Development and Preliminary Evaluation of the Tinnitus Severity Short Form

Madelyn R. Frumkin[a*], Dorina Kallogjeri[b], Jay F. Piccirillo[b], Eldre W. Beukes[c,d], Vinaya Manchaiah[c,e], Gerhard Andersson[f,g], & Thomas L. Rodebaugh[a]

[a]Washington University in St. Louis, Department of Psychological & Brain Sciences

[b]Washington University in St. Louis, Department of Otolaryngology – Head & Neck Surgery

[c]Department of Speech and Hearing Sciences, Lamar University, Beaumont, Texas, USA

[d]Psychology and Sport Sciences, Anglia Ruskin University, Cambridge, United Kingdom

[e]Department of Speech and Hearing, School of Allied Health Sciences, Manipal University, Karnataka, India

[f]Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

[g]Department of Clinical Neuroscience, Division of Psychiatry, Karolinska Institute, Stockholm, Sweden

*Correspondence regarding this article should be addressed to Madelyn R. Frumkin, Department of Psychological and Brain Sciences, Washington University in St. Louis, One Brookings Drive, Campus Box 1125, St. Louis, MO 63105. Phone: (314) 935-8627. Email: mfrumkin@wustl.edu.

Conflict of Interest: No relevant conflicts of interest to report.

Funding: This work was supported by the National Institutes of Health [grant numbers 5R01DC017451-02, UL1 RR024992, TL1 RR024995, DC00022, UL1 TR000448, P30 CA091842, AT007064, R21DC017214] and the Department of Homeland Security [grant number EMW-2010-FP-00601]. Additional funding was received from the Doris Duke Clinical Research Foundation, the Stanford University Medical Scholars Fellowship, and the Mallinckrodt Institute of Radiology at the Washington University School of Medicine. These funding sources had no role in any aspect of the research conducted

Abstract

Purpose: Tinnitus, or the perception of sounds that occur without an external sound source, is a prevalent condition worldwide. For a subset of adults, tinnitus causes significant distress and impairment. Several patient-reported outcome measures have been developed to assess severity of tinnitus distress. However, at present, the field lacks a brief measure that is sensitive to treatment change. The purpose of the current study was to develop and preliminarily validate a brief questionnaire for tinnitus severity from two existing measures of tinnitus-related distress, the Tinnitus Handicap Inventory (THI) and Tinnitus Functional Index (TFI).

Method: Using data from nine study samples in the United States and United Kingdom, we conducted exploratory and confirmatory factor analyses to identify a short measure with good psychometric properties. We also assessed sensitivity to treatment-related change by examining associations with change in the TFI and THI. Finally, we conducted a confirmatory factor analysis of the final short questionnaire in a new sample of adults seeking treatment for tinnitus-related distress.

Results: We identified 10 items from the THI and TFI that exhibited limited loadings on secondary factors. The resulting Tinnitus Severity Short Form (TS-SF) achieved good to excellent fit, including in a unique sample of individuals seeking online treatment for tinnitus, and appeared sensitive to treatment-related change.

Conclusions: The TS-SF developed in the current study may be a useful tool for the assessment of subjective severity and distress associated with tinnitus, especially when patient burden is a concern. Further research is necessary to fully validate the questionnaire for the assessment of treatment-related change.

Keywords: Tinnitus, Measurement, Outcome instruments, Confirmatory Factor Analysis

Development of a Short Questionnaire for Tinnitus Severity

Tinnitus is the perception of constant or persistent transient sounds (i.e., ringing, buzzing) generally noticed in the ears and/or head in the absence of a corresponding external sound in the environment. It is a prevalent condition globally with at least 10-20% of adults reporting frequently hearing tinnitus (e.g. Shargorodsky, Curhan, & Farwell, 2010). For most individuals, tinnitus is not bothersome. Some, however, have strong reactions to tinnitus. Results of the National Study of Hearing in England found that of the general population surveyed (N = 48, 313), 10.1% reported any tinnitus, 2.8% reported moderately annoying tinnitus, 1.6% reported severely annoying tinnitus, and 0.5% were unable to lead a normal life due to the severity of the tinnitus (Davis & Refaie, 2020). Chronic distressing tinnitus is associated with lower quality of life and greater prevalence of psychological problems, including anxiety and depression, and may in a small minority lead to suicidal inclinations (Aazh & Moore, 2018; Geocze et al., 2013; Pattyn et al., 2016).

In the absence of a clear cause of tinnitus, a cure remains elusive (McFerran et al., 2019). There are numerous tinnitus management strategies, such as sound-based therapies, aiming to reduce the perception of tinnitus (Cima et al., 2019; Searchfield et al., 2017; Tunkel et al., 2014). In addition, psychological treatments including mindfulness-based therapies (Rademaker et al., 2019), have been developed to address unhelpful reactions and behaviors associated with hearing tinnitus (Andersson & Lyttkens, 1999). The tinnitus management strategy with the most scientific evidence to date is Cognitive Behavioral Therapy (Andersson, 2002; Hesser et al., 2011; Landry et al., 2019; Martinez-Devesa et al., 2007). In this psychological treatment, the focus is not on eliminating the tinnitus, but on reducing the resulting difficulties and associated distress through cognitive and behavioral strategies. Results from systematic reviews and meta-

analyses support the efficacy of this approach (e.g., Andersson, 2002; Andersson & Lyttkens, 1999; Hesser et al., 2011; Landry et al., 2019; Martinez-Devesa et al., 2007).

Measurement of Tinnitus

There is currently no objective way to determine the presence or severity of tinnitus (Jackson et al., 2019). Consequently, tinnitus practice guidelines recommend quantifying tinnitus severity or intervention-related change over time using self-report outcome measures (Tunkel et al., 2014). To determine which interventions are most effective for tinnitus, pre- and post-outcome measure scores are compared (Hall, 2017). There are at least twelve English-language tinnitus questionnaires (Fackrell et al., 2014; Hall et al., 2016) that aim to assess tinnitus severity or its burden to an individual's life. Not all of the questionnaires have been validated to assess treatment-related change (Henry et al., 2016; Jacquemin et al., 2019; Meikle et al., 2007, 2008). The Tinnitus Handicap Inventory (THI; Newman et al., 1996) and Tinnitus Functional Index (TFI; Meikle et al., 2012) are the two most commonly used English-language primary outcome instruments (Hall et al., 2016). The THI was developed as a diagnostic tool with defined categories of tinnitus severity (Newman et al., 1998). Although not initially intended for this purpose, the THI has demonstrated an adequate ability to assess treatment outcomes (Zeman et al., 2011). The TFI was later developed as a measure of tinnitus severity and was specifically validated to be responsive to change (Meikle et al., 2012).

Ensuring that the outcome measure is reliable, valid, responsive, and can be appropriately interpreted is essential when assessing the outcome of an intervention (Terwee et al., 2007). The TFI and THI have high convergent validity (r = .86; Meikle et al., 2012), and their psychometric properties have been evaluated in numerous studies. The THI items are grouped into three subscales: functional, emotional, and catastrophic facets of tinnitus distress. These subscales

were identified based on face validity (Newman et al., 1996). Exploratory and confirmatory factor analysis is often used in psychometrics research to empirically examine subscales (c.f., Brown, 2015). In factor analysis, observed variables (e.g., questionnaire items) load onto unobservable latent factors (e.g., facets of tinnitus distress). The central notion of factor analysis is that responses to specific items are due to such latent variables, and part of the purpose of the analysis is to determine how many such latent variables are involved. More than one latent variable implies that the scale is not unitary, but rather has subscales. Ideally, the corresponding items load strongly onto only one latent factor, indicating discriminability between facets.

In a study examining the Danish version of the THI using exploratory factor analysis, Zachariae and colleagues (2000) found that a three-factor solution was suitable. However, the factors differed from those initially suggested by Newman and colleagues (1996). Using principal components analysis, Baguley and Andersson (2003) found support for a unifactorial structure of the THI, with only one item loading on to a separate factor without double loadings. To our knowledge, Kleinstäuber and colleagues (2015) have presented the only confirmatory factor analytic assessment of the THI to date. In this study of 373 adults with chronic tinnitus, the authors found partial support for the three-factor model proposed by Newman and colleagues (1996), as this was a better fit than a unifactorial model. However, fit indices did not indicate a uniformly excellent fit.

The more recently-developed TFI has eight subscales: intrusiveness, sense of control, cognition, sleep, auditory, relaxation, quality of life, and emotional distress (Meikle et al., 2012). These subscales were based on exploratory factor analysis of a set of 30 items, after which five items were removed from inclusion in the final measure (Meikle et al., 2012). Notably, a major difference from the THI is the exclusion of any items referring to catastrophic feelings (e.g.,

desperation, fear). Fackrell and colleagues (2016, 2018) found in at least two studies that the hypothesized eight-factor structure resulted in a sub-optimal fit due to cross-loadings and lack of contribution by the auditory subscale to the overall construct of the functional impact of tinnitus. As these validation studies used a research volunteer population rather than a clinical population, concerns have been raised about the applicability of these results for a clinical population (Folmer, 2016; Henry et al., 2017).

For both the THI and TFI, researchers typically sum the scales despite the measures having been developed to assess specific subfactors related to tinnitus distress (e.g., Hesser et al., 2012; Krings et al., 2015; Robinson et al., 2008; Shekhawat et al., 2014). This practice results in a single, unifactorial assessment of tinnitus severity. Due to the preference for a unified measure assessing tinnitus severity as a singular construct, determining which items most clearly assess this construct is important. Also, in regard to utility, it is worth noting that both measures are frequently administered together in treatment studies (e.g., Bauer, Berry, & Brozoski, 2017; Beukes et al., 2018; Folmer et al., 2015; Kimball et al., 2018; Roland et al., 2015), presumably in order to compare findings across studies and increase sensitivity to detect treatment-related change. This practice may be burdensome for patients, as each measure contains 25 items. Consequently, it is unlikely that the measures will be administered more than twice. Furthermore, depending on the type and length of interventions, simple pre- and post-intervention assessment may not be enough to accurately and sensitively assess changes in tinnitus severity.

Current Study

Careful monitoring of tinnitus severity during the course of interventions is critical for the development and validation of evidence-based treatments. To lessen patient burden, a shorter

tinnitus outcome measure with a single total would be preferable to longer, multi-faceted measures for these purposes. Although a 10-item short form for the THI, the Tinnitus Handicap Inventory-Screening version (THI-S; Newman et al., 2008), exists and correlates highly with the THI, it was not designed to be responsive to change. The THI-S also contains only three response options (yes, sometimes, and no), which is not sensitive enough to small changes in tinnitus severity over shorter periods of time. A short form that is responsive to treatment change in addition to assessing tinnitus severity is thus required.

To create such a measure, more information regarding the psychometric properties of both the TFI and THI is needed. Existing evaluations have only been based on a single population, being either a research or clinical population from a single country. Thus, examining the psychometric properties using a wider population is desirable. The current study was undertaken to address these gaps in measurement of tinnitus severity. First, we sought to thoroughly examine the psychometric properties of both the TFI and THI. We assessed whether the two scales measure similar constructs, as well as how well those constructs are measured, including across cultures (the United States and United Kingdom). Furthermore, we examined whether a subset of items from both scales can be used to estimate tinnitus severity and change in tinnitus after treatment, thereby reducing patient burden and allowing for cross-study comparisons. Finally, we examined the psychometric properties of the newly developed Tinnitus Severity Short Form (TS-SF) in a separate sample of individuals receiving psychological treatment for tinnitus.

Method

Data Collection

To include a diverse range of tinnitus populations, data sets were sought where adults with bothersome tinnitus had completed at least some items from both the THI and TFI at least once. Ideally, participants would have completed the TFI and THI on at least two occasions. Data from both research and clinical populations from two English-speaking countries were sought, namely the United States (US) and the United Kingdom (UK). Existing de-identified datasets from nine studies conducted by the authorship team were included as summarized in Table 1.

An additional data set was collected once the short form was developed in order to assess the psychometric properties of this measure in a unique sample. The developed short form was administered to individuals in the US seeking online treatment for tinnitus prior to beginning treatment. Ethical approval (IRB-FY17-209) was obtained from the Institutional Review Board at Lamar University, Beaumont, Texas, US.

Measures of Tinnitus Severity

Two commonly used measures of tinnitus severity, the THI and the TFI, were the focus on the current study. The THI-S (i.e., screening version of the THI) was also used in some studies. Information about these measures is provided below.

Tinnitus Handicap Inventory

The THI is a 25-item scale with three subscales designed to assess functional, and emotional, and catastrophic response components of perceived tinnitus severity. Each item is answered on a 3-level response scale, with a “yes” response receiving four points, “sometimes” receiving two points, and “no” receiving zero points. The THI has good internal validity with Cronbach’s Alpha of 0.96 (Newman et al., 1996) and test-retest validity of $r = 0.92$ (Newman et

al., 1998). It is scored between 0-100 with lower scores indicating less severity. Of note, only 51 distinct values are possible as the THI is not a fully dimensional scale.

Tinnitus Handicap Inventory – Screening Version

In addition to the full THI, a shortened screening version has been developed that includes ten items drawn from the original scale. Items were selected based on item-total correlations, representativeness of content domains thought to underlie THI total scores, and face validity (Newman et al., 2008). The THI-S has been shown to correlate highly with the full THI (*r*s = 0.90-0.95); however, principal components analysis of the THI-S suggests cross-loadings of its items between the emotional and functional factors (Lee et al., 2014; Newman et al., 2008).

Tinnitus Functional Index

The TFI contains 25 items designed to assess eight components of tinnitus severity: intrusiveness, sense of control, cognition, sleep, auditory, relaxation, quality of life, and emotional distress. Each item is answered on a 10-point scale. Total scores range from 0-100 after a raw total score is divided by 25 (or the number of items answered) and multiplied by 10. All integers between 0 and 100 are possible, and higher scores indicate greater tinnitus severity. The TFI has high test-retest reliability, and meaningful change has been defined as a score reduced by 13 points or more (Meikle et al., 2012).

Data Analytic Plan

The data analysis aimed to assess how well constructs from the THI and TFI are measured and to identify a subset of items that can be used to assess treatment-related change. A detailed data analytic plan was pre-registered on Open Science Framework (Frumkin et al., 2020) and outlined below. Notably, preliminary results required us to deviate from the pre-registered plan in ways that are also detailed on Open Science Framework. The following data

analytic plan reflects the steps that were ultimately conducted. All analyses were conducted in Mplus version 8 (Muthén & Muthén, 2017).

Evaluating the Questionnaire Constructs

Confirmatory factor analyses were used to assess the factorial validity of the THI and TFI and to determine whether the two questionnaires measure similar constructs of tinnitus severity. The full sample was randomly split in half to allow for exploratory and confirmatory work as needed. In the first half of the data, we examined the TFI and THI separately to determine the best-fitting model. For each measure, three models were compared: (1) a single-factor model; (2) the most strongly supported model of the TFI (see Fackrell et al., 2016) and THI (see Kleinstäuber et al., 2015); (3) a bifactor version of the previous model, with all items loading on a single "distress" factor in addition to secondary loadings.

A bifactor model is one in which items are assumed to arise from at least two separate sources: a general factor representing shared variance among all or most items, and more specific factors representing shared variance among smaller groups of items (Reise, 2012). In this case, a bifactor model seemed plausible under the assumption that at least some items assess general distress or severity associated with tinnitus, as well as specific aspects of the tinnitus experience that are not necessarily always coincident with tinnitus severity. For example, difficulty falling or staying asleep might reflect sleep disturbance due to distress about tinnitus but might also reflect sleep disturbance due to other issues.

For models with only continuous variables, maximum likelihood estimation robust to departures from normality (using the MLR estimator) in Mplus was used to handle missing data (Muthén & Muthén, 2017). For models with categorical variables (i.e., range of less than seven), multiple imputation was conducted using the Multivariate Imputation by Chained Equations

(mice) package in R (van Buuren & Groothuis-Oudshoorn, 2011). Of the entire sample, 59% (n = 298) completed the THI in full, whereas others completed only a subset of items that comprise the THI-S. Missing data was present in <1% of cases (n = 1) for the TFI.

Model fit was assessed using the Tucker–Lewis incremental fit index (TLI; Tucker & Lewis, 1973), the comparative fit index (CFI; Bentler, 1990), and the root mean square error of approximation (RMSEA; for further explanation of fit statistics, see Muthén & Muthén, 2017). When models included only continuous data, we evaluated comparative model fit using the Akaike Information Criterion (AIC; Akaike, 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978), with lower values indicating better fit. For the TLI and CFI, values of .90 and above were considered adequate, whereas values of .95 or above were considered very good. For the RMSEA, values of .08 and below were considered adequate and .05 or below very good (Hu & Bentler, 1999). If no model achieved good fit, exploratory factor analysis was used to determine an appropriate factor structure. Similar steps have been previously undertaken with a wide variety of measures (e.g., Fernandez et al., 2014; Levinson et al., 2017; Rodebaugh, 2009; Vodanovich et al., 2005). However, as described below, these steps could not produce a well-fitting model in this case. Exploratory and confirmatory analyses were conducted in order to generate a model with nearly excellent fit.

Combining the THI and TFI

Next, we determined the best-fitting joint factor structure of the THI and TFI through similar competing models (i.e., single factor vs. best fitting structures as determined through initial testing of individual measures). Where bifactor models were indicated, we also tested a bifactor model with a single higher-order “distress” factor in order to assess the degree to which the two measures were capturing similar constructs.

Generating the Tinnitus Severity Short Form (TS-SF)

We then generated a short form combining items from the TFI and THI. We focused on items that demonstrated high loadings on the factor of interest, as well as limited loadings on secondary factors. Items were first selected within each individual scale and then combined into a joint short-form model. The intent was for all items to load on to a single factor. However, competing factor models were investigated when indicated.

Comparing Factor Structure across Cultures

We next examined whether the joint short form we developed was appropriately invariant across US and UK samples. Given the categorical nature of the THI items, we followed recommendations from Wu & Estabrook (2016) to test four aspects of metric invariance in the following order: (1) configural invariance (i.e., whether the groups have the same general factor structure); (2) scalar invariance (i.e., invariance of item thresholds); (3) metric invariance (i.e., invariance of item loadings); and (4) residual variance invariance. Model fit was assessed through change in CFI and RMSEA (Chen, 2007; Cheung & Rensvold, 2002). Although the chi-square difference test has historically been used to assess the impact of imposing invariance constraints, several studies have shown that this statistic is sensitive to sample size and inflates Type I error (Bagozzi, 1977; Bentler & Bonett, 1980; Yuan & Chan, 2016). Based on recommendations by Chen (2007), invariance is indicated by the more constrained (invariant) model having not that much lower a CFI (< .005 change) and not that much higher an RMSEA (< .01 change). An acceptably partially invariant model would have invariant loadings (configural invariance), and metric, scalar, and residual variance would affect fewer than half of the items.

Assessing Treatment-Related Change using the TS-SF

Treatment data was used to assess how well the TS-SF accounted for changes in the TFI, THI, and THI-S. Initially, latent trajectory analyses were conducted to determine how well the short form captured change across treatment shown in other measures. Simple regression analyses were also used to determine how much variance the short form predicted in the observed scores of the TFI, THI, and THI-S.

Evaluating the TS-SF in a Separate Sample

Finally, factor structure of the short form was assessed in a unique sample. All items were placed on a 0-10 response scale. A single-factor structure was tested, as well as models involving method factors where indicated. Unfortunately, an instruction was added to the short form that resulted in confusion for participants as to how to respond to the lone reverse-scored items, so this item was removed from analyses. Together, these analyses provided further understanding of the utility of the proposed short form compared to existing measures.

Results

Participant Characteristics

Existing datasets from nine studies (N = 502) were used. Of the full sample, 42% of participants (n = 213) completed the questionnaires in the US and 58% of participants (n = 289) completed the questionnaires in the UK. Demographic and clinical characteristics are provided in Table 2. Once the short form was developed, it was administered to 164 individuals in the US seeking online treatment for tinnitus prior to beginning treatment.

Evaluating the Questionnaire Constructs

As seen in Table 3, the eight-factor and bifactor models of the TFI achieved acceptable to good fit. The bifactor model appeared to fit marginally better and was retained for further testing. When testing the Kleinstauber (2015) three-factor model of the THI, the emotional and

299 catastrophizing factors were so highly correlated that the model produced an improper estimate
300 (an $r > 1$). Combining these two factors resulted in some improvement in fit (see Table 3, THI
301 Two Factor). Fit was further improved by including a bifactor distress factor. As seen in Table 3,
302 fit was similar for a bifactor model of the THI with three additional factors and with only one
303 additional factor retained (the functional factor). As such, the more parsimonious bifactor model
304 was retained for further testing.

305 For both the TFI and THI, fit of the most successful models was still less than optimal.
306 Exploratory factor analyses did not suggest additional factors were indicated; nor did they
307 suggest plausible alternative factor structures. As such, the most parsimonious bifactor models
308 were retained for analysis of the joint TFI and THI factor structure. The joint bifactor model
309 achieved poor fit (CFI = .705, TLI = .677, RMSEA = .063), as did a non-bifactor version of the
310 same model (CFI =.692, TLI = .675, RMSEA = .065). When a single bifactor distress factor (that
311 all items loaded on) was included, fit continued to worsen (CFI = .649, TLI = .619, RMSEA =
312 .071), suggesting important discrepancies in the underlying constructs assessed by the two
313 measures despite their theoretical overlap.

314 At this point we departed from our pre-registered analysis plan because it became clear
315 that there were problems with the item pool. That is, although intended to measure a unified
316 construct, no joint factor structure could be found that fit well, although bifactor models showed
317 close to good fit in the individual measures. The pattern of results suggested a situation in which
318 most of the items assessed multiple constructs (e.g., tinnitus severity together with another
319 element partially related to the tinnitus experience), but assessment of the secondary elements
320 was not strong enough to produce good fit with a bifactor model.

321 **Generating the TS-SF**

Creating a short form from both the TFI and THI was approached using items with limited loadings on secondary factors. Based on the acceptable bi-factor model of the TFI we had already tested, we selected the item from each factor with the highest loading on the general "distress" factor. This resulted in an 8-item short form that achieved acceptable to excellent fit (CFI = .959, TLI = .943, RMSEA = .08). However, consistent with previous findings (e.g., Fackrell et al., 2016, 2018), the item from the auditory factor loaded less well. Removing this item resulted in a 7-item short form of the TFI with good, although still not excellent fit (CFI = .969, TLI = .953, RMSEA = .08).

For the THI, we selected items from the more parsimonious bifactor model. We chose items that loaded on to the general distress but not the functional factor in order to generate a model with a single factor. We then removed items with low loadings (i.e., below 0.6). The resulting unifactorial model with 11 items showed excellent fit according to the CFI and TLI (.954 and .943, respectively); however, the RMSEA remained high (.105).

A joint bifactor model was then fit utilizing the items selected from each measure. All items loaded on to a general distress factor, and the TFI items loaded on to an additional method factor. The five items from each scale with the highest loadings on the joint factor (TFI items 5, 9, 17, 20, and 24; THI items 10, 14, 16, 21, and 23) were retained. The resulting bifactor model fit well (CFI = .960, TLI = .941, RMSEA = .069). Final items and factor loadings are displayed in Table 4. A total score for the combined short form was generated by summing the five items from each scale and dividing that sum by the total possible points (50 for the TFI, 20 for the THI). The two scores were then added together and multiplied by 50 for a total possible score ranging from 0-100 (although having fewer than 100 possible scores). As seen in Table 5, the

short form correlated highly with observed scores on the full TFI and THI. Internal consistency for the short form was excellent ($\alpha = 0.92$).

The combined short form was then tested for invariance across the US and UK samples. Although item thresholds appeared to be invariant across groups, constraining loadings to be equal across groups caused worse fit (ΔCFI = .157, ΔTLI = .157, ΔRMSEA = .057). This result suggests a lack of metric invariance such that individuals may be responding differently to the items across the two groups**.** Freeing some loadings to impose partial metric invariance did not improve model fit to the point that we could be confident in moving forward treating the data as part of one group. Thus, further analyses regarding treatment data were conducted only in the UK sample, in which a larger amount of treatment data was available (n = 271).

Assessing Treatment-related Change using the TS-SF

Latent trajectory analyses indicated a non-linear slope. Because the nonlinear slope would complicate further analyses, treatment-related change was assessed using change scores rather than latent trajectory models. To maximize the data available, change pre- to post-treatment was examined. Change in the combined short-form correlated highly with change in TFI (r = .84), THI (r = .76), and the THI-S (r = .84). As seen in Table 6, regression analyses suggested that change in the combined short-form accounted for unique variance in THI and TFI change over and above change in the THI-S. Notably, change in the THI-S did not account for unique variance in change in the TFI, over and above the joint short form.

Evaluating the TS-SF in a Separate Sample

The combined short form was administered on a 0-10 scale to individuals seeking treatment for tinnitus distress in the US (see Appendix A). A single-factor model achieved below adequate fit (CFI = .881, TLI = .847, RMSEA = .161). However, a bifactor model with TFI and

THI items indicated by two separate method factors achieved excellent fit in terms of two of the three fit indices (CFI = .978, TLI = .958, RMSEA = .092, see Figure 1 for item loadings). Notably, RMSEA is likely to be a poor indicator of model fit due to the relatively small sample size (*n* = 164, Kenny, Kaniskan, & McCoach, 2014).

Discussion

The aim of this study was to compare the psychometric properties of two widely used measures of tinnitus severity, the THI and TFI. We further sought to develop a short measure of subjective tinnitus severity that is responsive to change and can be used to monitor those with tinnitus at regular intervals particularly when undertaking tinnitus interventions.

Constructs of the THI and TFI

Psychometric analyses suggest that the THI and TFI in their original forms do not measure precisely the same construct, as evidenced by poor fit of a unified bifactor model (and, indeed, any joint model that was attempted). Our results echo previous findings (e.g., Fackrell et al., 2016, 2018; Kleinstäuber et al., 2015), which have questioned the hypothesized factor structure of these original measures. Furthermore, our results suggest that the tinnitus severity constructs assessed by each questionnaire as a whole should not be compared to one another, as they do not appear to measure the same underlying construct, although the constructs they do measure overlap considerably. Of course, one solution to this problem is to use both measures, but this solution is impractical for studies in which tinnitus severity is to be measured multiple times. Thus, a short form combining the two measures is necessary for assessing tinnitus severity in a way that can be compared to prior research in the field while minimizing patient burden.

Short Form for Monitoring Tinnitus Severity

Due to the lack of a responsive short form to monitor changes in tinnitus severity over time, such a form was developed by combining the TFI and THI, focusing on items with limited loadings on secondary factors. The 10-item TS-SF developed here has the advantage of being a brief measure, which reduces patient burden. The new short form also has good psychometric properties. Confirmatory factor analysis suggests good to excellent fit of the initial combined measure, which exceeded the fit of confirmatory models assessing factor structure of the TFI and THI. Placing all items on the same 0-10 scale continued to indicate good to excellent fit. This administration format has the advantage of increasing continuity and variability, both of which should serve to make the measure more sensitive to treatment-related change.

The TS-SF correlates highly with the TFI, THI, and THI-S, both at baseline and when assessing treatment-related change. Regression analyses suggest that change in the TS-SF accounts for significant variance in change in both the THI and TFI. Change in the THI-S was not a significant predictor of treatment-related change in the TFI over and above the TS-SF. These results suggest that the 10-item TS-SF, which was developed in the current study by combining items from the TFI and THI, is a valuable tool for assessing both tinnitus severity and treatment-related changes in tinnitus severity.

Study Limitations

The following limitations of the current study need to be considered. There was a lack of measurement invariance found between the US and UK samples. There are multiple potential reasons for this outcome. First, participants in the UK sample were all offered treatment, whereas not all of the US studies provided treatment to participants. A second possibility is that cultural differences between the UK and US might lead individuals with tinnitus to vary in the way they rate the severity of their tinnitus or willingness to endorse distress. Thus, further research is

needed to examine the suitability of the TS-SF across countries and cultures to determine whether our current results are due to differences in sampling or differences in cultures. Notably, the TS-SF did achieve good fit in the additional sample of individuals seeking treatment for tinnitus in the US.

Ideally, the TS-SF would be unifactorial to allow for simple interpretation. However, our results suggest that method factors are needed even when placing all items on the same scale, likely because the THI items are designed to elicit extreme responses (e.g., yes or no), whereas the TFI was specifically designed to avoid extreme or catastrophic reactions (Fackrell et al., 2016). Additionally, it is unclear from our analysis whether it is useful to include one reverse-scored item. Finally, further research is needed to assess sensitivity to treatment-related change. Although our results suggest that change in the TS-SF correlates highly with change in the TFI and THI, we have not yet established how much of a change in the TS-SF score is clinically important (i.e., minimal clinically important difference). Research that compares change in the TS-SF with clinician assessment of treatment-related change will be critical in assuring that the measure is appropriately sensitive as an outcomes instrument. Until such research is completed, the TS-SF should not replace existing full-length measures that have shown to be sensitive to treatment-related change.

Conclusions

The aim of this study was to develop and preliminarily validate a brief measure of tinnitus severity. The resulting 10-item TS-SF combines items from the TFI and THI, two existing widely used measures of tinnitus severity. The TS-SF appears to effectively assess the construct of tinnitus severity, as demonstrated by its strong psychometric properties. The TS-SF also appears sensitive to treatment-related change based on associations with existing measures

used for this purpose. Further research is necessary to establish a threshold of clinically important change in the TS-SF and confirm its sensitivity to treatment-related change. As psychological treatments for tinnitus continue to be developed, we urge researchers to carefully consider their outcome measures and to utilize repeated assessments wherever possible. Repeating a 10-item short form several times is much more feasible and involves much less burden than repeating either the THI or TFI alone or (even more so) in combination. These practices will greatly increase the reliability of results and support the development of research on treatment of tinnitus severity.

References

Aazh, H., & Moore, B. C. J. (2018). Thoughts about suicide and self-harm in patients with tinnitus and hyperacusis. *Journal of the American Academy of Audiology*, *29*(03), 255–261. https://doi.org/10.3766/jaaa.16181

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Andersson, G. (2002). Psychological aspects of tinnitus and the application of cognitive–behavioral therapy. *Clinical Psychology Review*, *22*(7), 977–990. https://doi.org/10.1016/S0272-7358(01)00124-6

Andersson, G., & Lyttkens, L. (1999). A meta-analytic review of psychological treatments for tinnitus. *British Journal of Audiology*, *33*(4), 201–210. https://doi.org/10.3109/03005369909090101

Bagozzi, R. P. (1977). Structural equation models in experimental research. *Journal of Marketing Research*, *14*(2), 209–226. https://doi.org/10.1177/002224377701400209

Baguley, D. M., & Andersson, G. (2003). Factor analysis of the Tinnitus Handicap Inventory. *American Journal of Audiology*, *12*(1), 31–34. https://doi.org/10.1044/1059-0889(2003/007)

Bauer, C. A., Berry, J. L., & Brozoski, T. J. (2017). The effect of tinnitus retraining therapy on chronic tinnitus: A controlled trial. *Laryngoscope Investigative Otolaryngology*, *2*(4), 166–177. https://doi.org/10.1002/lio2.76

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of

covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Beukes, E. W., Allen, P. M., Manchaiah, V., Baguley, D. M., & Andersson, G. (2017). Internet-Based Intervention for Tinnitus: Outcome of a Single-Group Open Trial. *Journal of the American Academy of Audiology*, *28*(04), 340–351. https://doi.org/10.3766/jaaa.16055

Beukes, E. W., Andersson, G., Allen, P. M., Manchaiah, V., & Baguley, D. M. (2018). Effectiveness of guided internet-based cognitive behavioral therapy vs face-to-face clinical care for treatment of tinnitus. *JAMA Otolaryngology–Head & Neck Surgery*, *144*(12), 1126. https://doi.org/10.1001/jamaoto.2018.2238

Beukes, E. W., Baguley, D. M., Allen, P. M., Manchaiah, V., & Andersson, G. (2018). Audiologist-guided internet-based cognitive behavior therapy for adults with tinnitus in the United Kingdom. *Ear and Hearing*, *39*(3), 423–433. https://doi.org/10.1097/AUD.0000000000000505

Brown, T. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cima, R. F. F., Mazurek, B., Haider, H., Kikidis, D., Lapira, A., Noreña, A., & Hoare, D. J. (2019). A multidisciplinary European guideline for tinnitus: Diagnostics, assessment, and treatment. *HNO*, *67*(S1), 10–42. https://doi.org/10.1007/s00106-019-0633-7

Das, S. K., Wineland, A., Kallogjeri, D., & Piccirillo, J. F. (2012). Cognitive speed as an objective measure of tinnitus. *The Laryngoscope*, *122*(11), 2533–2538. https://doi.org/10.1002/lary.23555

Davis, A., & Refaie, A. E. (2020). The epidemiology of tinnitus. In R. Tyler (Ed.), *The Handbook of Tinnitus* (pp. 1–23). Singular.

Fackrell, K., Hall, D. A., Barry, J. G., & Hoare, D. J. (2016). Psychometric properties of the Tinnitus Functional Index (TFI): Assessment in a UK research volunteer population. *Hearing Research*, *335*, 220–235. https://doi.org/10.1016/j.heares.2015.09.009

Fackrell, K., Hall, D. A., Barry, J. G., & Hoare, D. J. (2018). Performance of the Tinnitus Functional Index as a diagnostic instrument in a UK clinical population. *Hearing Research*, *358*, 74–85. https://doi.org/10.1016/j.heares.2017.10.016

Fackrell, K., Hall, D. A., Barry, J., & Hoare, D. J. (2014). Tools for tinnitus measurement: Development and validity of questionnaires to assess handicap and treatment effects. In F. Signorelli & F. Turjman (Eds.), *Tinnitus: causes, treatment and short and long-term health effects* (pp. 13–60). Nova Science Publishers Inc.

Fernandez, K. C., Langer, J. K., Rodebaugh, T. L., Jakatdar, T. A., Heimberg, R. G., & Lenze, E. J. (2014). Validation of the Ambivalent and Purposeful Engagement – Trait Measure. *Anxiety, Stress, & Coping*, *27*(3), 317–334. https://doi.org/10.1080/10615806.2013.838228

Folmer, R. L. (2016). Reply to: Psychometric properties of the Tinnitus Functional Index (TFI): Assessment in a UK research volunteer population. *Hearing Research*, *335*, 236. https://doi.org/10.1016/j.heares.2016.02.011

Folmer, R. L., Theodoroff, S. M., Casiana, L., Shi, Y., Griest, S., & Vachhani, J. (2015). Repetitive transcranial magnetic stimulation treatment for chronic tinnitus: A randomized

clinical trial. *JAMA Otolaryngology - Head and Neck Surgery*, *141*(8), 716–722. https://doi.org/10.1001/jamaoto.2015.1219

Frumkin, M.R., Kallogjeri, D., Piccirillo, J.F., Beukes, E., Manchaiah, V., Andersson, G., & Rodebaugh, T.L. (2020). Development and preliminary evaluation of the Tinnitus Severity Short Form preregistration. Retrieved from https://osf.io/gstzm/

Geocze, L., Mucci, S., Abranches, D. C., Marco, M. A. de, & Penido, N. de O. (2013). Systematic review on the evidences of an association between tinnitus and depression. *Brazilian Journal of Otorhinolaryngology*, *79*(1), 106–111. https://doi.org/10.5935/1808-8694.20130018

Hall, D. A. (2017). Designing clinical trials for assessing the effectiveness of interventions for tinnitus. *Trends in Hearing*, *21*, 233121651773668. https://doi.org/10.1177/2331216517736689

Hall, D. A., Haider, H., Szczepek, A. J., Lau, P., Rabau, S., Jones-Diette, J., Londero, A., Edvall, N. K., Cederroth, C. R., Mielczarek, M., Fuller, T., Batuecas-Caletrio, A., Brueggemen, P., Thompson, D. M., Norena, A., Cima, R. F. F., Mehta, R. L., & Mazurek, B. (2016). Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. *Trials*, *17*(1), 270. https://doi.org/10.1186/s13063-016-1399-9

Henry, J. A., Griest, S., Thielman, E., McMillan, G., Kaelin, C., & Carlson, K. F. (2016). Tinnitus Functional Index: Development, validation, outcomes research, and clinical application. *Hearing Research*, *334*, 58–64. https://doi.org/10.1016/j.heares.2015.06.004

Henry, J. A., Thielman, E., & Zaugg, T. (2017). Reply to: Psychometric properties of the Tinnitus Functional Index (TFI): Assessment in a UK research volunteer population. *Hearing Research*, *350*, 222–223. https://doi.org/10.1016/j.heares.2016.10.022

Hesser, H., Gustafsson, T., Lundén, C., Henrikson, O., Fattahi, K., Johnsson, E., Westin, V. Z., Carlbring, P., Mäki-Torkko, E., Kaldo, V., & Andersson, G. (2012). A randomized controlled trial of internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. *Journal of Consulting and Clinical Psychology*, *80*(4), 649–661. https://doi.org/10.1037/a0027021

Hesser, H., Weise, C., Westin, V. Z., & Andersson, G. (2011). A systematic review and meta-analysis of randomized controlled trials of cognitive-behavioral therapy for tinnitus distress. *Clinical Psychology Review*, *31*(4), 545–553. https://doi.org/10.1016/j.cpr.2010.12.006

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jackson, R., Vijendren, A., & Phillips, J. (2019). Objective measures of tinnitus: A systematic review. *Otology & Neurotology*, *40*(2), 154–163. https://doi.org/10.1097/MAO.0000000000002116

Jacquemin, L., Mertens, G., Van de Heyning, P., Vanderveken, O. M., Topsakal, V., De Hertogh, W., Michiels, S., Van Rompaey, V., & Gilles, A. (2019). Sensitivity to change and convergent validity of the Tinnitus Functional Index (TFI) and the Tinnitus Questionnaire (TQ): Clinical and research perspectives. *Hearing Research*, *382*, 107796. https://doi.org/10.1016/j.heares.2019.107796

Kallogjeri, D., Piccirillo, J. F., Spitznagel, E., Hale, S., Nicklaus, J. E., Hardin, F. M., Shimony, J. S., Coalson, R. S., & Schlaggar, B. L. (2017). Cognitive training for adults with bothersome tinnitus a randomized clinical trial. *JAMA Otolaryngology - Head and Neck Surgery*, *143*(5), 443–451. https://doi.org/10.1001/jamaoto.2016.3779

Kimball, S., Johnson, C., Baldwin, J., Barton, K., Mathews, C., & Danhauer, J. (2018). Hearing aids as a treatment for tinnitus patients with slight to mild sensorineural hearing loss. *Seminars in Hearing*, *39*(02), 123–134. https://doi.org/10.1055/s-0038-1641739

Kleinstäuber, M., Frank, I., & Weise, C. (2015). A confirmatory factor analytic validation of the Tinnitus Handicap Inventory. *Journal of Psychosomatic Research*, *78*(3), 277–284. https://doi.org/10.1016/j.jpsychores.2014.12.001

Krings, J. G., Winel, A., Kallogjeri, D., Rodebaugh, T. L., Nicklaus, J., Lenze, E. J., & Piccirillo, J. F. (2015). A novel treatment for tinnitus and tinnitus-related cognitive difficulties using computer-based cognitive training and D-cycloserine. *JAMA Otolaryngology - Head and Neck Surgery*, *141*(1), 18–26. https://doi.org/10.1001/jamaoto.2014.2669

Landry, E. C., Sandoval, X. C. R., Simeone, C. N., Tidball, G., Lea, J., & Westerberg, B. D. (2019). Systematic review and network meta-analysis of Cognitive and/or Behavioral Therapies (CBT) for tinnitus. *Otology & Neurotology*, *41*(2), 1. https://doi.org/10.1097/MAO.0000000000002472

Lee, J. H., Ra, J. J., & Kim, Y. H. (2014). Adequacy of the simplified version of the Tinnitus Handicap Inventory (THI-S) to measure tinnitus handicap and relevant distress. *Korean Journal of Audiology*, *18*(1), 19–27. https://doi.org/10.7874/kja.2014.18.1.19

Levinson, C. A., Rodebaugh, T. L., Lim, M. H., & Fernandez, K. C. (2017). The core extrusion schema-revised: Hiding oneself predicts severity of social interaction anxiety. *Assessment*, *24*(1), 83–94. https://doi.org/10.1177/1073191115596568

Martinez-Devesa, P., Waddell, A., Perera, R., & Theodoulou, M. (2007). Cognitive behavioural therapy for tinnitus. *Cochrane Database of Systematic Reviews*, *1*. https://doi.org/10.1002/14651858.CD005233.pub2

McFerran, D. J., Stockdale, D., Holme, R., Large, C. H., & Baguley, D. M. (2019). Why is there no cure for tinnitus? *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00802

Meikle, M., Henry, J. A., Griest, S. E., Stewart, B. J., Abrams, H. B., McArdle, R., Myers, P. J., Newman, C. W., Sandridge, S., Turk, D. C., Folmer, R. L., Frederick, E. J., House, J. W., Jacobson, G. P., Kinney, S. E., Martin, W. H., Nagler, S. M., Reich, G. E., Searchfield, G., … Vernon, J. A. (2012). The Tinnitus Functional Index. *Ear and Hearing*, *33*(2), 153–176. https://doi.org/10.1097/AUD.0b013e31822f67c0

Meikle, M., Stewart, B., Griest, S., Martin, W. H., Henry, J. A., Abrams, H. B., McArdle, R., Newman, C. W., & Sandridge, S. A. (2007). Assessment of tinnitus: Measurement of treatment outcomes. *Progress in Brain Research*, *166*, 511–521. https://doi.org/10.1016/S0079-6123(07)66049-X

Meikle, M., Stewart, B. J., Griest, S. E., & Henry, J. A. (2008). Tinnitus outcomes assessment. *Trends in Amplification*, *12*(3), 223–235. https://doi.org/10.1177/1084713808319943

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide, Eighth Edition*. Muthén & Muthén.

Newman, C. W., Jacobson, G. P., & Spitzer, J. B. (1996). Development of the tinnitus handicap inventory. *Archives of Otolaryngology - Head and Neck Surgery*, *122*(2), 143–148. https://doi.org/10.1001/archotol.1996.01890140029007

Newman, C. W., Sandridge, S. A., & Bolek, L. (2008). Development and psychometric adequacy of the screening version of the tinnitus handicap inventory. *Otology and Neurotology*, *29*(3), 276–281. https://doi.org/10.1097/MAO.0b013e31816569c4

Newman, C. W., Sandridge, S. A., & Jacobsont, G. P. (1998). Psychometric adequacy of the

Tinnitus Handicap Inventory (THI) for evaluating treatment outcome. *J Am Acad Audiol, 9,* 153–160.

Pattyn, T., Van Den Eede, F., Vanneste, S., Cassiers, L., Veltman, D. J., Van De Heyning, P., & Sabbe, B. C. G. (2016). Tinnitus and anxiety disorders: A review. *Hearing Research, 333,* 255–265. https://doi.org/10.1016/j.heares.2015.08.014

Rademaker, M. M., Stegeman, I., Ho-Kang-You, K. E., Stokroos, R. J., & Smit, A. L. (2019). The effect of mindfulness-based interventions on tinnitus distress: A systematic review. *Frontiers in Neurology, 10,* 1135. https://doi.org/10.3389/fneur.2019.01135

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Robinson, S. K., Viirre, E. S., Bailey, K. A., Kindermann, S., Minassian, A. L., Goldin, P. R., Pedrelli, P., Harris, J. P., Mcquaid, J. R., & Robinson, S. (2008). A randomized controlled trial of cognitive-behavior therapy for tinnitus. *International Tinnitus Journal, 14*(2), 119–126.

Rodebaugh, T. L. (2009). Hiding the self and social anxiety: The core extrusion schema measure. *Cognitive Therapy and Research, 33*(1), 90–109. https://doi.org/10.1007/s10608-007-9143-0

Roland, L. T., Lenze, E. J., Hardin, F. M., Kallogjeri, D., Nicklaus, J., Wineland, A. M., Fendell, G., Peelle, J. E., & Piccirillo, J. F. (2015). Effects of Mindfulness Based Stress Reduction Therapy on Subjective Bother and Neural Connectivity in Chronic Tinnitus. *Otolaryngology–Head and Neck Surgery*, *152*(5), 919–926. https://doi.org/10.1177/0194599815571556

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–

464. https://doi.org/10.1214/aos/1176344136

Searchfield, G. D., Durai, M., & Linford, T. (2017). A state-of-the-art review: Personalization of tinnitus sound therapy. *Frontiers in Psychology*, *8*(SEP), 1599. https://doi.org/10.3389/fpsyg.2017.01599

Shargorodsky, J., Curhan, G. C., & Farwell, W. R. (2010). Prevalence and characteristics of tinnitus among US adults. *American Journal of Medicine*, *123*(8), 711–718. https://doi.org/10.1016/j.amjmed.2010.02.015

Shekhawat, G. S., Searchfield, G. D., & Stinear, C. M. (2014). Randomized Trial of Transcranial Direct Current Stimulation and Hearing Aids for Tinnitus Management. *Neurorehabilitation and Neural Repair*, *28*(5), 410–419. https://doi.org/10.1177/1545968313508655

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

Tunkel, D. E., Bauer, C. A., Sun, G. H., Rosenfeld, R. M., Chandrasekhar, S. S., Cunningham, E. R., Archer, S. M., Blakley, B. W., Carter, J. M., Granieri, E. C., Henry, J. A., Hollingsworth, D., Khan, F. A., Mitchell, S., Monfared, A., Newman, C. W., Omole, F. S., Phillips, C. D., Robinson, S. K., … Whamond, E. J. (2014). Clinical practice guideline: Tinnitus. *Otolaryngology–Head and Neck Surgery*, *151*(2), S1–S40. https://doi.org/10.1177/0194599814545325

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained

equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Vodanovich, S. J., Wallace, J. C., & Kass, S. J. (2005). A confirmatory approach to the factor structure of the Boredom Proneness Scale: evidence for a two-factor short form. *Journal of Personality Assessment*, *85*(3), 295–303. https://doi.org/10.1207/s15327752jpa8503_05

Wilson, M. B., Kallogjeri, D., Joplin, C. N., Gorman, M. D., Krings, J. G., Lenze, E. J., Nicklaus, J. E., Spitznagel, E. E., & Piccirillo, J. F. (2015). Ecological momentary assessment of tinnitus using smartphone technology: a pilot study. *Otolaryngology--Head and Neck Surgery*, *152*(5), 897–903. https://doi.org/10.1177/0194599815569692

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*(4), 1014–1045. https://doi.org/10.1007/s11336-016-9506-0

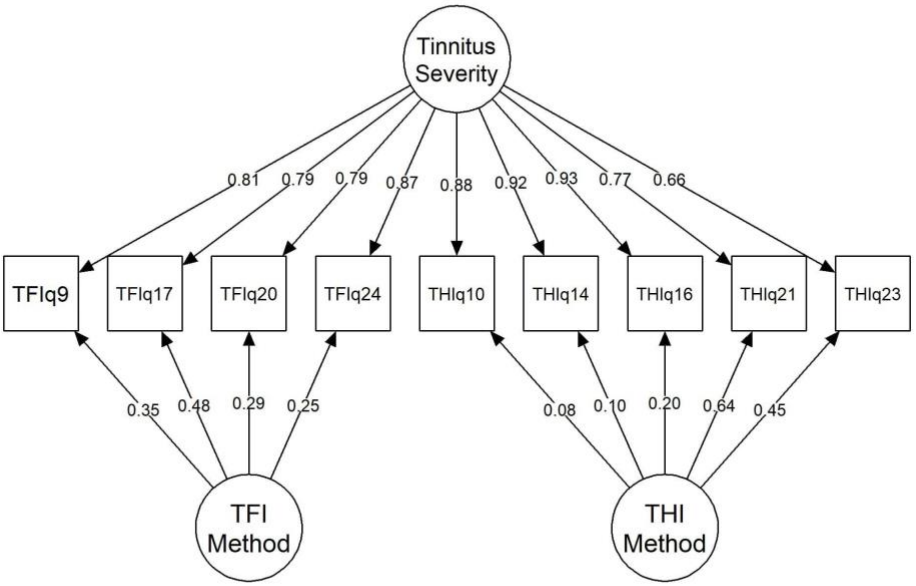
Yuan, K. H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests. *Psychological Methods*, *21*(3), 405–426. https://doi.org/10.1037/met0000080

Zachariae, R., Mirz, F., Johansen, L. V., Andersen, E., Bjerring, P., Pedersen, C. B., & Andersen, S. E. (2000). Reliability and validity of a Danish adaptation of the Tinnitus Handicap Inventory. *Scandinavian Audiology*, *29*(1), 37–43. https://doi.org/10.1080/010503900424589

Zeman, F., Koller, M., Figueiredo, R., Aazevedo, A., Rates, M., Coelho, C., Kleinjung, T., de Ridder, D., Langguth, B., & Landgrebe, M. (2011). Tinnitus Handicap Inventory for evaluating treatment effects. *Otolaryngology–Head and Neck Surgery*, *145*(2), 282–287. https://doi.org/10.1177/0194599811403882

673

674



675

676 *Figure 1*. Bifactor model of Tinnitus Severity Short Form administered to a new sample of

677 individuals seeking treatment for tinnitus. Loadings are standardized. TFI = Tinnitus Functional

678 Index; THI = Tinnitus Handicap Inventory. Error terms are not shown for the sake of simplicity.

679

Table 1. Data collected

Country	Reference	*N*	Outcome measures	Number of assessment time points	Population type	Administration format
United States	(Das et al., 2012)	100	TFI, THI	1	Research population seeking treatment for tinnitus. Study did not offer intervention.	Paper
	(Krings et al., 2015)	35	TFI, THI	2	Research population seeking treatment for tinnitus.	Online
	(Wilson et al., 2015)	20	TFI, THI	1	Research population seeking treatment for tinnitus. Study did not offer intervention.	Online
	(Kallogjeri et al., 2017)	40	TFI, THI	2	Research population seeking treatment for tinnitus.	Online
	(Roland et al., 2015)	18	TFI, THI	1	Research population seeking treatment for tinnitus.	Paper
United Kingdom	(Beukes et al., 2017)	31	TFI, THI-S	2	Research population seeking CBT treatment for tinnitus.	Online
	(Beukes, Baguley, et al., 2018)	146	TFI, THI-S	2	Research population seeking CBT treatment for tinnitus.	Online
	(Beukes, Andersson, et al., 2018)	92	TFI, THI, THI-S	2	Clinical population seeking treatment for tinnitus.	Online
	(Beukes, Baguley, et al., 2018)	48	TFI, THI-S	1	Research population seeking treatment for tinnitus but not undergoing the intervention (out of criteria or decided not to proceed).	Online

Note: TFI = Tinnitus Functional Index; THI = Tinnitus Handicap Index; THI-S = Tinnitus Handicap Index – Screening Version

Table 2. Baseline participant characteristics of existing data sets

Characteristic	Overall (*n* = 502)	United Kingdom (*n* = 289)	United States (*n* = 213)
Gender, % female (n)	57% (*n* = 288)	58% (*n* = 168)	56% (*n* = 120)
Age, M (SD)	53.96 (11.28)	54.76 (12.84)	52.82 (8.47)
TFI, M (SD)	48.12 (22.10)	55.73 (21.51)	37.75 (18.39)
THI, M (SD)	39.37 (20.24)	45.85 (21.83)	36.48 (18.84)
THI-S, M (SD)	18.41 (9.14)	21.00 (9.19)	14.85 (7.80)

Note: TFI = Tinnitus Functional Index; THI = Tinnitus Handicap Index; THI-S = Tinnitus Handicap Index – Screening Version; M = mean; SD = standard deviation

Table 3. Fit Statistics of Confirmatory Factor Analytic models.

Model	CFI	TLI	RMSEA	AIC	BIC
TFI Single Factor	0.581	0.543	0.177	27427.798	27691.908
TFI Eight Factor (Fackrell, 2016)	0.943	0.935	0.067	25023.429	25326.275
TFI Bifactor	0.945	0.934	0.067	25016.856	25372.524
THI Single Factor	0.874	0.863	0.095		
THI Three Factor (Kleinstauber, 2015)	0.896	0.884	0.088		
THI Two Factor	0.899	0.889	0.090		
THI Bifactor – 4 factors total	0.906	0.886	0.087		
THI Bifactor – 2 factors total	0.901	0.886	0.087		
Joint model – TFI Eight Factor + THI Two Factor	0.692	0.675	0.065		
Joint Bifactor	0.705	0.677	0.063		
Joint Bifactor – single distress factor	0.649	0.619	0.071		
Tinnitus Severity Short Form (joint short form)	0.960	0.941	0.069		

Note. TFI = Tinnitus Functional Index; THI = Tinnitus Handicap Inventory; CFI = comparative fit index; TLI = Tucker–Lewis incremental fit index; RMSEA = root mean square error of approximation; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

Table 4. Standardized loadings of Tinnitus Severity Short Form items.

Item	Factor 1	Factor 2
TFI5. Over the PAST WEEK...How easy was it for you to COPE with your tinnitus?	0.64	0.52
TFI9. Over the PAST WEEK, how much did your tinnitus interfere with... Your ability to FOCUS ATTENTION on other things besides your tinnitus?	0.65	0.49
TFI17. Over the PAST WEEK, how much has your tinnitus interfered with... Your ability to RELAX?	0.61	0.51
TFI20. Over the PAST WEEK, how much has your tinnitus interfered with... Your ENJOYMENT OF LIFE?	0.69	0.51
TFI24. Over the PAST WEEK... How BOTHERED or UPSET have you been because of your tinnitus?	0.73	0.44
THI10. Because of your tinnitus, do you feel frustrated?	0.76	
THI14. Because of your tinnitus, do you find that you are often irritable?	0.75	
THI16. Does your tinnitus make you upset?	0.82	
THI21. Because of your tinnitus, do you feel depressed?	0.87	
THI23. Do you feel that you can no longer cope with your tinnitus?	0.81	

Note. TFI = Tinnitus Functional Index; THI = Tinnitus Handicap Inventory. To estimate a short form score from the TFI total, use the following equation: $Y_{\text{Short Form}} = -4.31 + 0.97*TFI_{\text{Total}}$. To estimate a short form score from the THI total, use the following equation: $Y_{\text{Short Form}} = -4.44 + 1.09*THI_{\text{Total}}$. If the THI and TFI score are both available, the following formula can be used for greater precision: $Y_{\text{Short Form}} = -9.88 + 0.48*TFI_{\text{Total}} + 0.68*THI_{\text{Total.}}$ To estimate the TFI total from a short form score, use the following equation: $TFI_{\text{Total}} = 14.08 + 0.80*Short_{\text{Total}}$. To estimate the THI total from a short form score, use the following equation: $THI_{\text{Total}} = 10.81 + 0.76*\ Short_{\text{Total}}$.

Table 5. Means, standard deviations, and correlations among pre-treatment measures and post-treatment change scores

Variable	*M*	*SD*	Tinnitus Severity Short Form	TFI Pre	THI Pre	THI-S Pre	Tinnitus Severity Short Form Change	TFI Change	THI Change
Tinnitus Severity Short Form	52.28	21.76							
TFI Pre	56.20	19.64	.85 [.82, .89]						
THI Pre	49.71	14.79	.80 [.74, .87]	.68 [.58, .77]					
THI-S Pre	20.86	8.91	.91 [.89, .93]	.78 [.73, .83]	.83 [.77, .90]				
Tinnitus Severity Short Form Change	18.14	20.42							
TFI Change	18.63	20.26					.84 [.79, .88]		
THI Change	13.79	14.71					.76 [.67, .84]	.61 [.50, .72]	
THI-S Change	7.29	8.88					.85 [.80, .89]	.70 [.62, .78]	.79 [.71, .86]

Note. M = mean; *SD* = standard deviation; TFI = Tinnitus Functional Index; THI = Tinnitus Handicap Inventory; THI-S = Tinnitus Handicap Inventory Short Form. All values are averaged across results from 40 data sets multiply imputed using mice. Values in square brackets indicate the 95% confidence interval for each correlation. All correlations are statistically significant ($p < .001$).

Table 6. Short form predictors of treatment-related change in TFI and THI

	TFI Δ			THI Δ		
Predictor	b	SE	*p*	b	SE	*p*
Tinnitus Severity Short Form	0.83	0.09	<0.001	0.251	0.09	0.003
THI short form Δ	-0.08	0.23	0.728	0.974	0.20	<0.001

Note. TFI = Tinnitus Functional Index; THI = Tinnitus Handicap Inventory

Appendix A. Tinnitus Severity Short Form

Instructions: The purpose of this questionnaire is to identify difficulties that you may be experiencing because of your tinnitus. Please rate the following based on your experiences during the past week:

All on the following rating scale:

0	1	2	3	4	5	6	7	8	9	10
Not										Very
At										Much
All										So

1 [TFI5]. How easy was it for you to cope with your tinnitus?

2 [TFI9]. How much did your tinnitus interfere with your ability to focus attention on other things besides your tinnitus?

3 [TFI17]. How much has your tinnitus interfered with your ability to relax?

4 [TFI20]. How much has your tinnitus interfered with your enjoyment of life?

5 [TFI24]. How bothered have you been because of your tinnitus?

6 [THI10]. Because of your tinnitus, did you feel frustrated?

7 [THI14]. Because of your tinnitus, did you find that you were often irritable?

8 [THI16]. Did your tinnitus make you upset?

9 [THI21]. Because of your tinnitus, did you feel depressed?

10 [THI23]. Did you feel that you could no longer cope with your tinnitus?

Note: Items are adapted from the Tinnitus Handicap Inventory (THI; Newman et al., 1996) and Tinnitus Functional Index (TFI; Meikle et al., 2012). Item source appears in brackets. Instructions are adapted from THI and implied instructions of TFI (e.g., past week instruction is present on each item of the TFI). Scale is adapted from TFI. The short questionnaire administered to a new sample in the current study included an instruction to select lower scores when an aspect has not been a problem and higher scores when an aspect has been a problem. We believe this caused confusion for participants, as the first item should be reverse scored. To generate a total score, subtract the score of the first item from 10 (e.g., reverse score this item) and sum the result with the rest of the item scores.