A Systematic Review of the Psychometric Properties of the Geriatric Anxiety Inventory*

Alexandra Champagne,[1] Philippe Landreville,[1,2] and Patrick Gosselin[3,4]

RÉSUMÉ

Le Geriatric Anxiety Inventory (GAI) et sa forme courte (GAI-SF) sont des échelles utilisées internationalement pour évaluer les symptômes anxieux chez les aînés. L'objectif de cette étude était de conduire la première revue critique des propriétés psychométriques de ces outils. Les études pertinentes (n = 31) des deux versions du GAI ont été extraites de bases de données électroniques ainsi que d'une recherche à la main. La qualité des études a été évaluée par la grille COSMIN. Le GAI et le GAI-SF présentaient une consistance interne ainsi qu'une fidélité test-retest adéquates. La validité convergente présentait des corrélations élevées avec des mesures d'anxiété généralisée alors que de faibles corrélations étaient retrouvées avec celles incluant des symptômes somatiques. Un chevauchement important a été trouvé avec des mesures des symptômes dépressifs. Alors qu'il n'y a pas de consensus quant à la structure factorielle du GAI, le GAI-SF est unidimensionnel. Malgré de bonnes sensibilité et spécificité pour détecter l'anxiété, les scores-frontières recommandés variaient considérablement. Le GAI et le GAI-SF sont des instruments présentant des propriétés psychométriques satisfaisantes. Afin d'élargir leur utilisation, certaines d'entre elles nécessitent toutefois un examen plus approfondi. Cette revue souligne l'importance de porter attention quant à certaines lacunes méthodologiques qui ont été retrouvées dans les études.

ABSTRACT

The Geriatric Anxiety Inventory (GAI) and its short form (GAI-SF) are self-reported scales used internationally to assess anxiety symptoms in older adults. In this study, we conducted the first critical comprehensive review of these scales' psychometric properties. We rated the quality of 31 relevant studies with the COSMIN checklist. Both the GAI and GAI-SF showed adequate internal consistency and test-retest reliability. Convergent validity indices were highest with generalized anxiety measures; lowest with instruments relating to somatic symptoms. We detected substantial overlap with depression measures. While there was no consensus on the GAI's factorial structure, we found the short version to be unidimensional. Although we found good sensitivity and specificity for detecting anxiety, cut-off scores varied. The GAI and GAI-SF are relevant instruments showing satisfactory psychometric properties; to broaden their use, however, some psychometric properties warrant closer examination. This review calls attention to weaknesses in the methodological quality of the studies.

[1] School of Psychology, Université Laval, Québec

[2] Centre d'excellence sur le vieillissement de Québec, Québec

[3] Department of Psychology, Université de Sherbrooke, Sherbrooke

[4] Institut universitaire de première ligne en santé et services sociaux – Centre intégré universitaire en santé et services sociaux de l'Estrie - CHUS (CIUSSS de l'Estrie- CHUS), Sherbrooke

* The authors wish to thank Alexandra Michel in the searching and assessment of methodological quality of studies.

Funding Sources None.

Conflict of Interest Authors have no conflicts of interest to declare.

Manuscript received: / manuscrit reçu : 01/05/2019

Manuscript accepted: / manuscrit accepté : 22/04/2020

Mots-clés: vieillissement, Geriatric Anxiety Inventory, propriétés psychométriques, validité, fidélité, COSMIN

Keywords: aging, Geriatric Anxiety Inventory, psychometric properties, validity, reliability, COSMIN

La correspondance et les demandes de tirés-à-part doivent être adressées à : / Correspondence and requests for offprints should be sent to:

Alexandra Champagne, Psy.D.
School of Psychology
Université Laval
2325, rue des Bibliothèques
Québec, QC G1V 0A6
(alexandra.champagne@outlook.com)

Introduction

The Geriatric Anxiety Inventory (GAI; Pachana et al., 2007) is one of the few appropriate screening measures for assessing anxiety in elderly individuals (Creighton, Davison, & Kissane, 2018; Therrien & Hunsley, 2012). It is a self-report scale composed of 20 items designed to assess anxiety symptoms over the past week using a dichotomous (yes/no) response format. Prior to the final 20-item version of the GAI, scale developers generated a pool of items, either formulated de novo or adapted from existing anxiety scales (Pachana, Byrne, et al., 2007). Items were chosen by the developers if they reflected general anxiety and covered primary domains of existing scales (e.g., fearfulness, worry, cognitions about anxiety). Developers were careful to limit the inclusion of somatic symptoms that may overlap with symptoms of general medical conditions. The developers reduced the pool of items by consulting with a reference group and psychometric testing with pilot samples. The GAI was normed with samples of community-dwelling seniors and older adults receiving psychiatric services. Total score on the GAI ranges between 0 and 20 and higher scores indicate greater anxiety symptoms. Cut-off scores of 11 and 9 and above were recommended to detect generalized anxiety disorder (GAD) and other anxiety disorders respectively. In the original study, the GAI presented sound psychometric properties with internal consistency coefficients of .91 and .93 for the non-clinical and psychogeriatric samples respectively (Pachana, Byrne, et al., 2007). Convergent validity varied between .44 to .70 for measures that assessed anxiety or related constructs. Retest reliability was .91 for the psychogeriatric sample.

A 5-item short form (GAI-SF; items 1, 6, 8, 10 and 11 of the GAI long form; Byrne & Pachana, 2011) was further developed to make the use of the instrument more practical in primary care and acute geriatric medical settings. The short-form items were chosen among those of the standard GAI based on parameters such as scale cohesion (item-total correlation), endorsement rate, and ability to distinguish participants with GAD from those without the diagnosis. A cut-off score of 2/3 out of a maximum total score of 5 on the GAI-SF proved optimal for the identification of GAD. The GAI-SF presented satisfactory results with respect to internal consistency (α = .81), retest reliability (r = .80), and convergent validity indices (r = .48–.88) (Byrne & Pachana, 2011). The popularity of the GAI is well reflected by the fact that it has been translated into more than 24 languages (Pachana & Byrne, 2012). This includes English and French for Canada, which makes it a convenient tool for use at a national level.

Since the publication of Pachana, Byrne, et al.'s (2007) first study on the GAI, much research has been conducted on the psychometric properties of the GAI and GAI-SF in various populations and in different languages. Reviews of this literature (Balsamo, Cataldi, Carlucci, & Fairfield, 2018; Creighton et al., 2018; Dissanayaka, Torbey, & Pachana, 2015; Edelstein et al., 2008; Lin et al., 2016; Pachana & Byrne, 2012; Therrien & Hunsley, 2012) generally conclude that these measures present sound psychometric properties despite some disparities in the results (e.g., factorial structure, convergent validity) and possible problematic issues (e.g., problematic items, divergent validity, cross-cultural issues). However, none of the reviews evaluated the methodological quality of the studies they examined, which makes it difficult to determine the appropriateness of the findings and to offer guidelines to improve research in this area. Moreover, each review examined only a small portion of studies in light of the available empirical evidence on the GAI and GAI-SF. This is not surprising because most reviews are not dedicated exclusively to the GAI; some are specific to certain populations or settings (e.g., Parkinson's disease; residential aged care facilities), and new studies have only been published recently.

The goals of this review were to summarize existing evidence on the psychometric properties of the GAI and GAI-SF, to assess the methodological quality of the studies and to provide guidance for future psychometric validation studies. To our knowledge, this was the first systematic review that examined the methodological quality of studies that were conducted on the psychometric properties of these instruments.

Methods

This review targeted published studies that reported data on psychometric properties of the GAI and GAI-SF in older adults. The methodology was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines (Moher et al., 2015; Moher, Liberati, Tetzlaff, & Altman, 2009).

Hunsley and Mash (2008) have provided criteria for rating some psychometric results. Internal consistency is considered "adequate" with Cronbach α values of .70–.79, "good" with α values of .80–-.89, and "excellent" with α values equal or greater than .90. Test-retest reliability is considered "adequate" when correlations are of at least .70 over a period of several days to several weeks, "good" when they are of at least .70 over a period of several months, and "excellent" when at least they are of .70 over a period of a year or longer.

Search Strategy

We conducted a literature search using Pubmed, PsycINFO, CINAHL, EMBASE, and Google Scholar as

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

these databases are representative of the literature published on this topic. We made additional efforts to locate relevant studies through a handsearching process. The keywords "Geriatric Anxiety Inventory" in the title and abstract section of the databases was what we used to filter relevant studies. We decided on this approach after conducting different tests (e.g., with broader keywords like "anxiety" or "assessment" or by including them in the "any field" section), which considerably broadened the number of non-relevant articles retrieved. The search was restricted to articles published between January 1, 2007 (the GAI was developed in 2007 [Pachana et al., 2007]) and December 31, 2018. We retained articles according to the following criteria: (a) written in English or French, (b) presented original empirical research, and (c) expressed the primary objective of exploring the psychometric properties of the GAI and/or the GAI-SF. We excluded the following types of articles because either the information provided was limited or the articles were frequently non-peer reviewed: unpublished manuscripts, editorials, dissertations, theses, randomized controlled trials, case reports, and published abstracts. The first author and a research assistant independently screened the titles and abstracts of the retrieved studies to determine their eligibility. When a disagreement emerged between the two reviewers, a discussion ensued in order to reach a consensus. When necessary, a third reviewer made the decision.

Quality Assessment

We assessed the methodological quality of the included studies with the "COnsensus-based Standards for the selection of health status Measurement Instruments" (COSMIN) checklist (Mokkink, Terwee, Knol, et al., 2010; Mokkink, Terwee, Patrick, et al., 2010). The COSMIN checklist consists of eight boxes that each refer to a specific measurement property (i.e., internal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, cross-cultural validity, and responsiveness). Each box contains 5 to 18 items that assess methodological standards, and items are scored on a 4-point rating scale (i.e., poor, fair, good, or excellent) using specific criteria. For example, the fifth item on the internal consistency box assesses whether the unidimensionality of the scale was verified. Criteria proposed by the COSMIN checklist for rating this item follow: a factor analysis was performed in the study population (excellent); the authors refer to another study in which factor analysis was performed in a similar study population (good); authors refer to another study in which factor analysis was performed but not in a similar study (fair); factor analysis was not performed and contains no reference to another study (poor) (for more information on rating, see https://www.cosmin.nl/). For each measurement property, an overall score is determined by taking the lowest rating of any of the box items (worst score counts method; Terwee et al., 2012). Quality assessment of studies was independently performed by the first author and a research assistant. Discrepancies were resolved through a discussion. When necessary, a third reviewer made the decision.

Data Collection Process

Data extraction was conducted by the first author. Extracted data included basic information about study demographics (e.g., publication year, country in which the study was conducted, language in which the instrument was administered) as well as sample characteristics (e.g., type of sample, sample size, mean age). When available, we collected data on the different measurement properties (e.g., results, statistical methods used, time interval, comparator instruments) defined in the COSMIN checklist. More specifically, these properties were: internal consistency, test-retest reliability, measurement errors, content validity, structural validity (factor analysis), hypothesis testing, cross-cultural validity, criterion validity, and responsiveness.

Results

Search Results

As shown in Figure 1, the database search retrieved a total of 485 articles. Duplicates (n = 232) were removed and of the 253 remaining records, we excluded 222. The main reasons for exclusion were that the GAI or GAI-SF was not the topic of interest (n = 179) or that the article was not presented as an original published manuscript (i.e., conference proceeding; n = 21). Four articles were excluded based on language (i.e., were not written in English or in French). Thus, we retained a final list of 31 articles for the purpose of the current review.

Methodological Quality of the Included Studies

The results of COSMIN ratings for the 31 studies retained are displayed in Table 1. The studies assessed an average of 2.7 psychometric properties out of the nine COSMIN criteria. Most of the COSMIN boxes were rated as having "poor" (43.5%) or "fair" (40%) quality. The most frequent reasons for these ratings were low sample size or a lack of information concerning the number of missing items and how they were handled. This information corresponds to key criteria because it is assessed in almost all COSMIN boxes. Only 11.8 per cent of the rated boxes were rated as having "good" quality, and 4.7 per cent as having "excellent" quality.

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

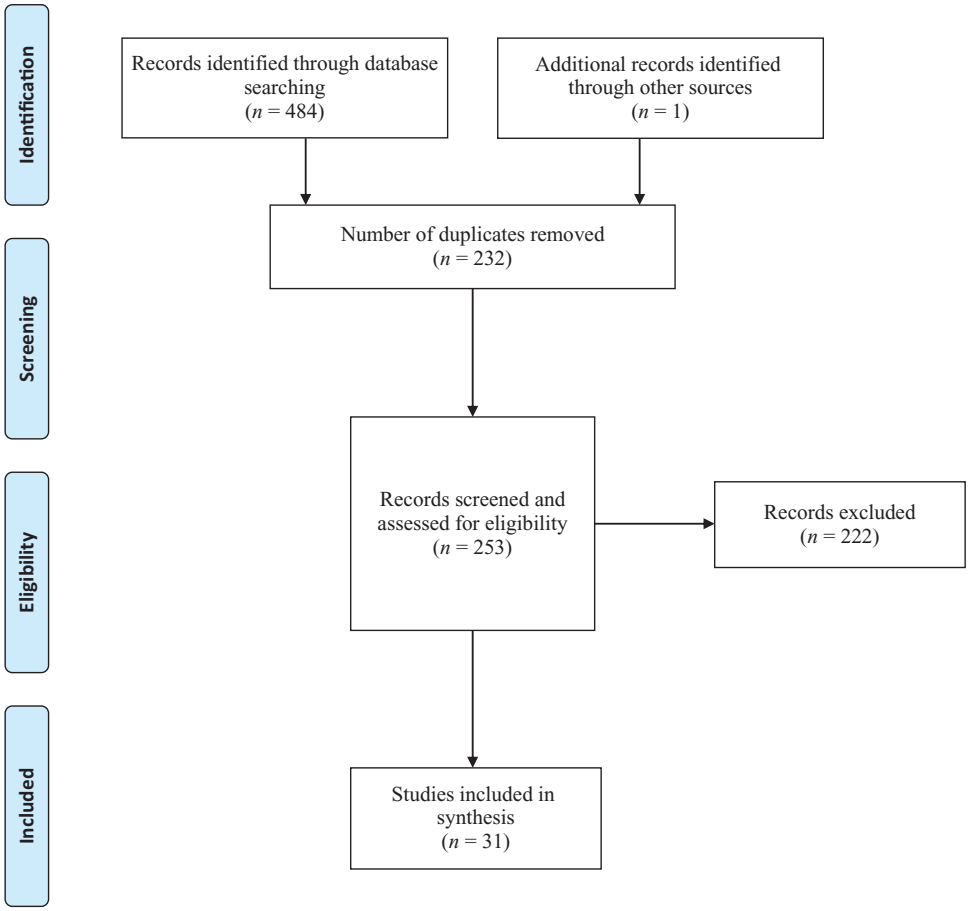

Figure 1: Flow diagram of study selection

Study and Participant Characteristics

Basic characteristics of the studies retained for the current review and their samples are presented in Table 2. Psychometric properties of the GAI were examined by 22 studies, while only one study investigated the properties of the GAI's short form, and eight studies examined both forms. The latter studies generally extracted GAI-SF scores from the GAI. We examined psychometric properties of 15 versions of the GAI: Brazilian Portuguese, Chinese, Czech (long and short forms), English (long and short forms), French Canadian (long and short forms), Italian (long and short forms), Norwegian (long and short forms), Portuguese (long and short forms) and Spanish.

The 31 retained studies provided data for 8,174 patients who completed the GAI and/or the GAI-SF. Sample sizes ranged from 32 to 1,318 patients. Most studies had samples composed mainly of women (on average, 64.9% of the samples were composed of women). Participants were aged between 52 and 94 years old, excluding participants in the study by Matheson et al. (2012) that included young adults aged 37 years old and older. Mean age of the participants was 72.5 years. Sample recruitment source was categorized as either non-clinical (e.g., community-dwelling seniors), psychiatric (e.g., in-patient, outpatient, or institutionalized patients, or individuals with a psychiatric diagnosis), medical (i.e., having a medical diagnosis or receiving medical care), or mixed (i.e., different sources of recruitment in the sample). Of the 31 selected studies, nine used mixed samples (non-clinical and/or psychiatric and/or medical). Other studies' recruitment sources were for the most part exclusively non-clinical (n = 9), medical (n = 9), or, in a smaller proportion, psychiatric (n = 4 studies). Mean scores on the GAI varied between .58 to 16.3, with a mean of 5.5. Those for the GAI-SF ranged between .17 to 3.64, with a mean of 1.8.

Reliability

Internal Consistency. The alpha coefficient of the GAI ranged between .71 and .97 with a mean of .91, and between .61 to .84 with a mean .80 for the GAI short form (see Table 2). According to the COSMIN checklist results, internal consistency was mostly (63%) rated as poorly assessed. The items rated as "poor" referred

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

Table 1: Methodological quality of each study per measurement property

Study	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypothesis Testing	Cross–Cultural Validity	Criterion Validity	Responsiveness
Brazilian Portuguese Version (GAI-BR)									
Massena et al. (2015)	Poor	Poor	–	–	–	Fair	–	Fair	–
Chinese version (GAI-CV)									
Yan et al. (2014)	Good	–	–	–	Good	–	Poor	–	–
Guan (2016)	Fair	–	–	–	Fair	–	Poor	–	–
Dow et al. (2018)	Poor	–	–	–	–	–	Poor	–	–
Czech version									
Heissler et al. (2018)	Fair	–	–	–	–	–	Poor	–	–
English version									
Cheung (2007)	Poor	–	–	–	–	–	–	–	–
Pachana et al. (2007)	Poor	Fair	–	Excellent	–	–	–	Fair	–
Boddice and Byrne (2008)	–	–	–	–	–	–	–	–	–
Diefenbach, Tolin, Meunier, and Gilliam (2009)	Poor	Poor	–	–	–	–	–	Fair	–
Byrne et al. (2010)	Poor	–	–	–	–	–	–	Fair	–
Cheung et al. (2012)	Poor	–	–	–	–	–	–	Fair	–
Matheson et al. (2012)	Poor	Poor	–	–	–	–	–	Fair	–
Bradford et al. (2013)	Poor	–	–	–	–	Poor	–	Fair	–
Gerolimatos, Gregg, and Edelstein (2013)	Poor	–	–	–	–	Good	–	Good	–
Diefenbach et al. (2014)	Fair	–	–	–	Fair	Fair	–	–	
Gould et al. (2014)	Poor	–	–	–	–	–	–	–	–
Ball et al. (2015)	Poor	–	–	–	–	–	–	–	Poor
Johnco, Knight, Tadic, and Wuthrich (2014)	Good	–	–	–	Good	–	–	Good	–
Gould et al. (2016)	Poor	–	–	–	–	–	–	–	–
Kneebone et al. (2016)	Poor	Poor	Fair	–	–	Fair	–	Fair	–
Creighton et al. (2018)	Poor	–	–	–	–	–	–	Fair	–
English version – short form (GAI-SF)									
Byrne and Pachana (2011)	Poor	Poor	–	–	–	–	–	Excellent	–
French Canadian version (GAI-FC)									
Champagne et al. (2016)	Fair	Fair	–	–	Fair	Fair	–	–	–
Italian version (GAI-It)									
Rozzini et al. (2009)	Poor	Fair	–	–	–	–	Poor	–	–
Ferrari et al. (2017)	Poor	–	–	–	–	Fair	–	Fair	–
Norwegian version									
Bakkane Bendixen et al. (2016)	Excellent	Poor	–	–	Excellent	–	–	–	–
Molde et al. (2017)	Fair	–	–	Poor	Fair	–	Fair	Fair	–
Portuguese version (GAI-PT)									
Ribeiro et al. (2011)	Good	Poor	–	–	Good	–	Poor	Good	–
Silva et al. (2016)	Poor	Poor	–	–	–	–	–	–	–
Spanish version									
Marquez-Gonzalez et al. (2012)	Fair	–	–	–	Fair	–	Poor	–	–
Mababu and RuizSánchez (2016)	Fair	–	–	–	Fair	–	–	–	–

Table 2: Characteristics of the retained studies on the GAI and GAI-SF and their reliability coefficients

Version	Study	Country	Type of Sample	Subjects (*n*)	Gender Female (%)	Mean Age & Range (years)	Mean Score (*SD*)	Internal Consistency (α)	Test-retest Reliability
Brazilian Portuguese version (GAI-BR) Long form	Massena et al. (2015)	Brazil	Mixed	Mixed sample from community and outpatient psychogeriatric clinic ($n = 72$)	82	72.2	8.77	.91	.85 1 week
Chinese version (GAI-CV) Long form	Yan et al. (2014)	Beijing	Nonclinical	Community-dwelling seniors ($n = 1,047$)	59.4	70.8	2.17 (4.19)	.94	—
	Guan (2016)	Beijing	Nonclinical	Community-dwelling seniors ($n = 1,318$)	59.4	71.4	—	.94	—
	Dow et al. (2018)	Australia	Nonclinical	Community-dwelling Chinese immigrants ($n = 87$)	66	76.9 60–92	—	.95	—
Czech version Long form	Heissler et al. (2018)	Czech Republic	Nonclinical	Community-dwelling seniors ($n = 485$)	52	75.5	Men: 2.27 (2.85) Women: 3.44 (3.74)	.85	—
Czech version Short form	Heissler et al. (2018)	Czech Republic	Nonclinical	Community-dwelling seniors ($n = 485$)	52	75.5	Men: .64 (1.11) Women: 1.08 (1.37)	.75	—
English version (GAI) Long form	Cheung (2007)	New Zealand	Psychiatric	Geriatric psychiatry patients ($n = 32$)	63	75.5 66–85	7.59 (6.5)	—	—
	Pachana et al. (2007)	Australia	Mixed – Nonclinical	Community-dwelling ($n = 452$)	64.4	71.7 60–90	2.3 (3.8)	.91	—
			– Psychiatric	Patients attending a psychogeriatric service ($n = 46$)	74	78.8 66–94	5.22 (5.83)	.83	.91 1 week
	Boddice and Byrne (2008)	Australia	Mixed – Nonclinical	Community-dwelling seniors ($n = 31$)	52	75.8	—	—	—
			– Medical	Older adults living in nursing homes ($n = 27$)	62.9	82.8	2.3 (4.2)	—	—
	Diefenbach et al. (2009)	United States	Medical	Older home care recipients ($n = 66$); data on the GAI available only for a subset of the sample ($n = 35$)	83.3	76.6 6592	4.63 (5.57)	.93	.95 1 to 2 weeks
	Byrne et al. (2010)	Australia	Nonclinical	Community-residing older women ($n = 286$)	100	71.7 60–86	2.33 (4.05)	.92	—
	Cheung et al. (2012)	New Zealand	Medical	Older adults with chronic obstructive pulmonary disease ($n = 55$)	44	72.7	3.3 (4.6)	.92	—

Continued

Table 2: Continued

Version	Study	Country	Type of Sample	Subjects (*n*)	Gender Female (%)	Mean Age & Range (years)	Mean Score (*SD*)	Internal Consistency (α)	Test-retest Reliability
	Matheson et al. (2012)	Australia	Medical	Parkinson's disease patient (*n* = 58)	41	66.2 37–88	5.03 (6.06)	.95	.99 2 weeks
	Bradford et al. (2013)	United States	Medical	Patients with mild to moderate dementia (*n* = 41)	23	79	5.8 (5.7)	.92	—
	Gerolimatos et al. (2013)	United States	Medical	Nursing home residents (*n* = 75)	52	69.6 52–94	7.92 (6.17)	.92	—
	Diefenbach et al. (2014)	United States	Mixed	Older adults with mild dementia (*n* = 45)	55.6	76.7	5.71 (5.37)	.91	—
			– Medical	Elderly individuals with cognitive impairment but no dementia (*n* = 55)	70.9	70.1	5.73 (5.64)	.92	—
			– Nonclinical	Nonclinical group (*n* = 50)	56	69.5	1.42 (2.44)	.83	—
	Gould et al. (2014)	United States	Mixed	Total sample (*n* = 110); data on the GAI available only for a subset of the sample (*n* = 74)	57.3	75.2	1.51 (3.07)	.89	—
			– Nonclinical	Nonclinical anxiety (*n* = 67)		—	1.19 (2.42)		
			– Psychiatric	Current anxiety disorder (*n* = 7)		—	4.57 (6.16)		
	Ball et al. (2015)	United States	Psychiatric	Elderly patients with generalized anxiety disorder (*n* = 291)	78	71.6	14.3 (4.1)	.81	—
	Johnco et al. (2014)	Australia	Mixed	Total sample (*n* = 256)	62.8	67.5 60–88	—	.93	—
			– Psychiatric	Clinical geriatric participants with co-morbid anxiety and unipolar mood disorder (*n* = 197)	60.4	67.5 61–88	11.08 (4.86)	.85	—
			– Nonclinical	Nonclinical control group (*n* = 59)	71.2	67.6 60–86	0.58 (1.32)	.71	—
	Gould et al. (2016)	United States	Medical	Older veterans who attended a geriatric primary care outpatient clinic (*n* = 50)	0	78.5	1.94 (2.77)	.82	—
	Kneebone et al. (2016)	England	Medical	In-patients with stroke (*n* = 81)	48.1	Mdn: 79 (IQR = 14.5)	4.41 (5.80)	.95	.53
	Creighton et al. (2018)	Australia	Medical	Nursing home residents (*n* = 180)	66.7	85.4	4.46 (5.79)	.95	—

Continued

Table 2: Continued

Version	Study	Country	Type of Sample	Subjects (*n*)	Gender Female (%)	Mean Age & Range (years)	Mean Score (*SD*)	Internal Consistency (α)	Test-retest Reliability
English version (GAI-SF) Short form	Byrne and Pachana (2011)	Australia	Nonclinical	Community-residing older women (*n* = 284)	100	72.2 60–87	2.5 (4.2)	.81	.80 1 week
	Gerolimatos et al. (2013)	United States	Medical	Nursing home residents (*n* = 75)	52	69.6 52–94	2.23 (1.72)	.73	—
	Diefenbach et al. (2014)	United States	Mixed	Older adults with mild dementia (*n* = 45)	55.6	76.7	2.09 (1.81)	.77	—
			– Medical	Elderly individuals with cognitive impairment but no dementia (*n* = 55)	70.9	70.1	2.05 (1.85)	.80	—
			– Nonclinical	Nonclinical group (*n* = 50)	56	69.5	0.56 (0.99)	.61	—
	Johnco et al. (2015)	Australia	Mixed	Total sample (*n* = 256)	62.8	67.5 60–88	—	.84	—
			– Psychiatric	Clinical geriatric participants with co-morbid anxiety and unipolar mood disorder (*n* = 197)	60.4	67.5 61–88	3.46 (1.48)	.67	—
			– Nonclinical	Nonclinical control group (*n* = 59)	71.2	67.6 60–86	0.17 (0.62)	.72	—
French Canadian version (GAI-FC) Long form	Champagne et al. (2016)	Canada	Nonclinical	Community-dwelling seniors (*n* = 331)	73.5	74.6	4.10 (5.51)	.94	.89 1 week
French Canadian version (GAI-FC-SF) Short form	Champagne et al. (2016)	Canada	Nonclinical	Community-dwelling seniors (*n* = 331)	73.5	74.6	1.31 (1.66)	.83	.85 1 week
Italian version (GAI-It) Long form	Rozzini et al. (2009)	Italy	Mixed – Medical	Total sample: outpatients with mild cognitive impairment (*n* = 57)	56	71.2	3.2 (3.8)	.76	.86 1 week
			– Psychiatric	Patients with anxiety (*n* = 44)	62	69	11.5 (1.2)	—	—
			– Nonclinical	Patients without anxiety (*n* = 13)	55	71.8	2.4 (2.5)	—	—
	Ferrari et al. (2017)	Italy	Mixed – Psychiatric – Medical	Mixed sample of outpatients from psychiatric services and from a clinic for cognitive disorders (*n* = 76)	60.5	72.7 65–91	11.3 (6.5)	.93	—
Italian version (GAI-It SF) Short form	Ferrari et al. (2017)	Italy	Mixed – Psychiatric – Medical	Mixed sample of outpatients from psychiatric services and from a clinic for cognitive disorders (*n* = 76)	60.5	72.7 65–91	3.1 (2.1)	.77	—

Continued

Table 2: Continued

Version	Study	Country	Type of Sample	Subjects (*n*)	Gender Female (%)	Mean Age & Range (years)	Mean Score (*SD*)	Internal Consistency (α)	Test-retest Reliability
Norwegian version (GAI) Long form	Bakkane Bendixen et al. (2016)	Norway	Mixed	Total sample: patients who were admitted to a department of geriatric psychiatry (*n* = 428)	67	75.7	8.5 (6.6)	.92 for the 1st factor .85 for the 2nd factor	—
			– Psychiatric	Patients with a diagnosis of depression (*n* = 220)	67.6	75.6	11.1 (6.0)	—	—
				Patients with a diagnosis of nonorganic psychosis (*n* = 68)	64.3	76.9	5.9 (6.1)	—	—
			– Medical	Patients with a diagnosis of dementia (*n* = 140)	70.6	73.3	5.5 (5.9)	—	—
	Molde et al. (2017)	Norway	Psychiatric	Psychogeriatric mixed in-and-out patient sample (*n* = 543)	67.9	75.7 62–78	8.2 (6.5)	.94	—
Norwegian version Short form	Molde et al. (2017)	Norway	Psychiatric	Psychogeriatric mixed in-and-out patient sample (*n* = 543)	67.9	75.7 62–78	—	.84	—
Portuguese version (GAI-PT) Long form	Ribeiro et al. (2011)	Portugal	Mixed	Total sample (*n* = 217)	—	—	—	.96	—
			– Nonclinical	Community-dwelling seniors (*n* = 152)	56.6	73.9 59–92	With PD: 16.3 (4.9) Without PD: 4.1 (5.4)	.97	.99 (ICC) 2 weeks
			– Psychiatric	Patients with depression (*n* = 32)	71.9	70.5 55–85	15.2 (5.58)	—	
				Patients with anxiety disorders (*n* = 23)	47.8	72.3 56–89	With AD: 14.8 (4) With GAD: 16.1 (4.7)	—	—
			– Medical	Patients with an early Alzheimer's disease (*n* = 10)	80	74.6 63–88	11.9 (5.7)	—	—
	Silva et al. (2016)	Brazil	Medical	Patients registered in a primary care setting (*n* = 55)	78.2	72.8 60–91	9.2 (4.89)	.89	.58 30 weeks
Portuguese version (GAI-SF) Short form	Silva et al. (2016)	Brazil	Medical	Patients registered in a primary care setting (*n* = 55)	78.2	72.8 60–91	3.04 (1.44)	.62	.97 30 weeks
Spanish version Long form	Marquez-Gonzalez et al. (2012)	Spain	Nonclinical	Community-dwelling seniors (*n* = 302)	75.5	71.7	Female: 8.30 (5.62) Male: 6.43 (6.12)	.91	—
	Mababu and RuizSánchez (2016)	Spain	Nonclinical	Community-dwelling seniors (*n* = 652)	61	67.6 60–89	—	.83	—

Note. **AD = anxiety disorder; GAD = generalized anxiety disorder; IQR = interquartile range; MDN = median; PD = psychological distress.**

mostly to the absence of information on missing items and unidimensionality of the scale.

Test-Retest Reliability. Test-retest coefficients (mostly Pearson's *r* and the intraclass correlation [ICC]) ranged between .53 to .99 with a mean of .79 for the GAI, and between .80 to .97 with a mean of .90 for the short form (see Table 2). Aside from the two lowest coefficients of the long form (*r* = .53 and .58; Kneebone, Fife-Schaw, Lincoln, & Harder, 2016; Silva et al., 2016), the lowest coefficient was .85. These large differences in the coefficients obtained are difficult to explain, and authors did not comment on their results. In general, the interval of time between the two administrations of the scale was one to two weeks, except in the study of Silva et al. (2016) in which the interval was 30 weeks. Surprisingly, this longer interval generated a coefficient of *r* = .58 for the GAI and the highest coefficient for the short form (*r* = .97).

Test-retest reliability was chiefly rated as poorly assessed (72.7%) according to the COSMIN checklist because of a lack of information concerning missing items, stability of participants, and similarity of test conditions between the two administrations. The COSMIN checklist asks whether there were any important flaws in the study design or method, and in light of certain retest research recommendations, there are several other weaknesses present. In the majority of the studies, little information was provided on sampling and rationale for major decisions that were made (e.g., length of the retest interval). Polit (2014) has suggested that seeking input from patients or experts regarding the stability of the construct being assessed can help support decisions regarding retest interval. Park, Kang, Jang, Lee, and Chang (2018) have recommended that the sample size be about five times the number of items, which was not the case for any of the studies since they generally assessed the retest reliability on a subgroup of the sample. Moreover, the attrition rate for the retest assessment was rarely reported although there is evidence that high rates of attrition can depress reliability estimates (Polit, 2014). Although the COSMIN checklist prioritizes the use of the ICC to analyze retest reliability, Vaz, Falkmer, Passmore, Parsons, and Andreou (2013) made a case to consider measurement error indices such as the coefficient of repeatability (CoR) or the smallest real difference (SRD) over coefficients like the Pearson's *r* and the ICC.

Validity

Content Validity. Only two studies addressed content validity. This very low number could be explained by the fact that validation studies may have assumed that items of the GAI and GAI-SF are relevant and comprehensive. As the developers of the GAI, Pachana, Byrne, et al. (2007) thoroughly evaluated the content of this scale. Molde et al. (2017) performed a content analysis on data retrieved from a panel of older adults and a group of clinical psychologists and psychiatrists who were invited to comment on the items. Content validity was rated as "excellent" for the study by Pachana, Byrne, et al. (2007) and "poor" for the one by Molde et al. (2017), according to COSMIN criteria. The latter study obtained such a rating because it is not clear whether all items were assessed to determine whether they comprehensively covered the construct of interest in regard to its theoretical foundation, and whether items were relevant to the purpose of the instrument.

Convergent Validity. Convergent validity has been established between the GAI and GAI-SF and a variety of other instruments that also assess anxiety and related constructs (e.g., symptoms of GAD, worry, general anxiety, or both anxiety and symptoms of depression at the same time). As shown in Table 3, correlations vary between .25 to .86 for the GAI and between .55 to .79 for the short form. Convergent validity with GAD scales appear to be the highest (*r* = .65 to .86). Data are scarce on the association between the GAI and measures that assess other anxiety disorders. Available evidence reveals only a moderate relationship (*r* = .56) with a measure of post-traumatic symptoms (Gould et al., 2014). The weakest associations were found for scales that contain somatic items such as the Hamilton Anxiety Scale (HAMA) (*r* = .25; Ball, Lipsius, & Escobar, 2015) and the Beck Anxiety Inventory (BAI) (*r* = .28; Gould et al., 2014). In contrast, the GAI focuses predominantly on psychological symptoms. Another low correlation was found with the State-Trait Anxiety Inventory [STAI]-subscale state) (*r* = .28; Massena, de Araújo, Pachana, Laks, & de Pádua, 2015). The authors explained this result as due to a possible bias in the formulation of the questions of the STAI-state, where symptoms were assessed according to participants' feelings at the time of the interview rather than those experienced over the past week.

Convergent validity was not evaluated in depth with the COSMIN checklist because only two items referred to it in the hypothesis test box. For the purpose of this review, we rarely used these items to assess convergent validity since most of the retained studies did not provide hypotheses to test. Despite this, the general trend was that studies provided a poor description of the constructs measured by the comparator instrument. In addition, it was not always clear whether the comparator instrument was an established and validated instrument for use with elderly individuals.

Divergent Validity. We assessed divergent validity in some studies by examining the association with a measure of depression symptoms. Correlations ranged between .28 to .86 for the GAI and between .37 to .63

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

for the short form (see Table 3). The lowest correlations (r = .28) between the GAI and the Hospital Anxiety and Depression Scale – Depression Scale (HADS-D) are explained by the fact that patients with major depressive disorder were excluded (Ball et al., 2015) and by the low prevalence of depression symptoms (Kneebone et al., 2016). These results suggest that there may exist different patterns of divergent validity where highly uniform samples with low rates of depression symptoms could facilitate distinction from anxiety symptoms assessed with the GAI and GAI-SF.

Diefenbach, Bragdon, and Blank (2014) and Bakkane Bendixen, Hartberg, Selbæk, and Engedal (2016) shed new light on the association between the GAI and GAI-SF and measures of depression. Diefenbach et al. (2014) found that depressive symptoms were more strongly correlated with the "central nervous system hyperarousal" factor and to a lesser extent with "gastrointestinal symptoms". These results suggest that there may be a certain response pattern in patients with greater co-morbid depressive symptoms. Bakkane Bendixen et al. (2016) found that in comparison to those with dementia or psychosis, a group of patients with depression present a different pattern of results on the GAI; that is, with a higher total score and a higher endorsement of 18 of the 20 items (except items 3 and 18).

Factorial Validity. The GAI was first described as being unidimensional although no factor analysis was presented to support this assumption (Byrne & Pachana, 2011; Pachana, Byrne, et al., 2007). Ten studies investigated the factorial validity of the GAI and half of them confirmed the one-factor structure (Champagne, Landreville, Gosselin, & Carmichael, 2016; Johnco, Knight, Tadic, & Wuthrich, 2014; Molde et al., 2017; Ribeiro, Paul, Simoes, & Firmino, 2011; Yan, Xin, Wang, & Tang, 2014).

The other five studies that investigated the factorial validity of the GAI found a two-factor structure (Bakkane Bendixen et al., 2016), a three-factor structure (Guan, 2016; Mababu & RuizSánchez, 2016; Marquez-Gonzalez, Losada, Fernandez-Fernandez, & Pachana, 2012), and a four-factor structure (Diefenbach et al., 2014) (see Table 4). The identified factors can be grouped into three categories: (a) cognitive symptoms (includes the following factors: worries, excessive worry symptoms, decision-making symptoms, and mental anxiety), (b) physical symptoms of anxiety (includes the following factors: central nervous system hyperarousal, arousal and somatic symptoms), and (c) negative anxiety. Cognitive and physical symptoms of anxiety were found across all five studies. In contrast, negative anxiety, which refers to the motives and behaviours related to anxiety disorders, was found only by Guan (2016). Most of the GAI items were not consistently associated with the same symptom category. Only items 1 – "I worry a lot of the time" – and 2 – "I find it difficult to make a decision" – were always related to cognitive symptoms and items 12 – "I get an upset stomach due to my worrying" – and 18 – "I sometimes feel a great knot in my stomach" – were always associated with physical symptoms. This variability may be due to the type of sample (i.e., three studies used non-clinical samples; one, a mixed sample of psychiatric and medical patients; and one, composed of elderly people with cognitive impairment) and cultural differences because four versions were used (Norwegian, English, Spanish, and Chinese).

Four studies investigated the factor structure of the GAI-SF and all confirmed its unidimensionality (Champagne et al., 2016; Diefenbach et al., 2014; Johnco et al., 2014; Molde et al., 2017). Most items with high factor loadings referred to cognitive symptoms of anxiety.

Criterion Validity. At first, Pachana, Byrne, et al. (2007) recommended a GAI cut-off score of 9 for the identification of any anxiety disorder and of 11 for the detection of GAD. Further studies suggested cut-off scores that varied between 3 and 13 out of 20 for the identification of an anxiety disorder (see Table 5). Multiple factors can explain this variability such as the type of sample (non-clinical vs. clinical), the proportion of patients who actually met the criteria for an anxiety disorder, cultural differences in the expression of anxiety, and the external criterion used for the diagnosis. The much lower cut-off score of 3 found by Cheung, Patrick, Sullivan, Cooray, and Chang (2012) may be attributable to differences in the nature of the sample as their participants had chronic obstructive pulmonary disease; the mean score on the GAI was low (M = 3.3; SD = 4.6) as was the proportion of participants with an anxiety disorder (25.5%). Test sensitivity values for the GAI ranged between 30 and 100 per cent; while specificity values ranged between 43 and 100 per cent. The area under the ROC curve (AUC) ranged between 79 and 98.1.

For the GAI short form, a score of 3 or more was originally found to be optimal for the detection of GAD in a non-clinical sample (Byrne & Pachana, 2011). Results of subsequent studies were similar with optimal thresholds at 2 to 3 out of 5 for the identification of an anxiety disorder. Sensitivity varied between 72 and 100 per cent and specificity ranged between 35 and 98.3 per cent. The AUC ranged between 78 and 95.4.

With regard to the different diagnostic parameters, the performance of the standard and short forms of the GAI seemed quite comparable. According to the COSMIN checklist, we largely rated criterion validity as "fair" for different reasons (e.g., no information on how missing items were handled; unclear if the criterion was a "gold standard").

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

Table 3: Convergent and divergent validity of the GAI and GAI-SF

Convergent Validity	*r* GAI	*r* GAI-SF	Study
Anxiety Measure			
ASI = Anxiety Inventory Status	.85		Rozzini et al. (2009)
BAI = Beck Anxiety Inventory	.28–.75	.58	Diefenbach et al. (2009); Gould et al. (2014); Massena et al. (2015); Pachana et al. (2007); Silva et al. (2016); Yan et al. (2014)
GADS = Goldberg Anxiety and Depression Scale – Anxiety scale	.57		Pachana et al. (2007)
GAI = Geriatric Anxiety Inventory		.77–.94	Byrne and Pachana (2011); Champagne et al. (2016); Gerolimatos et al. (2013); Heissler et al. (2018); Johnco et al. (2015); Silva et al. (2016)
GAI-SF = Geriatric Anxiety Inventory – Short Form	.77–.94		Champagne et al. (2016); Gerolimatos et al. (2013); Heissler et al. (2018); Johnco et al. (2015); Silva et al. (2016)
GAS = Geriatric Anxiety Scale	.60–.82		Cheung (2007); Pachana et al. (2007) Gould et al. (2014)
HADS-A = Hospital Anxiety and Depression Scale – Anxiety Scale	.51–.71	.61	Ball et al. (2015); Creighton et al. (2018); Dow et al. (2018); Ferrari et al. (2017); Kneebone et al. (2016)
HAMA = Hamilton Anxiety Scale	.25–.47		Ball et al. (2015); Gould et al. (2014)
RAID = Rating Anxiety in Dementia Scale	.61		Creighton et al. (2018)
SAS = Self-Rating Anxiety Scale	.52		Yan et al. (2014)
SRQ-20 = Self-Reporting Questionnaire	.74	.55	Silva et al. (2016)
STAI = State-Trait Anxiety Inventory	.61–.69		Cheung (2007); Massena et al. (2015); Matheson et al. (2012); Ribeiro et al. (2011)
STAI-S = State-Trait Anxiety Inventory – subscale state	.28–.80	.48–.50	Byrne and Pachana (2011); Byrne et al. (2010); Ferrari et al. (2017); Massena et al. (2015); Pachana et al. (2007)
STAI-T = State-Trait Anxiety Inventory – subscale trait	55	.53	Ferrari et al. (2017); Massena et al. (2015)
Anxiety/Depression Measure			
GHQ = General Health Questionnaire	.76		Ribeiro et al. (2011)
Generalized Anxiety Disorder Measure			
GAD-7 = Generalized Anxiety Inventory-7	.86	.79	Champagne et al. (2016)
GADQ-IV: Generalized Anxiety Disorder Questionnaire for DSM-IV	.65		Diefenbach et al. (2009)
GADSS = Generalized Anxiety Disorder Severity Scale	.84		Diefenbach et al. (2009)
Intolerance to Uncertainty			
IUI = Intolerance of Uncertainty Inventory	.62	.58	Champagne et al. (2016)
Neuroticism Measure			
NEO-N = NEO Five-Factor Inventory- neuroticism	.63		Byrne et al. (2010)
Posttraumatic Stress Disorder Measure			
PCL-C = Posttraumatic stress disorder checklist-civilian version	.56		Gould et al. (2014)
Worry Measure			
BMWS = Brief Measure of Worry Severity	.78		Diefenbach et al. (2009)
GWS = Geriatric Worry Scale	.86		Diefenbach et al. (2009)
PSWQ = Penn State Worry Questionnaire	.70–.79		Pachana et al. (2007) Diefenbach et al. (2009); Gould et al. (2014)
PSWQ-A = Penn State Worry Questionnaire – Abbreviated	.60–.79	.56–.79	Champagne et al. (2016); Diefenbach et al. (2009); Johnco et al. (2015)
WSOA-R = Worry Scale for Older Adults Revised	.53	.53	Champagne et al. (2016)

Continued

Table 3: Continued

Divergent Validity	r		Study
	GAI	GAI-SF	
Depression Measure			
BDI-II = Beck Depression Inventory-II	.49		Gould et al. (2014)
CES-D = Center for Epidemiologic Studies – Depression scale	.56		Marquez-Gonzalez et al. (2012)
GDS = Geriatric Depression Scale	.31–.86	.37-.47	Diefenbach et al. (2014); Johnco et al. (2015); Ribeiro et al. (2011)
GDS-A = Geriatric Depression Scale – Abbreviated	.38–.79	.37–.63	Byrne and Pachana (2011); Champagne et al. (2016); Diefenbach et al. (2009); Gerolimatos et al. (2013); Massena et al. (2015)
HADS-D = Hospital Anxiety and Depression Scale – Depression Scale	.28		Ball et al. (2015); Kneebone et al. (2016)
Negative Affect Measure			
PANAS = Positive and Negative Affect Schedule – Negative	.58		Pachana et al. (2007)
Positive Affect Measure			
PANAS = Positive and Negative Affect Schedule – Positive	–.34		Pachana et al. (2007)

Sensitivity to Change and Responsiveness

To our knowledge, only Ball et al. (2015) explicitly assessed sensitivity to treatment of the GAI in a clinical controlled trial. They concluded that the GAI is a useful tool for monitoring the outcome of treatment. According to the COSMIN checklist, responsiveness was rated as "poor" because no analyses were conducted between the score on the GAI and the gold standard to demonstrate the good performance of the former. Although it wasn't their primary aim, there are studies that support the sensitivity to change of the GAI in the treatment monitoring of anxiety or specific phobia (Pachana, Woodward, & Byrne, 2007; Welch et al., 2010).

Problematic Issues

Cross-Cultural Adaptation. Simple translation of a questionnaire is insufficient if it is to be used with a population from another country, culture, or language. In such cases, cross-cultural adaptation of the instrument is recommended (Gjersing, Caplehorn, & Clausen, 2010). Studies conducted with other language versions of the GAI generally report information on the translation process but most do not report having culturally adapted the instrument.

Potentially Problematic Items. Some authors have pointed out that item 12 – "I get an upset stomach due to my worrying" – is problematic and not precise enough (Champagne et al., 2016; Molde et al., 2017; Yan et al., 2014). This item may present a limitation due to its cultural validity. Alternatively, it may be difficult for older adults to consider an upset stomach as a consequence of worrying. It is also possible that this symptom is endorsed only by some individuals because it refers to a more severe level of anxiety.

Other authors have questioned whether certain items truly assess the construct of interest. For example, item 2 – "I find it difficult to make a decision" – was reported as being insufficiently precise and having low corrected item-total correlations (Gould et al., 2014; Yan et al., 2014). Researchers have proposed that this item may instead assess decision-making abilities that aren't necessarily related to worry. For the same reasons, authors have questioned whether item 18 "I sometimes feel a great knot in my stomach" assesses anxiety or a symptom of a general medical condition (Gould et al., 2014; Heissler, Kopecek, & Stepankova Georgi, 2018; Yan et al., 2014). Gould et al. (2014) have also indicated that item 7 – "I often feel like I have butterflies in my stomach" – and 14 – "I always anticipate the worst will happen" – may be endorsed for reasons unrelated to anxiety.

Results from Rasch models in Molde et al. (2017) identified substantial item overlap between item-pairs 10 ("I often feel nervous") and 15 ("I often feel shaky inside"),

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

Table 4: Factors associated with each item of the GAI

Item	Bakkane Bendixen et al. (2016)	Diefenbach et al. (2014)	Marquez-Gonzalez et al. (2012)	Mababu and RuizSánchez (2016)	Guan (2016)
1	Worries	Excessive worry symptoms	Cognitive symptoms	Cognitive symptoms	Mental anxiety
2	Worries	Decision-making symptoms	Cognitive symptoms	Cognitive symptoms	Mental anxiety
3	Physical symptoms	NA	Cognitive symptoms	Cognitive symptoms	Mental anxiety
4	Worries	CNS hyperarousal symptoms	Arousal symptoms	Arousal symptoms	Mental anxiety
5	Worries	CNS hyperarousal symptoms	Cognitive symptoms	Cognitive symptoms	Mental anxiety
6	Worries	Excessive worry symptoms	Arousal symptoms	Arousal symptoms	Mental anxiety
7	Physical symptoms	Gastrointestinal symptoms	Somatic symptoms	Somatic symptoms	Mental anxiety
8	Worries	Excessive worry symptoms	Cognitive symptoms	Cognitive symptoms	Physical anxiety
9	Worries	Gastrointestinal symptoms	Cognitive symptoms	Cognitive symptoms	Mental anxiety
10	Physical symptoms	CNS hyperarousal symptoms	Arousal symptoms	Arousal symptoms	Mental anxiety
11	Worries	CNS hyperarousal symptoms	Cognitive symptoms	Cognitive symptoms	Mental anxiety
12	Physical symptoms	Gastrointestinal symptoms	Somatic symptoms	Somatic symptoms	Physical anxiety
13	Worries	Gastrointestinal symptoms	Arousal symptoms	Arousal symptoms	Mental anxiety
14	Worries	Excessive worry symptoms	Cognitive symptoms	Cognitive symptoms	Negative anxiety
15	Physical symptoms	Gastrointestinal symptoms	Somatic symptoms	Somatic symptoms	Negative anxiety
16	Worries	Excessive worry symptoms	Cognitive symptoms	Cognitive symptoms	Negative anxiety
17	Worries	Excessive worry symptoms	Cognitive symptoms	Cognitive symptoms	Negative anxiety
18	Physical symptoms	Gastrointestinal symptoms	Somatic symptoms	Somatic symptoms	Physical anxiety
19	Worries	Excessive worry symptoms	Cognitive symptoms	Cognitive symptoms	Negative anxiety
20	Physical symptoms	CNS hyperarousal symptoms	Arousal symptoms	Arousal symptoms	Mental anxiety

Note. **CNS = central nervous system; NA = not available.**

7 ("I often feel like I have butterflies in my stomach") and 18 ("I sometimes feel a great knot in my stomach"), 10 ("I often feel nervous") and 13 ("I think of myself as a nervous person"), 6 ("Little things bother me a lot") and 9 ("I can't help worrying about even trivial things"), 1 ("I worry a lot of the time") and 8 ("I think of myself as a worrier"), and 16 ("I think that my worries interfere with my life") and 17 ("My worry often overwhelms me"). To a certain extent, the same phenomenon of redundancy was observed for items 8, 10, and 11 of the GAI short version. Thus, the detected item overlap suggests that there may be redundant items in the GAI and the GAI-SF that do not provide any additional information because of their similar content (Molde et al., 2017).

Floor Effects. Some authors have made assumptions about the possible presence of floor effects in the GAI and GAI-SF. Yan et al. (2014) and Johnco et al. (2014) hypothesized the presence of floor effects when they observed that the GAI may be less suitable for elderly people with low-level anxiety as this would mean that they would not endorse several items suggesting high-level anxiety and serious outcomes.

Discussion

Since their development, the GAI and, to a lesser extent, the GAI-SF, have undergone extensive psychometric testing in a wide range of populations and countries. These tools have been the subject of different reviews (Balsamo et al., 2018; Creighton et al., 2018; Dissanayaka et al., 2015; Edelstein et al., 2008; Lin et al., 2016; Pachana & Byrne, 2012; Therrien & Hunsley, 2012). However, these reviews were mostly dedicated to the GAI long form and examined in specific populations or settings (e.g., Parkinson's disease; residential aged care facilities). Moreover, these reviews examined only a small sample of studies. The study that reviewed the largest number of articles on the GAI included only 18 (Balsamo et al., 2018), a number that is almost half of what our own search retrieved.

The goals of our study were to summarize existing evidence on the psychometric properties of the GAI and GAI-SF, to assess the methodological quality of the studies and to provide guidance for future psychometric validation studies. For the current review, we identified 31 studies that purposely studied the psychometric properties of these scales. As in reviews by other researchers (Balsamo et al., 2018; Creighton et al., 2018; Dissanayaka et al., 2015; Edelstein et al., 2008; Lin et al., 2016; Pachana & Byrne, 2012; Therrien & Hunsley, 2012), we generally found appropriate psychometric properties for the GAI and GAI-SF among various clinical and non-clinical populations of older adults. However, we also mostly found low levels of methodological quality in the studies retained for this review.

To our knowledge, this is the first systematic review to examine the methodological quality of research conducted on the psychometric properties of the GAI and GAI-SF. The majority of the COSMIN boxes were rated as "poor" or "fair" (83.5% of all boxes) for the different psychometric properties evaluated. The results of

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

Table 5: Criterion validity and cut-off point of the GAI and GAI-SF

Study	Cut-off	Sn	Sp	PPV	NPV	AUC	% correct	Criterion
Brazilian Portuguese Version (GAI-BR)								
Massena et al. (2015)	13 to detect GAD	83.3	84.6	—	—	90	—	MINI
English version								
Diefenbach et al. (2009)	9 to detect GAD or an anxiety disorder NOS	87.5	95.5	87.5	95.5	89.5	—	ADIS-IV
Byrne et al. (2010)	9 to detect GAD	—	—	—	—	93	93	MINI-V
Cheung et al. (2012)	3 to detect an anxiety disorder	85.7	78	57.1	94.1	83	—	MINI
Matheson et al. (2012)	7 to detect an anxiety disorder	87.5	85.7	—	—	91	—	MINI-Plus
Bradford et al. (2013)	8 to detect an anxiety disorder	58	93	94	56	69.1	—	MINI
	10 to detect an anxiety disorder (when rated by collaterals)	62	93	94	57	80.9	—	
Gerolimatos et al. (2013)	9 to detect an anxiety disorder	100	60	—	—	79	65.3	Psychiatric diagnosis based on an assessment that is part of routine care.
Johnco, Knight, Tadic, and Wuthrich (2014)	9 to detect an anxiety disorder	69.5	100	—	—	98.1	—	ADIS-IV
Kneebone et al. (2016)	7 to detect anxiety	88	84	—	—	84	—	SCID-I-RV
Creighton et al. (2018)	9 to detect GAD	90	86.3	45	98.6	93	86.7	MINI
English version – short form (GAI-SF)								
Byrne and Pachana (2011)	3 to detect GAD	75	86.8	—	—	80	86	MINI-V
Gerolimatos et al. (2013)	2 to detect an anxiety disorder	100	46.2	—	—	78	53.3	Psychiatric diagnosis based on an assessment that is part of routine care.
Johnco et al. (2014)	3 to detect an anxiety disorder	78.1	98.3	—	—	95.4	—	ADIS-IV
Italian version (GAI-It)								
Ferrari et al. (2017)	Unspecified	30	57	31	55	—	—	DSM-IV diagnosis after clinical psychiatric evaluation
		74	43	36	79	—	—	ICD-10 diagnosis after clinical psychiatric evaluation
		76	88	95	54	—	—	STAI-Trait
		74	81	93	46	—	—	STAI-State
		55	71	60	67	—	—	HADS-Anxiety
Italian version- short form (GAI-It-SF)								
Ferrari et al. (2017)	Unspecified	70	35	45	69	—	—	DSM-IV diagnosis after clinical psychiatric evaluation
		75	44	36	81	—	—	ICD-10 diagnosis after clinical psychiatric evaluation
		74	76	89	52	—	—	STAI-Trait
		72	69	88	44	—	—	STAI-State
		82	71	69	83	—	—	HADS-Anxiety
Norwegian version								
Molde et al. (2017)	Unspecified	77	79	—	—	86	78	
Norwegian version- short form								
Molde et al. (2017)	Unspecified	79	68	—	—	82	74	

Continued

Table 5: Continued

Study	Cut-off	Sn	Sp	PPV	NPV	AUC	% correct	Criterion
Portuguese version (GAI-PT) Ribeiro et al. (2011)	9 to detect clinically significant anxiety	88.8	80.4	–	–	91.6	–	Clinical diagnosis established by a psychiatrist and/or by a clinical psychologist
	9 to identify patients with any psychiatric disorder (clinical samples) versus community group without psychological distress	84.4 to 95.7	80.4	–	–	–	–	
	11 to identify patients with psychological distress versus community group without psychological distress	90	85.9	–	–	92.7	–	

Note. **ADIS-IV = Anxiety Disorder Interview Schedule for DSM-IV; AUC = area under the curve; GAD = generalized anxiety disorder; ICD-10 = *International Classification of Diseases*, 10th Edition; MINI = Mini International Neuropsychiatric Interview; NOS = not otherwise specified; NPV = negative predictive value; PPV = positive predictive value; SCID-I-RV = Structured Clinical Interview for DSM-IV Axis 1 Disorders: Research; Sn = sensitivity; Sp = specificity.**

studies with low methodological quality were not ignored in this review. A "poor" or "fair" methodological quality score indicates a certain risk of bias in the results but does not necessarily mean that findings are in fact biased or invalid. For example, the most frequent reasons for these ratings were low sample size and a lack of information on the number of missing items and how they were handled. Even though information on missing items was lacking, it still could have been correctly managed in an effort to avoid introducing bias into the study results. It is also important to note that results of studies that presented a "poor" versus "good" or "excellent" level of methodological quality according to the COSMIN checklist generally presented comparable results in this review. Nonetheless, our main finding – that the GAI and GAI-SF generally show adequate psychometric properties based on methodologically weak studies – is important.

We recommend the use of these tools and emphasize the need for better designed research on the validity and reliability of the GAI and GAI-SF. Researchers should always assess the psychometric properties of the instruments they use. Psychometric properties of an instrument are specific to the population and purpose for which it was intended and thus, they are not the properties of the instrument per se (Hunsley & Mash, 2008).

Not all psychometric properties received the same level of attention, which in turn influences the confidence that can be placed in the results (e.g., internal consistency was evaluated in 30 studies vs. two for content validity). Internal consistency was assessed in the majority of the studies and coefficients mostly fell above the acceptable threshold recommended by Hunsley and Mash (2008) for evidence-based assessment. Since the alpha values are influenced by the number of items of a scale, it is not surprising that lower coefficients were found for the GAI-SF (Streiner, 2003). Retest reliability was assessed in only 11 studies. Apart from the two lowest coefficients found for the long form (r = .53 and .58; Kneebone et al., 2016; Silva et al., 2016), the lowest coefficient was .85. The two studies that presented the lowest coefficients also exhibited poor methodological quality for the reliability assessment according to the COSMIN checklist. However, other studies also presented poor quality but higher retest coefficients. Most studies presented a retest coefficient that was considered as acceptable according to the recommendations of Hunsley and Mash (2008).

Results concerning test-retest reliability as well as convergent and criterion validity suggest that the GAI and GAI-SF both assess rather stable components of anxiety. The good convergent and criterion validity found for measures of GAD symptoms suggest a capacity to evaluate trait anxiety (stable tendency to experience

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

anxiety), and a low correlation found with a measure of state anxiety supports this hypothesis. The higher retest reliability of the short form versus the long form found in Silva et al. (2016) also leads us to question whether the items of the short version refer to more stable anxiety symptoms.

Inconsistency in divergent, factorial, and criterion validity was found across studies. Whereas some authors have concluded that moderate to high correlations with a measure of depressive symptoms are evidence of poor divergent validity, others have argued that it has not been well established that anxiety and depression are completely independent disorders in the elderly population considering the overlap of symptoms (Cassidy, Lauderdale, & Sheikh, 2005). The associations found are not specific to the GAI and GAI-SF, but are rather characteristic of other measures commonly used to assess anxiety in the elderly population (Therrien & Hunsley, 2012).

Further, conflicting results were found for the factorial structure of the GAI concerning the unidimensionality and multidimensionality of the scale. However, the fact that three of the four studies that presented the highest methodological quality (either "good" or "excellent") for structural validity concluded to the unidimensionality of the scale leads us to support this result as well. Mababu and RuizSánchez (2016) suggested that the different factorial structures found in previous studies for the GAI could be due in part to the dichotomous response format. Among the possible impacts of a dichotomous scale are a decrease in the percentage of explained variance and lower loadings (Lozano, García-Cueto, & Muñiz, 2008; Velicer, DiClemente, & Corriveau, 1984). Molde et al. (2017) also proposed different explanations for the lack of factorial consistency: different cultural response styles, differences in semantics due to translation processes, different sample characteristics, and true cultural differences in the structure of anxiety across countries.

The unidimensionality of the GAI raises the question as to whether it reflects all manifestations of anxiety in a context where the GAI was designed to assess a range of anxiety presentations (Pachana, Byrne, et al., 2007). There is currently a consensus on the unidimensionality of the GAI-SF, which is not surprising for a 5-item scale. An obvious issue when designing the short form of an instrument is to ensure that the target content domain is still adequately represented despite the reduced number of items (Smith, McCarthy, & Anderson, 2000). This does not seem to be the case with the GAI-SF as it is composed largely of items that relate to cognitive symptoms. Evidence on criterion validity shows that the GAI and the GAI-SF can screen for probable cases of anxiety disorders. However, no specific cut-off score for the detection of an anxiety disorder can be established because of significant variability in the results.

Some psychometric properties of the GAI were sometimes found to be slightly better than those of the GAI-SF, but most authors concluded that the results were nevertheless comparable. Most studies that assessed the psychometric properties of the short form had extracted data from the GAI long form. Although we can only speculate about the consequences of this procedure, it is possible that the psychometric properties of the GAI-SF differ when administered independently because of context, primacy and recency, and warm-up effects. When validating a brief scale, it should be considered as a completely new measure and thus submitted to independent validation procedures (Smith et al., 2000). Until further data from an independent assessment of the short version become available, the findings of this review suggest that psychometric properties are not a major issue when choosing between the GAI and the GAI-SF. The choice of one instrument or the other depend on the user's needs. For example, the short form may be the best option in specific situations such as when time is limited, when there is a demanding clinical context (e.g., acute geriatric settings), with elderly people who are easily fatigued or distracted, when multiple questionnaires are administered to patients, or when patients are frequently monitored.

Suggestions for Future Research

Future efforts to validate the GAI and GAI-SF should include paying particular attention to the previously identified problems and aiming to achieve a higher degree of methodological quality. Since content validity is considered to be the most important psychometric property according to COSMIN and that it was hardly tested in previous research, more studies should address this situation. Researchers should not assume that the culture and scales' content are equivalent. Also, the ability of the GAI and GAI-SF to distinguish between anxiety and depression symptoms is limited. Therefore, it would be interesting to further examine this issue by going beyond standard correlational analyses (e.g., by using the heterotrait-monotrait ratio of the correlation method, by comparing the answer profiles of depressed and non-depressed elderly individuals, or by identifying specific items that spark confusion as to the true nature of symptoms [i.e., related to depression or anxiety]).

The appropriateness of the GAI and GAI-SF for monitoring treatment change also requires further attention. Considering that the GAI and GAI-SF were developed to assess a range of anxiety disorders rather than a

https://doi.org/10.1017/S0714980820000185 Published online by Cambridge University Press

specific disorder, it would also be interesting to define cut-off scores to identify different severity levels of anxiety symptoms (i.e., no symptoms; mild, moderate, and severe anxiety). The usefulness of the GAI as a screening tool for various anxiety disorders should also be documented. Some previously described results (e.g., high alpha values [Panayides, 2013], possible floor effects, the overrepresentation of cognitive symptoms in factorial structures, problematic items due to item overlap) lead us to think that the GAI and GAI-SF may not comparably cover the various areas of anxiety. Thus, a review of the performance of each item is required to identify those that lack precision and those that adequately measure anxiety. A reformulation of items may also be necessary in order to improve the scale's validity.

Strengths and Limitations

Our findings are subject to several limitations. Despite great effort having been made to identify all relevant studies for this review, it is still possible that some were not included. Exclusion of non-English and non-French papers may have introduced a selection bias. Although a notable strength is that the literature search and the quality assessments were conducted by individuals working independently by using the COSMIN checklist, data extraction was completed by the first author only. Some notable strengths of this review are that the methodological quality of the research was assessed and that the coverage of the review is the broadest to date.

Conclusion

To our knowledge, this was the first systematic review to be conducted on the psychometric properties of the GAI and GAI-SF. The data provided by this review support the recommendation of these tools as screening measures for anxiety in older adults, especially in countries and in versions wherein psychometric properties were shown to be adequate. However, this review also highlights weaknesses in the methodological quality of research in this area. Researchers are encouraged to continue to assess the psychometric properties of the GAI and GAI-SF and apply standards such as the COSMIN checklist for study design and result reporting.

References

Bakkane Bendixen, A., Hartberg, C. B., Selbæk, G., & Engedal, K. (2016). Symptoms of anxiety in older adults with depression, dementia, or psychosis: A principal component analysis of the geriatric anxiety inventory. *Dementia and Geriatric Cognitive Disorders*, *42*(5–6), 310–322.

Ball, S. G., Lipsius, S., & Escobar, R. (2015). Validation of the geriatric anxiety inventory in a duloxetine clinical trial for elderly adults with generalized anxiety disorder. *International Psychogeriatrics*, *27*(9), 1533–1539. https://doi.org/10.1017/S1041610215000381

Balsamo, M., Cataldi, F., Carlucci, L., & Fairfield, B. (2018). Assessment of anxiety in older adults: a review of self-report measures. *Clinical Interventions in Aging*, *13*, 573–593. https://doi.org/10.2147/CIA.S114100

Boddice, G., & Byrne, G. J. (2008). The clinical utility of the geriatric anxiety inventory in older adults with cognitive impairment. *Nursing Older People*, *20*(8), 36–39. doi:10.7748/nop2008.10.20.8.36.c6809

Bradford, A., Brenes, G. A., Robinson, R. A., Wilson, N., Snow, A. L., Kunik, M. E., … Amspoker, A. B. (2013). Concordance of self- and proxy-rated worry and anxiety symptoms in older adults with dementia. *Journal of Anxiety Disorders*, *27*(1), 125–130. doi:10.1016/j.janxdis.2012.11.001

Byrne, G. J., Pachana, N. A., Goncalves, D. C., Arnold, E., King, R., & Khoo, S. K. (2010). Psychometric properties and health correlates of the Geriatric Anxiety Inventory in Australian community-residing older women. *Aging Mental Health*, *14*(3), 247–254. doi:10.1080/13607861003587628

Byrne, G. J., & Pachana, N. A. (2011). Development and validation of a short form of the Geriatric Anxiety Inventory - the GAI-SF. *International Psychogeriatrics*, *23*(1), 125–131. https://doi.org/10.1017/S1041610210001237

Cassidy, E. L., Lauderdale, S., & Sheikh, J. I. (2005). Mixed anxiety and depression in older adults: Clinical characteristics and management. *Journal of Geriatric Psychiatry and Neurology*, *18*(2), 83–88. https://doi.org/10.1177/0891988705276060

Champagne, A., Landreville, P., Gosselin, P., & Carmichael, P.H. (2016). Psychometric properties of the French Canadian version of the Geriatric Anxiety Inventory. *Aging & Mental Health*, *22*(1), 40–45. doi: https://doi.org/10.1080/13607863.2016.1226767

Cheung, G., Patrick, C., Sullivan, G., Cooray, M., & Chang, C. L. (2012). Sensitivity and specificity of the Geriatric Anxiety Inventory and the Hospital Anxiety and Depression Scale in the detection of anxiety disorders in older people with chronic obstructive pulmonary disease. *International Psychogeriatrics*, *24*(1), 128–136. https://doi.org/10.1017/S1041610211001426

Cheung, G. (2007). Concurrent validity of the Geriatric Anxiety Inventory in late-life depression. *International Psychogeriatrics*, *19*(2), 333–335. doi:10.1017/S1041610206004340

Creighton, A. S., Davison, T. E., & Kissane, D. W. (2018). The assessment of anxiety in aged care residents: A systematic review of the psychometric properties of commonly used measures. *International Psychogeriatrics*, *30*(7), 967–979.

Diefenbach, G. J., Bragdon, L. B., & Blank, K. (2014). Geriatric anxiety inventory: Factor structure and associations with cognitive status. *The American Journal of Geriatric Psychiatry*, *22*(12), 1418–1426. https://doi.org/10.1016/j.jagp.2013.04.014

Diefenbach, G. J., Tolin, D. F., Meunier, S. A., & Gilliam, C. M. (2009). Assessment of anxiety in older home care recipients. *Gerontologist*, *49*(2), 141–153. doi:10.1093/geront/gnp019

Dissanayaka, N. N. W., Torbey, E., & Pachana, N. A. (2015). Anxiety rating scales in Parkinson's disease: A critical review updating recent literature. *International Psychogeriatrics*, *27*(11), 1777–1784. https://doi.org/10.1017/S1041610215000885

Dow, B., Lin, X., Pachana, N. A., Bryant, C., LoGiudice, D., Goh, A. M., & Haralambous, B. (2018). Reliability, concurrent validity, and cultural adaptation of the Geriatric Depression Scale and the Geriatric Anxiety Inventory for detecting depression and anxiety symptoms among older Chinese immigrants: an Australian study. *International Psychogeriatrics*, *30*(5), 735–748. doi:10.1017/S1041610217002332

Edelstein, B. A., Woodhead, E. L., Segal, D. L., Heisel, M. J., Bower, E. H., Lowery, A. J., & Stoner, S. A. (2008). Older adult psychological assessment: Current instrument status and related considerations. *Clinical Gerontologist: The Journal of Aging and Mental Health*, *31*(3), 1–35. https://doi.org/10.1080/07317110802072108

Ferrari, S., Signorelli, M., Cerrato, F., Pingani, L., Massimino, M., Valente, S., … De Ronchi, D. (2017). Never too late to be anxious: validation of the Geriatric Anxiety Inventory, Italian version. *La Clinica terapeutica*, *168*(2), e120. doi: 10.7417/CT.2017.1992

Gerolimatos, L. A., Gregg, J. J., & Edelstein, B. A. (2013). Assessment of anxiety in long-term care: Examination of the Geriatric Anxiety Inventory (GAI) and its short form. *International Psychogeriatrics*, *25*(09), 1533–1542. doi: 10.1017/S1041610213000847

Gjersing, L., Caplehorn, J. R., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. *BMC Medical Research Methodology*, *10*(1), 13.

Gould, C. E., Beaudreau, S. A., Gullickson, G., Tenover, J. L., Bauer, E. A., & Huh, J. W. (2016). Implementation of a brief anxiety assessment and evaluation in a Department of Veterans Affairs geriatric primary care clinic. *Journal of Rehabilitation Research and Development*, *53*(3), 335–344. doi:10.1682/jrrd.2014.10.0258

Gould, C. E., Segal, D. L., Yochim, B. P., Pachana, N. A., Byrne, G. J., & Beaudreau, S. A. (2014). Measuring anxiety in late life: A psychometric examination of the Geriatric Anxiety Inventory and Geriatric Anxiety Scale. *Journal of Anxiety Disorders*, *28*(8), 804–811. http://dx.doi.org/10.1016/j.janxdis.2014.08.001

Guan, M. (2016). Factor structure of the Chinese version of the geriatric anxiety inventory. *Annals of General Psychiatry*, *15*, 4. https://doi.org/10.1186/s12991-016-0092-4

Heissler, R., Kopecek, M., & Stepankova Georgi, H. (2018). Geriatric Anxiety Inventory (GAI) and its short form GAI-SF: Czech normative study. *Ceskoslovenska Psychologie*, *62*(5), 462–476.

Hunsley, J., & Mash, E. J. (2008). *A guide to assessments that work*. NY, United States: Oxford University Press.

Johnco, C., Knight, A., Tadic, D., & Wuthrich, V. (2014). Psychometric properties of the Geriatric Anxiety Inventory (GAI) and its short-form (GAI-SF) in a clinical and non-clinical sample of older adults. *International Psychogeriatrics*, *11*, 1–9. https://doi.org/10.1017/S1041610214001586

Kneebone, I. I., Fife-Schaw, C., Lincoln, N. B., & Harder, H. (2016). A study of the validity and the reliability of the Geriatric Anxiety Inventory in screening for anxiety after stroke in older inpatients. *Clinical Rehabilitation*, *30*(12), 1220–1228.

Lin, X., Haralambous, B., Pachana, N. A., Bryant, C., LoGiudice, D., Goh, A., & Dow, B. (2016). Screening for depression and anxiety among older Chinese immigrants living in Western countries: The use of the Geriatric Depression Scale (GDS) and the Geriatric Anxiety Inventory (GAI). *Asia-Pacific Psychiatry*, *8*(1), 32–43.

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, *4*(2), 73–79.

Mababu, R., & RuizSánchez, G. (2016). Factorial validity of the Spanish version of the Geriatric Anxiety Inventory (GAI): Empirical analysis of its structure and dimensions. *Revista de Psicopatología y Psicología Clínica*, *21*(3), 201–208.

Marquez-Gonzalez, M., Losada, A., Fernandez-Fernandez, V., & Pachana, N. A. (2012). Psychometric properties of the Spanish version of the Geriatric Anxiety Inventory. *International Psychogeriatrics*, *24*(1), 137–144. https://doi.org/10.1017/S1041610211001505

Massena, P. N., de Araújo, N. B., Pachana, N., Laks, J., & de Pádua, A. C. (2015). Validation of the Brazilian Portuguese version of Geriatric Anxiety Inventory—GAI-BR. *International Psychogeriatrics*, *27*(7), 1113–1119. https://doi.org/10.1017/S1041610214001021

Matheson, S. F., Byrne, G. J., Dissanayaka, N. N., Pachana, N. A., Mellick, G. D., O'Sullivan, J. D., … Marsh, R. (2012). Validity and reliability of the Geriatric Anxiety Inventory in Parkinson's disease. *Australasian Journal on Ageing*, *31*(1), 13–16. https://doi.org/10.1111/j.1741-6612.2010.00487.x

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1.

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., … de Vet, H. C. W. (2010). The COSMIN checklist for evaluating the methodological

quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, *10*(1), 22.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … 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.

Molde, H., Hynninen, K., Torsheim, T., Bendixen, A., Engedal, K., Pachana, N., & Nordhus, I. (2017). A bifactor and item response analysis of the geriatric anxiety inventory. *International Psychogeriatrics*, *29*(10), 1647–1656.

Pachana, N. A., & Byrne, G. J. (2012). The Geriatric Anxiety Inventory: International use and future directions. *Australian Psychologist*, *47*(1), 33–38. https://doi.org/10.1111/j.1742-9544.2011.00052.x

Pachana, N. A., Byrne, G. J., Siddle, H., Koloski, N., Harley, E., & Arnold, E. (2007). Development and validation of the Geriatric Anxiety Inventory. *International Psychogeriatrics*, *19*(1), 103–114. doi:10.1017/S1041610206003504

Pachana, N. A., Woodward, R. M., & Byrne, G. J. A. (2007). Treatment of specific phobia in older adults. *Clinical Interventions in Aging*, *2*(3), 469–476.

Panayides, P. (2013). Coefficient alpha: Interpret with caution. *Europe's Journal of Psychology*, *9*(4), 687–696.

Park, M. S., Kang, K. J., Jang, S. J., Lee, J. Y., & Chang, S. J. (2018). Evaluating test-retest reliability in patient-reported outcome measures for older people: A systematic review. *International Journal of Nursing Studies*, *79*, 58–69.

Polit, D. F. (2014). Getting serious about test–retest reliability: A critique of retest research and some recommendations. *Quality of Life Research*, *23*(6), 1713–1720.

Ribeiro, O., Paul, C., Simoes, M. R., & Firmino, H. (2011). Portuguese version of the Geriatric Anxiety Inventory: Transcultural adaptation and psychometric validation. *Aging & Mental Health*, *15*(6), 742–748. https://doi.org/10.1080/13607863.2011.562177

Rozzini, L., Chilovi, B. V., Peli, M., Conti, M., Rozzini, R., Trabucchi, M., & Padovani, A. (2009). Anxiety symptoms in mild cognitive impairment. *International Journal of Geriatric Psychiatry*, *24*(3), 300–305. doi:10.1002/gps.2106

Silva, L. d. S. V. e., Oliveira, G. M. d., Yokomizo, J. E., Saran, L. F., Bottino, C. M. d. C., & Yassuda, M. S. (2016). The Geriatric Anxiety Inventory in primary care: Applicability and psychometric characteristics of the original and short form. *Archives of Clinical Psychiatry (São Paulo)*, *43*, 103–106.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102–111. https://doi.org/10.1037/1040-3590.12.1.102

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651–657.

Therrien, Z., & Hunsley, J. (2012). Assessment of anxiety in older adults: A systematic review of commonly used measures. *Aging & Mental Health*, *16*(1), 1–16. https://doi.org/10.1080/13607863.2011.602960

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test–retest reliability. *PloS One*, *8*(9), e73990. https://doi.org/10.1371/journal.pone.0073990

Velicer, W. F., DiClemente, C. C., & Corriveau, D. P. (1984). Item format and the structure of the Personal Orientation Inventory. *Applied Psychological Measurement*, *8*(4), 409–419.

Welch, T., Welch, M., Baer, J., Dias, J., Gurney, C., Van Dale, B., … Psiurski, S. (2010). Removed but not out of reach: Seniors with depression in smaller center, rural, and remote communities. *Psychiatric Annals*, *40*(12), 616–623.

Yan, Y., Xin, T., Wang, D., & Tang, D. (2014). Application of the Geriatric Anxiety Inventory-Chinese version (GAI-CV) to older people in Beijing communities. *International Psychogeriatrics*, *26*(3), 517–523. https://doi.org/10.1017/S1041610213002007