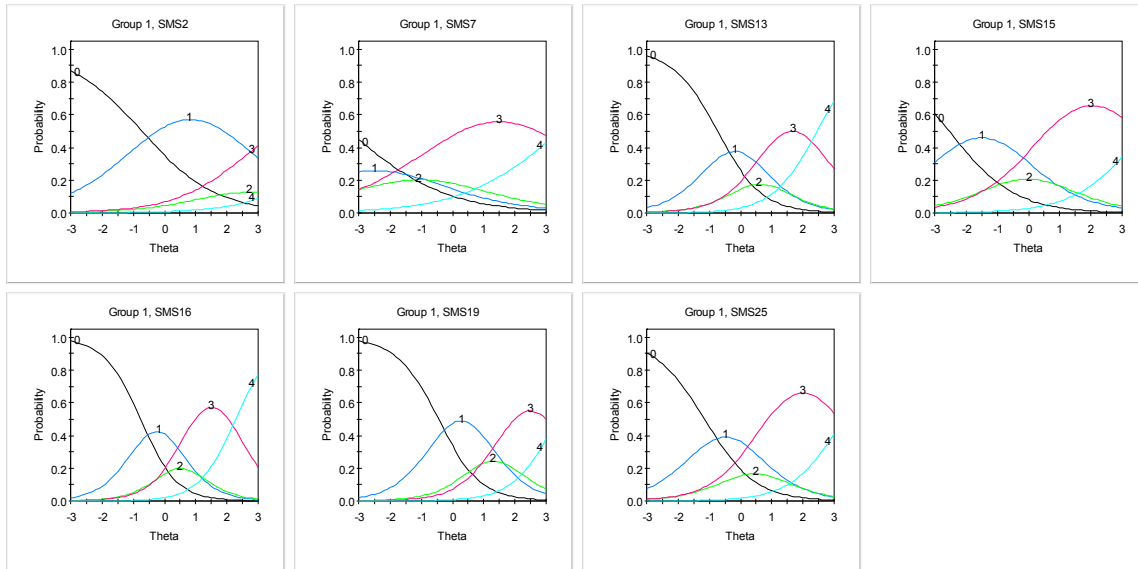
Sample 2

Figure S15. Item characteristic curves (ICC) of the 7 final items included in the protective self-monitoring scale. Sample 2 (*N* = 709) items scored using a polytomous response format.
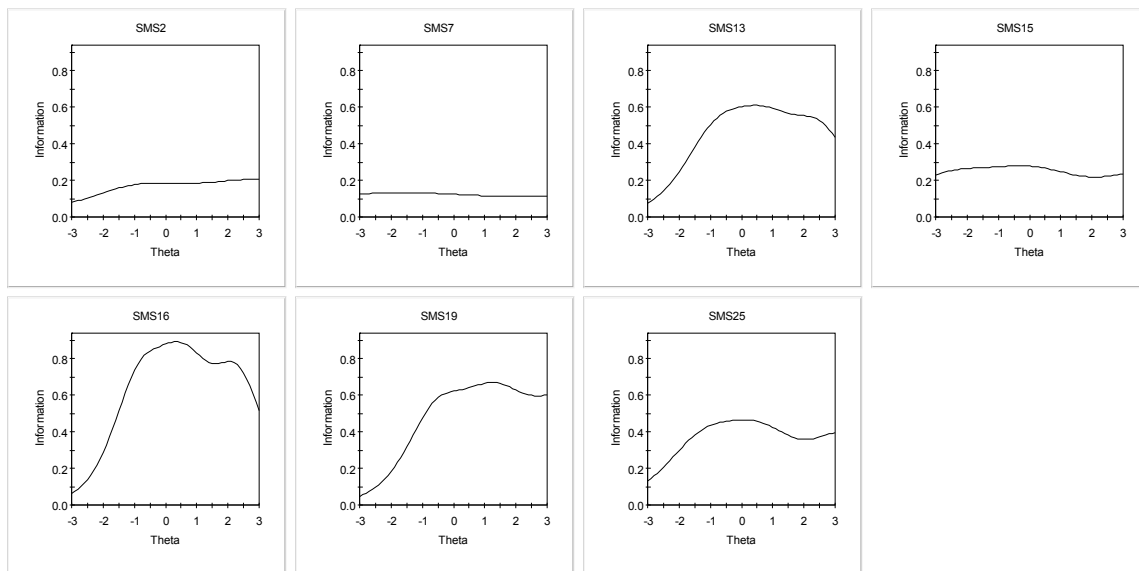
Sample 2

Figure S16. Item information functions (IIF) of the 7 final items included in the protective self-monitoring scale. Sample 2 (*N* = 709) items scored using a polytomous response format.