REPLY

Out of the Armchair and Into the Streets: Measuring Mindfulness Advances Knowledge and Improves Interventions: Reply to Grossman (2011)

Kirk Warren Brown  
Virginia Commonwealth University

Richard M. Ryan  
University of Rochester

Tamara M. Loverich  
Eastern Michigan University

Gina M. Biegel  
Stressed Teens Program, San Jose, California

Angela Marie West  
Central Michigan University

We address 3 critiques raised by Grossman (2011) of self-report measures of mindfulness and the Mindful Attention Awareness Scale (MAAS) and Mindful Attention Awareness Scale—Adolescent (MAAS–A) in particular. Grossman questioned whether self-report measures actually assess mindfulness, whether the construct of mindfulness can be understood apart from mindfulness training, and whether there is empirical evidence to support the validity of mindfulness measures. In response we discuss established theory that attention (and secondarily meta-awareness) is core to the meaning of mindfulness and is the central feature of the MAAS and MAAS–A. We then argue that mindfulness is an inherent capacity that varies between and within persons and is not, as Grossman claimed, a concept applicable to only a trained few. Further, as assessed by the MAAS and MAAS–A, mindfulness is associated with the same variety of outcomes as mindfulness training is theorized to yield. Finally, we provide considerable evidence that the MAAS and MAAS–A are valid instruments. We conclude that although construct measurement is inevitably imperfect, such efforts are critical to building basic knowledge and refining effective interventions.

Keywords: Mindful Attention Awareness Scales, mindfulness construct, validity

We welcome the opportunity to respond to Grossman’s (2011) critique of self-report measures of mindfulness and the Mindful Attention Awareness Scales—MAAS and MAAS—Adolescent (MAAS–A)—in particular. It is common in science for new developments to attract criticism, whether that be based on fidelity to a construct or a variety of other reasons. Indeed, criticism surrounding new advances is not limited to science. Instructive in this regard is the public criticism that attended the introduction of the phonograph in the late 19th century. Perhaps most prominently, John Philip Sousa pointed out that not only were recordings inferior to live music but that because music could now be heard privately in the home and the same “performance” could be heard over and over, the listening experience was cheapened. From the perspective of the 21st century we know that both sides of this debate were correct on specific points. The detractors of recorded music were correct that early musical recordings were inferior to live music; such recordings were not “true to life,” as a 1916 advertisement for the Victrola phonograph put it. Yet the proponents of recorded music were also correct in claiming value in those early efforts, which have in many ways extended the reach of music and facilitated its learning and appreciation beyond a privileged few.

Similarly, Grossman (2011) argued that the assessment of mindfulness through self-report measures has distorted, denatured, and cheapened the mindfulness construct. In fact, most of Grossman’s complaints boil down to two central concerns, one methodological and one substantive. First, he is concerned that researchers are muddying the mindfulness phenomenon by applying impoverished indicators and overly simplified indicators. He illustrated his concern with a casual content analysis of the scale. In this, Grossman appeared to put “face validity” at the center of the science of assessment, which
we see as methodologically regressive, and he failed to acknowledge the greater importance of construct validity and the normative network of findings that underpins it. Second, Grossman does not think mindfulness should be assessed in laypersons; instead, he would restrict mindfulness research to focus only on individuals with extensive mindfulness training. In this argument Grossman assumed a substantive view of mindfulness as only a product of training, rather than a natural and varying capacity of humans, and this too is a position with which we take issue.

In what follows we address the three key questions entailed in Grossman’s (2011) argument (p. 1035) that underpin his objections to current measures:

1. Are self-report scales inventing their own definitions of mindfulness?
2. Is the process leading to a denaturing and decontextualization of the Buddhist construct?
3. What is the empirical basis for the measurement of mindfulness?

We address each of these questions in turn, before turning to our own conclusions about the current status and future of mindfulness research.

Are Self-Report Scales Inventing Their Own Definitions of Mindfulness?

Grossman (2011) presented four descriptions of mindfulness from esteemed Buddhist scholars and teachers and then presented five items from the MAAS in an attempt to show that these items do not reflect the richly nuanced nature of mindfulness and that, on the face of it, the items composing the scale are redefining mindfulness. We see two issues with the contrast that Grossman made. First, it is paradoxical that in an outlet specializing in modern assessment and methodology, face validity is made such a central issue in construct validity. In Grossman’s critique, items on a measure must appear to fit expert descriptions, explicitly and straightforwardly. In contrast we suggest that the MAAS and MAAS–A—collectively, MAAS(–A)—as well as other instruments, cannot be judged merely by appearances as to whether they adequately assess a construct. This is better determined by other, statistical forms of validity evidence that we address shortly.

The second issue with Grossman’s claim that self-report measures of mindfulness are reinventing mindfulness is that, as we noted in our article on the MAAS–A, there is considerable scholarly and clinical literature discussing mindfulness as a construct that primarily concerns attention and secondarily concerns meta-awareness. These properties appeared in a recent content analysis of the modern (post-1950) scholarly literature in Theravada Buddhism, which devotes considerable attention to mindfulness (Brown & Goodman, 2011). They are the very properties tapped by the MAAS(–A), as well. Also characteristic of modern Buddhist discourse and practice is a focus on simplicity of expression of mindfulness in daily life (McMahon, 2008); this is also seen in the MAAS(–A). In our development of the MAAS it was important that the measure not contain items about compassion, empathy, lack of bias, and other attributes that are often associated with more mindful states. Indeed, we appreciate these attributes as typically following from or precipitated within open and receptive attention. To include them as explicit items potentially creates criterion contamination, confusing predictors and outcomes, and could ultimately hold back rather than advance our understanding of these processes.

Is the Process Leading to a Denaturing and Decontextualization of the Buddhist Construct?

Grossman’s concerns about the reinvention of mindfulness are related to his argument that self-report measures of mindfulness are denaturing and decontextualizing the Buddhist construct. Grossman (2011) noted that “mindfulness, within [the] Buddhist perspective, is an active, investigative practice or process that inherently involves cognitive, attitudinal, affective, and even social and ethical dimensions” (p. 1035). This argument implies that mindfulness is not an ordinary state of mind and cannot be disassociated either from the practice of it or from the training in it. He further noted that the MAAS and other scales rely on self-reports made during ordinary states of awareness by individuals who have not necessarily acquired any form of mindfulness training, who are not performing a deliberate act of paying attention, . . . and who may be involved in a very different kind of paying attention, marked by high levels of judgmentalness and low levels of patience, tolerance, or kindness. (Grossman, 2011, p. 1035)

We agree with Grossman (2011) that self-report measures of mindfulness have been decontextualized from the Buddhist origins of mindfulness training, which was primarily done through monastic and retreat residency marked by the social, ethical, and other dimensions that Grossman suggests are inherent to mindfulness. But is mindfulness a rarified state open only to those undergoing such training? We suggest not, on three counts. First, we argue—as have others cited in our target article (Brown, West, Loverich, & Biegel, 2011)—that mindfulness is an inherent human capacity, one in which there are meaningful individual differences. We certainly agree with Grossman that this capacity can be enhanced and cultivated through training, but this is in no way inconsistent with seeing it as a naturally occurring phenomenon as well.

Mindfulness, as Jon Kabat-Zinn has suggested, is Buddhist in the same way that gravity is Newtonian. That is, Buddhism has pointed to rather than invented a human capacity. Whether mindfulness is indeed an inherent capacity, available to regular people and “beginners,” or can be understood only by those extensively and traditionally trained may be a matter of debate. It is, however, worth noting that the mindfulness-based interventions (MBIs) that Grossman (2011) discussed (and himself conducts) represent decontextualized training, in that they are completely secular in nature and carry none of the traditional Buddhist ethical, moral, or cultural dimensions. When it comes to capacities for mindfulness, we all start somewhere.

A second reason we argue that mindfulness is not a rarified state is that measures of mindfulness, such as the MAAS(–A), designed for untrained respondents, correlate with the very criteria that mindfulness theory and practices indicate are the expected outcomes of this quality of attention. Included in such outcomes are several of the psychological features that Grossman, in the quote cited earlier in this section, indicated are associated only with mindfulness practice. In particular, the MAAS has been linked
with higher levels of acceptance (less judgmentality; Brown & Ryan, 2004), more tolerance (Niemiec et al, 2010), greater self-kindness (Shapiro, Brown, Thoresen, & Plante, 2011), and more empathy for others (Beitel, Ferrer, & Cecero, 2005).

These are not simply bivariate correlations. For example, Niemiec et al. (2010) showed that the more tolerant and compassionate behaviors that were associated with MAAS-assessed mindfulness in their experiments resulted from a more receptive processing of momentary threat. Thus, the deeper experiential processing associated with higher MAAS scores helped to explain how the outcomes of compassion and tolerance emerged. The MAAS, that is, predicted the processes as well as outcomes expected in mindfulness theory.

Finally, although much of both Buddhist and popular literature on mindfulness focuses on its training, the real issue concerning mindfulness and its value for individuals and society is whether it can be manifest in everyday life. If there were no expression of this quality in daily life, which is what the MAAS(–A) and other mindfulness scales are designed to tap, there would be little interest in the phenomenon. Indeed, this distinction is important because it has implications for the definitions and descriptions of mindfulness. It is not difficult to imagine that mindfulness as expressed in the protected environment of a training session will be different than how it is expressed “on the street.” In the former, the focus is typically on internal experience—observing the arising and passing of thoughts and emotions, for example. But in day-to-day life, attention is directed both inward and outward at different times, and the MAAS(–A) measures were designed, in part, to reflect this contextually bound variation in the direction of mindfulness. Can we be mindful while driving the car or doing the laundry? One hopes the answer is yes, more or less. Can we measure this? The answer is yes, with good validity, as we detail next.

Are the MAAS and MAAS–A Valid Instruments?

Grossman (2011) raised numerous concerns about the validity of the MAAS(–A), and we address each in turn.

Appropriateness of Samples

To begin, we respond to the point that the scale was validated on inappropriate samples. This point reflects Grossman’s (2011) abiding concern that assessments of mindfulness in use for untrained populations do not capture the deeper meaning of the phenomenon. We readily admit that the MAAS and MAAS–A, like other published measures of mindfulness in common use, were constructed for use with individuals in normative and clinical populations, including those untrained in mindfulness. They are intended for the common person. Because the MAAS(–A) measures were designed primarily for respondents untrained in mindfulness, the meaning of mindfulness assessed is rather basic, as we noted in our target article. Grossman (2011) expressed confusion that respondents can score at the upper end of the Likert scale range, but this should not be surprising. Higher scores reflect good psychometric practice, in which scores on each item vary across the full range. However, these scores must be understood as reflecting the basic form of mindfulness that the scales measure. As we noted in our target article, it is important that measures of mindfulness be developed for use among populations highly trained in mindfulness to better capture the upper range of expression of this quality of attention. However, we think that even for advanced Buddhist practitioners, the items of the MAAS(–A) still apply, and even the most developed meditators will vary in what the scales measure over time.

Content Validity

Grossman (2011) argued that the MAAS(–A) items inadequately cover the range of expression of mindfulness, as understood in theory and practice. This issue harkens back to the point, already made, that the scales were designed for the layperson, not mindfulness experts, so the more subtle manifestations of mindfulness cannot be directly or explicitly assessed. It is notable, however, that the development of the MAAS was rooted in Buddhist and MBI writings on mindfulness, received input from teachers and senior students of mindfulness, and began with a large item pool (N = 184). As is typical in scale validation efforts, this pool was reduced to those items that performed most effectively with the target population. Full details on the validation of the MAAS are available in Brown and Ryan (2003).

Moreover, there is a difference between having a nuanced and detailed definition of a construct and believing that you best measure it by including those nuances explicitly in self-report items. There is also confusion in Grossman’s (2011) definitions, which often conflate the outcomes and products of mindfulness with open and receptive attention per se. Although a more extended argument might be warranted, the central problem goes back to an excessive reliance on face validity in Grossman’s critique. Both he and we understand that mindfulness is complex, but Grossman expects complexity to be directly apparent in the items used to tap the construct. But since the 1950s psychometricians have distinguished content and construct validity, knowing that it is the latter upon which a science is built. In our view, returning to the days of face validity as a central criterion for measurement is a step back in time and in psychological science.

Construct Validity

Grossman did, however, raise several points concerning the construct validity of the MAAS(–A). Most pointedly, in a discussion of external referents for the scales, he argued that “there lack clear external referents, or gold-standard measures, with which to define a mindful person. Therefore, no possibility currently exists to assess whether these questionnaire measures accurately reflect mindfulness or something else” (Grossman, 2011, p. 1035). We agree with Grossman that there are no pure “external referents” available for this psychological state, leaving only two possible courses of action: (a) give up on the validation of mindfulness measures until such referents are available or (b) let go of the perceived importance of unambiguous external referents, for the time being, to establish the construct validity of mindfulness measures. The second course of action is more tenable and has clear precedent in the history of self-report scale validation. As psychometrics expert Phil Shaver noted (personal communication, April 30, 2010):

The concept of “construct validity” (e.g., Cronbach & Meehl, 1955) was developed for theoretical constructs for which there is no hard
and fast “criterion,” as is the case with many psychological and social scientific constructs. Over the years, many self-report measures devised to tap theoretical constructs have proven very useful in tests and elaboration of key theories (e.g., about self-efficacy, depression, attachment security, and locus of control). Very few research-generating theories and measures in psychology have a single “gold standard” criterion; if they did, the criteria could be used directly and the self-report measures wouldn’t be necessary.

It is important to recognize that all scales, including ours, imperfectly measure the constructs at which they aim. As Loevinger (1957) long ago argued, such measures are tools of a developing science, never finished products. But we fear that if Grossman’s philosophy of science were to be broadly applied, we would have no measures of complex constructs in psychology at all. Instead we would all be waiting for external referents of psychological states to materialize. Yet it is through the very imperfect attempts of researchers who climb out of their armchairs and do the hard work of validation and measurement that we learn more and more about phenomena such as mindfulness, and it is only through such efforts that we will, over time, refine both our basic science and our efforts to apply it for the benefit of all.

Also note that the use of self-report measures of mindfulness does not preclude other assessment strategies, including others’ reports on a subject’s level of mindfulness (e.g., Carlson, Livingston, & Vazire, 2011), computer-mediated attention tasks, and so on. But given the subjective nature of the phenomenon, self-reports of mindfulness offer a useful window into individual states of mind. A parallel example is the assessment of pain, which, like mindfulness, is a largely subjective phenomenon. After years of work in this domain, a gold standard for the measuring of pain is still self-report (e.g., Melzack, 1975), and studies of self-reported pain have uncovered correlates with social, physiological, and behavioral measures. As already discussed, this is occurring in the realm of mindfulness assessment as well.

Convergent Validity

Grossman (2011) noted that the MAAS is rather inconsistently related to other mindfulness scales. Yet the evidence indicates that we are not in a “blind men and the elephant” scenario in which each measure is describing something different. Articles validating mindfulness scales that include the MAAS have quite consistently shown convergence with this measure (e.g., Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006). Given the varying theoretical bases and dimensionalities of these scales, it is not surprising that the degree of correlation varies from scale to scale. Intercorrelations between measures of the same construct have varied widely in other, established domains of inquiry as well (e.g., self-report measures of self-esteem; Bosson, Swann, & Pennebaker, 2000). Unquestionably, intercorrelations between measures of mindfulness could be higher. But all the extant measures represent early attempts at assessing mindfulness, and first-generation measures can be expected to show lower than desirable convergent validity.

Criterion Validity

Grossman (2011) argued that the MAAS lacks appropriate criterion validity, as it has not been shown to predict behavioral outcomes. Yet, as we pointed out in our article, the MAAS has been shown to predict a number of theoretically consistent behavioral outcomes. In fact, over the past 8 years of research on the MAAS, it has been associated with subjective, objective behavioral, physiological, and neurological outcomes in ways highly consistent with the theoretical position that mindfulness supports more adaptive functioning. Such diverse evidence supports the criterion validity of the MAAS, and the triangulation of findings across subjective, behavioral, and biological measures offers confidence that the instrument reflects veridical experience. The evidence also helps to refute Grossman’s claim that the MAAS assesses a construct indistinguishable from attention lapse. He did not specify how this latter construct is operationalized, but measures such as the Attention-Related Cognitive Errors Scale (ARCES; Cheyne, Carriere, & Smilek, 2006), which purportedly assesses attention lapse, have mainly been linked with attention-based outcomes (e.g., memory, boredom proneness), not the wide range of outcomes that the MAAS has been shown to predict.

Semantic Item Interpretation Across Groups

Grossman (2011) argued that it is unlikely that items on self-report scales of mindfulness are understood the same way across populations of interest. Grossman’s claim rests on armchair conjecture, as he offered no evidence for the presumed inequivalence of semantic item interpretation across groups. However, the MAAS was validated with a variety of populations within the United States. Recent evidence (Christopher, Charoensuk, Gilbert, Neary, & Pearce, 2009) indicates semantic equivalence of the MAAS items across American and Thai populations (the latter being a culture with greater exposure to mindfulness and its practice), and we know that more efforts devoted to cross-cultural comparisons of mindfulness measures are under way. But our point is that these armchair claims by Grossman are testable, and that is so only because we have a measure.

Semantic Item Interpretation Across Time

Grossman (2011) stated that those trained in mindfulness will understand the items on mindfulness scales differently than will those who are untrained. Again, he provided no evidence to support this claim; it simply comes from his personal speculations. But here again we believe it is an empirical question that deserves testing. Fortunately, we have instruments that could be tested for such equivalence across trained and untrained groups, whereas if we were not to develop measures, there would be no opportunities to rule in or rule out such speculations.

Grossman’s conviction that mindfulness is accessible only to the trained underlies this aspect of his critique. Beyond the dichotomous view of mindfulness he espouses (some people have it and some do not), there are other issues with this position. In psychological research to date, what we call individuals “trained” in mindfulness are those who complete an MBI (typically, 8 weeks of training). Such people have just begun to learn about the application of mindfulness, and there is little reason to suspect that the meaning of the scale items will have changed in such a short period, especially considering that mindfulness training can extend over years, even decades. Grossman’s argument begs the question, Where is the cutoff between trained practitioners and persons to whom mindfulness measures should not be applied a priori?
In our target article we too have called for measures that apply to advanced practitioners of mindfulness, not because we think mindfulness is exclusively in their possession but because we may be able to learn more by studying this exemplary population with more detailed instruments. Nonetheless, the MAAS has been successfully applied to experienced students of meditation, as well as naive persons (see, e.g., Brown & Ryan, 2003). Again, if the measure works differently in trained people, this is something to be demonstrated and would itself be a matter of interest in how mindfulness develops over time. Instead, Grossman would have his speculation preclude such inquiries while we wait for external referents to befall us.

**Scale Scores Conflation of Competencies With Valuations of Importance**

This question reflects Grossman’s (2011) suggestion that participants recently completing an MBI are more likely to endorse items on the MAAS(–A) than at baseline. Once again, Grossman provided no evidence to support this claim. Yet, it is common for researchers to use self-report measures pre- and postpsychosocial intervention. Moreover, participants do not necessarily know what each scale they complete actually measures because the scales are administered as part of a battery of psychological measures, some of which are quite similar to each other. Also, even if participants were showing some social desirability or other response bias (see Brown & Ryan, 2003, for evidence to the contrary), it seems hard to explain why in Study 2 of the target article, and in other articles (e.g., Shapiro et al., 2011), the elevations in mindfulness scores are sustained from posttreatment to treatment follow-up points months (even a year) later, when any enthusiastic flush from the training has likely gone. It is also notable that scores on self-report scales can differ significantly at follow-up from those at posttest in an unanticipated direction, as we found recently on the Freiburg Mindfulness Inventory measure of mindfulness (Brown, Kasser, Linley, Ryan, & Orzich, 2009). This suggests to us that without evidence to support a competing claim, participants appear to be completing scales without significantly biasing their responses in some way.

Finally, Grossman’s (2011) claim concerning response biases toward valuations of importance does not address the fact that changes in mindfulness in Study 2 of the target article, as in other studies (e.g., Brown & Ryan, 2003, Study 5; Shapiro et al., 2011), were related to changes in mental health outcomes over time in unexpected directions. We think it unlikely that adult and adolescent participants were manipulating not only MAAS scale responses but also the relations between the MAAS and other scale responses.

In sum, it is very common for participants in psychosocial interventions to complete psychological measures before and after training, and there is expectation that at posttest people will have more insight into their behavior that may affect responses to those measures. This is not technically an issue of semantic inequivalence but rather of so-called secular (time and repeated administration) trends in construct validation (Loevinger, 1957). We have found that MAAS scores are quite strongly correlated between pre- and postintervention time points (e.g., Shapiro et al., 2011), which suggests that people interpret the scale items similarly across time; there is also evidence that change scores have systematically predicted subsequent outcomes.

**Conclusions: Can and Should Mindfulness Be Assessed?**

We have argued, as have others, that the capacity for mindfulness is inherent to the human psyche and that this capacity can be validly (albeit imperfectly) measured by the MAAS(–A). The MAAS has been shown to predict numerous psychological outcomes, and research is accumulating to show that the scale predicts overt behavior related to attention and behavior regulation. Research is also uncovering the cognitive processes and neural correlates that may help to explain the benefits of MAAS-assessed dispositional mindfulness. The study of mindfulness has become a full-fledged scientific enterprise within the past 30 years, and the assessment of this construct through self-reports has, we believe, been helpful for the science to progress. Without such assessment researchers and clinicians cannot know whether mindfulness is in fact trained in MBIs, whether mindfulness improves over time, and whether mindfulness is related to outcomes of interest, be they biological or functional in nature.

Grossman (2011) suggests that qualitative and other approaches be taken to delineate the psychological mechanisms and characteristics of mindfulness practice, which may result in the uncovering of novel effects associated with mindfulness and the interventions designed to enhance it. But he suggests that until concrete “external” criteria for mindfulness emerge, any form of quantitative assessment of the construct is unhelpful and misleading. Eliminating such efforts, however, makes it impossible to know whether mindfulness is an active ingredient in MBI and other mindfulness training programs, even when imperfectly measured. It also eliminates the body of knowledge gained as researchers attempt to define, operationalize, and quantitatively compare psychological constructs of interest.

Yet we share Grossman’s concern about fidelity to the concept of mindfulness in measurement development. Measurement of a new phenomenon often begins crudely, while both thinking about the phenomenon and methods used to assess it develop over time. We note at the beginning of this article that opinions about the recording of music in the late 19th century were strongly divided. To detractors, the reproduction was crude, and allowing people to take music into their living rooms only cheapened the experience of music listening. Improvements in recording technology have made music recordings much truer to life than in their early days, and yet people still appreciate live music. In like fashion we suggest that the assessment of mindfulness has room for significant improvement (we may have a Victrola!), but the evidence for the validity of the MAAS(–A) presented here and elsewhere indicates that efforts to develop this measure have already had payoffs in terms of our scientific understanding of processes and outcomes associated with mindfulness.

**References**


Received March 17, 2011
Revision received June 22, 2011
Accepted August 15, 2011