The self-reflection and insight scale: applying item response theory to craft an efficient short form

By: Paul J. Silvia

Silvia, P.J. The self-reflection and insight scale: applying item response theory to craft an efficient short form. *Current Psychology* (2021). https://doi.org/10.1007/s12144-020-01299-7

Abstract:

The human ability for self-consciousness—the capacity to reflect on oneself and to think about one's thoughts, experiences, and actions—is central to understanding personality and motivation. The present research examined the psychometric properties of the Self-reflection and Insight Scale (SRIS), a prominent self-report scale for measuring individual differences in private self-consciousness. Using tools from Rasch and item response theory models, the SRIS was evaluated using responses from a large sample of young adults (n = 1192). The SRIS had many strengths, including essentially zero gender-based differential item functioning (DIF), but a cluster of poor performing items was identified based on item misfit, high local dependence, and low item difficulty and discrimination. Based on the IRT analyses, a concise 12-item scale, evenly balanced between self-reflection and insight, was crafted. The short SRIS showed strong dimensionality, reliability, item fit, and local independence as well as essentially no gender DIF. Taken together, the many psychometric strengths of the SRIS support its popularity, and the short form will be useful for research and applied contexts where an efficient, concise version is needed.

Keywords: Self-reflection | Insight | Self-awareness | Private self-consciousness | Psychometrics | Item response theory

Article:

The human ability for self-awareness—the capacity to reflect on oneself and to think about one's thoughts, experiences, and actions—is central to understanding personality and motivation (Carver, 2003; Carver & Scheier, 1990; Duval & Silvia, 2001; Silvia & Duval, 2001). In personality psychology, individual differences in self-focused attention have been explored for nearly 50 years. The self-consciousness scales (Fenigstein et al., 1975), the first tools for measuring dispositional self-awareness, contained a scale measuring private self-consciousness, a tendency to focus on inner states. The original and revised (Scheier & Carver, 1985) private self-consciousness scales have been widely used (Buss, 1980; Smári et al., 2008). A challenge for researchers, however, was the scale's inconsistent factor structure. The private self-consciousness scale often split into two facets, each with different patterns of relationships (Anderson et al., 1996; Bernstein et al., 1986; Britt, 1992; Cramer, 2000; Nystedt &

Ljungberg, 2002), but the items composing the facets varied as well, so the scale's factor structure was difficult to pin down.

In the second generation of research on dispositional self-awareness, researchers developed new self-report scales from scratch (Smári et al., 2008). One prominent scale is the Rumination-Reflection Questionnaire (RRQ; Trapnell & Campbell, 1999), which distinguishes between ostensibly adaptive and dysfunctional forms of self-focused attention. At around the same time, Grant et al. (2002) developed the Self-reflection and Insight Scale (SRIS), a refined measure of private self-consciousness. This popular scale is widely used as an alternative to the original private self-consciousness scale, and it been translated into several languages (Aşkun & Cetin, 2017; S. Y. Chen et al., 2016; DaSilveira et al., 2015; Naeimi et al., 2019; Sauter et al., 2010) and modified for younger ages (Sauter et al., 2010).

The SRIS comes from a coaching tradition and is grounded in models of metacognition and personal development (Grant, 2001, 2003). It proposes that private self-consciousness can be represented with two factors. *Self-reflection* is defined as "the inspection and evaluation of one's thoughts, feelings, and behavior" (Grant et al., 2002, p. 821). *Insight*, in contrast, is defined as "the clarity of understanding of one's thoughts, feelings, and behavior" (Grant et al., 2002, p. 821). Table 1 displays the items and their features. The SRIS contains 20 items: 12 for self-reflection, 8 for insight. The self-reflection scale can be further divided into correlated facets that emphasize "engagement in self-reflection" (sr1-sr6 in Table 1) and "need for self-reflection" (sr7-sr12), but most research emphasizes the overall self-reflection and insight scores. Consistent with its underlying metacognitive model, the items are relatively neutral and emphasize metacognitive goals and experiences, such as wanting and trying to understand one's thoughts and experiencing inner states as clear or confusing. This metacognitive emphasis distinguishes the SRIS from other scales that are more emotionally charged, such as the RRQ (Silvia et al., 2005; Trapnell & Campbell, 1999).

In contrast to the private self-consciousness scale, the proposed factor structure of the SRIS has been consistently supported by exploratory and confirmatory factor analysis (Grant et al., 2002; Roberts & Stark, 2008). The correlation between the self-reflection and insight subscales varies across studies but is usually modest. Typical correlations are within $r = \pm.10$ (e.g., Lyke, 2009; Stein & Grant, 2014), although somewhat larger positive (e.g., $r = .25$; Cowden & Meyer-Weitz, 2016) and negative ($r = -.31$; Grant et al., 2002) correlations have been found. One reason for the popularity of the scales has been their differentiated relationships with other outcomes, particularly markers of well-being, resilience, and psychopathology symptoms (Cowden & Meyer-Weitz, 2016; Harrington & Loffredo, 2010; Lyke, 2009; Nakajima et al., 2017, 2018, 2019; Silvia & Phillips, 2011; Stein & Grant, 2014). Viewed broadly, self-reflection tends to have negative correlations with markers of well-being, whereas insight has positive correlations.

Table 1. Self-reflection and insight scale: item-level statistics

Item	Text	*M* (*SD*)	*a* (slope)	*b* (difficulty)	Infit	Outfit	RMSD	Local dependence
Self-reflection scale								
sr1	I don't often think about my thoughts (R)	5.23 (1.53)	.30	−1.51	1.01	1.14*	.056*	sr2, sr4
sr2	I rarely spend time in self-reflection (R)	5.02 (1.49)	.49	−1.18	1.02	1.06	.042	sr1, sr4
sr3	I frequently examine my feelings	4.76 (1.43)	.93	−.63	1.02	1.04	.033	
sr4	I don't really think about why I behave in the way that I do (R)	5.05 (1.53)	.37	−1.17	1.01	1.11*	.051*	sr1, sr2
sr5	I frequently take time to reflect on my thoughts	4.83 (1.38)	1.25	−.61	1.02	1.01	.028	sr6
sr6	I often think about the way I feel about things	5.10 (1.31)	1.00	−.90	1.03	1.02	.037	sr5
sr7	I am not really interested in analyzing my behavior (R)	5.11 (1.51)	.64	−.95	1.03	1.08*	.039	
sr8	It is important to me to evaluate the things that I do	4.99 (1.37)	1.08	−.78	1.02	1.04	.027	
sr9	I am very interested in examining what I think about	4.74 (1.38)	1.68	−.52	1.01	1.00	.019	
sr10	It is important to me to try to understand what my feelings mean	4.99 (1.48)	1.78	−.63	1.00	.97	.022	
sr11	I have a definite need to understand the way that my mind works	4.69 (1.54)	1.19	−.51	1.00	1.00	.022	sr12
sr12	It is important to me to be able to understand how my thoughts arise	4.64 (1.45)	1.35	−.45	1.01	1.00	.025	sr11
Insight scale								
ins1	I am usually aware of my thoughts	5.34 (1.22)	.20	−2.49	1.01	1.03	.032	ins3, ins8
ins2	I'm often confused about the way that I really feel about things (R)	4.19 (1.65)	.68	−.12	1.00	1.03	.024	
ins3	I usually have a very clear idea about why I've behaved in a certain way	4.85 (1.38)	.38	−.98	1.01	1.02	.034	ins1, ins8
ins4	I'm often aware that I'm having a feeling, but I often don't quite know what it is (R)	4.16 (1.52)	.54	−.13	1.00	1.02	.029	
ins5	My behavior often puzzles me (R)	4.79 (1.66)	.93	−.55	1.02	1.04	.027	ins7
ins6	Thinking about my thoughts makes me more confused (R)	4.73 (1.68)	1.12	−.50	1.03	1.02	.030	
ins7	Often I find it difficult to make sense of the way I feel about things (R)	4.67 (1.59)	1.60	−.44	1.02	1.00	.017	ins5
ins8	I usually know why I feel the way I do	4.98 (1.39)	.58	−.93	1.02	1.02	.032	ins1, ins3

Note. n = 1192. The items were completed using a 7-point scale. For Infit and Outfit, items with an asterisk have statistically significant misfit. RMSD values greater than .50 indicate "medium" misfit and are flagged with an asterisk. The local dependence column indicates items pairs with aQ_3 correlations greater than |.25|

Taken together, the SRIS has proven to be an effective tool for advancing the study of individual differences in self-awareness. In the present research, the scale receives a close psychometric examination using tools from the family of Rasch and item response theory (IRT) models (Bond et al., 2020; DeMars, 2010; Ostini & Nering, 2006). It appears that these tools have not yet been applied to the SRIS, yet they can illuminate the behavior of items and scales in incisive and practical ways. In addition to shedding light on the strengths and weaknesses of the SRIS, a key goal of the present research is to use the results of this psychometric examination to craft a concise, shorter form of the SRIS. Item response theory is particularly useful for identifying items that are relatively inefficient, such as items that contribute little information, impair unidimensionality, or bias the scores toward some respondents. A brief, efficient version of the SRIS would be useful for researchers conducting studies where brevity is important, such as online surveys or field studies, as well as for applied users of the scale, such as practitioners in counseling and coaching contexts where brief tools can be more easily worked into assessments and activities. In the present research, then, an item response analysis of SRIS responses from a sample of nearly 1200 young adults was conducted to illuminate the behavior of the scale and its items. Based on these results, a concise, 12-item scale—6 for self-reflection, 6 for insight—was crafted and evaluated.

Method

Participants

A total of 1192 adults—864 women, 328 men—completed the SRIS as part of one of many research projects conducted from 2005 to the present. All participants provided informed consent. The participants were virtually all college-aged young adults enrolled in psychology courses at a regional public university in the Southeastern United States.

Method and Analysis Approach

The data were collated from a series of studies that examined the role of self-reflection and insight in motivation and emotion (Silvia et al., 2011, 2013; Silvia & Phillips, 2011; Silvia et al., 2018), many additional studies that measured the SCIS as an incidental measure (Harper et al., 2018; Silvia, 2012; Silvia et al., 2014; Silvia et al., 2018; Silvia et al., 2020a, b; Sperry et al., 2018), and several unpublished datasets. Participants completed the 20 SRIS items using a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*) in either a pencil-and-paper or electronic format.

The SRIS responses were analyzed with R 4.0.3 (R Core Team, 2020) using the packages *TAM* 3.6.20 (Robitzsch et al., 2020) and *psych* 2.0.9 (Revelle, 2020). The self-reflection scale and the insight scale were modeled as separate unidimensional scales because some researchers use only one of them. The first set of analyses scrutinized the full 20-item SRIS to appraise its strengths and weaknesses. Based on these results, items were trimmed to create a brief scale, which was in turn evaluated to examine its psychometric qualities. The scales were modeled with a generalized partial credit model (GPCM; Ostini & Nering, 2006), which had better fit than a simpler model (a Rasch partial credit model; Bond et al., 2020) according to information theory

fit statistics. The IRT models were estimated in *TAM* using marginal maximum likelihood and case-centering, which centers the estimated underlying trait scores on zero.

Results

Descriptive Statistics

Table 1 shows descriptive statistics for the SRIS; Fig. 1 displays the distribution of scale responses for all 20 items. Figure 1 reveals that the items are, in IRT terms, "easy": endorsement rates were high, and responses tended to pile up at the high end of the response scale. The hardest item—the one with the lowest mean—was Item 4 on the insight scale, and its mean (4.16) was nevertheless above the scale's midpoint. The self-reflection and insight scores, formed via item averages, correlated weakly, $r = .07$ [.01, .13], $p = .04$, as in past research. The relationship is shown in Fig. 2.

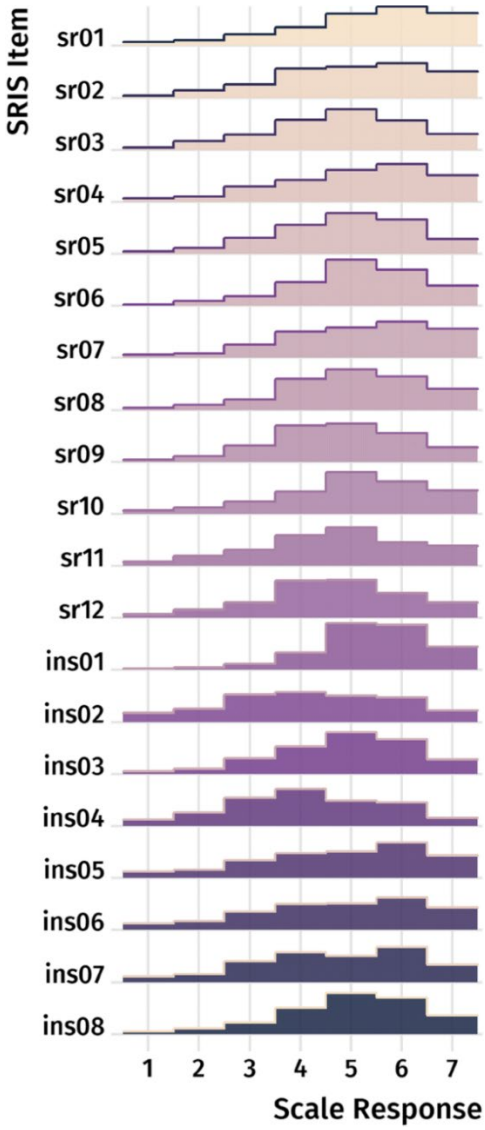

Fig. 1. Distribution of scale responses for all 20 items in the self-reflection and insight scale

Fig. 2. Relationship between the full self-reflection and insight scales

Reliability and Dimensionality

The reliability of the self-reflection and insight subscales was very good. Table 2 displays several reliability coefficients: Cronbach's alpha (α), omega-hierarchical (ω_H; degree to which the items are saturated by the general, common factor), and the reliability of the IRT-estimated expected a posteriori (EAP) trait scores.

Table 2. Subscale features and statistics

	Self-reflection		**Insight**	
	Original	**Short**	**Original**	**Short**
Number of items	12	6	8	6
Cronbach's α	.90	.87	.80	.83
Omega ω_H	.75	.77	.64	.75
EAP reliability	.92	.88	.86	.85

Because item response theory assumes at least essential unidimensionality—a single dominant factor with ignorable minor factors (Slocum-Gori & Zumbo, 2011)—the dimensionality of the self-reflection and insight scales were examined using several criteria: Velicer's MAP criterion, the ratio of first to second eigenvalues (e.g., at least 3:1), and parallel analysis (Hayton et al., 2004; Slocum-Gori & Zumbo, 2011; Velicer, 1976). Consistent with past work, both subscales were essentially unidimensional. Velicer's MAP criterion suggested 1 factor for both subscales, and the first eigenvalue was over 8 times larger (self-reflection) and 5 times larger (insight) than the second. The scree plots for the actual and resampled values from the parallel analysis (Fig. 3) suggested that factors beyond the first were at most modest. Taken together, the evidence suggests that essential unidimensionality is credible, but the scree plots suggest that unidimensionality could probably be strengthened further.

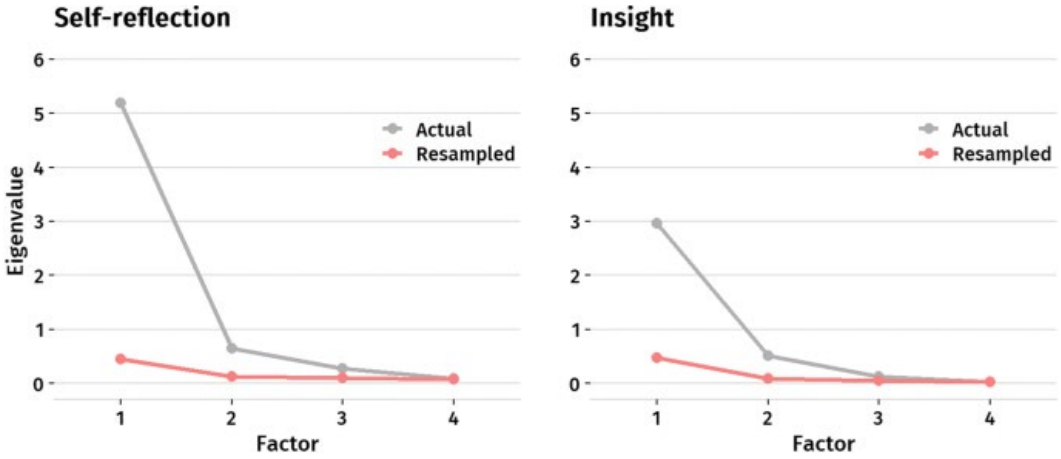

Fig. 3. Parallel analysis scree plots for the full self-reflection and insight scales. (Note. For clarity, only the first four factors are depicted)

To explore dimensionality further, local dependence between item pairs was examined. In a unidimensional IRT model the items should be conditionally independent, in that they should be uncorrelated after the shared influence of the latent variable is accounted for (Chen & Thissen, 1997). Items with a notable residual correlation display local dependence, often due to superficial item features (e.g., similar wording). Local dependence was quantified using the adjusted Q_3 (aQ_3; Marais, 2013) statistic, a bias-corrected form of the traditional Q_3 statistic (Yen, 1984). A cut-off of |.25| was used to flag the most notably dependent item pairs (Christensen et al., 2017).

Table 1 notes the item pairs with noteworthy residual correlations. For the self-reflection scale, a group of items (items 1, 2, and 4) all clustered together; in addition, notable residual correlations appeared for items 5 and 6 and for items 11 and 12. For the insight scale, a group of items (1, 3, and 6) clustered together, and items 5 and 7 had a notable residual correlation. Most of these apparently reflect overlaps in item wording or relatively redundant item meaning. Regardless of the reason, these patterns of local dependence underlie the small, minor factor shown in Fig. 3, so selectively trimming locally dependent items would strengthen the unidimensionality of these scales.

Item Fit and IRT Parameters

Item Fit

The fit of the items to the IRT model was evaluated with Infit and Outfit statistics. Values greater than 1 reflect increasingly noisy scores (Bond et al., 2020). As Table 2 shows, several self-reflection items showed significantly noisy misfit on the Outfit metric (items 1, 4, and 7). No items were flagged for the insight scale. Additional insight into item fit comes from the root mean square deviation (RMSD) statistic, which reflects the deviation between the expected and observed item response functions (Buchholz & Hartig, 2019). Values over .20 reflect “small” misfit, and values over .50 reflect “medium” misfit (Köhler et al., 2020). For the self-reflection scale, two items had RMSD values over .50 (items 1 and 4, which both had poor Outfit values). No items on the insight scale were flagged for high RMSD values. In brief, the analyses of item

fit suggest that a small number of items on the self-reflection scale are possible candidates for trimming.

Item Difficulty

In the context of self-report measures of personality, the IRT difficulty parameter (*b*) represents "endorsability": how easy or hard it is to agree or disagree with the item. Higher numbers reflect "harder" items for which it takes higher trait levels to endorse. Table 1 reports the difficulty values for the items. Both the self-reflection scale and the insight scale were quite "easy." Figure 4 displays the difficulty values sorted from the easiest to the hardest item. The self-reflection scale's difficulty values are within a fairly small range on the easy end of the scale. The insight scale, in contrast, shows a notable outlier—item 1 is far too easy to endorse to be informative for these respondents.

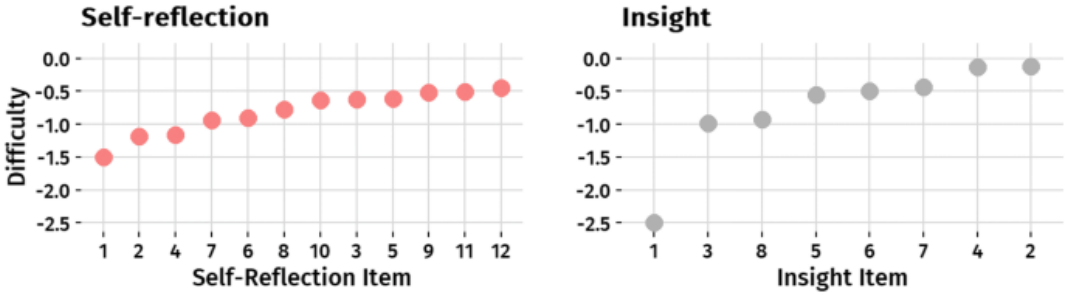

Fig. 4. IRT difficulty parameters for the full self-reflection and insight scales

Item Discrimination

The IRT discrimination parameter (*a*; also known as the slope parameter) represents how quickly the probability of giving a higher item response changes as the underlying trait changes. Higher values indicate that items more closely track the underlying trait and thus more precisely distinguish between participants. Discrimination values are conceptually akin to loadings in a confirmatory factor analysis. Table 1 lists the discrimination values; Fig. 5 displays the values sorted from the least to the most discriminating item. Two patterns stand out. First, the self-reflection scale, as a whole, has better item discrimination than the insight scale, which has only two items with *a* values greater than 1. Second, each scale has a handful of weak items. There are not firm cut-offs for discrimination values for polytomous self-report scales, but values under .50 are clearly poor. These items provide relatively little information about the underlying trait and thus poorly discern the respondents' different trait levels.

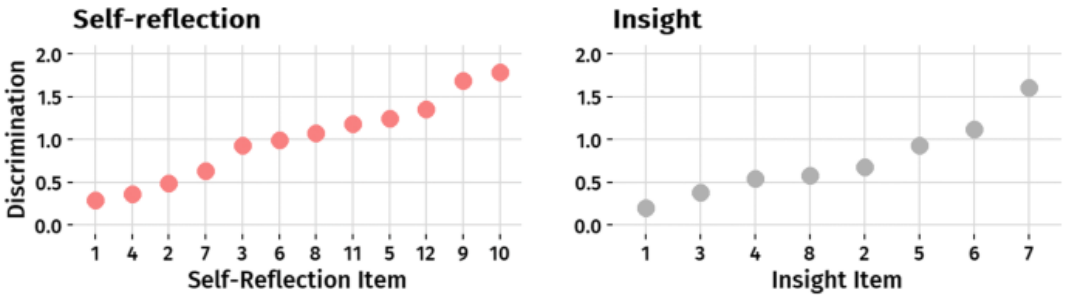

Fig. 5. IRT discrimination parameters for the full self-reflection and insight scales

Test Information

Given the difficulty and discrimination values of their items, scales provide more information at some levels of the trait than others (Bond et al., 2020; DeMars, 2010). Figure 6 shows the test information functions for the self-reflection and insight scales. The insight scale provides relatively less information because it has fewer items and contains many items with relatively poor discrimination values. For both scales, the test information peak is centered at the lower end of the trait—around −1.30 for self-reflection and −.60 for insight. Because the items are all relatively easy to endorse, the scale more reliably sorts between participants who are low in self-reflection and in insight but provides less reliable information about participants who are higher in the traits.

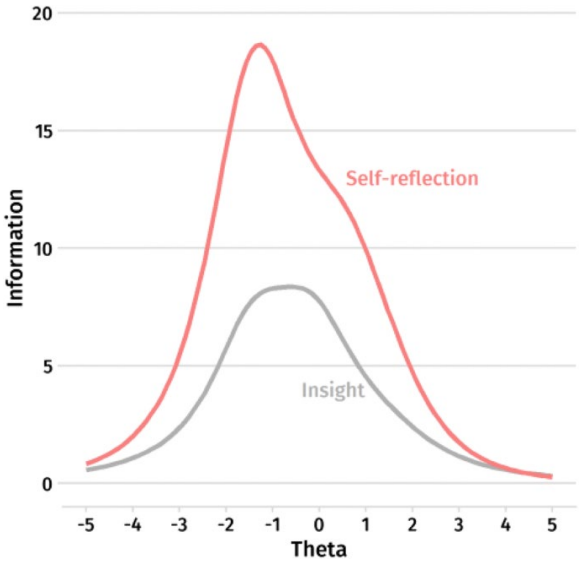

Fig. 6. Test information functions for the full self-reflection and insight scales

Differential Item Functioning

As the final step, the SRIS items were evaluated for gender-based item bias via the analysis of differential item functioning (DIF; Osterlind & Everson, 2009). When members of different groups have different item responses, the differences should be due to variation in the underlying trait, not to unintended, construct-irrelevant factors. In the case of gender, for example, men and women with the same trait level should have the same predicted response. When the expected response varies for groups with the same trait level, the item is said to show DIF and is hence biased in favor of one group (Penfield & Camilli, 2006). As a backdrop, gender differences were modest in the observed scores, as Table 3 shows. In the Cohen's *d* metric, gender had a very small effect on self-reflection scores (women slightly higher: $d = .07$ [−.06, .19]) and a small effect on insight scores (women slightly lower: $d = -.19$ [−.32, −.06]).

Analyses of item bias have apparently not been conducted thus far for the SRIS. To see if any of the items showed gender-based bias, DIF analyses were conducted using the logistic ordinal regression approach implemented in *lordif* (Choi et al., 2011), which uses IRT-based trait scores

and iterative purification to identify items displaying DIF. As the criterion, we used effect size metrics, namely McFadden's R^2 (Menard, 2000), as an intuitive way to express the amount of DIF, if any (Meade, 2010). A stringent criterion of $R^2 > .01$ was used for the initial flagging of items for DIF. Remarkably, no items for either the self-reflection or insight scale were flagged for gender-based DIF. This lack of gender bias is a strength of the scale and should lend confidence in the conclusions drawn about gender based on SRIS scores.

Table 3. Gender differences in self-reflection and insight

	Original scales	**Short scales**
Self-reflection	.07 [−.06, .19]	.14 [.02, .27]
Insight	−.19 [−.32, −.06]	−.18 [−.31, −.05]

Note. Cohen's *d* and its 95% confidence intervals are displayed. Gender is coded 0 (men, $n = 328$) and 1 (women, $n = 864$), so positive *d* values reflect relatively higher scores for women

Sharpening the SRIS into Shorter Scales

IRT offers a wealth of information about items and their behavior that is useful for developing short forms. To sift through the items and determine which should be omitted, several statistical criteria were used: (1) poor item fit (Infit, Outfit, RMSD), which flags items which generate scores that are relatively poorly predicted by the IRT model; (2) local dependence (aQ_3), which impairs unidimensionality; (3) unreasonable difficulty levels (IRT *b*), which are either too hard or too easy for the target population; and (4) weak slope/discrimination values (IRT *a*), which indicates that an item provides relatively little information for reliably rank-ordering participants. Finally, as practical criteria, (1) the brief self-reflection and insight scales should be the same length, and (2) the brief self-reflection scale should have equal numbers of items from the "engagement in" and "need for" facets to ensure balanced coverage. Although the brief scale would not be ideal for researchers interested in facet-level scores, balancing the facets ensures that the brief self-reflection scale has similar construct coverage.

These criteria resulted in a 12-item brief form of the SRIS. Table 4 shows the items selected for the brief scales. For the self-reflection scale, a handful of items were obvious low-hanging fruit to pick for omission because they were poor on several indicators. Items 1, 4, and 7 had poor fit on Outfit, RMSD, or both. In addition, Items 1 and 4 had widespread local dependence with other items, so omitting them removed ill-fitting items that impaired unidimensionality. Items 2 and 7 were among the easiest and least discriminating items, so they were good candidates for omission. Finally, Items 11 and 12 had a strong aQ_3 residual correlation, so omitting these would improve unidimensionality. This yielded a 6-item brief self-reflection scale that was balanced between 3 "engagement in" items and 3 "need for" items.

For the insight scale, items 1 and 3 were omitted. Item 1 had the lowest difficulty value and the lowest discrimination value. Item 3 had the second-lowest discrimination value, and had strong local dependence with Items 1 and 8. Omitting these two items thus removed two low-discriminating items and eliminated substantial local dependence. This yielded a 6-item brief insight scale.

Table 4. Short form of the self-reflection and insight scale: item-level statistics

New and old item number	Text	*a* (slope)	*b* (difficulty)	Infit	Outfit	RMSD
Short Self-reflection Scale						
1 (sr3)	I frequently examine my feelings	.98	−.64	1.02	1.03	.036
2 (sr5)	I frequently take time to reflect on my thoughts	1.50	−.62	1.02	1.00	.028
3 (sr6)	I often think about the way I feel about things	1.20	−.89	1.03	1.01	.032
4 (sr8)	It is important to me to evaluate the things that I do	1.05	−.82	1.02	1.02	.028
5 (sr9)	I am very interested in examining what I think about	1.37	−.57	1.01	1.00	.024
6 (sr10)	It is important to me to try to understand what my feelings mean	1.50	−.68	1.01	.98	.022
Short insight scale						
1 (ins2)	I'm often confused about the way that I really feel about things (R)	.67	−.12	1.00	1.02	.024
2 (ins4)	I'm often aware that I'm having a feeling, but I often don't quite know what it is (R)	.57	−.13	1.00	1.02	.028
3 (ins5)	My behavior often puzzles me (R)	.93	−.56	1.03	1.03	.028
4 (ins6)	Thinking about my thoughts makes me more confused (R)	1.16	−.50	1.03	1.01	.030
5 (ins7)	Often I find it difficult to make sense of the way I feel about things (R)	1.64	−.44	1.02	1.00	.018
6 (ins8)	I usually know why I feel the way I do	.51	−.98	1.02	1.02	.033

n = 1192. The items were completed using a 7-point scale. No items showed statistically significant misfit for the Infit and Outfit item fit statistics. No items exceeded the .50 "medium misfit" threshold for the RMSD item fit statistic. There were no aQ_3 correlations greater than |.25|

Psychometric Properties of the Short Scales

Because the item omission was grounded in a set of IRT-based statistics, the short scales showed excellent psychometric properties, both relative to their length and to the full scales. Table 1 reports the short scales' reliability coefficients, which were similar to, and sometimes higher than, the full scales' coefficients. Both short scales had higher ω_H values than the full scales, which reflects their improved unidimensionality and a more dominant common factor. Regarding dimensionality, the subscales were even more strongly unidimensional. MAP again indicated a single factor, and the first eigenvalues were many times larger than the second for both self-reflection (12.4 times) and insight (20.8 times). When the items were averaged to form subscale scores, the brief self-reflection and insight scales continued to correlate modestly with each other, $r = -.07$ [−.13, .00], $p = .04$, as in past work (see Fig. 7). As one would expect, the short and long forms correlated very highly for the both the self-reflection ($r = .94$ [.93, .95], $p < .001$) and the insight ($r = .97$ [.96, .97], $p < .001$) scales.

As before, an IRT GPCM was estimated for each short scale. Table 4 reports the results. Regarding item fit, no items on either scale were flagged for notable misfit on Infit, Outfit, or RMSD, so the items' fit to the IRT model was much improved. Regarding local dependence, there were no aQ_3 residual correlations greater than |.25| for any pair of items. And as before, no items showed gender-based DIF, using McFadden's $R^2 > .01$ as a criterion.

Figure 8 illustrates the difficulty (*b*; top panel) and discrimination (*a*; bottom panel) parameters for the short scales. The items remain relatively easy to endorse overall, and the self-reflection scale continues to be more discriminating than the insight scale, but the brief scales no longer have the worst performing items that were far too easy or too low in discrimination.

Fig. 7. Relationship between the short self-reflection and insight scales

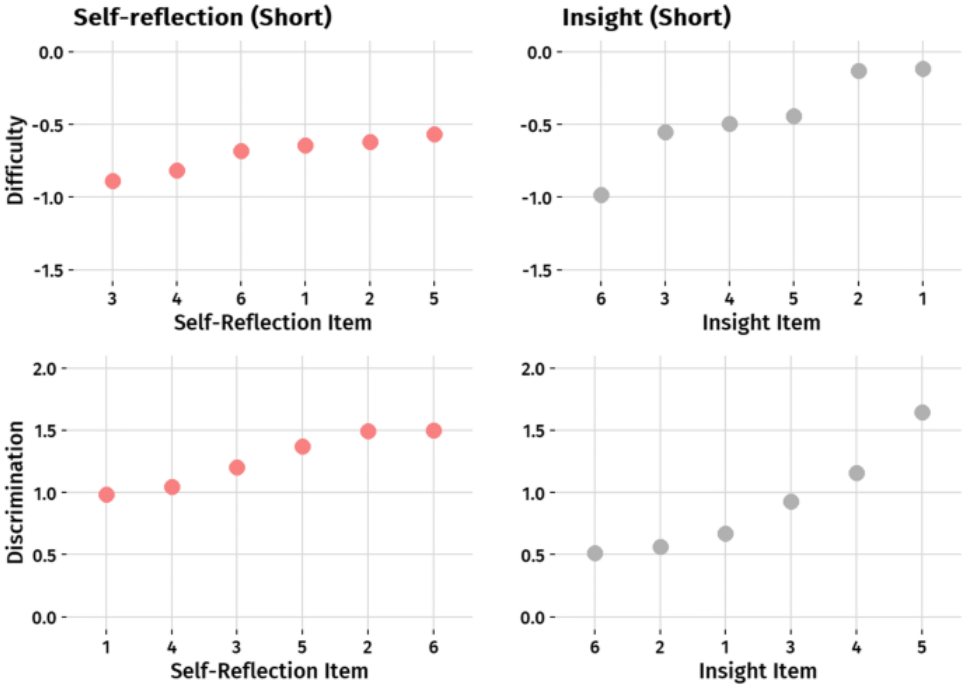

Fig. 8. IRT difficulty (top) and discrimination (bottom) parameters for the short self-reflection (left) and insight (right) scales

Finally, the test information functions for the brief scales are shown in Fig. 9. As before, the scales provided the most information at the lower end of the scale. Compared to the long scales, the test information peaked at similar locations for the short self-reflection (around −1.40) and short insight (around −.50) scales.

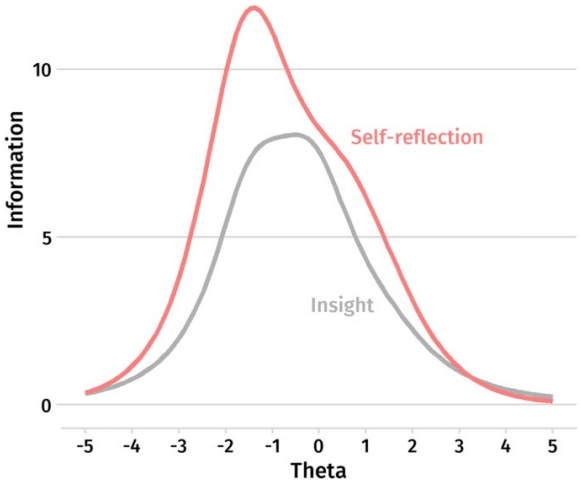

Fig. 9. Test information functions for the short self-reflection and insight scales

Discussion

Self-focused attention is a key concept in meta-cognition, motivation, and self-regulation (Carver & Scheier, 1998; Silvia, 2014). The present research took a close look at the Self-Reflection and Insight Scale (SRIS), a prominent measure of individual differences in two facets of self-focus (Grant et al., 2002). Using tools from the family of Rasch and IRT models, the analyses identified many strengths of the scale: the two scales have solid unidimensionality and strong reliability, consistent with past research. In addition, the analyses showed, for the first time, an essential lack of gender-based item bias for the SRIS, which is an important strength for a measure of personality and individual differences.

At the same time, the analyses revealed some problematic qualities, such as (1) the "easy" quality of all the items, reflecting high endorsement rates, and (2) a handful of misbehaved items in terms of item misfit, low discrimination, excessively low difficulty, or high residual correlations. Using the IRT-based statistics, the items were whittled down to a concise, 12-item scale that was evenly balanced between self-reflection and insight. The short scales preserved and enhanced the full scale's strengths. The subscales were more strongly unidimensional, and reliability was either minimally reduced or enhanced. The short scale showed excellent item properties in terms of item fit and local independence, and they retained the lack of gender-based DIF.

One limitation of this work is its focus on young, college-aged adults. Although this is a common target population for the SRIS, the scale is widely used in a range of populations for both basic and applied purposes. The recent growth in translations of the SRIS (Aşkun & Cetin, 2017; DaSilveira et al., 2015; Naeimi et al., 2019; Nakajima et al., 2017) both illustrates its popularity and highlights the need to understand cultural aspects of self-reflection and insight. Applying Rasch and IRT methods to translated forms of the scales and evaluating the merit of the proposed short form is an important goal for future psychometric research on the SRIS.

Data Availability

The data and R code are publicly available at Open Science Framework: https://osf.io/qsa5w/.

Ethics declarations

Ethics Approval. This research was evaluated, approved, and monitored by the Institutional Review Board at the University of North Carolina at Greensboro.

Conflicts of Interest/Competing Interests. The author has no conflicts or competing interests to declare.

References

Anderson, E. M., Bohon, L. M., & Berrigan, L. P. (1996). Factor structure of the private self-consciousness scale. *Journal of Personality Assessment, 66*, 144–152.

Aşkun, D., & Cetin, F. (2017). Turkish version of self-reflection and insight scale: A preliminary study for validity and reliability of the constructs. *Psychological Studies, 62*(1), 21–34.

Bernstein, I. H., Teng, G., & Garbin, C. P. (1986). A confirmatory factoring of the self-consciousness scale. *Multivariate Behavioral Research, 21*, 459–475.

Bond, T. G., Yan, Z., & Heine, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Abingdon, UK: Routledge.

Britt, T. W. (1992). The self-consciousness scale: On the stability of the three-factor structure. *Personality and Social Psychology Bulletin, 18*, 748–755.

Buchholz, J., & Hartig, J. (2019). Comparing attitudes across groups: An IRT-based item-fit statistic for the analysis of measurement invariance. *Applied Psychological Measurement, 43*(3), 241–250.

Buss, A. H. (1980). *Self-consciousness and social anxiety*. San Francisco: Freeman.

Carver, C. S. (2003). Self-awareness. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 179–196). New York: Guilford.

Carver, C. S., & Scheier, M. F. (1990). Principles of self-regulation: Action and emotion. In R. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (Vol. 2, pp. 3–52). New York: Guilford.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.

Chen, W. H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265–289.

Chen, S. Y., Lai, C. C., Chang, H. M., Hsu, H. C., & Pai, H. C. (2016). Chinese version of psychometric evaluation of self-reflection and insight scale on Taiwanese nursing students. *Journal of Nursing Research, 24*(4), 337–346.

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*(8), 1–30.

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*, 178–194.

Cowden, R. G., & Meyer-Weitz, A. (2016). Self-reflection and self-insight predict resilience and stress in competitive tennis. *Social Behavior and Personality, 44*(7), 1133–1149.

Cramer, K. M. (2000). Comparing the relative fit of various factor models of the self-consciousness scale in two independent samples. *Journal of Personality Assessment, 75*, 295–307.

DaSilveira, A., DeSouza, M. L., & Gomes, W. B. (2015). Self-consciousness concept and assessment in self-report measures. *Frontiers in Psychology, 6*, 930.

DeMars, C. (2010). *Item response theory*. Oxford: Oxford University Press.

Duval, T. S., & Silvia, P. J. (2001). *Self-awareness and causal attribution: A dual systems theory*. Boston: Kluwer Academic Publishers.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522–527.

Grant, A. M. (2001). Rethinking psychological mindedness: Metacognition, self-reflection, and insight. *Behaviour Change, 18*, 8–17.

Grant, A. M. (2003). The impact of life coaching on goal attainment, metacognition and mental health. *Social Behavior and Personality, 31*, 253–264.

Grant, A. M., Franklin, J., & Langford, P. (2002). The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality, 30*, 821–836.

Harper, K. L., Silvia, P. J., Eddington, K. M., Sperry, S. H., & Kwapil, T. R. (2018). Conscientiousness and effort-related cardiac activity in response to piece-rate cash incentives. *Motivation and Emotion, 42*, 377–385.

Harrington, R., & Loffredo, D. A. (2010). Insight, rumination, and self-reflection as predictors of well-being. *The Journal of Psychology, 145*(1), 39–57.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205.

Köhler, C., Robitzsch, A., & Hartig, J. (2020). A bias-corrected RMSD item fit statistic: An evaluation and comparison to alternatives. *Journal of Educational and Behavioral Statistics, 45*(3), 251–273.

Lyke, J. A. (2009). Insight, but not self-reflection, is related to subjective well-being. *Personality and Individual Differences, 46*, 66–70.

Marais, I. (2013). Local dependence. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 111–130). London: Wiley.

Meade, A. W. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *Journal of Applied Psychology, 95*, 728–743.

Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician, 54*, 17–24.

Naeimi, L., Abbaszadeh, M., Mirzazadeh, A., Sima, A. R., Nedjat, S., & Hejri, S. M. (2019). Validating self-reflection and insight scale to measure readiness for self-regulated learning. *Journal of Education and Health Promotion, 8*(1), 150.

Nakajima, M., Takano, K., & Tanno, Y. (2017). Adaptive functions of self-focused attention: Insight and depressive and anxiety symptoms. *Psychiatry Research, 249*, 275–280.

Nakajima, M., Takano, K., & Tanno, Y. (2018). Contradicting effects of self-insight: Self-insight can conditionally contribute to increased depressive symptoms. *Personality and Individual Differences, 120*, 127–132.

Nakajima, M., Takano, K., & Tanno, Y. (2019). Mindfulness relates to decreased depressive symptoms via enhancement of self-insight. *Mindfulness, 10*(5), 894–902.

Nystedt, L., & Ljungberg, A. (2002). Facets of private and public self-consciousness: Construct and discriminant validity. *European Journal of Personality, 16*, 143–159.

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning*. Thousand Oaks, CA: Sage.

Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: Sage.

Penfield, R. D., & Camilli, G. (2006). Differential item functioning and item bias. *Handbook of Statistics, 26*, 125–167.

R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org

Revelle, W. (2020). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. https://CRAN.R-project.org/package=psych

Roberts, C., & Stark, P. (2008). Readiness for self-directed change in professional behaviours: Factorial validation of the self-reflection and insight scale. *Medical Education, 42*, 1054–1063.

Robitzsch, A., Kiefer, T., & Wu, M. (2020). *TAM: Test analysis modules*. https://CRAN.R-project.org/package=TAM

Sauter, F. M., Heyne, D., Blöte, A. W., Widenfelt, B. M., & Westenberg, P. M. (2010). Assessing therapy-relevant cognitive capacities in young people: Development and psychometric properties of the self-reflection and insight scale for youth. *Behavioural and Cognitive Psychotherapy, 38*, 303–317.

Scheier, M. F., & Carver, C. S. (1985). The self-consciousness scale: A revised version for use with general populations. *Journal of Applied Social Psychology, 15*, 687–699.

Silvia, P. J. (2012). Mirrors, masks, and motivation: Implicit and explicit self-focused attention influence effort-related cardiovascular reactivity. *Biological Psychology, 90*, 192–201.

Silvia, P. J. (2014). Self-striving: How self-focused attention affects effort-related cardiovascular activity. In G. H. E. Gendolla, M. Tops, & S. Koole (Eds.), *Handbook of biobehavioral approaches to self-regulation* (pp. 301–314). New York: Springer.

Silvia, P. J., & Duval, T. S. (2001). Objective self-awareness theory: Recent progress and enduring problems. *Personality and Social Psychology Review, 5*, 230–241.

Silvia, P. J., & Phillips, A. G. (2011). Evaluating self-reflection and insight as self-conscious traits. *Personality and Individual Differences, 50*(2), 234–237.

Silvia, P. J., Eichstaedt, J., & Phillips, A. G. (2005). Are rumination and reflection types of self-focused attention? *Personality and Individual Differences, 38*, 871–881.

Silvia, P. J., Jones, H. C., Kelly, C. S., & Zibaie, A. (2011). Masked first name priming increases effort-related cardiovascular reactivity. *International Journal of Psychophysiology, 80*(3), 210–216.

Silvia, P. J., Kelly, C. S., Zibaie, A., Nardello, J. L., & Moore, L. C. (2013). Trait self-focused attention increases sensitivity to nonconscious primes: Evidence from effort-related cardiovascular reactivity. *International Journal of Psychophysiology, 88*, 143–148.

Silvia, P. J., Nusbaum, E. C., Eddington, K. M., Beaty, R. E., & Kwapil, T. R. (2014). Effort deficits and depression: The influence of anhedonic depressive symptoms on cardiac autonomic activity during a mental challenge. *Motivation and Emotion, 38*, 779–789.

Silvia, P. J., Sizemore, A. J., Tipping, C. J., Perry, L. B., & King, S. F. (2018). Get going! Self-focused attention and sensitivity to action and inaction effort primes. *Motivation Science, 4*, 109–117.

Silvia, P. J., Eddington, K. M., Harper, K. L., Burgin, C. J., & Kwapil, T. R. (2020a). Appetitive motivation in depressive anhedonia: Effects of piece-rate cash rewards on cardiac and behavioral outcomes. *Motivation Science, 6*(3), 259–265. https://doi.org/10.1037/mot0000151.

Silvia, P. J., McHone, A. N., Mironovová, Z., Eddington, K. M., Harper, K. L., Sperry, S. H., & Kwapil, T. R. (2020b). RZ interval as an impedance cardiography indicator of effort-related cardiac sympathetic activity. *Applied Psychophysiology and Biofeedback*. https://doi.org/10.1007/s10484-020-09493-w.

Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research, 102*, 443–461.

Smári, J., Ólason, D., & Ólafsson, R. P. (2008). Self-consciousness and similar personality constructs. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (Vol. 1, pp. 486–505). Thousand Oaks, CA: Sage.

Sperry, S. H., Kwapil, T. R., Eddington, K. M., & Silvia, P. J. (2018). Psychopathology, everyday behaviors, and autonomic activity in daily life: An ambulatory impedance cardiography study of depression, anxiety, and hypomanic traits. *International Journal of Psychophysiology, 129*, 67–75.

Stein, D., & Grant, A. M. (2014). Disentangling the relationships among self-reflection, insight, and subjective well-being: The role of dysfunctional attitudes and core self-evaluations. *The Journal of Psychology, 148*(5), 505–522.

Trapnell, P. D., & Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology, 76*, 284–304.

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

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*, 125–145.