



Development and Validation of the Awareness Outcomes Measure (AOM) Using Rasch Approach

Anna Sutton¹ · Oleg N. Medvedev¹

Accepted: 9 December 2022
© The Author(s) 2023

Abstract

Objectives Awareness is a key component of concepts related to well-being, such as mindfulness and authenticity. Similarly, interventions to enhance mindfulness and well-being often focus on developing awareness. But measuring the effect of awareness development represents a challenge due to the lack of reliable and valid measures focused specifically on awareness outcomes. This study aimed to develop and validate the Awareness Outcomes Measure (AOM) using modern Rasch methodology.

Method The measure was developed from Self-Awareness Outcomes Questionnaire (SAOQ) items, drawn from previous research with awareness-development experts. A partial credit Rasch model was applied to examine the psychometric properties of the AOM with a combined sample of 713 participants from three English-speaking countries.

Results The 21-item AOM met expectations of the unidimensional Rasch model. It is a reliable and psychometrically sound instrument, invariant across sex, country, and age, designed to measure the outcomes of awareness development. Person-item thresholds demonstrated excellent coverage of awareness outcomes, and we developed an algorithm for ordinal-to-interval transformations presented in a table to further enhance precision of the AOM.

Conclusions In this study, we have developed and validated the AOM, providing researchers and practitioners with a robust measure of awareness outcomes that is suitable for use in a range of populations.

Keywords Mindfulness · Awareness · Awareness Outcomes Measure · Rasch analysis · Reliability · Validity · Assessment

The development of both mindfulness and authenticity is proposed to be important in the improvement of well-being (Querstret et al., 2020; Sutton, 2020) and awareness of the self is a key element of both of these constructs. Mindfulness, for example, is defined in terms of an awareness that operates on “thought, feeling, and other contents of consciousness” (Brown & Ryan, 2003, p. 823). Similarly, authenticity is commonly defined in terms of self-awareness and genuine self-expression (Kernis & Goldman, 2006). Interventions to improve awareness, such as mindfulness training or talking therapies, therefore have significant potential to improve individual well-being.

When researchers and practitioners wish to evaluate the effectiveness of such awareness-building interventions, they frequently attempt to measure changes in levels of awareness

itself or choose related outcomes such as well-being, engagement, or work performance. While valuable, there are some drawbacks to these approaches. First, they involve the paradox of attempting to measure self-awareness by directing the individual’s attention towards their awareness. Second, the different measures of awareness are not interchangeable and may be associated with different outcomes. Third, the measurement of non-overlapping outcomes in different studies reduces the possibility of effective comparisons between interventions. We explore each of these drawbacks in more detail next and demonstrate how the Self-Awareness Outcomes Questionnaire (SAOQ; Sutton, 2016), a measure focused on evaluating the outcomes associated with awareness rather than awareness itself, provides a potential solution.

A significant, but often overlooked, confounding issue with measuring awareness is that it necessarily involves metacognitive awareness of one’s own self-awareness. The well-known Dunning-Kruger effect describes how individuals with less awareness of their own skills can overestimate their abilities, while developing more accurate awareness

✉ Anna Sutton
anna.sutton@waikato.ac.nz

¹ School of Psychology, University of Waikato, Hillcrest, Private Bag 3105, Hamilton 3240, New Zealand

can paradoxically reduce their self-ratings (Kruger & Dunning, 1999). Recent research has shown that those individuals who show evidence of the Dunning-Kruger effect in the reasoning domain also report greater self-efficacy and overestimate their use of effective decision-making styles (Christopher et al., 2021). Thus, when practitioners attempt to aid people to develop their awareness, there is likely to be at least a sub-group who are less able to accurately judge their abilities. Grossman (2019) has suggested exactly this as a challenge to mindfulness studies using self-report questionnaires. Ironically, even if an awareness-developing intervention is successful, participants may report reduced awareness in post- than in pre-intervention measures as their metacognitive abilities have improved. There is some evidence that this may be the case in mindfulness development as Baer et al. showed that meditators scored higher on all scales of the Five Facet Mindfulness Questionnaire (FFMQ) except Acting with Awareness (Baer et al., 2008). A measure that assesses outcomes associated with greater awareness, rather than expecting participants to have accurate metacognitive awareness, can avoid this potential confound.

The second challenge facing measures of awareness is that commonly used self-rating instruments use facets or subscales drawn from a range of mindfulness or authenticity questionnaires. For example, the FFMQ (Baer et al., 2006) assesses awareness with its Acting with Awareness facet (Krägeloh et al., 2019) drawn from the Mindful Attention and Awareness Scale (MAAS; Brown & Ryan, 2003). The Authenticity Scale evaluates the extent to which a person's awareness and expression of their states, emotions, and cognitions is congruent (Wood et al., 2008) while the Integrated Authenticity Scale assesses self-awareness as a subscale (Knoll et al., 2015). While each measure or index of awareness is of course valuable in its own right, the distinct definitions and conceptualizations mean that they are not equivalent and thus the ability to compare the effectiveness of different interventions to improve awareness is diminished. A focus on outcomes that are associated with general increases in awareness across a range of interventions and conceptualizations would enable this kind of comparison.

A third drawback of these common approaches to evaluating awareness-building interventions is that they do not measure “downstream” changes in a coherent or standardized manner. That is, they do not evaluate the effect of increased awareness on a specified range of individual outcomes that are theoretically and demonstrably linked to awareness. Instead, the effectiveness of mindfulness training may be evaluated in terms of general outcomes such as individual well-being, stress, and physical symptoms through work-related outcomes such as job satisfaction or client satisfaction (Jamieson & Tuckey, 2017). Similarly, workplace awareness-building interventions, such as 360-degree

feedback, may be evaluated solely in terms of their impact on work performance measures (Fletcher & Bailey, 2003). There are very few measures available to researchers or practitioners who wish to assess the effect of awareness on a standardized range of individual outcomes and thereby enable comparative evaluations of different interventions.

The SAOQ was designed to address these issues in awareness measurement (Sutton, 2016). It avoids the metacognitive issues of measuring awareness by focusing on reporting the frequency with which individuals experience certain states or engage in specific behaviors. Second, as a single measure for a wide range of outcomes, it allows for the testing of theoretical distinctions between existing awareness concepts by considering their differential effects on sets of outcomes. And third, it provides a means to evaluate awareness-building interventions using a standard array of changes people experience when awareness improves.

The SAOQ was developed through a three-part process. First, qualitative analysis of responses to awareness-building workshops identified an initial list of outcomes (Sutton et al., 2015). Second, expert focus groups expanded on this list and developed thematic categories with example items. The resulting list of items was administered to a stratified sample of respondents with higher and lower awareness as measured by several awareness-related questionnaires. Finally, items were revised through factor analysis, resulting in a 38-item measure assessing awareness outcomes on four subscales. The reflective self-development subscale (RSD, 11 items, $\alpha=0.87$) captures the development of continuous attention to the self, with a focus on conscious, reflective, and balanced learning. The acceptance of self and others subscale (Acceptance, 11 items, $\alpha=0.83$) measures outcomes such as improved self-image and a deeper understanding of other people. A third scale was developed specifically to assess proactivity at work (Proactive at Work, 9 items, $\alpha=0.81$) and measures the extent to which an individual has an objective and proactive approach to dealing with work. And finally the Emotional Costs scale (7 items, $\alpha=0.77$) includes items to capture the potential negative impacts identified by people engaged in awareness development, such as guilt, vulnerability, and fear.

The SAOQ has proven valuable in distinguishing between conceptualizations of dispositional awareness, identifying their unique patterns of associations with theoretically relevant constructs (Sutton, 2016). It provided further confirmation that dispositional rumination (as measured by the RRQ, Trapnell & Campbell, 1999) is a generally damaging form of awareness as it was found to predict reduced acceptance and proactivity, as well as increased emotional costs. The SAOQ subscales also helped to distinguish between the concepts of dispositional reflection as measured by the RRQ and reflection as measured by the Self-Reflection and

Insight Scale (SRIS; Grant et al., 2002): While the former was not associated with any of the 4 SAOQ subscales, the SRIS self-reflection scale positively predicted both reflective self-development and proactivity outcomes. Insight into the self, the second subscale of the SRIS (Grant et al., 2002), was a positive predictor of acceptance only. Mindfulness, as measured by the MAAS (Brown & Ryan, 2003), was associated with increased proactivity as well as emotional costs. Somewhat similarly, increased frequency of engagement in awareness-building practices (such as meditation, talking therapy, or mindfulness practice) was positively associated with reflective self-development and acceptance but also increased emotional costs. In evaluating the effects of awareness-building interventions, therefore, the SAOQ provides a measure of both positive and negative outcomes associated with different conceptualizations of awareness.

The SAOQ is somewhat unique in its inclusion of a subscale measuring the potentially negative outcomes associated with improved awareness. These negative outcomes have received some attention in the literature (Brown et al., 2007; Zhang et al., 2013) but rarely have researchers actively tried to measure these outcomes or practitioners attempted to mitigate them. The SAOQ has demonstrated some utility in providing a balanced assessment of positive and negative outcomes in longitudinal research designed to evaluate interventions to increase participant awareness. For example, pre-post-intervention research to evaluate the impact of group coaching (Sutton & Crobach, 2022) and emotional intelligence workshops (Mounce & Culhane, 2021) both showed that carefully designed interventions are indeed able to improve the positive outcomes of awareness while protecting participants from the emotional costs.

Despite these promising results, there has also been some confusion in the research literature regarding the use of the SAOQ, with some studies attempting to use it to measure awareness itself rather than the anticipated outcomes of awareness. For example, items from the SAOQ were adapted for a study of students building a personal brand through Instagram (Wiryananta et al., 2021) and to measure student self-awareness in the use of reference management software (Avidiansyah & Kurniajaya, 2020). Although these interpretations of the measure were flawed, psychometrically, the subscales demonstrated good convergent and discriminant validity and Cronbach alphas above 0.70.

Thus far, the SAOQ has proven useful from both a theoretical and practical standpoint, addressing a need for psychometrically sound instruments to measure the effects of awareness-building practices. However, the SAOQ is fairly lengthy, with a total of 38 items. Many research projects seek to measure several variables at once, and the need for short yet robust measures is well-known (Stewart-Brown et al., 2009). This becomes even more critical in research designs

using longitudinal measures, where participants are asked to complete the same measures two or more times. To continue the development of our understanding of awareness, especially with regards to the effect of awareness-building activities, a robust instrument that can precisely measure the latent construct is needed.

A further challenge in using the SAOQ is that, in common with many psychological measures, it uses an ordinal-level measurement and hence does not satisfy fundamental assumptions of many parametric tests. Specifically, each item in the questionnaire explains a different quantity of information in the latent traits being measured, meaning that it may not be an accurate measure of that latent concept (Stucki et al., 1996). In addition, conversion of ordinal scores to interval level scales is advantageous in comparing scores across different instruments.

Rasch analysis uses a probabilistic model that accounts for unique contribution of individual items to the overall latent construct and is well suited to address these challenges (Rasch, 1960; Tennant & Conaghan, 2007). A Rasch model can improve the precision of measurement by allowing researchers to derive interval-level scores from ordinal-level responses, taking account of both the item difficulty and the person's ability (Bond & Fox, 2013). These models are increasingly used in analyzing mindfulness measures (Medvedev et al., 2016, 2017): researchers were able to refine and improve widely used measures of mindfulness by using a combination of item subsets to resolve local dependencies and discarding items that do not fit the model adequately. The increased precision of these kinds of measures has been demonstrated by comparing ordinal with Rasch-transformed scores (Norquist et al., 2004). In addition, Rasch models typically have lower standard deviations for person-parameter estimates allowing for a more fine-grained estimate of changes (Tennant & Conaghan, 2007).

In this study, therefore, we used Rasch analysis of the SAOQ to develop a short, unidimensional measure of awareness outcomes (the Awareness Outcomes Measure or AOM). The use of Rasch analysis will ensure that this measure is robust across contexts and cultures and well suited to a range of applications, including intervention and longitudinal studies. In addition, the study aims to produce ordinal-to-interval transformation tables to increase precision of measurement in intervention research.

Method

Participants

Researchers who had contacted the first author to request permission to use the SAOQ were emailed with a request to share their data for the purposes of improving and

validating the measure. Datasets which recorded participant age, sex, and country were then included in this study, resulting in a total sample of 713 participants from three countries (New Zealand, Philippines, and UK). To enable investigation of differential item functioning (DIF) in Rasch analysis, original age categories in the datasets were combined into two approximately equal-sized age categories: 18–24 and > 25 years old. Demographics are summarized in Table 1.

Measures

As noted in the introduction, the SAOQ items were developed in a three-stage process (Sutton, 2016). First, qualitative responses describing the effects a self-awareness development intervention were analyzed to produce a pool of 61 items. Second, expert focus groups categorized and expanded on this pool of items to produce a total of 83 potential outcomes of developing awareness. Finally, these items were arranged into a questionnaire, refined through factor analysis and tested for reliability and validity, to produce the 38-item SAOQ.

The SAOQ employs a 5-point Likert scale format from 1 = *never* to 5 = *almost always*, with a N/A option. There are no negatively scored items and total scores are calculated by taking the mean of each subscale, excluding the N/A items. Higher scores correspond to a higher frequency of experiencing the self-awareness outcomes. For the purposes of this Rasch analysis, all N/A scored data was coded as missing and cases with > 5% missing data were excluded. In the current dataset, Cronbach's $\alpha = 0.86$ [95% CI 0.85, 0.88] and McDonald's $\omega = 0.85$ [95% CI 0.83, 0.87], but items from the emotional costs subscale correlated negatively with the scale overall, indicating further need for development and refinement.

Data Analyses

RUMM2030 software (Andrich et al., 2009) was used to conduct Rasch analysis following the main steps outlined elsewhere (Medvedev et al., 2016, 2017). An acceptable Rasch model fit requires that individual item fit residuals do not exceed a range of -2.50 to $+2.50$ and that chi-square

statistics for the interaction between items and the latent variable, reflecting deviation of the scale data from the Rasch model, should be non-significant ($p > 0.05$). In addition, individual items should not be locally dependent, which requires evaluation of a residuals correlation matrix. Typically, a correlation magnitude of 0.20 with reference to the mean of all residual correlations is interpreted as a sign of local dependency (Christensen et al., 2016). Local dependency is often referred to as a method effect and can be addressed by creating super-items by combining individual item scores (Wainer & Kiely, 1987). Rasch models also require scale invariance across personal factors such as age, gender, or country, which is reflected by having no DIF.

Reliability in Rasch analysis is examined using the person separation index (PSI), which assesses how well a construct defined by its items discriminate between individuals' scores spread out along scale continuum. Dimensionality in Rasch analysis is typically investigated using the method developed by (Smith, 2002), which involves independent-samples *t*-test comparisons of person-estimates between the groups items with the lowest and highest factor loadings on the first principal component of the residuals. Unidimensionality requires that the percentage of significant *t*-test comparisons is not statistically significant. In other words, the number of significant *t*-test comparisons should be less than 5% or the lower bound of a binominal confidence interval calculated for the number of tests should overlap 5% (Tennant & Pallant, 2006).

Results

Rasch model fit of the initial analysis statistics including item location, fit residual, and chi-square for goodness of fit for the original 38 SAOQ items are shown in Table 2 and the overall model fit summarized in Table 3 (A1). Sixteen items demonstrated significant misfit to the Rasch model and the overall fit was also unacceptable as evidenced by significant chi-square of item-trait interaction and lack of unidimensionality. In Step 2, 6 items (4, 8, 16, 20, 24, and 28) showing significant misfit with extremely high fit residuals (> 4.00) were deleted. As this represented 6 of the 7 items contributing to the EC scale, the final item from this scale (Item 12) was also deleted to ensure internal construct validity. This resulted in increased reliability of the scale and improvement of the overall model fit but the interaction chi-square was still significant indicating a deviation from the expectations of the Rasch model (Analysis A2 in Table 3). In Step 3, further misfitting items were identified and removed (3, 6, 7, 11, 14, 23, 25, 26, 27, 29). This resulted in acceptable overall fit and good reliability but no evidence of unidimensionality, as the percentage of significant *t*-tests was 18.5% (Analysis A3 in Table 3). The residual correlation matrix was therefore examined and residual correlations

Table 1 Participant demographic data (frequencies)

		New Zealand	Philippines	UK	Total
Age	18–24	159	279	20	458
	> 25	82	0	173	255
Sex	Female	196	194	145	535
	Male	38	85	48	171
	Other/unspecified	7	0	0	7
Total		241	279	193	713

Table 2 Summary of individual item fit statistics for the initial Rasch analyses of the AOM ($n=713$)

Item	Content	Location	Fit Resid	Chi Sq
1	I learn about myself and how I see the world	-0.60	-1.85	36.05*
2	I understand my emotions	-0.02	-0.65	16.11
3	I am content with my work situation	0.53	2.37	25.99
4	I find it scary to try something new or step out of what I know	0.65	9.92*	285.23*
5	I focus on ways of amending my behaviour that would be useful	-0.19	-1.28	25.76*
6	I have fun	-0.05	1.27	9.40
7	I recognize the stress and worry in my current work	-0.53	0.48	11.65
8	I feel vulnerable	0.60	8.55*	207.65*
9	I reassess my own and others' responsibilities	-0.30	-1.32	20.04
10	I have compassion and acceptance for others	-0.69	-0.85	12.21
11	I see my work life as something I have power to affect	0.15	0.99	7.52
12	I feel my emotions deeply	-0.13	0.85	14.19
13	I'm aware of my abilities and limitations	-0.36	-0.89	15.00
14	I am objective	-0.09	1.10	5.19
15	I understand how I work within a team	-0.34	-0.25	13.54
16	I have had to revisit difficult past experiences	0.44	4.31*	48.64*
17	I "observe" myself	-0.22	-2.47	36.31*
18	I understand myself well	0.10	-1.23	28.70*
19	I can "take a step back" from situations to understand them better	-0.17	-1.99	41.05*
20	I feel exposed	0.83	6.64*	167.30*
21	I feel generally positive about self-awareness	-0.10	-1.82	37.94*
22	I am consistent in different situations or with different people	0.49	0.17	17.12
23	I think about how my personality fits with my work role	-0.18	-0.39	18.88
24	I find making changes is difficult and scary	0.62	8.62*	223.26*
25	I have insight into myself	-0.23	-2.64*	71.17*
26	I stop and think before judging	-0.24	-0.54	8.41
27	I have changed the way I work	0.29	-0.09	26.55*
28	I feel guilty for criticizing others	0.20	4.30*	52.54*
29	I look at why people act the way they do	-0.40	-1.19	14.58
30	I am confident	0.72	1.32	10.33
31	I take control of my work	-0.11	-0.29	8.49
32	I am continuing to work on and develop myself	-0.81	-2.93*	46.79*
33	I interact well with colleagues or peers	-0.14	-0.08	9.47
34	I think about how as colleagues or peers we interact with each other	-0.17	-1.66	37.95*
35	I am realistic about myself	-0.12	-1.16	22.75
36	I feel on the whole very comfortable with the way I am	0.38	-0.73	15.31
37	I am reflective	-0.43	-3.03*	36.60*
38	I have a good self-image	0.64	-0.65	30.01*

Fit Resid, item fit residual; *Chi Sq*, chi-square test for goodness of fit, * $p < 0.05$

Table 3 Summary of the overall fit statistics for the main stages of Rasch analyses of the AOM ($n=713$)

	Person mean		Goodness of fit		PSI	Unidimensionality <i>t</i> -tests	
Analyses	Value/SD		χ^2 (df)	<i>p</i>		%	Lower bound
Initial (A1)	0.98	0.54	1715.68 (342)	<0.001	0.86	38.43	36.8 (NO)
Analysis (A2)	1.20	0.82	588.74 (279)	<0.001	0.90	21.3	19.7 (NO)
Analysis (A3)	1.26	0.98	204.65 (189)	0.207	0.88	18.5	16.9 (NO)
Final (A4)	1.41	1.04	162.27 (153)	0.288	0.89	6.6	5.0 (YES)

between several items that exceeded the cut-off point of 0.20, indicating local dependency, were identified. These locally dependent items were combined into 3 super-items: 17, 31, and 36; 5 and 30; and 9 and 18. After this minor modification, the final analysis (A4) demonstrated the best fit to the Rasch model with non-significant interaction chi-square, excellent reliability, and unidimensionality. DIF was also evaluated and the scale proved invariant across individual factors including sex, age, and country. Finally, the ordinal-to-interval conversion table was constructed based on person estimates of the Rasch model to convert ordinal scale scores into interval-level data (Table 4).

Figure 1 presents the person-item plot from the final analysis, showing good coverage of sample awareness outcomes with no significant ceiling or floor effects. There are no gaps between item thresholds, reflecting excellent reliability of the scale in discriminating between awareness outcome levels of the sample. The sample abilities are relatively high reflected by person mean of 1.37 logits above the item mean set to zero, which would be expected for the healthy population sampled for the current study.

Discussion

Awareness is a central component of several psychological concepts related to well-being, including mindfulness and authenticity. Development of awareness is often a key element of interventions and training programs aiming to improve well-being, yet measurement of the effects of awareness has been hampered by the lack of a suitable measure. This study used Rasch analysis to validate a robust, unidimensional measure of these awareness outcomes. The resulting Awareness Outcomes Measure (AOM; see Supplementary Information) is an internally structurally valid and reliable measure that allows researchers and practitioners to measure awareness outcomes with improved precision by using the transformation table provided here.

The AOM consists of 21 items measuring the original three positive areas of awareness outcomes (reflective self-development, acceptance of self and others, and proactivity). These outcomes are theoretically and empirically linked to higher awareness (Mounce & Culhane, 2021; Sutton, 2016; Sutton & Crobach, 2022) and the new AOM provides researchers with short scale that can be used to evaluate the effectiveness of awareness-building interventions across contexts and cultures. In some recent studies, such as the Rasch analysis of the MAAS, disordered thresholds needed to be corrected by uniform rescoring to address limitations of this scale (Medvedev et al., 2016). However, among the AOM items, there were no significantly disordered thresholds, supporting the utility of the measure's current response format without the need for correction. The AOM is also invariant across sex, age, and

Table 4 Ordinal-to-interval Rasch conversion algorithm

Ordinal Scores	Interval		Ordinal Scores	Interval	
	Logits	Scale		Logits	Scale
21	-5.21	21.00	64	-0.08	59.39
22	-4.44	26.80	65	-0.01	59.91
23	-3.92	30.65	66	0.06	60.42
24	-3.58	33.20	67	0.13	60.95
25	-3.32	35.11	68	0.20	61.48
26	-3.12	36.67	69	0.27	62.02
27	-2.94	37.97	70	0.34	62.56
28	-2.79	39.11	71	0.42	63.12
29	-2.66	40.12	72	0.49	63.68
30	-2.53	41.03	73	0.57	64.25
31	-2.42	41.87	74	0.65	64.83
32	-2.32	42.64	75	0.73	65.43
33	-2.22	43.37	76	0.81	66.03
34	-2.13	44.06	77	0.89	66.65
35	-2.04	44.71	78	0.97	67.27
36	-1.96	45.33	79	1.06	67.91
37	-1.88	45.92	80	1.15	68.57
38	-1.80	46.50	81	1.24	69.24
39	-1.73	47.06	82	1.33	69.92
40	-1.66	47.60	83	1.42	70.62
41	-1.58	48.13	84	1.52	71.34
42	-1.52	48.65	85	1.62	72.07
43	-1.45	49.17	86	1.72	72.84
44	-1.38	49.67	87	1.82	73.61
45	-1.31	50.16	88	1.93	74.41
46	-1.25	50.66	89	2.04	75.24
47	-1.18	51.14	90	2.15	76.09
48	-1.12	51.63	91	2.27	76.98
49	-1.05	52.11	92	2.39	77.90
50	-0.99	52.59	93	2.52	78.86
51	-0.93	53.07	94	2.66	79.86
52	-0.86	53.54	95	2.80	80.93
53	-0.80	54.02	96	2.95	82.06
54	-0.73	54.50	97	3.11	83.27
55	-0.67	54.97	98	3.29	84.59
56	-0.61	55.45	99	3.48	86.03
57	-0.54	55.93	100	3.70	87.65
58	-0.48	56.42	101	3.95	89.53
59	-0.41	56.90	102	4.25	91.79
60	-0.35	57.39	103	4.64	94.68
61	-0.28	57.89	104	5.20	98.89
62	-0.22	58.38	105	6.02	105.00
63	-0.15	58.89			

three countries (New Zealand, Philippines, and UK), meaning that items do not function differently for these population groups. The AOM therefore provides a robust and promising

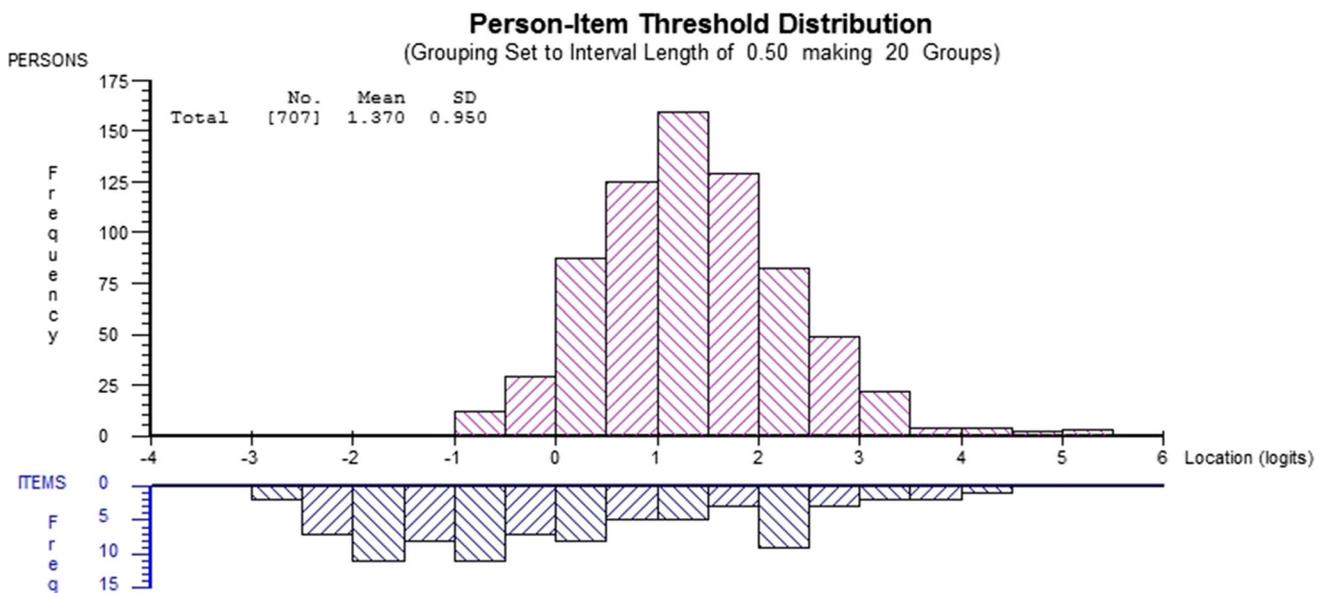


Fig. 1 Person-item threshold distribution

instrument that can be used for cross-cultural comparisons and interventions with a range of populations.

While awareness is central to constructs such as mindfulness, and interventions to improve mindfulness or well-being often attempt to develop individual awareness, until now researchers have had little recourse should they wish to evaluate the range of outcomes associated with awareness. Similarly, the development of awareness is central to many therapeutic interventions, such as person-centered counseling (Rogers, 1951) or workplace coaching, and is particularly important as a pre-requisite of self-determined behavior (Deci & Ryan, 1980). Yet the effect of awareness-building efforts on awareness outcomes themselves is rarely considered. Instead, researchers and practitioners usually seek to measure changes in mindfulness itself or in more distal outcomes such as well-being or performance. The AOM offers a robust measure of a range of outcomes associated with improved awareness, providing immediate insight into the beneficial impacts of these types of interventions.

The use of a measure focused on outcomes, i.e., the frequency of experiences or behaviors, rather than self-report of awareness, avoids the potential confounding effects of low metacognitive awareness (e.g., Dunning-Kruger effect; Dunning, 2011). It also allows researchers to conduct comparative evaluations of interventions designed to improve awareness from a range of theoretical backgrounds, whether that be mindfulness training or talking therapies. Mindfulness training, for example, is often badged and delivered in different ways, meaning that it becomes difficult to identify the mechanisms by which mindfulness may have its effects. Training may be

pitched as a way to develop mindfulness, reduce anxiety and stress, or improve well-being, and the training itself can vary in terms of the time commitment or delivery methods (Bartlett et al., 2019). The AOM can provide a means to evaluate the effectiveness of different mindfulness training programs in improving the key outcomes associated with increased awareness.

Model fit for the revised questionnaire was achieved by exclusion of some items (notably, the whole emotional costs subscale). These items were also noted as correlating negatively with other items when the overall internal consistency was calculated for this dataset. In the original SAOQ scale development, the emotional costs subscale showed substantially different relationships with measures of mindfulness and reflection (Sutton, 2016). For example, emotional costs were strongly positively correlated with both mindfulness and ruminative self-focus, while the positive SAOQ subscales showed the opposite relationships. Of some concern was the finding that increased frequency of awareness practices (such as meditation or journaling) was also associated with increased emotional costs. It has been important in previous research to identify these costs in order to ensure that interventions can demonstrate they do not cause harm and that individuals undergoing awareness-development are adequately supported by practitioners (Mounce & Culhane, 2021; Sutton & Crobach, 2022). It should therefore be noted that the AOM is a measure of the positive outcomes of awareness only and future work could consider the need for a similarly robust measure of the potential drawbacks or costs of developing awareness.

The AOM demonstrated internal structural validity, unidimensionality, and excellent reliability. These outstanding

psychometric properties met the expectations of the Rasch model, leading to the production of ordinal-to-interval transformation tables to further increase the precision of the measure. Individual AOM items were found to have varying degrees of “difficulty,” meaning that they contribute differently to the total score. This difference in individual items contributions is a common limitation of ordinal measures as they do not take account of it during total scale score calculation, thereby increasing measurement error (Bond & Fox, 2013). Therefore, using the conversion table presented in Table 4 enables the researcher to overcome limitations of ordinal scales, transform ordinal scores into interval-level data, and thus meet the assumptions for parametric statistical tests.

Limitations and Directions for Future Research

As noted above, the AOM does not include items from the original SAOQ which measure the emotional costs associated with increased awareness, and this should be borne in mind when evaluating awareness-building interventions. While the original SAOQ has been shown to be useful in distinguishing the outcomes associated with different conceptualizations of awareness and self-insight, the AOM instead focuses on evaluating a wide range of positive outcomes in a coherent manner. It includes items measuring outcomes as diverse as confidence, acceptance of others, and proactive behavior. The relationship of these outcomes with psychological concepts such as self-efficacy or conflict resolution strategies is a promising direction for future research. Additionally, as many of the items assess behaviors that may be observed by others, equivalent other-report measures may be also developed, giving researchers the opportunity to triangulate their data.

There was no ceiling or floor effect on the AOM for the sample used in this analysis, but the person mean was relatively high compared to the item mean (1.37 logits above the item mean set to zero). This could reflect an underlying interest in awareness on the part of the people who volunteered for this study, resulting in them reporting a relatively high level of experience with the positive outcomes associated with awareness. It would be beneficial to explore the properties of the scale in a clinical sample or a sample that perhaps does not equally value the importance of awareness.

Overall, the AOM has excellent reliability and meets the expectations of the unidimensional Rasch model. In addition, item-person thresholds indicate thorough coverage of awareness outcomes, reflecting excellent discrimination between awareness outcome levels. Given that the AOM was based on a thorough and wide-ranging conceptual development and has demonstrated good psychometric properties and invariance across three English-speaking samples, further work could be carried out to translate and validate the measure in other languages and populations and thereby extend its utility.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12671-022-02047-7>.

Acknowledgements The authors would like to thank Demi Amoore, Samantha Carey, and Jason Render for their assistance in data collection, as well as Niña Ubalde for generously sharing her data.

Author Contribution Anna Sutton: conceptualization; data curation; formal analysis; writing — original draft, review and editing. Oleg N. Medvedev: formal analysis; writing — review and editing.

Data Availability All data are available at the Open Science Framework <https://bit.ly/3FxlzDc>.

Declarations

Ethics Approval The study was conducted in compliance with the guidelines of the University of Waikato Human Research Ethics Committee.

Informed Consent This study utilized secondary data and authors of the primary studies indicated informed consent had been given prior to data collection.

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andrich, D. A., Sheridan, B., & Luo, G. (2009). *RUMM 2030*. RUMM Laboratory.
- Avidiansyah, Z., & Kurniajaya, J. F. (2020). Analysis of final-year students self-awareness using reference management software. *Record and Library Journal*, 6(1), 99. <https://doi.org/10.20473/rlj.v6-i1.2020.99-109>
- Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., Walsh, E., Duggan, D., & Williams, J. M. G. (2008). Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. *Assessment*, 15(3), 329–342. <https://doi.org/10.1177/1073191107313003>
- Bartlett, L., Martin, A., Neil, A. L., Memish, K., Otahal, P., Kilpatrick, M., & Sanderson, K. (2019). A systematic review and meta-analysis of workplace mindfulness training randomized controlled trials. *Journal of Occupational Health Psychology*, 24(1), 108–126. <https://doi.org/10.1037/ocp0000146>
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84(4), 822–848. <https://doi.org/10.1037/0022-3514.84.4.822>

- Brown, K. W., Ryan, R. M., & Creswell, J. D. (2007). Mindfulness: Theoretical foundations and evidence for its salutary effects. *Psychological Inquiry, 18*(4), 211–237. <https://doi.org/10.1080/10478400701598298>
- Christensen, K. B., Makransky, G., & Horton, M. (2016). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178–194. <https://doi.org/10.1177/0146621616677520>
- Christopher, K. I., Padmakumari, P., & Herbert, H. S. (2021). Presence or absence of Dunning-Kruger effect: Differences in narcissism, general self-efficacy and decision-making styles in young adults. *Current Psychology, 1*–12. <https://doi.org/10.1007/s12144-021-01461-9>
- Deci, E. L., & Ryan, R. M. (1980). Self-determination theory: When mind mediates behavior. *Journal of Mind and Behavior, 1*(1), 33–43.
- Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (vol. 44, pp. 247–296). Academic Press. <https://doi.org/10.1016/B978-0-12-385522-0.00005-6>
- Fletcher, C., & Bailey, C. (2003). Assessing self-awareness: Some issues and methods. *Journal of Managerial Psychology, 18*(5), 395–404. <https://doi.org/10.1108/02683940310484008>
- 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*(8), 821–835. <https://doi.org/10.2224/sbp.2002.30.8.821>
- Grossman, P. (2019). On the porosity of subject and object in 'mindfulness' scientific study: Challenges to 'scientific' construction, operationalization and measurement of mindfulness. *Current Opinion in Psychology, 28*, 102–107. <https://doi.org/10.1016/j.copsyc.2018.11.008>
- Jamieson, S. D., & Tuckey, M. R. (2017). Mindfulness interventions in the workplace: A critique of the current state of the literature. *Journal of Occupational Health Psychology, 22*(2), 180–193. <https://doi.org/10.1037/ocp0000048>
- Kernis, M. H., & Goldman, B. M. (2006). A multicomponent conceptualization of authenticity: Theory and research. In P. Z. Mark (Ed.), *Advances in Experimental Social Psychology, 38*, 283–357. [https://doi.org/10.1016/S0065-2601\(06\)38006-9](https://doi.org/10.1016/S0065-2601(06)38006-9). Academic Press.
- Knoll, M., Meyer, B., Kroemer, N. B., & Schröder-Abé, M. (2015). It takes two to be yourself. *Journal of Individual Differences, 36*(1), 38–53. <https://doi.org/10.1027/1614-0001/a000153>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Medvedev, O. N., Siegert, R. J., Feng, X. J., Billington, D. R., Jang, J. Y., & Krägeloh, C. U. (2016). Measuring trait mindfulness: How to improve the precision of the Mindful Attention Awareness Scale using a Rasch model. *Mindfulness, 7*(2), 384–395. <https://doi.org/10.1007/s12671-015-0454-z>
- Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2017). Improving the precision of the Five Facet Mindfulness Questionnaire using a Rasch approach. *Mindfulness, 8*(4), 995–1008. <https://doi.org/10.1007/s12671-016-0676-8>
- Mounce, M., & Culhane, N. (2021). Utilization of an emotional intelligence workshop to enhance student pharmacists' self-awareness. *Currents in Pharmacy Teaching and Learning, 13*(11), 1478–1483. <https://doi.org/10.1016/j.cptl.2021.09.001>
- Norquist, J. M., Fitzpatrick, R., Dawson, J., & Jenkinson, C. (2004). Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Medical Care, 42*(1), 25. <https://doi.org/10.1097/01.mlr.0000103530.13056.88>
- Querstret, D., Morison, L., Dickinson, S., Cropley, M., & John, M. (2020). Mindfulness-based stress reduction and mindfulness-based cognitive therapy for psychological health and well-being in nonclinical samples: A systematic review and meta-analysis. *International Journal of Stress Management, 27*(4), 394–411. <https://doi.org/10.1037/str0000165>
- Rasch, G. (1960). *Probabilistic models for some intelligence tests and attainment tests*. Danish Institute for Educational Research.
- Rogers, C. R. (1951). *Client-centered therapy - Its current practice, implications, and theory*. Houghton Mifflin Company.
- Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205–231. <http://www.ncbi.nlm.nih.gov/pubmed/12011501>
- Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. *Health and Quality of Life Outcomes, 7*(1), 15. <https://doi.org/10.1186/1477-7525-7-15>
- Stucki, G., Daltroy, L., Katz, J. N., Johannesson, M., & Liang, M. H. (1996). Interpretation of change scores in ordinal clinical scales and health status measures: The whole may not equal the sum of the parts. *Journal of Clinical Epidemiology, 49*(7), 711–717. [https://doi.org/10.1016/0895-4356\(96\)00016-9](https://doi.org/10.1016/0895-4356(96)00016-9)
- Sutton, A. (2016). Measuring the effects of self-awareness: Construction of the Self-Awareness Outcomes Questionnaire. *Europe's Journal of Psychology, 12*(4), 645–658. <https://doi.org/10.5964/ejop.v12i4.1178>
- Sutton, A. (2020). Living the good life: A meta-analysis of authenticity, well-being and engagement. *Personality and Individual Differences, 153*, 109645. <https://doi.org/10.1016/j.paid.2019.109645>
- Sutton, A., & Crobach, C. (2022). Improving self-awareness and engagement through group coaching. *International Journal of Evidence Based Coaching and Mentoring, 20*(1), 35–49. <https://doi.org/10.24384/dqft-9x16>
- Sutton, A., Williams, H. M., & Allinson, C. W. (2015). A longitudinal, mixed method evaluation of self-awareness training in the workplace. *European Journal of Training and Development, 39*(7), 610–627. <https://doi.org/10.1108/EJTD-04-2015-0031>
- Tennant, A., & Pallant, J. F. (2006). Unidimensionality matters! (a tale of two Smiths?). *Rasch Measurement Transactions, 20*, 1048–1051.
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism, 57*(8), 1358–1362. <https://doi.org/10.1002/art.23108>
- 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*(2), 284–304. <https://doi.org/10.1037/0022-3514.76.2.284>
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24*(3), 185–201. <https://doi.org/10.1111/j.1745-3984.1987.tb00274.x>
- Wiryananta, K., Safitri, R., & Prasetyo, B. D. (2021). The importance of self-awareness in building personal branding of Brawijaya University students through Instagram. *Technium Social Sciences Journal, 24*, 218–227. <https://doi.org/10.47577/tssj.v24i1.4790>
- Wood, A. M. L., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. *Journal of Counseling Psychology, 55*(3), 385–399. <https://doi.org/10.1037/0022-0167.55.3.385>
- Zhang, J., Ding, W., Li, Y., & Wu, C. (2013). Task complexity matters: The influence of trait mindfulness on task and safety performance of nuclear power plant operators. *Personality and Individual Differences, 55*(4), 433–439. <https://doi.org/10.1016/j.paid.2013.04.004>