

Construct validity

1

Construct validity and vulnerability to anxiety: A cognitive interviewing study of the revised Anxiety Sensitivity Index

Gary P. Brown, Ph.D.

Royal Holloway University of London

Nick C. Hawkes, DClInPsy

Barnet, Enfield and Haringey Mental Health NHS Trust

Philip Tata, CPsychol

Central and Northwest London Mental Health NHS Trust and Royal Holloway

University of London

Running Head: CONSTRUCT VALIDITY AND VULNERABILITY TO ANXIETY

Keywords: Anxiety; Vulnerability; Construct Validity; Cognitive Therapy

Address for correspondence: Gary Brown, Psychology Department, Royal Holloway University of London, Egham, Surrey, TW20 0EX, United Kingdom.
Gary.Brown@rhul.ac.uk. Telephone: +44-1784-443-522; Fax: +44-1784-434-347

Abstract

The Anxiety Sensitivity Index (ASI; Reiss, Peterson, Gursky, & McNally, 1986) is probably the most widely used measure of cognitive vulnerability to anxiety. However, there have been periodic doubts expressed about whether it measures beliefs about the negative consequences of anxiety symptoms, as it purports to, or actual anxiety experiences. The present study investigated the construct validity of the revised ASI using a cognitive interviewing approach. Sixteen outpatients with anxiety problems responded to ASI-R items and items from another measure of vulnerability to anxiety, the Anxiety Attitude and Belief Scale, while thinking aloud. The resulting verbal protocols were coded according to the apparent cognitive processes respondents engaged in when answering the items. Responses to the revised ASI-R more often entailed retrieval of past episodes of anxiety, and participants more frequently formulated their responses based on judgments of the occurrence or intensity of feelings rather than on the appraisal of anticipated consequences of what was described in the items. These findings potentially have significant implications for interpretation of results from the large body of literature using the different versions of the ASI.

Research into cognitive behavioral therapy (CBT) and its model of psychopathology (Beck, 1976; Beck & Emery, 1985) has embraced a wide variety of methodologies but, in common with much of psychology, has had a particular affinity for questionnaire-based research. This has been the basis of a number of criticisms (Brown, MacLeod, Tata, & Goddard, 2002; Coyne & Gotlib, 1983; Hammen & Krantz, 1985; Hollon & Bemis, 1981; Segal, 1988). To different degrees, these criticisms share a concern about whether scores on putative cognitive measures have been shown conclusively to reflect variations in the underlying target phenomena or, indeed, whether they actually measure cognition. This focus echoes recent trends in measurement theory (Embretson, 1983; Borsboom, Mellenbergh, & Van Heerden, 2004) that seek to address the limitations of the accepted notions of construct validity. In particular, these authors have questioned sufficiency of the prevailing paradigm, within which the validity of an instrument rests on establishing a pattern of findings that is consistent with the relevant theory (epitomized by Cronbach and Meehl's, 1955, so-called "nomological net"). Meanwhile, within the growing methodology subfield known as cognitive aspects of survey methodology (CASM; Lessler, Tourangeau, & Salter, 1989) the technique of cognitive interviewing has been developed to identify the mechanisms involved in responding to self-report instruments to help ensure that these plausibly reflect the intended underlying target phenomena.

The central concern of the critics of the classical notion of construct validity is its susceptibility to inferential ambiguity, particularly lack of a basis for distinguishing

the preferred interpretation of a pattern of associations from a less preferred one that explains the pattern just as well. Within the CBT literature, it was precisely criticisms along these lines that were at the heart of an energetic debate concerning the Anxiety Sensitivity Index (ASI; Reiss, Peterson, Gursky, & McNally, 1986), one of the most widely used measures of cognitive vulnerability to anxiety. Anxiety sensitivity is defined as the fear of anxiety symptoms based on beliefs about their harmful consequences. The psychometric soundness of the ASI has been amply documented (Cox, Borger, & Enns, 1999). It correlates with and predicts measures of fear and panic; for example, it was shown to predict the development of panic attacks following stressful military training (Schmidt, Lerew, & Jackson, 1997). The ASI is associated in particular with agoraphobia and panic disorder, in which the experience of the fear reaction itself is the major source of distress (Reiss, 1991). However, in the course of interchanges with ASI researchers, Lilienfeld and colleagues made a number of cogent criticisms of the ASI that raise significant questions about its validity:

1. Questionable face validity. Lilienfeld, Jacob, and Turner (1989, p. 100) noted that “inspection of the item content of the ASI reveals that virtually all of the items appear to tap fear of anxiety and of anxiety symptoms (e.g., ‘It scares me when I am nervous,’ ‘When I notice that my heart is beating rapidly, I worry that I might have a heart attack’), rather than beliefs concerning the negative consequences of anxiety, as claimed by the test’s developers. [The authors] in effect make the unsubstantiated assumption that individuals who fear the consequences of anxiety

necessarily possess cognitions that anxiety has harmful consequences.”

2. Content overlap with criterion variables. “[B]ecause the word “scare” or “scares” appears in eight of the ASI’s 16 items, a more parsimonious explanation for the partial correlation between the ASI and the [Fear Survey Schedule II] is that one measure of fear is highly associated with another measure of fear.” (Lilienfeld, Turner, & Jacob, 1993., p. 168). With regard to comparisons of ASI scores between panic and other diagnostic groups, they state “because many of the ASI’s items assess symptoms that are already known to be prevalent among panic disordered patients, such comparisons are not especially informative vis-a-vis the ASI’s construct validity and provide little or no new information regarding panic disorder” (Lilienfeld, Turner, & Jacob, 1996, p. 413).

3. Double-barreled items. There is no way of knowing from a low score on an ASI item if the respondent does not often experience the sensation in question or if they experience it but are not scared of it. “This could produce a spurious correlation between the ASI and panic disorder (as well as similar criteria), because panic disorder patients are more likely than other subjects to experience anxiety-related symptoms... Indeed, because many of the items on the ASI refer explicitly to panic symptoms, it may be this shared content, rather than the AS construct per se, that is primarily responsible for the ASI’s relation to panic disorder and related criteria” (Lilienfeld et al., 1993, pp. 166-167).

Arguments as to the dimensionality of the ASI and its relationship to trait anxiety also featured in this debate. However, the narrower criticisms summarized above pertaining to the basis upon which ASI scores vary are more fundamental and would potentially make these further considerations moot.

The response of some ASI researchers has been largely to reassert the intended purpose of the ASI, namely to assess beliefs about the consequences of anxiety rather than the frequency of anxiety, without providing evidence that the ASI is a valid means for meeting this aim (Taylor, 1996, p. 433; Reiss, 1997, p. 208).

However, other ASI researchers have acknowledged the aptness of these criticisms. Thus, McNally (1999, p. 10) has concurred that the ASI does not, on its face, appear to assess beliefs, and that whether or not it does so in practice is a valid empirical question. Similarly, Cox et al. (1999) have stated that “the question remains as to whether the ASI items assess beliefs independent of occurrences of relevant fear experiences” (p. 143).

The present study was undertaken to investigate these aspects of the construct validity of an expanded version of the ASI, the ASI-R. The ASI-R (Taylor & Cox, 1998a) is a superset of the original ASI consisting of ten of the original 16 ASI items along with 26 novel items. The authors' motivation in developing the ASI-R was to have available a large enough item set to settle the question of the dimensionality of the AS construct. The instructions and the structure and wording of the items remained the same as for the original ASI. The same can be said of

the more recent ASI-3 (Taylor et al., 2007), a verbatim subset of the ASI-R which also retains the same instructions. As such, issues and criticisms related to the ASI are equally relevant to the ASI-R and ASI-3.

While originally developed to explain panic disorder, the ASI-R is now more typically used across the range of anxiety disorders and, indeed, Axis I disorders in general (Schmidt et al, 2007; Taylor et al., 2007). As such, the sixteen outpatients included in the present study presented with a variety of anxiety-related problems. Similarly, consistent with the considerable attention paid in the ASI literature to the question of uniformity of measurement across symptomatic and asymptomatic populations (Deacon, Abramowitz, Woods, & Tolin, 2003; Taylor et al., 2007), participants ranged across different stages of treatment. Participants responded to ASI items and items from another measure of vulnerability to anxiety, the Anxiety Attitude and Belief Scale (AABS; Brown, Craske, Rassovsky, Tata, & Tsao, 2000), a scale developed with the aim of indexing beliefs independent of affect. A cognitive interviewing procedure was used, and the resulting verbal protocols were coded according to the apparent cognitive processes respondents engaged in when answering the items. The information provided permitted a direct test of the central assumption underlying the validity of the ASI, namely that it is a measure of beliefs rather than predominantly a measure of intensity and frequency of prior anxiety experiences. The methodology also permitted a more general examination of whether participants were responding as intended to the ASI instructions.

Participants

The 16 participants (10 female, 6 male) were patients of an adult psychology outpatient department and were included if their primary presenting problem was anxiety. ICD-10 diagnostic classifications given by their clinician were: agoraphobia without panic disorder (N = 1), agoraphobia with panic disorder (N = 2), social phobia (N = 4), other anxiety disorder (N = 3), panic disorder (N = 4), generalized anxiety disorder (N = 1), and obsessive compulsive disorder (N = 1). Participants were interviewed after assessment (N = 7), or during (N = 4) or after (N = 5) individual or group treatment. The sample was mostly White (N = 12) with one Black participant and three declining to state their ethnicity). Mean age was 37.5 (SD = 12.8, range = 21 to 64 years). Potential participants were excluded if their English was not fluent or if they had cognitive deficits (head injury or learning disability). Participation was voluntary. Participants were not compensated, although travel expenses were reimbursed.

Measures

Anxiety Sensitivity Index – Revised (ASI-R). According to Reiss's expectancy theory, anxiety sensitivity (AS) is the "fear of fear," said to arise from the belief that the experience of anxiety has negative consequences, including illness, embarrassment or additional anxiety (Reiss, 1991). The Anxiety Sensitivity Index aims to index such beliefs. AS is supposed to amplify fear and anxiety reactions, playing a role in the etiology and maintenance of anxiety disorders, especially panic disorder and agoraphobia. The large body of research using the ASI has

been extensively reviewed (Cox et al, 1999; McNally, 1999; Taylor, 1996). The ASI-R is an expanded Anxiety Sensitivity Index, consisting of 36 items. The respondent is asked to rate how applicable the item is to them in terms of their agreement, from *very little* to *very much*, with what the item states.

Anxiety Attitude and Belief Scale (AABS). The AABS is a 58-item scale that aims to index beliefs and attitudes which constitute a cognitive vulnerability to anxiety (Brown et al., 2000). Items were derived from the literature on cognitive-behavioral approaches to anxiety disorders and refined by a poll of 17 researchers prominent in the area. Items are worded so that they do not refer to or presuppose the occurrence of anxiety symptoms. Participants indicate endorsement on a seven point Likert scale from *Totally disagree* to *Totally agree*. Brown et al. found three underlying dimension of the AABS: Vigilance-Avoidance, Catastrophizing, and Imagination. The scale and factor subscales were reliable, with alpha coefficients ranging between .72 and .82 for the factors and .87 for the scale as a whole.

Procedure

Sampling of Items. The time-intensive nature of the think aloud protocols collected in the course of cognitive interviewing precluded having all participants answer all items. Therefore, in order that each item was completed by four participants, each participant completed approximately one quarter of the items: nine (of the 36) randomly selected ASI-R items and 15 (of the 58) AABS items. This produced a total of 144 item protocols for the ASI-R and 240 item protocols

for AABS.

Collection of cognitive interviews

The interview schedule was constructed based on recommendations for conducting cognitive interviews (Willis, 2005) and think-aloud protocol analysis (Green & Gilhooly, 1996). Participants were first familiarized and permitted to become comfortable with the task of thinking aloud. They were randomly assigned to complete either ASI-R or AABS items first. Interviewer input was limited to prompting to think aloud if silent, and the interviewer sat out of view to avoid influencing participants. Interviews were recorded and the recordings were transcribed for analysis. Following the think aloud task, participants were debriefed and further information was gathered about their experience of answering the questionnaire items.

Analytic approach

As a framework for analyzing the resulting protocols, Chi's (1997) verbal analysis method was adapted. According to Chi, "this quantitative-based qualitative approach basically operationalizes one's subjective impression by coding the verbal evidence for that impression and comparing the frequencies of the codes quantitatively" (Chi, 1997, p. 277). Chi's approach offers a generalization of Ericsson and Simon's (1984) protocol analysis method to less highly structured, "real world" contexts. She outlines eight basic steps that are followed to a greater or lesser extent or omitted depending on the aims and subject matter of the research in question. The following three of Chi's steps were seen as relevant to

the present study: segmenting the protocols, developing or choosing a coding scheme or formalism, and operationalizing evidence in the coded protocols. Each step is described in detail, below.

Segmenting of protocols. Verbal protocols were transcribed and the entire protocol for each participant was segmented. Segmentation was based on the identification of separate thoughts, ideas, or cognitive processes within the protocols. Independent blind coders segmented transcripts for five respondents using initial criteria. Four undergraduate students served as raters and followed a basic set of written instructions directing them to identify segments within the protocols representing separate ideas or cognitive processes. Mean inter-coder segmentation agreement was 81%. Disagreements were reconciled by one of the authors (NCH), the criteria were clarified, and a single coder segmented the remaining transcripts accordingly.

Developing the coding scheme

Using content analysis (Krippendorff, 1980), as applied to think-aloud protocols (Ericsson & Simon, 1993; Green & Gilhooly, 1996), an iterative procedure was followed to develop category codes for describing the cognitive processes reported. The starting point was the standard set of stages employed within the cognitive interviewing approach to analyze the process of responding to self report questions: comprehension, retrieval, judgment formation, response selection, and verification/editing (Tourangeau, 1984). Likewise, the initial codes applied were

drawn from the conventional nomenclature within the CASM paradigm (Tourangeau, 1984) and from existing cognitive interviewing codes used in a previous questionnaire evaluation study (Bickhart & Felcher, 1996). The codes were initially applied to the first two participants' protocols. Two or more codes could be applied to a single segment when more than one distinct process was clearly indicated in that segment. Disagreements and uncoded segments were discussed and codes added or adjusted accordingly. Existing concepts and labels were used if they matched well with those obtained from the data; however, novel sub-codes were developed where necessary to better capture the specific anxiety-related content. Resulting codes were re-applied to the first two protocols and refined further. The final set of codes was then applied to the entire set of protocols from all participants. Disagreements were resolved in discussion between the two raters where possible, and a third rater arbitrated unresolved disagreements.

Categories were denoted by a three-digit code. The first digit referred to the stage of processing taking place in the segment, the second to the specific process being employed, and the third to the content of the segment. The first two digits reflect the standard CASM codes and the last the more refined categorization related to the specific content of the questionnaires. Thus, a code of 221 would be used to refer to a segment in which the stage of processing was retrieval (2 in the first digit related to stage of processing), an episode from memory was being recalled (2 in the second digit related to the specific process) and that episode concerned the

occurrence of anxiety related to the self (1 in the third digit related to content). A code of 241 would also pertain to retrieval of anxiety information about the self (first code and last codes again 2 and 1), but the 4 rather than 2 for the middle digit denoted recall of general information rather than recall of a specific episode. All codes are listed in Table 1.

Insert Table 1 around here

Mean inter-coder reliability for the final codes for the initial blind coding was 71% agreement. This was felt to reflect a sufficient consensus between raters in light of the fact that agreement was defined strictly as complete agreement on all codes, when there were up to 47 possible codes, and it was possible to apply more than one to each segment. If segments where both coders used the same code but one also used an additional code were counted as agreements, reliability rose to 82%. Agreement rate on process only (with potential disagreement on the content of that process) was 87%. Agreement was 91% on stage of processing. The two initial coders resolved 98% of codes, and the remaining 2% were arbitrated by a third rater.

Results

Operationalizing evidence in the coded protocols

Comprehension. Certain responses can indicate potential difficulties for respondents in understanding or complying with what is being requested of them. Three of these were focused upon. Responses categorized as “recasting” entailed respondents changing the wording of an item before answering. For example, one respondent before responding to the ASI-R item “When I feel dizzy, I worry there is

something wrong with my brain” stated “No. I worry there’s something wrong with my blood pressure.” In “specifying” the respondent applies limiting conditions to their response: “I totally agree with that one. *For myself, anyway.*” Finally, a respondent might note in some way that an item is ambiguous to them. Neither the ASI-R or AABS showed excessive comprehension problems, with recasting occurring in 10.4% and 8.3% of protocols, respectively, and either specifying or ambiguity occurring in 12.5% of protocols for each scale.

A further evaluation of the comprehension of and adherence to the ASI-R instructions was carried out through examination of the retrieval codes “inapplicable” (210) and “hypothetical” (codes in the 23x range). According to the ASI-R instructions, if an item describes a situation that is not applicable to the respondent, they are to answer hypothetically, as if the situation in question was applicable. Twelve of the 16 respondents at some point stated that an ASI-R item was not applicable to them (this did not occur for any AABS protocols), and this took place in 24 of the 144 ASI-R item protocols (16.7%). Hypothetical codes co-occurred with inapplicable codes for 17 (70.8%) of the 24 protocols, suggesting that respondents were following instructions for the most part. However, it also suggests that a small but appreciable percentage of responses (N = 7, or about 5% of all ASI-R protocols) were not being made on any apparent valid basis, as the item was inapplicable but a hypothetical response was not verbalized.

Retrieval. The prior debates regarding the validity of the ASI reviewed in the

introduction reflect a consensus on both sides that a scale operationalizing the cognitive component of the CBT theory of emotional disorders should mainly vary on the basis of representation and appraisal processes and not on the basis of retrieval of experiences of an affective state, especially if that affective state is meant to be predicted by the scale in question. In fact, to access and report a belief, it should not be necessary to retrieve instances from autobiographical episodic memory, and such retrievals should arguably only occur in a minority of responses. In this regard, one of the authors of the ASI has stated, "...anxiety sensitivity predicts future fearfulness based on the degree of endorsement of the beliefs assessed by the ASI, regardless of the frequency or the intensity of anxiety experiences in the past...'past experiences of anxiety' and 'beliefs about the consequences of anxiety' are different phenomena..." (Reiss, 1997, p. 208). To evaluate the ASI-R and AABS in this respect, self-relevant anxiety retrievals (codes 221 and 241) were tallied, representing segments in which respondents recalled specific episodes or summary knowledge of their own past anxiety experiences. Twenty-five ASI-R protocols (17.4%) and 15 AABS protocols (6.3%) contained anxiety retrievals. The difference in proportions was significant, $\chi^2(1, 384) = p < .001$, corrected $p = .021$.¹ Twenty (56%) of the 36 ASI-R items and 11 (19%) of the 58 AABS items produced self-relevant anxiety retrievals; this difference in proportions was also significant ($\chi^2(1, 94) = 7.37, p < .01$).

Judgment Formation. Judgment formation refers to whatever processing is carried out to formulate the required response following comprehension of the item and

retrieval of any material from memory. Even after retrieving memories of anxiety episodes, ASI respondents might still, as ASI researchers assume, largely rely on appraisals about the consequences of anxiety to develop their responses.

Appraisals identified in the current study fell into three categories: reasoning (i.e., logical inferences, code 311), imperatives (e.g., “should” type judgments, code 312), and arbitrary conclusions (propositional statements made without any specified basis, code 313). The remaining non-appraisal judgment formations were based on the reported occurrence, intensity, or frequency of cognitions (340 codes), behaviors (320 codes), or feelings (340 codes) within the respondent’s experience. In practice, only feelings produced intensity-based judgment formations. Examples of segments assigned each judgment formation code are shown in Table 2.

Insert Table 2 about here

The distinction between appraisal and cognition occurrence codes primarily hinged on whether the respondent made a verifiable appraisal while answering the question (e.g., “I don’t think there is anything wrong with me”) or was relating the past occurrence of a cognition (“I don’t usually think, ‘There is something wrong with me.’”). The majority of AABS protocols (52.9%) but only a minority of ASI-R protocols (20.8%) contained appraisals (see Table 3), suggesting that the AABS is more accurately characterized as a measure of appraisals than the ASI-R.

Cognition occurrence codes were used for 10.4% of the AABS protocols and 20.8% of the ASI-R protocols. Either appraisal or cognition codes were used in 59.2% of the AABS protocols and 38.9% of the ASI-R protocols.

Judgment formations apart from those intended by the designers of a scale may or may not be relevant to a scale's validity. For example, judgments made on the basis of the perceived occurrence of behaviors would not represent an obvious confound for a scale meant to predict anxiety, especially if these did not predominate. Consistent with this, respondents relied in modest proportions on estimation of the occurrence of behaviors in forming their answers (3.5% for the ASI-R and 6.3% for the AABS). In contrast, a substantial proportion of feeling-based judgments in a scale intended to predict future affect, even if these do not predominate, confounds predictor and criterion and therefore seriously threatens validity. Such was the case for the ASI-R, with feelings-based judgment formations (both occurrence and intensity) being identified in the majority (56.9%) of protocols as compared to a minority (12.1%) of AABS protocols. The ASI-R instructions do not specify whether frequency or intensity is intended to be the basis for respondent judgments of how true an item is for them; however, according to the breakdown in Table 3, it is clear that respondents largely based their judgments on the perceived frequency of the occurrence of feelings. Twenty of 58 AABS items produced feeling-based judgments, whereas nearly all (34 of 36) ASI-R items did. The difference in proportions was significant ($X^2(1,94) = 32.67, p < .001$).

As McNally (1999, p. 10) has noted, AS researchers assume that statements such as "It scares me when my heart beats rapidly," which appear to be concerned with

affective reactions, imply beliefs about the negative consequences of the phenomenon in question (here, rapid heartbeat). Although explicit appraisals occurred in only 30 of 144 ASI-R protocols, AS researchers might assume that feelings-based judgment formations indirectly reflect implicit appraisals of this sort. As such, associations between feelings-based judgment formations and explicit appraisals were examined, under the assumption that the rate of explicit appraisals might mirror the rate of implicit appraisals. In fact, explicit appraisals were found to be just as likely to co-occur with feelings-based judgment formations (16 of 30 protocols) as with their absence (14/30), suggesting no apparent relationship between feeling-based judgment formations and stated or unstated appraisals.

Insert Table 3 about here

Response selection. While the operations involved in answering self-report items are conceived of as forming successive stages, these stages do not necessarily occur in the assumed sequence of comprehension, retrieval, judgment, and response selection, and particular stages are sometimes omitted. Which stages appear in protocols and their order can provide further information about an instrument. For instance, relatively late response selection may reflect effortful processing that does not draw on immediately available online knowledge and can indicate problems with wording, comprehensibility, or applicability (e.g., Bassili & Scott, 1996). Conversely, it is argued within the protocol analysis and cognitive interviewing literatures (Conrad, Blair & Tracy, 1999) that beliefs and appraisals draw on semantic knowledge and so should be directly reportable without retrieval from episodic memory. Therefore, the sequencing of response selection relative to

other codes was considered for the ASI-R and AABS-R. As shown in Table 4, response selection was typically the last operation carried out for ASI-R items (54.2% of protocols), following all other operations. For the AABS, response selections more frequently occurred either as the sole reported cognitive process (25.4% of protocols) or prior to all other codes (39.2% of protocols), $\chi^2(3, N = 384) = 28.23$, corrected $p = .017$, which is more consistent with the report of beliefs drawn from semantic memory than the retrieval of experiences from episodic memory.

Insert Table 4 about here

Discussion

The ASI is a popular measure, widely used beyond its initially intended purpose of assessing fear of fear. However, periodic concerns have been raised about its construct validity and, in particular, whether it measures what it purports to measure, namely beliefs about the negative consequences of anxiety symptoms. In contrast to the usual strategy taken to analyzing construct validity, in which this is inferred from a pattern of associations, the present study examined in a more directly empirical fashion the reports of the thought processes actually engaged in by respondents answering the scale items in order to develop an idea of what accounts for variation in ASI-R scores. Compared to the responses to the AABS, a measure constructed with the aim of assessing beliefs independent of affect, responses to the ASI-R more often entailed retrieval of past episodes of anxiety. Moreover, ASI-R respondents frequently formulated their responses based on judgments of the occurrence or intensity of feelings rather than on the appraisal of

anticipated consequences of what was described in the items.

The authors of the ASI sought to develop a measure of individual differences in the appraisal of body sensations, particularly anxiety symptoms. However, rather than writing items that directly inquire about appraisals of these sensations (e.g., “A racing heart is a sign of something seriously wrong.”), the authors relied on items concerning the amplified reaction thought to follow from the appraisal (“It scares me when my heart beats rapidly”). The fact that the implied appraisal occurred somewhere in the response sequence was taken for granted. The self report of the amplified reaction that was assumed to result from the appraisal was seen as sufficient evidence of its operation. Indeed, it is generally reasonable to infer from someone’s reaction that they have appraised a situation as threatening, for why else would they be scared? It is the reasonableness of this type of inference that underpins the ASI-R’s face validity for appraisals. However, the root of the ASI-R’s potential shortcomings stems from the fact that it is simultaneously, and arguably more straightforwardly, face valid as a measure of affect.

Although it may be reasonable to attribute a person’s affective reaction on a given occasion to an implied appraisal, there is little basis for assuming that this is what ASI-R respondents are referring to when they provide their responses or that scores reflect the operation of such appraisal processes. To infer that ratings provided by respondents to the ASI reflect their global degree of belief in the implied appraisal would require an assumption that either (1) the rating of affective

response (how “scared” the respondent would be) is necessarily isometric with the respondent’s typical degree of belief in the underlying appraisal, (2) that the respondent intuits that they are to disregard the affective response highlighted by the item wording and to focus instead on whatever appraisals they typically make so as to make a rating of degree of belief rather than degree of affect, or (3) that relevant appraisals will arise spontaneously and inevitably in the course of responding as requested. The first two possibilities are implausible, with the second additionally contradicted by the current results: most often, respondents based their ASI-R responses not on the degree to which they held a belief about the dangerousness of the body sensation in question, but rather based on the intensity or perceived rate of occurrence of their affective (“scares me”) response. Of the listed alternatives, the current results offer qualified support for the third possibility. It does appear that appraisals played a role in an appreciable minority of ASI-R protocols. However, their appearance was sporadic and often secondary to the affect based judgment formations that appeared in nearly twice the number of protocols as appraisals.

Furthermore, it was found that respondents frequently formulated ratings not just on the basis of the intensity of their remembered affective reactions but also by judging how often such reactions occurred. On the basis of a review of the literature on emotional self report, Robinson and Clore (2002) concluded that global judgments such as these are likely to arise when the referent of an item (fainting, heart racing) is something with which the respondent has not had

immediate or frequent experience. Respondents asked to estimate emotional responses that are not sufficiently immediate do not base their responses on situation-specific beliefs tied to discrete memories (which is what the ASI-R assumes) but instead base them on identity-related beliefs more reflective of the individual's perception of their overall temperament and personality. In these circumstances, responses on the ASI-R are likely to be removed further still from any putative underlying appraisal processes.

The foregoing highlights the fact that items on the ASI-R presuppose previous experience of the body sensations in question or, in the absence of previous experience, the ability to evaluate these hypothetically. This contributes to a number of difficulties, particularly when items are inapplicable, as was the case for a substantial proportion of items even within the current anxiety disorder sample. The ASI-R instructions specifically direct respondents to answer inapplicable items hypothetically. In most cases in the current study an inapplicable code (210) was accompanied by a hypothetical retrieval code (230), suggesting that the respondents were following instructions in these cases for the most part (although whether the hypothetical pertained to the inapplicable aspect of the item was not verified). Still, 5% of items were inapplicable without an accompanying hypothetical. Although this is a small proportion, a mathematical simulation study has shown that it is large enough to produce serious distortions in psychometric analyses. Waller (1989) studied the general case in which items are inapplicable because they share an unmet prerequisite and found that as small an inapplicable

rate as 5% can lead to substantially inflated correlations between items and the resulting extraction of spurious or distorted factors.

The issue of greater or lesser applicability has potentially far-reaching consequences for all versions of the ASI. Schwarz and colleagues (see Schwarz, 1999 for a review) have shown that individuals responding to items that are less applicable to them are much more likely to be affected by the context, sequencing, and wording of items than individuals for whom the subject matter is more continually accessible, resulting in greater instability of measurement in the former populations and lack of measurement invariance across populations. Consistent with this, Deacon, et al. (2003) found in a factor analysis of the revised ASI-R cross-validated in two large undergraduate samples that items concerned with somatic sensations loaded on factors that appeared to differ on the basis of item wording (being scared of anxiety symptoms versus being worried about the potentially catastrophic consequences of such symptoms) as compared to Taylor and Cox's (1998b) original analysis in a clinical population, in which items loaded according to domain of somatic sensation (cardiovascular, gastrointestinal, etc.). Deacon et al. attribute this to the inapplicability of these items in a non-clinical sample: "it is possible that individuals without clinically significant anxiety symptoms may have difficulty identifying specific feared consequences of somatic sensations, even when these sensations are feared" (p. 1446). Indeed, it has proven difficult to identify a stable factor structure of the ASI-R within non-clinical populations, let alone one that is related to the structure found in clinical

populations (Zvolensky et al., 2003). The potential for the factor scores of the ASI variants to be unstable and substantially based on spurious statistical artifacts is a serious concern, and particularly so for the ASI-R, as arguments for its validity are in large part based upon studies of its factor structure (Taylor & Cox, 1998a; Taylor & Cox, 1998b; Taylor et al., 2007; Zinbarg, Barlow, & Brown, 1997).

The present study is based on the premise that it is important for a cognitive measure to reflect variance in cognitive mechanisms. As self evident as this might seem to most, it is reasonable to ask whether it is critical that actual cognitions are demonstrated to underlie responses to the ASI-R. Glass and Arnkoff (1997) have suggested that, whether or not self-report inventories yield a veridical report of actual internal dialogue, the evidence is that clients are still conveying something clinically meaningful about themselves (p. 917). In a similar vein, it could be argued that because a large body of literature appears to document that the ASI-R “works,” that is, that it shows relevant predictive validity, the details of its underlying mechanics are secondary. However, this is an argument based on the quantity of the evidence rather than on the sensibleness of the evidence. Furthermore, if the present findings are given credence, it would follow that much of the previous literature should be reconsidered in light of the fact that the ASI-R appears to be at least equally a measure of affective as cognitive processes.

A natural place to start a reconsideration of ASI research would be the study by Schmidt et al. (1997) of 1,014 US air force cadets in basic training, as this study

represented a watershed for the ASI. Schmidt et al found that the ASI predicted future panic attacks after accounting for panic attack history. It also predicted future anxiety after controlling for current levels of anxiety and trait anxiety. The latter finding was regarded as addressing one of the criticisms of the ASI put forth by Lillienfeld and colleagues with respect to the potential confounding of the ASI with trait anxiety. However, to regard these findings as evidence supporting the distinctness of AS from trait anxiety depends on acceptance of the preferred interpretation of the incremental variance explained by the ASI as being due to cognitive appraisal. The possibility that the incremental variance is due to anxiety measured by the ASI that is not shared with the trait anxiety measure is equally consistent with the pattern of results. A conclusive choice of one explanation over the other requires evidence from a separate source about the mechanisms underlying ASI responses. As Borsboom and colleagues (2004) state: "...tables of correlations between test scores and other measures cannot provide more than circumstantial evidence for validity. What needs to be tested is not a theory about the relation between the attribute measured and other attributes but a theory of response behavior. Somewhere in the chain of events that occurs between item administration and item response, the measured attribute must play a causal role in determining what value the measurement's outcomes will take; otherwise, the test cannot be valid for measuring the attribute" (p. 1062). The present study appears to contradict the prevailing assumptions about what processes underlie ASI-R scores.

The clearest potential limitation of the current study is that raters were not blind to the study hypothesis. This was necessary in order to establish the rating system. The raters strove to remain impartial; however, subsequent research using this approach will be more convincing if blind raters are used. In addition, the sample size employed, while actually fairly large for a protocol analysis study, could be seen as limiting generalizability. However, the magnitude of the effects found makes it unlikely that the results would be different with substantially different participants. For example, in order to reduce the nearly 60% affect judgment formation rate of the ASI-R to 25%, twelve participants who provided 100% cognitive judgment formations for ASI-R items would need to be added. To further reduce this to the 10% affect rate found for the AABS, 59 such participants would need to be added.

Another potential objection is to the make-up of the sample, which ranged across the anxiety disorders, with a minority of six of 16 participants receiving the ASI-R target diagnosis of panic disorder and some participants having advanced at least partially through psychotherapy. However, as noted in the introduction, a sample limited only to panic disorder would not represent the full range of populations in which ASI-R research is typically carried out. Whether responding affirmatively to having experienced a specific body sensation, as many with panic are likely to do, or responding negatively, as those with other disorders or, for example, undergraduates are likely to do, responses need to be made on a valid basis reflecting variance (including presence *and* absence) in the underlying

phenomenon of interest. Indeed, as discussed above, it is precisely when items are less applicable that the ASI-R's validity is most suspect. It is not evident how proscribing use of the ASI-R outside of a symptomatic panic disorder population would obviate these shortcomings.

The present study sought to examine the fundamental validity of the ASI-R, an instrument considered a gold standard in its area whose validity has typically been taken for granted. It is especially important to scrutinize a scale like the ASI-R in the manner of this study as the built-in confounding of predictor and criterion, as appears to be the case with the ASI, is a particularly insidious validity threat that will likely obscure rather than advance the research involved. The problems with all the versions of the ASI appear to stem from the unnecessarily indirect manner in which its authors chose to measure appraisal of body sensations. It is likely that a more straightforward approach will advance the field more unambiguously.

Footnotes

¹ Chi-square analyses involving protocols violated the usual assumption of independence of observations. The appropriate correction was therefore carried out for this and all subsequent analyses at the level of protocols (rather than at the level of participants or items) using the SPSS Complex Samples procedure. The stated statistics and significance levels reflect this correction.

References

- Bassili, J. N., & Scott, B. S. (1996). Response latency as a signal to question problems in survey research. *Public Opinion Quarterly*, *60*, 390-399.
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. New York: International Universities Press.
- Beck, A. T., & Emery, G. (1985). *Anxiety disorders and phobias: A cognitive perspective*. New York: Basic Books.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061-1071.
- Brown, G. P., Craske, M. G., & Rassovsky, Y., Tsao, J. C. I., & Tata, P. (2000). The Anxiety Attitude and Belief Scale: Initial psychometric properties in an undergraduate sample. *Clinical Psychology and Psychotherapy*, *7*, 230-239.
- Brown, G. P., MacLeod, A. K., Tata, P., & Goddard, L. (2002). Worry and the simulation of future outcomes. *Anxiety, Stress & Coping*, *15*, 1-17.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, *6*, 271-315.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319.
- Conrad, F., Blair, J., & Tracy, E. (1999). *Verbal reports are data! A theoretical approach to cognitive interviews*. Presented at 1999 Federal Committee on Statistical Methodology (FCSM) Research Conference. Washington DC: Office of Management and Budget.
- Cox, B. J., Borger, S. C., & Enns, M. W. (1999). Anxiety sensitivity and emotional

- disorders: Psychometric studies and their theoretical implications. In S. Taylor (Ed.), *Anxiety sensitivity: Theory, research, and treatment of the fear of anxiety. The LEA series in personality and clinical psychology* (pp. 115-148). Mahwah, NJ, US: Lawrence Erlbaum Associates.
- Cox, B. J., Parker, J. D. A., & Swinson, R. P. (1996). Anxiety sensitivity: Confirmatory evidence for a multidimensional construct. *Behaviour Research and Therapy, 34*, 591-598.
- Coyne, J. C., & Gotlib, I. H. (1983). The role of cognition in depression: A critical appraisal. *Psychological Bulletin, 94*, 472-505.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- Deacon, B. J., Abramowitz, J. S., Woods, C. M., & Tolin, D. F. (2003). The anxiety sensitivity index-revised: Psychometric properties and factor structure in two nonclinical samples. *Behaviour Research and Therapy, 41*, 1427-1449.
- Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179-197.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.
- Glass, C. R., & Arnkoff, D. B. (1997). Questionnaire methods of cognitive self-statement assessment. *Journal of Consulting and Clinical Psychology, 65*, 911-927.
- Green, C., & Gilhooly, K. (1996). Protocol analysis: Practical implementation. In J. Richardson (Ed.), *Handbook of Qualitative Research Methods for Psychology*

- and the Social Sciences* (pp. 55-74). British Psychological Society: Leicester.
- Hammen, C. L., & Krantz, S. E. (1985). Measures of psychological process in depression. In E.E. Beckham and W.R. Leber (Eds.), *Handbook of depression: Treatment, assessment, and research* (pp. 408-444). Homewood, IL: The Dorsey Press.
- Hollon, S. D., & Bemis, K. M. (1981). Self-report and the assessment of cognitive functions. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook* (pp. 125-174). New York: Pergamon.
- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Newbury Park, CA: Sage.
- Lessler, J., Tourangeau, R., & Salter, W. (1989). Questionnaire design in the cognitive research laboratory: Results of an experimental prototype. National Center for Health Statistics. *Vital Health Statistics*, 6. Washington, DC: US Government Printing Office.
- Lilienfeld, S. O., Turner, S., & Jacob, R. (1993). Anxiety sensitivity: An examination of theoretical and methodological issues. *Advances in Behaviour Research and Therapy*, 15, 147-183.
- Lilienfeld, S. O., Jacob, R. G., & Turner, S. M. (1989). Comment on Holloway and McNally's (1987) "Effects of anxiety sensitivity on the response to hyperventilation". *Journal of Abnormal Psychology*, 98, 100-102.
- Lilienfeld, S. O., Turner, S. M., & Jacob, R. G. (1996). Further comments on the nature and measurement of anxiety sensitivity: A reply to Taylor (1995b). *Journal of Anxiety Disorders*, 10, 411-424.

- McNally, R. J. (1999). Theoretical approaches to fear of anxiety. In S. Taylor (Ed.), *Anxiety Sensitivity: Theory, research, and treatment of the fear of anxiety* (pp. 3-16). Mahwah, New Jersey: Lawrence Erlbaum.
- Reiss, S. (1991). Expectancy theory of fear, anxiety, and panic. *Clinical Psychology Review, 11*, 141–153.
- Reiss, S. (1997). Trait anxiety: It's not what you think it is. *Journal of Anxiety Disorders, 11*, 201-214.
- Reiss, S., Peterson, R. A., Gursky, D. M., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behaviour Research and Therapy, 24*, 1-8.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*, 934–960
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.
- Taylor, S., Zvolensky, M. J., Cox, B. J., Deacon, B., Heimberg, R. G., Ledley, D. R., et al. (2007). Robust dimensions of anxiety sensitivity: Development and initial validation of the Anxiety Sensitivity Index-3. *Psychological Assessment, 19*, 176-88.
- Taylor, S. (1996). Nature and measurement of anxiety sensitivity: Reply to Lilienfeld, Turner, and Jacob (1996). *Journal of Anxiety Disorders, 10*, 425-451.
- Taylor, S., & Cox, B. J. (1998). An expanded anxiety sensitivity index: Evidence for

a hierarchic structure in a clinical sample. *Journal of Anxiety Disorders*, 12, 463-483.

Taylor, S., & Cox, B. J. (1998b). Anxiety sensitivity: Multiple dimensions and hierarchic structure. *Behaviour Research and Therapy*, 36, 37-51.

Tourangeau, R. (1984). Cognitive Sciences and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau, (Eds). *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp 73-100). Washington DC: National Academy Press.

Waller, N. G. (1989). The effect of inapplicable item responses on the structure of behavioral checklist data: A cautionary note. *Multivariate Behavioral Research*, 24, 125-134.

Willis, G. (2005), *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage.

Zinbarg, R. E., Barlow, D. H., & Brown, T. A. (1997). Hierarchical structure and general factor saturation of the anxiety sensitivity index: Evidence and implications. *Psychological Assessment*, 9, 277-84.

Zvolensky, M. J., Arrindell, W. A., Taylor, S., Bouvard, M., Cox, B. J., & Stewart, S. H., et al. (2003). Anxiety sensitivity in six countries. *Behaviour Research and Therapy*, 41, 841-859.

Table 1

The coding system

Process	Content
<u>Stage: Comprehension</u>	
110 Restates question	
120 Difficulty assessment	121 Easy 122 Difficult
130 Thinking	
140 Changes question	141 Recasting 142 Specifying
150 Ambiguity	
160 Re-reads instructions	
<u>Stage: Retrieval</u>	
210 Inapplicable	
220 Recall episodes	221 Recall episodes, self, anxiety-relevant 222 Recall episodes, self, anything else 223 Recall episodes, others, anxiety-related 224 Recall episodes, others, anything else
230 Hypothetical	231 Hypothetical, anxiety-relevant 232 Hypothetical, anything else
240 General knowledge	241 General knowledge, self, anxiety-relevant 242 General knowledge, self, anything else 243 General knowledge, others, anxiety-relevant 244 General knowledge, others, anything else 245 General knowledge about anything else

Process	Content
<u>Stage: Judgment formation</u>	
310 Appraisal	311 Reasoning 312 Imperative 313 Arbitrary conclusion
320 Behavior occurrence	321 Behavior, non-zero frequency 322 Behavior, zero frequency 323 Behavior, positive qualitative frequency assessment 324 Behavior, negative qualitative frequency assessment
330 Feeling occurrence	331 Feeling, non-zero frequency 332 Feeling, zero frequency 333 Feeling, positive qualitative frequency assessment 334 Feeling, negative qualitative frequency assessment 335 Feeling, positive qualitative intensity assessment 336 Feeling, negative qualitative intensity assessment
340 Cognition occurrence	341 Cognition, non-zero frequency 342 Cognition, zero frequency 343 Cognition, positive qualitative frequency assessment 344 Cognition, negative qualitative frequency assessment
<u>Stage: Response selection</u>	
410 Select direction	411 Select direction, agreement 412 Select direction, disagreement
420 Select response	
<u>Stage: Verify/Edit</u>	
510 Judge social desirability	
520 Justify response	
530 Certainty assessment	531 Certain 532 Uncertain
600 Other	610 Miscellaneous 620 Uncodable

Table 2

Examples of judgment formations from the verbal protocols

Judgment formation (code)	Item	Verbal protocol (showing segmentation)
Appraisal: Reasoning (311)	There's a high probability of rejection in most social situations. (AABS).	"...Depends on how you approach the social situation. / <i>And if you went into a social situation imagining you'll be rejected, / then actually you could probably make that happen.</i> "
Appraisal: Imperative (312)	When I begin to sweat in a social situation I fear people will think negatively of me. (ASI-R)	"Yes, because I was when I was young I used to get a lot of anxiety sweating / um and I used to hate it / um and find it terribly embarrassing / and I don't yeah and I, / <i>I would not want people to see it / so I agree with that very much / although it doesn't tend to happen now but it used to.</i> "
Appraisal: Arbitrary conclusion (313)	It is important for me not to appear nervous. (ASI-R)	"Um, it is to a certain degree. / <i>Although, I've always been nervous, / I know I appear nervous / and it is actually fine / so although I would like to appear less nervous I think I'm quite comfortable appearing just the way I am, / which is a little um, a little or some</i>
Feeling occurrence (331)	When my chest feels tight, I get scared that I won't be able to breathe properly. (ASI-R)	"Um / I don't get scared that I won't be able to breathe properly, / <i>but I do get anxious. / So, I would agree some with that.</i> "
Feeling intensity (335)	It scares me when I become short of breath. (ASI-R)	"If I became short of breath for no reason / <i>I think I would be very fearful</i>
Behavior occurrence (322)	When I notice that my heart is beating rapidly, I worry I might have a heart attack. (ASI-R)	"Um/ <i>well I relax, I lie down / and I don't really worry about having a heart attack. / So a little bit. / I've got used to it.</i> "
Cognition occurrence (343)	I'd rather keep things the way they are than risk a disaster. (AABS).	"... / No. / I like, I like change. / Um. / It unnerves me, makes me feel anxious. / I disagree very much on that one. / I wouldn't say I totally disagree. / Uh, / 'cause sometimes there are things, / <i>you think "it's better to just stay the same."</i> /

Table 3

Bases for judgment formations in the ASI-R and AABS protocols, number of protocols (percent)

<u>Judgment</u>	<u>Presence/ Absence</u>	<u>Scale</u>		<u>χ^2 (1, N =384)</u>
		<u>ASI-R</u>	<u>AABS</u>	
Appraisals	Present	30 (20.8)	127 (52.9)	38.33 ^a
	Absent	114 (79.2)	113 (47.1)	
Cognition occurrence	Present	30 (20.8)	25 (10.4)	2.43 ^b
	Absent	114 (79.2)	215 (89.6)	
Behavior occurrence	Present	5 (3.5)	15 (6.3)	1.41 ^b
	Absent	139 (96.5)	225 (93.8)	
Feeling overall (occurrence or intensity)	Present	82 (56.9)	29 (12.1)	88.14 ^a
	Absent	62 (43.1)	211 (87.9)	
Feeling occurrence	Present	70 (48.6)	24 (10.0)	72.58 ^a
	Absent	74 (51.4)	216 (90.0)	
Feeling intensity	Present	24 (16.7)	10 (4.2)	17.43 ^a
	Absent	120 (83.3)	230 (95.8)	

Note. N = 144 protocols for the ASI-R and 240 for the AABS. Feeling occurrence and intensity do not sum to feeling overall as some protocols received both codes.

^a corrected $p < .001$; ^b ns

Table 4

Order of response selection relative to other think aloud codes, number of protocols (percent)

Scale	Order of response selection			
	None stated	Without any other codes	Prior to all other codes	After all other codes
AABS	13 (5.4)	61 (25.4)	94 (39.2)	72 (30.0)
ASIR	13 (9.0)	23 (16.0)	30 (20.8)	78 (54.2)

Note. N of protocols = 144 total for the ASI-R and 240 total for the AABS. $\chi^2(3, N = 384) = 28.23$, corrected $p = .017$.