What do Anxiety Scales Really Measure?

An Item Content Analysis of Self-Report Measures of Anxiety

Andy D. Wall, B.S.

Eric B. Lee, Ph.D.

Southern Illinois University

Department of Psychology

Corresponding Author:

Andy D. Wall

Southern Illinois University

andrew.wall@siu.edu

(618)-453-3551

SIU Psychology

Life Science II, Rm 268

1125 Lincoln Drive-MC 6502

Carbondale IL, 62901

United States

Abstract

Anxiety is a common experience and component of the human condition, but as a construct it is difficult to define and anxiety scales vary in which aspects are emphasized. In the present study, we analyzed and quantified the content overlap of self-report instruments used to measure non-specific or global anxiety. Categorization of the 313 items across thirteen commonly used anxiety scales resulted in 60 disparate categories that spanned a wide variety of symptoms. Mean overlap between all of the selected scales was generally low. On average, only 30.9.% of all symptoms were captured by any given measure. Results indicate that the scales are heterogenous in the content that they measure. Therefore, anxiety-related scales should not be assumed to be interchangeable and careful consideration should be taken when selecting measures.

Keywords: anxiety, symptom overlap, content analysis, measurement, scales

What do Anxiety Scales Really Measure?

An Item Content Analysis of Self-Report Measures of Anxiety

Anxiety is a common human experience and a prominent component of multiple mental health problems. Approximately one in six people will meet diagnostic criteria for an anxiety disorder in their lifetime (Baxter et al., 2014). Moreover, the costs associated with anxiety-related disorders impose a significant burden to society and the economy in addition to the distress and dysfunction it causes on a personal level (Wittchen, 2002). Accordingly, significant research effort is afforded to understanding, measuring, and treating anxiety. Despite this interest, there is a lack of agreement on how to properly define this somewhat nebulous construct. Due to the lack of consensus on this issue, researchers and clinicians alike are put in the situation of selecting one of the many available scales, whose disparate item content could potentially bias their conclusions.

Broadly speaking, anxiety can be defined as an apprehensive anticipation of some future threat that results in feelings of unease and a sense that things are uncontrollable or unpredictable (American Psychiatric Association, 2013). It is distinct from fear, which involves a fight-or-flight response due to a present danger rather than a potential misfortune (Antony, 2002). Anxiety is generally associated with tension, difficulty concentrating, negative affect, and a tendency to worry. However, there is significant disagreement as to which aspects of anxiety are most salient, resulting in the wide variety of scales commonly used to assess the construct.

Many such scales have been included in recent reviews that detail their reliability and validity, as well as other psychometric properties (Antony et al., 2001). Some scales, such as the Beck Anxiety Inventory (Beck et al., 1988), focus on somatic, panic-like symptoms, while others like the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983a) intentionally omit

somatic symptoms. Differences like these can have a negative impact on research and therapy, as different symptoms are measured but are then grouped into a single "anxiety" construct when included in systematic reviews and meta-analyses (Kagan, 2009). The differences in scales are particularly important when considering cultural disparities, such as how Asian clients tend to report more somatic symptoms than emotional symptoms compared to other groups (Antony & Barlow, 2020). The implication is that certain populations might systematically fail to meet criteria for treatment or inclusion in an experiment depending on the scale being used.

In addition, it is often assumed that said scales are all measuring the same construct and therefore, can be used interchangeably to assess anxiety. Meta-analyses that include anxiety as an outcome variable illustrate this problem well. For example, researchers conducted a meta-analysis of studies examining the effect of exercise on anxiety and pooled outcome measures that cover a wide spectrum of anxiety-related symptoms such as the Penn State Worry Questionnaire, the Liebowitz Social Anxiety Self- Report Scale (LSAS), the Beck Anxiety Inventory (BAI), the Depression Anxiety and Stress Scale (DASS), and the State-Trait Anxiety Inventory (STAI; Aylett et al., 2018). In another example, researchers performed a meta-analysis of smartphone interventions for anxiety and pooled the LSAS, BAI, DASS, STAI, Hamilton Anxiety and Depression Scales, and the Generalized Anxiety Disorder-7 (GAD-7; Firth et al., 2017). These measures differ significantly in anxiety-related content including domains such as, social, generalized, hyperarousal, somatic, cognitive, and stress. While most authors of meta-analyses point out the limitation of generalizability due to the heterogeneity of measured anxiety symptoms, the take away message is still often reduced down to a simple statement of effects on "anxiety," with little consideration for the complexity of the construct. This creates risk to the replicability and generalizability of anxiety research if variables of interest are measured using

scales that are potentially assessing different, but related, underlying constructs (Santor et al., 2006). In other words, heterogeneity among anxiety measurement could result in a fragmented literature if the tools we use to measure anxiety are assumed to be homogenous when they are not. This is further complicated by the fact that many researchers make a distinction between state and trait anxiety, although this distinction is not reflected in many of the most widely used anxiety scales (Elwood et al., 2012). Thus, the conclusions drawn in anxiety research are not only limited by addressing a single facet of anxiety, they also fail to differentiate between the temporal nature of said anxiety. If this issue was not already complex enough, it has been postulated that state and trait anxiety can be intertwined, as the level of state anxiety is dependent on both the stressful situation an individual finds themselves in as well as their level of trait anxiety (Endler & Kocovski, 2001).

Another complication of measurement relates to the various specific categories of anxiety disorders (e.g., generalized anxiety disorder, social anxiety disorder, panic disorder, specific phobia, etc.). The specific symptoms and presentations of each have led to the development of disorder-specific measures of anxiety in addition to the more broad, global measures. The advantages of this are clear: These measures allow for a more fine-tuned and specific assessment of unique anxiety-related behavior; however, they also further bifurcate the literature and our understanding of anxiety as a global construct. For example, the Agoraphobic Cognitions Questionnaire is a 14-item scale that assess a specific type of anxiety disorder in which individuals fear and avoid situations that might cause them to feel panicked and/or trapped (Chambless et al., 1984). While agoraphobia is distinct from global anxiety, agoraphobic symptoms appear in scales such as the Four-Dimensional Symptom Questionnaire that purportedly only measure the presence and frequency of non-specific general anxiety.

While the psychometric properties of anxiety scales have been reviewed elsewhere (e.g. Julian, 2011), no such reviews exist that study the amount of overlap between these scales, which is not captured by convergent validity. As pointed out by Visontay et al. (2019), it is possible that scales can fail to assess the same underlying construct, despite having high convergent validity. Although scales may all be assessing the same broad construct of anxiety, the different facets that they measure may not overlap well. This means that these scales could possess high convergent validity despite the lack of overlap, due to the underlying construct that they all attempt to tap into or because the constructs are related to each other. The goal of our present study is to explore this possibility by assessing and quantifying the content overlap of self-report instruments that purportedly measure general anxiety experiences. Previous studies have examined item content in measures of depression (Fried, 2017), mania (Chrobak et al., 2018), and youth OCD (Visontay et al., 2019). We aim to provide a similar item content analysis of self-report measures of non-specific or global anxiety.

Method

Measure Selection

In order to assess the extent of content overlap among common anxiety rating scales, we examined 13 frequently used measures: the 28-item Mood and Anxiety Symptom Questionnaire (MASQ; Watson & Clark, 1991), the two 20-item state and trait scales of the State Trait Anxiety Inventory (STAI; Spielberger, 1983), the 50-item Taylor Manifest Anxiety Scale (TMAS; Taylor, 1953), the 42-item State Trait Inventory for Cognitive and Somatic Anxiety (STICSA; Ree et al., 2000), the 12-item somatic, 7-item phobia, and 10-item anxiety subscales from the Symptom Checklist 90-R (SCL-90R; Derogatis, 1994), the 20-item Zung Self-Rating Anxiety Scale (SAS; Zung, 1971), the 9-item Costello-Comrey Depression and Anxiety Scale (CC-DAS;

Costello, 1967), the 12-item anxiety subscale and the 16-item somatic subscale from the Four-Dimensional Symptom Questionnaire (4DSQ; Terluin et al., 2004), the 7-item Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983), the 7-item anxiety subscale and the 7-item stress subscale of the Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1996), the 21-item Beck Anxiety Inventory (BAI; Beck et al., 1988), and the 25-item Clinical Anxiety Scale (CAS; Snaith et al., 1982).

We employed the same search methods as previous similar content analysis studies to compile a list of potentially relevant anxiety measures (Chrobak et al., 2018; Fried, 2017; Visontay et al., 2019). These included, Google Scholar searches (e.g., "anxiety AND measure" "anxiety AND scale"), reference sections of measures, psychometric comparisons of anxiety measures, systematic reviews and guides (e.g., Antony, 2002; Elwood et al., 2012), and conferring with experts in the area of anxiety disorders. In order to meet criteria for inclusion in the current study, scales or subscales needed to: a) be designed or used to measure the frequency and/or intensity of anxiety experiences at the state and/or trait level; b) be available in a self-report format; c) be published and available in English; d) be designed/adapted for adult populations; and e) be frequently used in the field of psychology/psychiatry within the past decade. In a similar fashion to the recent work done by Newson and colleagues on the heterogeneity of assessment tools in the mental health field, we used Google Scholar citation counts to estimate how frequently scales were being utilized for research and clinical purposes within the past decade (Newson et al., 2020).

Common reasons for exclusion included: a) low citation count over the past decade (i.e., low relevance); b) being a clinician measured rating of anxiety or not being available in a self-report format (e.g., Hamilton Anxiety Scale); c) it solely measured an anxiety-related construct

(e.g., worry); and/or d) it measured a specific anxiety disorder (e.g., social phobia). The anxiety-related subscales of the DASS, 4DSQ, and SCL-90R were included in the analysis. Each of these measures have subscales labeled "anxiety" as well as other anxiety-related subscales (i.e., "stress," "somatization," "somatic," and "phobia"). Moreover, we included both anxiety-related subscales of the MASQ (i.e., "anxious symptoms" and "anxious arousal") as it does not have a general "anxiety" subscale. Because their items had such a high degree of overlap with the other measures, excluding them would have decreased overall item overlap, thus increasing the perceived heterogeneity in the analyses. By and large, the item content included in these six anxiety-related subscales overlapped with the other measures included in the analysis. For example, only 3 of the 48 (6.3%) items included in these subscales contained unique content that did not overlap with another measure. Moreover, 35 of the 48 (72.9%) items included content shared by four or more of the 13 included measures. Therefore, we chose to include these subscales as part of our conservative approach to measure inclusion and item content categorization. Due to the focus of the review, psychometric properties of measures (e.g., factor structure, reliability, validity) were not considered with regard to inclusion/exclusion criteria. For the purposes of our study, we instead aimed to analyze the item content of measures that are frequently being used, regardless of their psychometric properties. An itemized list of scales that were excluded along with specific reasons as to why they did not meet criteria for this study can be found at https://osf.io/89fkc/

Item Selection and Symptom Categories

Altogether, a total of 313 items were identified and organized from the 13 scales that were selected to be included in our content analysis. All items of any scale or subscale that met criteria for the review were included in the analysis. First, scales that met criteria for inclusion

were reviewed and specific items were entered into a spreadsheet. These specific items were used to generate what we refer to as symptom categories, which served to organize items that assess the same specific aspect of anxiety. For example, “feeling fearful” and “I feel frightened” were both listed under the symptom category labeled as “fear.” Although anxiety is distinct from fear, items such as these that were not congruent with the typical definition of anxiety were still included in our categorization and analysis as the scales they were drawn from are reportedly used to measure anxiety. Similarly worded items, as well as reverse worded items, were systematically sorted into symptom categories unless one did not exist that fit that specific item, in which case one was created for it.

Throughout this process, we encountered items that could have been placed under multiple symptom categories simultaneously. For example, the item “I cannot concentrate without irrelevant thoughts intruding” measures both concentration difficulties as well as the presence of intrusive thoughts, both of which were distinct symptom categories in our study. Items such as these were placed in both symptom categories and labeled as a “compound” item, similar to the terminology used by Visontay et al., (2019) in their study of youth obsessive compulsive disorder scales. The majority of compound items were “double-barreled” (i.e., measured two or more distinct symptoms in a single item). For example, “I had hot or cold spells” clearly inquires about two separate experiences. A high rating on this item could indicate hot *or* cold spells and was therefore labeled as a compound item. However, others were not double-barreled and still met our criteria as a compound item because they still fit into more than one of our symptom categories. For example, the item “I feel steady” from the State and Trait Anxiety Inventory could be measuring someone being a “steady” person, or could be assessing the presence of physical shakiness. In addition to this, we also took note of specific items that

were idiosyncratic, meaning that across and among all scales included in this study, only one item existed that assessed a specific facet of anxiety. This means that idiosyncratic items fell under and were the sole member of their own symptom category.

Similar to other studies of this nature, we were conservative in our approach and attempted to only differentiate between items when they were clearly assessing different facets of anxiety. As it has been pointed out elsewhere, using a less conservative approach would likely overestimate heterogeneity among and between scales (Chrobak et al., 2018; Fried, 2017; Visontay et al., 2019). This means that when organizing items and deciding which symptom category a specific item should fall under or when to combine symptom categories into one, we erred on the side of similarity rather than assuming that they were distinct. This process is notably subjective and qualitative in nature. As it has been noted by others who have performed this type of analysis, there is no formalized or objective way to carry out this kind of work (Visontay et al., 2019).

Although it is not taken into account in the majority of the scales included in our analysis, many researchers suggest that there should be a distinction between state and trait anxiety. Trait anxiety is an individual's predisposition to experience anxiety and is a stable personality feature whereas state anxiety is a transitory emotion in response to adverse events, thoughts, feelings, etc. (Saviola et al., 2020). Others argue that most psychological constructs are both traits and states simultaneously; however, the conceptualization of anxiety as a state and as a trait is consistent (Endler & Kocovski, 2001; Kenny, 2019). In light of this distinction, we chose to treat the two subscales of the State and Trait Anxiety Inventory as two separate measures of anxiety, since they purportedly measure different anxiety constructs and are often used separately from one another. However, we did not differentiate between the trait and state scales of the State-

Trait Inventory for Cognitive and Somatic Anxiety in this way, as unlike the State and Trait Anxiety Inventory, there is complete item overlap among the two subscales. In addition to this, we did not consider items to be idiosyncratic if they were in both subscales of the State and Trait Anxiety Inventory, even if said items were not present on any other scales.

Statistical Analysis

Following in the footsteps of others whom have undertaken this type of analysis, the content overlap of the selected scales was estimated by performing a Jaccard Index (Chrobak et al., 2018; Fried, 2017; Visontay et al., 2019). The Jaccard Index is a coefficient of similarity that is used when examining binary data that ranges from 0 (no overlap among scales) to 1 (complete overlap). It is calculated by using the formula s/u1 + u2 + s, where s represents the number of shared items between two questionnaires, and u1 and u2 are the number of items that are unique to each of the two scales (Fried, 2017). As there are currently no well-cited guidelines for determining the strength or weakness of a Jaccard similarity coefficient, we have decided to follow suit of other researchers like Fried (2017) and used criteria based upon Evans' *Straightforward Statistics for the Behavioral Sciences* (Evans, 1996). According to his rules of interpretation, a 0.00-0.19 Jaccard coefficient is very weak, a 0.20-0.39 is weak, a 0.40-0.59 is moderate, 0.60-0.79 is strong, and a 0.80-1 is very strong. In addition to this, we also calculated the rate of specific (e.g., "perspired noticeably in the absence of high temperatures or physical exertion") and compound (e.g., "Sometimes when embarrassed, I break out in a sweat") items per scale. Additionally, we counted the number of idiosyncratic items per scale, which are items that are only present in one scale. The analyses were conducted using R (R Core Team, 2015) and the following packages: qgraph (Epskamp et al., 2012), ggplot2 (Wickham, 2009), data.table (Dowle et al., 2019), reshape2 (Wickham, 2007), psych (Revelle, 2017), ade4 (Dray & Dufour,

2007), and viridis (Garnier et al., 2018). The code was adapted from (Fried, 2017) and can be found at https://osf.io/89fkc/, along with our data and other supplementary materials.

Results

The process of narrowing all 313 items from 13 scales resulted in 60 disparate symptom categories that spanned a wide variety of content such as hyperarousal, somatic, cognitive, fear, stress, and worry (see Figure 1). Many symptoms are unique to a single scale or only found in a few scales. For example, no symptoms appear across all 13 scales, 21.7% of symptoms are unique to a single scale, and 55% of symptoms appear across three or fewer scales. See Table 1 for a full list of symptoms appearing across combinations of scales. The most common scale items are “nervous,” “relaxed,” “physical shakiness,” and “fear;” however, no scale captured each of these.

Across all scales, the computed Jaccard Index is 0.24. Although, the index has no well-established heuristics, 0.24 in this case likely indicates a weak mean overlap of all scales. Most Jaccard Indices between any two scales are quite low with only a few reaching values in the “moderate” range (0.40–0.59) and a single value in the “strong” range (0.65 between the STAI-T and STAI-S). Mean overlap with the other scales is highest, meaning they have more item content in common with the other scales, for the MASQ (0.33), BAI (0.31), and ZSAS (31). Overlap is lowest, meaning item content has less in common with the other scales, for the CC-DAS (0.18), STAI-T (0.19), and STICSA (0.21). Overall, no scales appear to have meaningful overlap with the other scales on average. See Table 2 for all Jaccard Indices.

The SCL, the 4DSQ and the TMAS each contain the most unique symptoms with three idiosyncratic items, and 6 of the 13 scales contain at least one unique symptom. Unique symptoms vary significantly in content and appear to mostly assess anxiety-adjacent symptoms

or behaviors such as substance use, compulsions, memory problems, and confusion. No scale contains a significant portion of the 60 symptoms. This is likely due to the relatively large number of symptoms captured across all 13 scales. On average across all 13 scales, only 30.9% of all symptoms are captured by any given measure. The TMAS captures the most at 51.7% while the CC-DAS and the HADS and capture the least, at 13.3%.

Regarding item content capturing anxiety disorder-related DSM-5 criteria, scales heavily favor content related to generalized anxiety disorder and panic disorder. All but one scale (BAI) captures some core DSM-5 diagnostic criteria for generalized anxiety disorder and all but two (the STAI-T and STAI-S) capture some core panic disorder criteria. Few scales capture any core OCD (4 scales), agoraphobia (3 scales), social anxiety disorder (2 scales), or specific phobia (1 scale) diagnostic criteria. Additionally, eight scales capture some core major depressive disorder symptom criteria with the TMAS capturing 55.5% of core DSM-5 symptoms. See Table 3 for full symptom characteristics by scale and Table 4 in the supplementary materials for the process of defining specific symptoms to diagnostic criteria.

Discussion

The analyses we conducted identified a total of 60 unique anxiety symptoms in 13 commonly used anxiety scales. The mean overlap of the item content among the scales was weak, indicated by the small Jaccard Index. This means that the symptoms captured by the scales were quite heterogenous. Moreover, the 60 symptoms we identified were not represented well by any single measure, with a low average of only 30.9% of the symptoms being captured by any of the 13 scales. The only scales that had high overlap were the STAI-T and the STAI-S, which were developed concurrently and besides temporality, are largely similar. In addition to this, 21.7% of the symptoms we identified were unique to only a single scale, meaning about a fifth of

the symptoms contained across scales were idiosyncratic. Taken together, the results indicate that the scales did not achieve meaningful overlap with each other on average and that as a whole, the scales vary significantly in item content.

At times this heterogeneity of content is problematic, as it is common practice in the field to use these measures as if they are interchangeable. This is likely done with the assumption that these "anxiety" scales measure the same construct given their respective convergent and divergent validity coefficients. However, measures can be highly correlated with one another and still meaningfully differ in the content they assess. The low content overlap of the scales suggests that they may not be tapping into the same construct or may be measuring different facets of a more general construct. Therefore, we caution researchers when comparing scores obtained from these scales and to consider the potential impact that content heterogeneity could have to the generalizability and replicability of their research. This is also clinically relevant, as the scales differed significantly with regard to measuring core DSM-5 symptom criteria of anxiety related disorders, OCD, and major depressive disorder. While the heterogeneity found among these scales in not necessary problematic, it does highlight the need for careful and considered scale selection by researchers and clinicians.

Where does the heterogeneity we have observed come from? We would argue that the conceptual breadth of anxiety and the lack of construct clarity are significant factors. The definition of anxiety can markedly differ depending on the context in which it is used. Ask psychotherapists, medical doctors, researchers, and laypeople what anxiety means and you are likely to get very different answers. Depending on people's area of training, expertise, and personal experience, they may emphasize or exclude different aspects of worry, restlessness, panic, somatic sensations, fear, etc. in their definition. This lack of agreement on what

constitutes anxiety is reflected in the scales that were analyzed and the on average low percentage of total symptom measurement. Of the 13 scales, the one that measured the greatest number of symptoms (the TMAS) only captured 51.7% of the symptom categories. For better or worse, the lack of a common, agreed upon conceptualization of what constitutes anxiety has fostered the development of a wide variety of what can only be called, "anxiety-related constructs". Clearly, some of these scales were developed to intentionally measure different aspects of anxiety. As previously mentioned, the State Trait Anxiety Inventory differentiates between experiences related to stability, threat, and negative emotions that are occurring in the present moment (i.e., state) and one's general tendency to experience anxiety (i.e., trait). Conversely, the Beck Anxiety Inventory emphasizes somatic and panic-related concerns, while the Hospital Anxiety and Depression Scale, intentionally omits somatic items in favor of emotional states like fear, worry, and happiness. Regardless of whether the heterogeneity originates from the nebulous nature of this construct, the intentional differentiation of types of anxiety when developing scales, or another source we have not considered, it is clear that this common human experience is in need of construct clarification and elaboration.

The lack of item overlap and heterogeneity among the scales is not inherently problematic, it is however, notable and worth consideration when selecting measure related to anxiety. For example, the TMAS captures more of the symptom categories than other measures included in this review, yet this does not mean that the TMAS is a particularly good scale. In fact, the length TMAS is now used less-frequently when compared with many other measures of anxiety included in our study despite capturing comparatively more symptoms in our analysis (Antony et al., 2001). Therefore, decisions relating to scale selection should not be solely based on content coverage. Other factors need to be taken into consideration when choosing a measure,

including: psychometric properties, treatment setting and cultural considerations, administration and format, performance characteristics (i.e., specificity and sensitivity), and the purpose and goals of the assessment (e.g., intervention planning, screening, outcome assessment). The purpose and goals of an assessment may be of particular importance, as a researcher or clinician may be interested in specific content and thus, a low overlap of content may not be as important.

Limitations

The most salient limitation of this study is the subjective nature of multiple aspects of the content analysis itself. Although we were conservative in our approach, there is currently no objective methodology for comparing or organizing items across and within scales to assess whether similarity worded items are measuring the same symptom or not. This led to a situation in which there were many points during the organization process where there could have been divergence. Other researchers may have categorized items differently than we have. Relatedly, although we used a set of criteria that determined if a scale would be included or not, the scale selection process was also somewhat subjective. For example, other researchers may have wished to include clinician rated scales that have not been adapted into a self-report format, such as the Hamilton Anxiety Scale. Furthermore, even working with the same criteria, there are alternative methods to selecting scales that could have been utilized. For example, other researchers may have determined the frequency of scale use by alternative means other than Google Scholar citation counts. Therefore, there is a need for a more objective process when undertaking this type of research, similar to the PRISMA and COSMIN guidelines for literature reviews and appropriate outcome measurement selection (Liberati et al., 2009; Mokkink et al., 2010). Finally, interpretation of the Jaccard coefficients is also a subjective process as there are no established guidelines. Although we have followed the precedent set by Fried (2017) by

following the rule from Evans (1996), he himself acknowledged that this was done in the absence of any well cited guidelines. Therefore, more accepted standards for interpreting this type of statistic could play an instrumental role in future studies that involve content analysis. Our materials and code are available so that others have the opportunity to categorize items differently, include/exclude scales or subscales, or run different types of analyses with our data. All materials, including data, code, and item review spreadsheets are available at https://osf.io/89fkc/.

Future Directions and Implications

In light of these findings, we urge researchers and clinicians to be thoughtful when selecting anxiety-related scales and to not simply use them interchangeably. Furthermore, results that are generated by using different anxiety scales should be compared and integrated with caution, as it is unlikely that they are capturing the same "anxiety". This is line with the trend that we have observed in the research of this type (see Chrobak et al., 2018; Fried, 2017; Visontay et al., 2019), which have shown that similar heterogeneity exists for self-report measures of depression, youth OCD, and hypomania. As new measures are created and validated and existing measures are updated or adapted, we urge developers to consider assessing content overlap alongside other psychometric properties that are regularly tested.

A potential solution to help with this problem that has been proposed by Visontay et al. (2019) is to use an item banking method, which involves gathering pools of items from scales and using them to develop measurement tools. After going through the process of item standardization, evaluating items from the perspective of both expert judges as well as consumers, and refining the questionnaire through Item Response Theory, a measure that captures a greater percentage of symptoms than any one of the measures on their own should

hopefully emerge. This, however, assumes that there is a latent "global anxiety" construct that is clinically useful. Furthermore, Fried's (2017) recommendations for depression scales are relevant to measures of anxiety. These include the development and use of scales that measure a wide variety of specific anxiety symptoms with multiple questions to control for measurement error and taking care to measure only a single symptom per item. For example, compound items that assess two opposite or unique symptoms (e.g., "hot or cold spells") should be separated into multiple items. Future work and measurement development related to anxiety remains necessary in order to better capture relevant components of anxiety.

Like many psychological constructs, anxiety is complicated and difficult to define. Its measurement therefore requires thoughtful consideration. We believe that content analysis of measures of psychopathology provides a useful lens to use when making measurement-related decisions. It is our hope that this review and analysis will assist researchers and clinicians in making more informed decisions when selecting anxiety-related measures as well as inform the development of future measures.

Acknowledgements

We would like to thank Eiko Fried for openly sharing his code and data and for so kindly helping troubleshoot our code.

Declaration of Interest

None

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.)* (2013-14907-000). Author; psyh. http://dist.lib.usu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-14907-000&site=ehost-live

Antony, M. M. (2002). Assessment of Anxiety and the Anxiety Disorders: An Overview. In M. M. Antony, S. M. Orsillo, & L. Roemer (Eds.), *Practitioner's Guide to Empirically Based Measures of Anxiety* (pp. 9–17). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47628-2_2

Antony, M. M., & Barlow, D. H. (Eds.). (2020). *Handbook of assessment and treatment planning for psychological disorders* (Third edition). The Guilford Press.

Antony, M. M., Orsillo, S. M., & Roemer, L. (2001). *Practitioner's guide to empirically based measures of anxiety*. Springer Science & Business Media.

Aylett, E., Small, N., & Bower, P. (2018). Exercise in the treatment of clinical anxiety in general practice–a systematic review and meta-analysis. *BMC Health Services Research*, *18*(1), 1–18.

Baxter, A., Vos, T., Scott, K., Ferrari, A., & Whiteford, H. (2014). The global burden of anxiety disorders in 2010. *Psychological Medicine*, *44*(11), 2363.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988a). An inventory for measuring anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893–897. https://doi.org/10.1037/0022-006X.56.6.893

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988b). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893–897. https://doi.org/10.1037/0022-006X.56.6.893

Chambless, D. L., Caputo, G. C., Bright, P., & Gallagher, R. (1984). Assessment of fear of fear in agoraphobics: The Body Sensations Questionnaire and the Agoraphobic Cognitions Questionnaire. *Journal of Consulting and Clinical Psychology*, *52*(6), 1090–1097. https://doi.org/10.1037/0022-006X.52.6.1090

Chrobak, A. A., Siwek, M., Dudek, D., & Rybakowski, J. K. (2018). Content overlap analysis of 64 (hypo) mania symptoms among seven common rating scales. *International Journal of Methods in Psychiatric Research*, *27*(3), e1737.

Costello, C. G. (1967). Costello Comrey depression and anxiety scales. *Measures for Clinical Practice: A Sourcebook. 2nd*, *2*, 155–160.

Derogatis, L. R. (1994). *SCL-90-R: Administration, scoring, and procedures manual*. Minneaplis, MN: National Computer Systems.

Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., & Parsonage, H. (2019). Package 'data. Table.' *Extension of 'data. Frame*.

Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20.

Elwood, L. S., Wolitzky-Taylor, K., & Olatunji, B. O. (2012). Measurement of anxious traits: A contemporary review and synthesis. *Anxiety, Stress & Coping*, *25*(6), 647–666.

Endler, N. S., & Kocovski, N. L. (2001). State and trait anxiety revisited. *Journal of Anxiety Disorders*, *15*(3), 231–245. https://doi.org/10.1016/S0887-6185(01)00060-3

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole Pub. Co.

Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017). Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *Journal of Affective Disorders*, *218*, 15–22.

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Garnier, S., Ross, N., Rudis, B., Sciaini, M., & Scherer, C. (2018). viridis: Default Color Maps from 'matplotlib.' *R Package Version 0.5*, *1*, 2018.

Julian, L. J. (2011). Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). *Arthritis Care & Research*, *63*(S11), S467–S472. https://doi.org/10.1002/acr.20561

Kagan, J. (n.d.). *Two Is Better Than One*. 2.

Kenny, D. A. (2019). Enhancing validity in psychological research. *American Psychologist*, *74*(9), 1018–1028. https://doi.org/10.1037/amp0000531

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*(jul21 1), b2700–b2700. https://doi.org/10.1136/bmj.b2700

Lovibond, S. H., & Lovibond, P. F. (1996). *Manual for the depression anxiety stress scales*. Psychology Foundation of Australia.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, *19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Newson, J. J., Hunter, D., & Thiagarajan, T. C. (2020). The Heterogeneity of Mental Health Assessment. *Frontiers in Psychiatry*, *11*, 76. https://doi.org/10.3389/fpsyt.2020.00076

R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ree, M. J., MacLeod, C., French, D., & Locke, V. (2000). State-trait inventory for cognitive and somatic anxiety (STICSA). *Australia: The University of Western Australia*.

Revelle, W. (2017). *psych: Procedures for personality and psychological research.* https://CRAN.R-project.org/package=psych

Santor, D. A., Gregus, M., & Welch, A. (2006). FOCUS ARTICLE: Eight Decades of Measurement in Depression. *Measurement: Interdisciplinary Research & Perspective*, *4*(3), 135–155. https://doi.org/10.1207/s15366359mea0403_1

Saviola, F., Pappaianni, E., Monti, A., Grecucci, A., Jovicich, J., & De Pisapia, N. (2020). Trait and state anxiety are mapped differently in the human brain. *Scientific Reports*, *10*(1), 11112. https://doi.org/10.1038/s41598-020-68008-z

Snaith, R. P., Baugh, S. J., Clayden, A. D., Husain, A., & Sipple, M. A. (1982). The clinical anxiety scale: An instrument derived from the Hamilton anxiety scale. *The British Journal of Psychiatry*, *141*(5), 518–523.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (form Y)(" self-evaluation questionnaire")*.

Taylor, J. A. (1953). *Manifest anxiety scale*. American Psychological Association Washington, DC.

Terluin, B., Rhenen, W. V., Schaufeli, W. B., & De Haan, M. (2004). The Four-Dimensional Symptom Questionnaire (4DSQ): Measuring distress and other mental health problems in a working population. *Work & Stress*, *18*(3), 187–207.

Visontay, R., Sunderland, M., Grisham, J., & Slade, T. (2019). Content overlap between youth OCD scales: Heterogeneity among symptoms probed and implications. *Journal of Obsessive-Compulsive and Related Disorders*, *21*, 6–12. https://doi.org/10.1016/j.jocrd.2018.10.005

Watson, D., & Clark, L. A. (1991). The mood and anxiety symptom questionnaire (MASQ). *Unpublished Manuscript, University of Iowa, Iowa City*.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20.

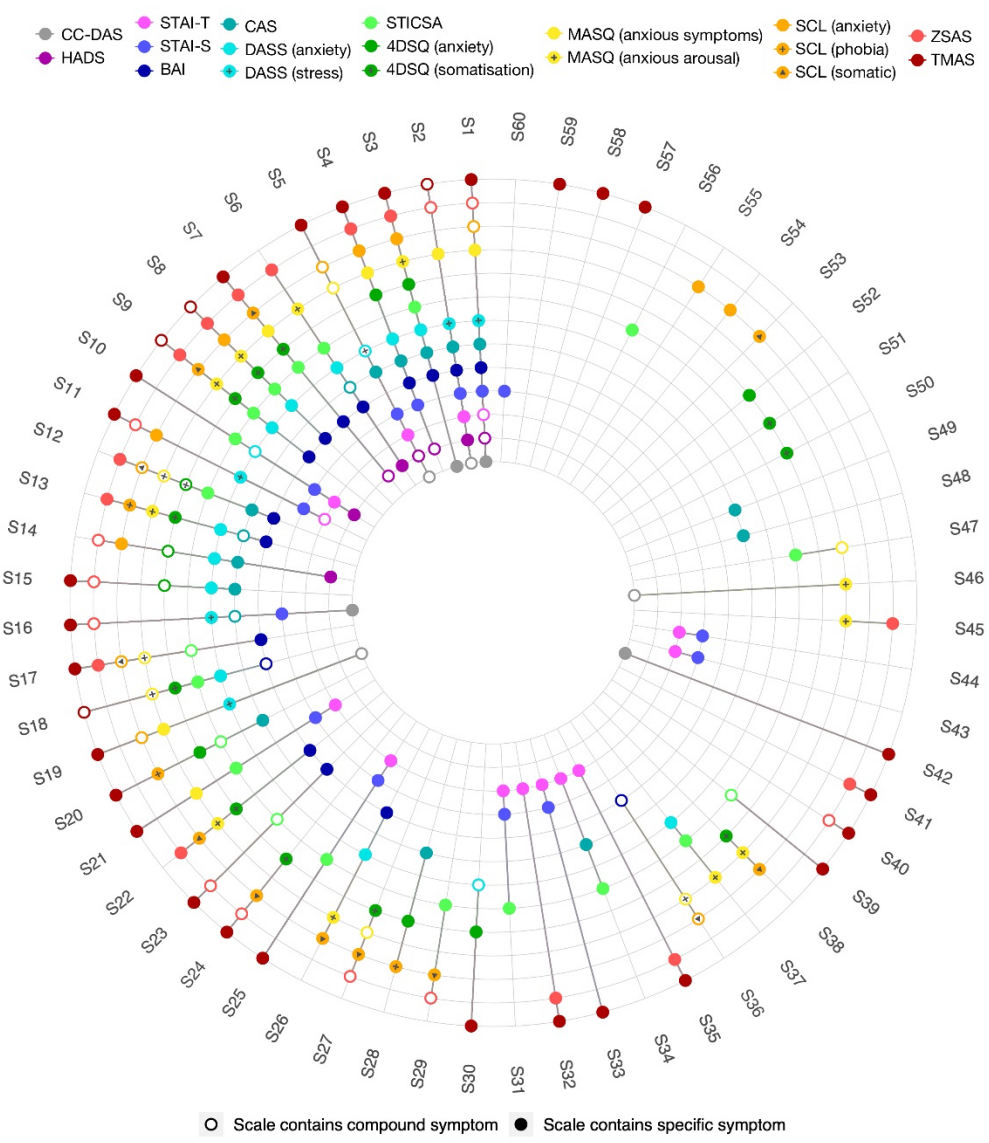
Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wittchen, H.-U. (2002). Generalized anxiety disorder: Prevalence, burden, and cost to society. *Depression and Anxiety*, *16*(4), 162–171. https://doi.org/10.1002/da.10065

Zigmond, A. S., & Snaith, R. P. (1983a). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370.

Zigmond, A. S., & Snaith, R. P. (1983b). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370.

Zung, W. (1971). *Zung Self-Rating Anxiety Scale (SAS)*.



1 Nervous
2 Relaxed
3 Physical shakiness
4 Fear
5 Tension (general)
6 Anticipated fear
7 Indigestion / nausea
8 Increased heart rate
9 Breathing difficulty
10 Worry
11 Restlessness
12 Dizziness
13 Faint / lightheaded
14 Panic
15 Anxiety
16 Upset / rattled
17 Hot body temperature
18 Perspiration
19 Irritable / on edge / annoyed
20 Avoidance
21 Negative feelings
22 Numbness / tingling
23 Flushed face / blushing
24 Headache
25 Self-worth
26 Choking / difficulty swallowing
27 Muscle pain
28 Agoraphobia
29 Physical weakness
30 Social fear
31 Indecision
32 Sleep difficulty
33 Positive feelings
34 Instrusive thoughts
35 Overwhelmed
36 Cold body temperature
37 Dry mouth / dry throat
38 Chest pain
39 Concentration
40 Tired
41 Nightmares / dreams
42 Sensitive
43 Secure
45 Frequent urination
46 Jumpy / startled
47 Muscle tension
48 Substance use
49 Optimism
50 Chest pressure
51 Vision problems
52 Compulsions
53 Physical heaviness
54 Pressure
55 Derealization
56 Memory
57 Crying
58 Hunger
59 Self-conscious
60 Confusion

Table 1
Number of symptoms that appear across scales

Symptoms	Scales	%
13	1	21.7
9	2	15.0
11	3	18.3
5	4	8.3
5	5	8.3
6	6	10.0
3	7	5.0
3	8	5.0
1	9	1.7
3	10	5.0
1	11	1.7
0	12	0
0	13	0

Note. 13 of the 60 symptoms are only present on a single scale and none of the 60 symptoms appear across all 13 scales.

Table 2
Jaccard Index overlap of item content

	MASQ	STAI-T	STAI-S	TMAS	STICSA	SCL	ZSAS	CC-DAS	4DSQ	HADS	DASS	BAI	CAS
MASQ	1	0.15	0.18	0.3	0.32	0.52	0.43	0.22	0.32	0.22	0.25	0.47	0.58
STAI-T	0.15	1	0.65	0.29	0.17	0.08	0.15	0.16	0	0.22	0.08	0.18	0.07
STAI-S	0.18	0.65	1	0.29	0.13	0.11	0.15	0.22	0.03	0.29	0.08	0.27	0.11
TMAS	0.3	0.29	0.29	1	0.31	0.27	0.44	0.22	0.24	0.18	0.35	0.39	0.27
STICSA	0.32	0.17	0.13	0.31	1	0.22	0.25	0.04	0.21	0.12	0.25	0.22	0.33
SCL	0.52	0.08	0.11	0.27	0.22	1	0.43	0.14	0.45	0.18	0.32	0.33	0.41
ZSAS	0.43	0.15	0.15	0.44	0.25	0.43	1	0.14	0.36	0.22	0.25	0.38	0.46
CC-DAS	0.22	0.16	0.22	0.22	0.04	0.14	0.14	1	0.04	0.23	0	0.29	0.14
4DSQ	0.32	0	0.03	0.24	0.21	0.45	0.36	0.04	1	0.12	0.33	0.3	0.33
HADS	0.22	0.22	0.29	0.18	0.12	0.18	0.22	0.23	0.12	1	0.33	0.35	0.26
DASS	0.47	0.18	0.27	0.39	0.22	0.33	0.38	0.29	0.3	0.35	1	0.27	0.4
BAI	0.58	0.07	0.11	0.27	0.33	0.41	0.46	0.14	0.33	0.26	0.23	1	0.4
CAS	0.24	0.15	0.2	0.21	0.16	0.28	0.32	0.26	0.29	0.33	0.33	0.4	1
Mean overlap	.33	.19	.22	.28	.21	.29	.31	.18	.22	.23	.23	.32	.31

Note. Values range from 0 (no overlap) to 1 (total overlap)

Table 3
Breakdown of Symptom Characteristics by Scale

	CC-DAS	HADS	STAI-T	STAI-S	BAI	CAS	DASS	STICSA	4DSQ	MASQ	SCL	ZSAS	TMAS
Symptoms captured	8	8	14	15	16	16	19	20	20	24	25	25	31
Unique symptoms (n)	0	0	0	1	0	2	0	1	3	0	3	0	3
Scale captures X% of all 60 items	13.3	13.3	23.3	23.3	26.7	26.7	31.7	33.3	33.3	41.7	41.7	41.7	51.7
Scale captures X% of 8 core DSM-5 GAD symptoms	12.5	12.5	37.5	25.0	0	12.5	37.5	50	12.5	25.0	25.0	50.0	75.0
Scale captures X% of 17 core DSM-5 panic disorder symptoms	5.9	11.8	0	0	64.7	17.6	47.1	41.2	52.9	64.7	64.7	17.1	35.3
Scale captures X% of 2 core DSM-5 specific phobia symptoms	0	0	0	0	0	100	0	0	0	0	0	0	0
Scale captures X% of 3 core DSM-5 social anxiety disorder symptoms	0	0	0	0	0	33.3	33.3	0	0	0	0	0	0
Scale captures X% of 8 core DSM-5 agoraphobia symptoms	0	0	0	0	0	75.0	0	0	37.5	0	62.5	0	0
Scale captures X% of 2 core DSM-5 OCD symptoms	0	0	50.0	0	0	50.0	0	50.0	50.0	0	0	0	0
Scale captures X% of 9 core DSM-5 major depressive disorder symptoms	0	11.1	33.3	11.1	0	0	0	11.1	0	0	11.1	22.2	55.5