Journal of Affective Disorders 375 (2025) 359–365

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Factor structure and measurement invariance of the GAD-7 across time, sex, and language in young adults

Teodora Riglea [a,b,*], Robert J. Wellman [c], Marie-Pierre Sylvestre [a,b], Catherine Sabiston [d], Jennifer O'Loughlin [a,b]

[a] École de Santé Publique, Université de Montréal, 7101 Avenue du Parc, Montréal, QC H3N 1X9, Canada
[b] Centre de recherche du Centre Hospitalier Universitaire de Montréal, 900 Rue Saint-Denis, Montréal, QC H2X 0A9, Canada
[c] Department of Population and Quantitative Health Sciences, Division of Preventive and Behavioral Medicine, UMass Chan Medical School, 368 Plantation Street, Worcester, MA 01605, United States of America
[d] Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, Ontario M5R 2W6, Canada

ARTICLE INFO

Keywords:
GAD-7
Anxiety
Young adults
Measurement invariance
Sex-specific analyses

ABSTRACT

Introduction: The Generalized Anxiety Disorder-7 Scale (GAD-7) is widely used to measure anxiety symptom severity. One-factor, two-factor, and bifactor latent structures are supported by previous research. Yet, measurement invariance of the GAD-7 across sex and language (i.e., between groups) and longitudinally (i.e., within group over time) is infrequently studied in population-based samples. The objective was to examine the factor structure of the GAD-7 and its measurement invariance across sex, language, and time in young adults.
Methods: Data were drawn from an ongoing longitudinal investigation in Canada that began in 1999–2000 at age 12. One-factor, two-factor, and bifactor (S-1) models were compared in a sample of 799 participants at age 30. Measurement invariance was tested using multigroup confirmatory factor analyses iteratively in four steps (i.e., configural, thresholds, thresholds and loadings/strong) across sex (male; female) and language of questionnaire completion (English; French). Invariance across time was tested among 633 participants with data at ages 30, 34 and 35.
Results: A one-factor model showed the best fit. Partial strong invariance across sex and full strong invariance across language of the one-factor model was demonstrated. Strong invariance across time was also demonstrated, indicating stability in parameters over time for the same participants ages 30 to 35.
Limitations: The results are restricted to young adults and may not generalize to wider age ranges. Participants are predominantly born in Canada and report high levels of education and employment.
Conclusion: The one-factor structure of the GAD-7 demonstrated measurement invariance across sex, language, and time in young adults.

1. Introduction

Generalized anxiety disorder (GAD) is characterized by "excessive anxiety and worry (apprehensive expectation)" about events and activities, and is distinguished from object-specific anxiety disorders such as phobias, panic attacks, obsessions and compulsions, and traumatic events (American Psychiatric Association, 1994). While GAD can reach the level of a diagnosable disorder, its symptoms (i.e., feeling nervous, worrying, difficulty relaxing, restlessness, annoyance and irritability, and fear that something terrible might happen) are common experiences. The seven-item GAD-7 questionnaire (Spitzer et al., 2006) assesses the severity of anxiety symptoms even in individuals who do not meet diagnostic criteria for the syndrome. Widely used in epidemiological and mental health research and in clinical practice since its publication in 2006, this scale permits description of trends and short-term temporal variations in anxiety symptoms in longitudinal studies, and comparison of scores across subgroups of individuals. In most studies, one summary score of the GAD-7 is computed, with higher scores representing higher frequency (i.e., higher severity) (Spitzer et al., 2006) of anxiety symptoms, and scores are often compared between groups of individuals. Given the experiential nature of anxiety and the absence of easily obtained biological markers, self-report

* Corresponding author at: École de santé publique de l'Université de Montréal, 7101 avenue du Parc, Montréal QC, H3N 1X9, Canada.
E-mail address: teodora.riglea@umontreal.ca (T. Riglea).

https://doi.org/10.1016/j.jad.2025.01.117
Received 4 June 2024; Received in revised form 21 January 2025; Accepted 22 January 2025
Available online 27 January 2025

instruments such as the GAD-7 are key for quantifying anxiety severity. Consequently, there is an imperative for assuring strong psychometric properties of the GAD-7 across contexts to enable accurate measurement of anxiety symptoms and for valid comparisons between groups.

Despite widespread use, four psychometric properties of the GAD-7 remain unclarified. First, the literature on the factor structure of the GAD-7 is inconsistent and provides conflicting evidence for the use of the scale scores. Several studies to date supported a one-factor structure, a situation where all scale items load onto one latent factor (Borgogna et al., 2021; Delamain et al., 2024; Romano et al., 2022; Saunders et al., 2023; Shevlin et al., 2022; Sriken et al., 2022; Teymoori et al., 2020) and where the use of a total scale score as a summary statistic is justified. However, several studies identified a two-factor solution, with four items of the GAD-7 loading onto "affective symptoms" (i.e., feeling nervous, cannot stop worrying, worrying too much, and feeling afraid) and three items loading onto "somatic symptoms" (i.e., trouble relaxing, restlessness, and easily irritable), with both latent factors highly correlated (i.e., correlation coefficient > 0.8) (Hinz et al., 2023; Moreno et al., 2019; Ong et al., 2022; Stochl et al., 2022). The evidence of a two-factor structure conflicts with that on a one-factor structure because it would indicate that affective and somatic anxiety indicators are best examined separately. No differences between analytical procedures or sample characteristics in these studies appear to underpin the competing factor structures. Furthermore, studies have also examined higher order solutions such as bifactor models and found similar goodness-of-fit to unidimensional and two-dimensional solutions of the GAD-7 (Brattmyr et al., 2022; De Man et al., 2021; Doi et al., 2018). Additional population-based studies are needed to compare and contrast the different factor structures.

Second, measurement invariance of the GAD-7 across subgroups of individuals has seldom been studied. Comparing scores between groups hinges on the assumption that scores in one subgroup are psychometrically equivalent to scores in another subgroup (Clark and Donnellan, 2021), such that the measured variable carries the same meaning. If not invariant, then observed differences may reflect an imperfect instrument rather than a meaningful difference in the level of the construct (Liu et al., 2017). Lack of invariance across sex for example, would impede comparison of scores between females and males. Moreno et al. found the GAD-7 to be invariant across sex in a sample of Spanish primary care patients (Moreno et al., 2019), and both Sriken et al. (2022) and Borgogna et al. (2021) reported similar findings in samples of American college and university students. Saunders et al. (2023) reported measurement invariance across sex in a large representative sample ($n = 173{,}578$) of British individuals seeking psychological therapy services. Although these studies are consistent, the findings have not yet been replicated in Canadian males and females. This is an important gap because the GAD-7 is used to assess anxiety in large national surveys of the general population, such as the Canadian Survey on COVID-19 and Mental Health (Statistics Canada, 2023).

Third, the GAD-7 is translated from its original English into numerous languages including French (Micoulaud-Franchi et al., 2016), and it is used frequently as a self-report questionnaire in large Canadian longitudinal studies (Romano et al., 2022; Sylvestre et al., 2022; Watkins-Martin et al., 2021). Yet, its measurement invariance when translated to French has not been thoroughly examined. Research conducted in 18 European countries has shown evidence of scale invariance across six linguistic groups (i.e., Dutch, English, Italian, Spanish, Finnish, and Norwegian; French was not included due to small sample size) (Teymoori et al., 2020), and other studies have replicated findings in Spanish (Mills et al., 2014) and Arabic (Sawaya et al., 2016). Invariance across language is important for the generalizability of the findings and for the comparison of anxiety across different countries and linguistic groups. Investigation of measurement invariance of the French GAD-7 questionnaire is especially important for research conducted within a Canadian context, given Canada's official bilingualism.

Finally, invariance across time (i.e., measurement invariance within the same group of individuals across time, known as longitudinal invariance) of the GAD-7 is not well studied. This is a critical gap because longitudinal invariance allows for appropriate inferences from longitudinal analyses and is central to research examining changes in anxiety symptoms over time. To our knowledge, four studies addressing this issue have been published to date, with participant ages ranging from 18 to 65 and follow-up periods ranging from 3 months to 6 years (Hinz et al., 2023; Moreno et al., 2019; Ong et al., 2022; Stochl et al., 2022). Although three studies established full invariance (Hinz et al., 2023; Moreno et al., 2019; Stochl et al., 2022), one found invariance at the configural level only, such that the two-factor model structure was invariant over time but the factor loadings of each item changed over time (Ong et al., 2022). None of the samples in these studies were comprised of Canadian individuals.

In this paper, we address these gaps by studying the factor structure of the GAD-7 and by examining its measurement invariance across sex, language, and time, in community-dwelling Canadian young adults.

2. Methods

2.1. Population and design

Data were drawn from the Nicotine Dependence in Teens (NDIT) study, an ongoing longitudinal investigation in Montreal, Canada, that began in 1999 (O'Loughlin et al., 2015). Its main objective was to describe the natural course of nicotine dependence in youth. Grade 7 students from ten high schools were invited to participate and 1294 of the 2325 eligible students (56 %) agreed. Data on sociodemographic, psychosocial, lifestyle, and physical and mental health characteristics were collected repeatedly in 25 data collection cycles to date. Participants could complete the questionnaires in their preferred language (English or French) at every cycle. Characteristics of NDIT participants at inception were similar to those of a provincially representative sample of Quebec youth (O'Loughlin et al., 2015). The NDIT study was approved by Centre de recherche du Centre Hospitalier Universitaire de Montréal research ethics board.

In this study, data from three consecutive study cycles during adulthood were used. Data collection for the three cycles took place, respectively, between January 2017 and June 2020 at age 30; between December 2020 and September 2021 at age 34; and between June 2022 and March 2023 at age 35.

2.2. Measures

Participants completed the GAD-7 by responding to the question "*In the past 2 weeks, how often have you been bothered by …?*" for seven anxiety symptoms: (1) feeling nervous, anxious, or on edge; (2) not being able to stop or control worrying; (3) worrying too much about different things; (4) trouble relaxing; (5) being so restless that it's hard to sit still; (6) becoming easily annoyed or irritable; and (7) feeling afraid, as if something terrible might happen. Response options were *not at all* (0), *several days* (1), *over half the days* (2), and *every day* (3), yielding a summary score with possible values ranging from 0 to 21. The reported sensitivity and specificity of the GAD-7 (against diagnoses by mental health professionals) are 0.89 and 0.82, respectively (Spitzer et al., 2006).

Participant and sociodemographic characteristics include sex (female; male), language of questionnaire completion (English; French), age (in years), born in Canada (yes; no), attended university (yes; no), is employed (yes; no), and household income >C$50,000 (yes; no).

2.3. Statistical analysis

Statistical analyses were conducted in two sub-studies. In study 1, the cross-sectional analysis, the factor structure of the GAD-7 was examined and measurement invariance across sex and language at age

30 was tested. In study 2, the longitudinal analysis, measurement invariance across time (i.e., longitudinal invariance) of the GAD-7 at ages 30, 34, and 35 was tested.

2.3.1. Study 1

We examined the factor structure of the GAD-7 in the sample of participants who completed survey cycle 23 (i.e., age 30) by comparing three confirmatory factor analysis (CFA) models: (i) a one-factor model; (ii) a two-factor model with correlated latent factors; and (iii) a bifactor (S-1) model (Fig. 1). A bifactor model can be appropriate in situations where the construct is conceptually thought to be primarily unidimensional, but there are also secondary factors of substantive interest that should be included (Kline, 2023). The bifactor (S-1) model was estimated with a general factor and one "affective symptoms" specific group (Eid et al., 2017). The "affective symptoms" group was added as a specific group because past studies found an "affective symptoms" latent factor to explain a larger proportion of the variance than a "somatic symptoms" latent factor and have higher factor loadings (De Man et al., 2021; Moreno et al., 2019). All models were run using diagonally weighted least squares (DWLS) and robust fit statistics, as recommended for ordered-categorical measures (Flora and Curran, 2004).

Fit evaluation followed the most up-to-date best practices for CFA (Kline, 2023; Svetina et al., 2020). We first examined the chi-square statistic and p value. If the result of the chi-square test was statistically significant ($p < 0.05$), it was deemed that the model failed the chi-square test and we investigated the covariance correlation matrix for values $> |0.10|$ and for possible sources of misfit using modification indices (i.e., which approximate the improvement in model fit if identified modification(s) were applied to the model) (Kline, 2023). If deemed appropriate, we allowed correlated error variances between items and re-examined model fit as previously described. We also report the omega hierarchical for each model, a reliability measure representing the proportion of total variance due to a particular factor (Flora, 2020). For additional qualitative model appraisal, we also report the CFI, RMSEA, and SRMR fit indices and the average variance extracted (AVE) (i.e., a measure of the explanatory power of the model computed as the average of the squared standardized loadings, recommended to explain the majority of the variance (i.e., $\geq$0.50)) (Fornell and Larcker, 1981; Kline, 2023). We referred to the following thresholds in fit indexes for qualitative appraisal: CFI $\geq$ 0.95, RMSEA $\leq$ 0.06, and SRMR $\leq$ 0.08 (Cheung and Rensvold, 2002; Hu and Bentler, 1999; Perry et al., 2015), however model selection was not based solely on the fit indexes due to their unclear application to ordered categorical variables (Xia and Yang, 2019). The selection of one of the three models for subsequent

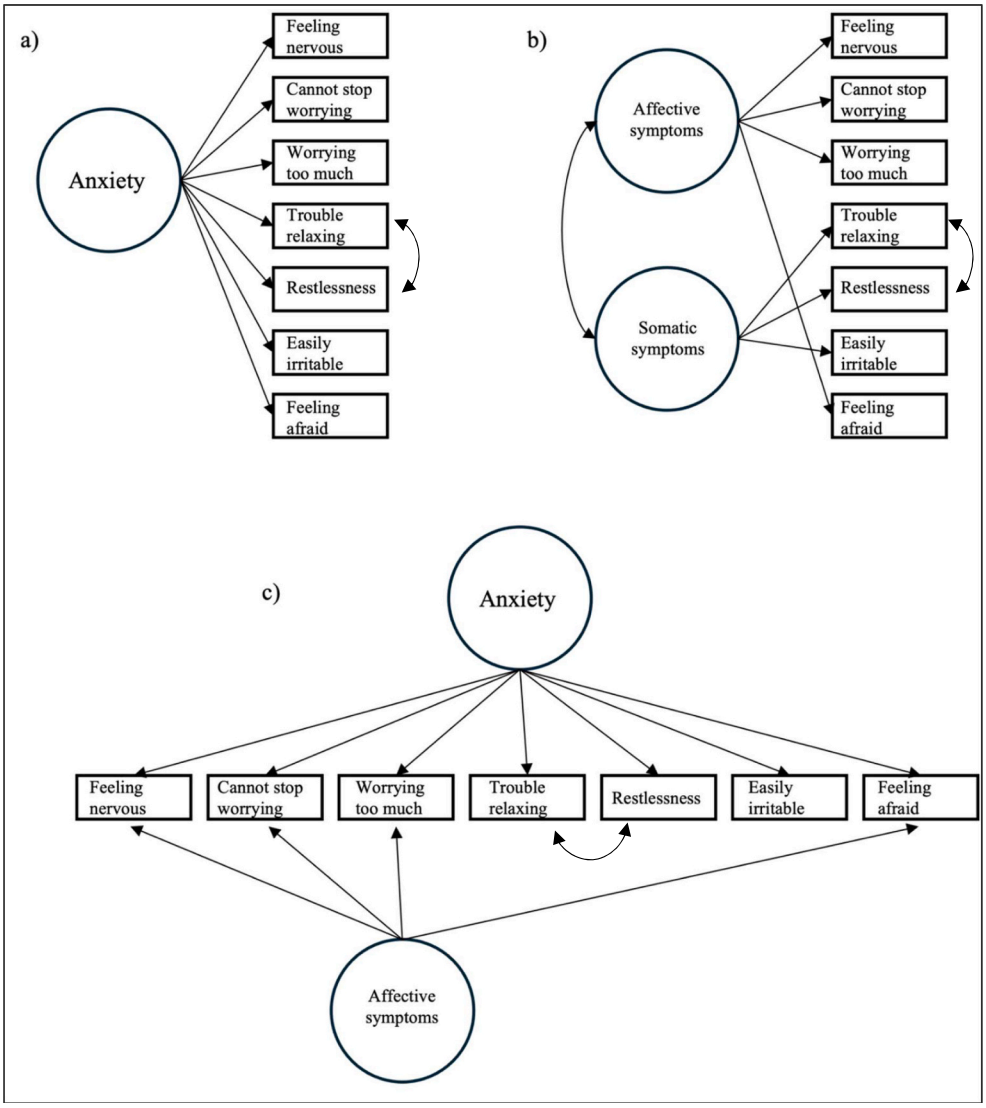

Fig. 1. Factor structure of the GAD-7 examined in the cross-sectional study in three competing models: a) one-factor model; b) two-factor model; c) bifactor (S-1) model.

measurement invariance was thus based on comparisons of chi-square statistics, residual covariances, AVE and fit indices, examination of factor loadings for each item, and the omega hierarchical.

We assessed measurement invariance across sex and language of questionnaire completion using multigroup CFA (Kline, 2023). The models were specified with increasingly stringent requirements on the equality constraints in an iterative process: (i) configural invariance, assessing equivalence of model form without any constraints; (ii) equality of thresholds; and (iii) equality of thresholds and loadings (corresponding to strong invariance). This procedure is proposed by Svetina et al. (2020) and Wu and Estabrook (2016) for examining invariance for categorical outcomes. Model fit at each stage was compared with the corresponding model at the previous stage (e.g., the model with constraints on the thresholds was compared to the configural invariance model) for model deterioration by comparing chi-square values, inspection of residual covariances (i.e., residuals > |0.10| would be deemed problematic (Kline, 2023)), and a qualitative appraisal of important changes in the CFI and RMSEA (90 %) values. If model fit was not appraised to be satisfactory, the procedure would end, and no further constrained models would be considered.

According to recent recommendations, evidence for at least partial strong invariance (i.e., equality of thresholds and loadings) is required to perform direct comparisons of scales scores across groups (Kline, 2023). Therefore, if tests for strong invariance were not successful (as described above) we tested for partial strong invariance, where we would accept violations of measurement invariance on some items (i.e., some, but not all, GAD-7 items would be equal over groups at the level of strong invariance). We used the Lagrange multiplier test to evaluate the changes in model fit by freely estimating, one at a time, a constrained parameter) (Kline, 2023).

2.3.2. Study 2

Finally, we examined measurement invariance over time using 'longitudinal measurement invariance.' The procedure considers inter-item correlations from one cycle to the next (i.e., dependent observations) (Widaman et al., 2010) and was tested in a single-group CFA (i.e., data from all three survey cycles included in one model). The same iterative process as explained above for study 1 was followed.

All statistical analyses were performed using R version 4.3.2 using the *lavaan (*Mackinnon et al., 2022*)* and *minvariance* packages. Coding script is openly available at: https://osf.io/6fbtq.

3. Results

3.1. Sample characteristics

A total of 799 participants had data available on the GAD-7 at age 30 and comprised the sample for the cross-sectional study. Missing data on items of the GAD-7 ranged from 2 to 7 participants. Mean (SD) age was 30.6 (1.0) years; 56 % were female; 31 % completed the questionnaire in French; and 62 % attended university (Table 1).

Table 1
Sociodemographic characteristics at baseline of the study participants included in the cross-sectional study (n = 799) and the longitudinal study (n = 633), Nicotine Dependence in Teens Study, 2017–2023.

	Cross-sectional sample (n = 799) Mean (SD) or n (%)	Longitudinal sample (n = 633) Mean (SD) or n (%)
Female, n (%)	448 (56.1)	376 (59.4)
Born in Canada, n (%)	749 (93.7)	596 (94.2)
Age, mean (SD)	30.6 (1.0)	30.5 (1.0)
Language of questionnaire completion, n (%)		
French	245 (30.7)	185 (29.3)
English	554 (69.3)	448 (70.7)
Attended university, n (%)	494 (61.8)	407 (64.3)
Household income >C$50,000, n (%)	581 (72.7)	463 (73.1)
Employed, n (%)	673 (84.2)	538 (85.1)
GAD-7 score, mean (SD)	4.6 (4.5)	4.7 (4.6)
median (IQR)	3.0 (6.0)	3.0 (6.0)
Levels of anxiety severity GAD-7 scale score, n (%)		
Minimal (0–4)	466 (59.5)	365 (59.1)
Mild (5–9)	203 (25.9)	157 (25.4)
Moderate (10–14)	80 (10.2)	68 (11.0)
Severe (15–21)	34 (4.3)	28 (4.5)

SD: standard deviation; IQR: interquartile range. Percentages exclude missing data from the denominator.

3.1.1. Study 1: cross-sectional study

3.1.1.1. Factor structure of the GAD-7. The one-factor model preliminarily failed the chi-square test (standard $\chi^2 = 39.172$, $p < 0.001$). Examination of residuals and modification indices suggested improved model fit by allowing a residual correlation between items 4 (trouble relaxing) and 5 (restlessness) in the one-factor model, but no important improvement in model fit (i.e., small change in chi-square values) in the two-factor or bifactor models. It was deemed reasonable to assume that the items "trouble relaxing" and "restlessness" might be interpreted as very similar symptoms. Table 2A presents the fit indices of the one-factor (with correlated residuals), two-factor, and bifactor (S-1) models. No correlations between residuals exceeded |0.10| across models. CFI, RMSEA, and SRMR values did not indicate any issues and, alongside the AVE, were comparable between models. The omega hierarchical was 0.90 for the one-factor model, 0.91 for the two-factor model, and 0.91 for the bifactor model.

In the two-factor model, the latent factors showed a very high correlation (i.e., standardized correlation coefficient: 0.96), which may indicate that no additional information would likely be provided from two latent factors compared to one. Furthermore, for the bifactor model, the omega hierarchical for the general factor was 0.88 and the specific "affective" factor in the bifactor model had an omega hierarchical of 0.06, indicating that most of the variance is attributable to the general factor in the bifactor model. Furthermore, factor loadings on the specific "affective" factor in the bifactor model were low, particularly for item 7 (factor loading = 0.04).

Based on model fit information, a one-factor model with correlated residuals for items 4 and 5 was retained for the subsequent measurement invariance analyses. Factor loadings ranged between 0.73 and 0.94.

3.1.1.2. Measurement invariance across sex. The one-factor model showed good fit in each group (for males: standard $\chi^2 = 11.55$, $p = 0.57$; for females: standard $\chi^2 = 13.09$, $p = 0.44$; all residuals $\leq |0.10|$) (Table 2B). When performing the MGCFA, the strong invariance model (i.e., equality of thresholds and loadings) had a statistically significant chi-square (standard $\chi^2 = 54.74$, $p = 0.04$) and was not retained. The Lagrange multiplier test suggested that allowing the first threshold of item 5 (restlessness) to be freely estimated improved model fit. This modification on one item out of seven (i.e., 14 %) led to a partial strong invariance model that was retained (standard $\chi^2 = 38.96$, $p = 0.38$). There were no important changes in fit indices compared to prior less constrained model (ΔCFI: 0.001; ΔRMSEA 0.015; ΔSRMR: 0, compared to the threshold model).

3.1.1.3. Measurement invariance across language. The one-factor model showed good fit in each sub-group of questionnaire language (for English: standard $\chi^2 = 15.34$, $p = 0.29$; for French: standard $\chi^2 = 12.36$, $p = 0.50$; all residuals $< |0.10|$). (Table 2C). The strong invariance model (equality of thresholds and loadings) showed good fit (standard $\chi^2 = 42.99$, $p = 0.27$, residuals $< |0.10|$) no important changes in fit indices compared to prior less constrained model (ΔCFI: 0.001; ΔRMSEA:

Table 2
Fit indices of the one-factor, two-factor, and bifactor CFA models at cycle 23 (n = 799), and results of the measurement invariance by sex and by language, Nicotine Dependence in Teens Study (2017–2023).

	Standard χ^2 (*p*)	Robust χ^2 (*p*)	df	CFI	RMSEA (90 % CI)	SRMR	AVE
A) Model fit indices at cycle 23							
One-factor[a]	20.080 (0.093)	48.701 (0.000)	13	0.997	0.059 (0.042; 0.077)	0.026	0.701
Two-factor	16.666 (0.163)	41.624 (0.000)	12	0.998	0.056 (0.038; 0.075)	0.024	0.718
Bifactor	10.521 (0.310)	28.087 (0.001)	9	0.998	0.052 (0.031; 0.074)	0.019	0.722
B) Measurement invariance by sex of the one-factor[a] model							
Males	11.551 (0.565)	29.386 (0.006)	13	0.997	0.060 (0.031; 0.089)	0.029	0.708
Females	13.090 (0.441)	30.256 (0.004)	13	0.997	0.055 (0.029; 0.080)	0.028	0.695
Configural	24.641 (0.539)	59.650 (0.000)	26	0.997	0.057 (0.038; 0.076)	0.029	
Thresholds	39.928 (0.158)	94.611 (0.000)	32	0.995	0.070 (0.054; 0.087)	0.029	
Thresholds + loadings	54.739 (0.039)	114.039 (0.000)	38	0.994	0.071 (0.056; 0.086)	0.030	
Partial thresholds + loadings[b]	38.958 (0.382)	81.677 (0.000)	37	0.996	0.055 (0.039; 0.071)	0.029	
C) Measurement invariance by language of the one-factor[a] model							
English	15.341 (0.287)	36.670 (0.000)	13	0.998	0.057 (0.036; 0.080)	0.026	0.706
French	12.360 (0.498)	30.400 (0.004)	13	0.995	0.074 (0.040; 0.109)	0.037	0.698
Configural	27.701 (0.373)	67.079 (0.000)	26	0.997	0.063 (0.045; 0.082)	0.030	
Thresholds	39.421 (0.172)	92.793 (0.000)	32	0.995	0.069 (0.053; 0.086)	0.030	
Thresholds + loadings	42.992 (0.266)	88.860 (0.000)	38	0.996	0.058 (0.042; 0.074)	0.030	

All residual correlations ≤ |0.10|.
[a] Model includes the specification for correlated residuals between items "trouble relaxing" and "restlessness".
[b] The specification item 5 | t1 makes an exception for the first threshold of item "restlessness".

0.011; ΔSRMR: 0, compared to the threshold model), and was retained.

3.1.2. Study 2: longitudinal study

3.1.2.1. Longitudinal measurement invariance. Of the 799 participants included in the cross-sectional study, 633 completed the study questionnaires at ages 34 and 35 and comprised the sample for the longitudinal analyses. There were missing data for 10 participants across the seven items of the GAD-7. Characteristics of participants in the longitudinal study were similar to those of participants in the cross-sectional study (Table 1).

Table 3 presents the fit indices of the one-factor model at each of the three consecutive study cycles (i.e., ages 30, 34, and 35), and the results of the longitudinal measurement invariance procedure. At ages 30 and 35, the models showed non-significant chi-square values, all residuals < |0.10|, and non-problematic fit indices values. At age 34, the one-factor model failed the chi-square test (standard $\chi^2 = 23.549$, $p = 0.04$). Examination of residuals and modification indices suggested improved model fit by allowing an additional residual correlation between items 2 (cannot stop worrying) and 3 (worrying too much) (standard $\chi^2 = 12.83$, $p = 0.38$). The similarity in wording in the symptoms ("cannot stop worrying" and "worrying too much") was deemed reasonable to explain the added residual correlation between the two items.

During the longitudinal measurement invariance procedure, all models showed statistically significant standard chi-square values. This was expected given that data from all three cycles were included in one model leading to an increase in the degrees of freedom. The strong invariance model (i.e., equality of thresholds and loadings) did not show evidence of model deterioration, compared to the previously less constrained model (ΔCFI: 0.001; ΔRMSEA: 0.003; ΔSRMR: 0.003, compared to the threshold model). Strong invariance (i.e., equality of thresholds and loadings) was deemed to be attained.

4. Discussion

Our results support a one-factor latent structure of the GAD-7 scale, providing evidence of unidimensionality in both whole sample and sex- and language-stratified CFA models. This aligns with previous studies showing the unidimensional GAD-7 structure to have satisfactory psychometric properties in other samples (Delamain et al., 2024; Saunders et al., 2023; Shevlin et al., 2022; Sriken et al., 2022). It is noteworthy that both the two-factor and bifactor models showed acceptable model fit, although we observed a high correlation between the two latent factors in the two-factor model, and insufficient psychometric evidence

Table 3
Fit indices of the one-factor CFA model at each of the three study cycles and results of the longitudinal measurement invariance (n = 633), Nicotine Dependence in Teens Study (2017–2023).

	Standard χ^2 (*p*)	Robust χ^2 (*p*)	df	CFI	RMSEA (90 % CI)	SRMR	AVE
A) Model fit indices at each cycle[a]							
Cycle 23	12.155 (0.515)	29.673 (0.005)	13	0.998	0.045 (0.024; 0.067)	0.022	0.704
Cycle 24[b]	12.832 (0.381)	28.882 (0.004)	12	0.998	0.047 (0.025; 0.070)	0.025	0.636
Cycle 25	18.654 (0.134)	44.961 (0.000)	13	0.996	0.063 (0.044; 0.083)	0.028	0.673
B) Longitudinal measurement invariance[c]							
Configural	188.992 (0.000)	226.863 (0.000)	119	0.994	0.038 (0.030; 0.045)	0.039	0.676
Threshold	188.992 (0.001)	244.171 (0.000)	131	0.994	0.037 (0.030; 0.044)	0.039	0.676
Threshold + loadings	225.508 (0.000)	246.002 (0.000)	143	0.995	0.034 (0.026; 0.041)	0.042	0.677

All residual correlations ≤ |0.10|.
[a] Models include the specification for correlated residuals between items "trouble relaxing" and "restlessness".
[b] Model at cycle 24 includes the additional specification for correlated residuals between items "cannot stop worrying" and "worrying too much".
[c] Models include the specifications for correlated residuals a) "trouble relaxing" and "restlessness" at each cycle and b)"cannot stop worrying" and "worrying too much" at cycle 24.

of the added value of a higher-order model such as the bifactor model. These findings replicate those of Stochl et al., who also reported highly correlated dimensions in their estimated two-factor solution (Stochl et al., 2022).

A key finding in this study is the evidence of partial strong measurement invariance across sex and full strong measurement invariance across language for the one-factor model, which is sufficient for whole sample as well as stratified analyses and allow the valid comparison of anxiety symptom severity across subgroups. This study in community-dwelling young Canadian adults complements past research who have demonstrated measurement invariance in other populations, including Canadian adolescents (Romano et al., 2022), in males and females seeking treatment for mental health disorders (Saunders et al., 2023), and in individuals with traumatic brain injury who came from six European linguistic groups, excluding French (Teymoori et al., 2020). Our findings of partial strong invariance across sex are in line with Borgogna et al. (2021), whom also demonstrated partial strong invariance across genders and sexual minority groups in a sample of young American adults (mean age: 23 years). On the other hand, Sriken et al. (2022) showed evidence of full strong invariance, while Moreno et al. (2019) and Saunders et al. (2023) showed evidence of residual/strict invariance of the GAD-7 (not tested in the current study), further demonstrating the invariance of the GAD-7 across sex and gender. This work deepens the knowledge surrounding the GAD-7 and should be appraised alongside existing theory, past and current studies, and its other demonstrated psychometric properties beyond measurement invariance.

The second key finding in this study is the evidence supporting the longitudinal invariance of the one-factor structure of the GAD-7. This indicates that anxiety symptoms represented the same underlying constructs across the time frame studied, (i.e., young adulthood), and that model deterioration with increasingly strict requirements was minimal. This is important for longitudinal studies employing the GAD-7 because longitudinal invariance enables the study of changes in summary scores over time. In the current study, participants were age 30 to 35 years. However, certain anxiety symptoms may be reported differently in older compared to younger adults. A study of 375 patients in a clinical sample reported that the older age group (i.e., $\geq$60 years) had higher levels of reported worries about community/world affairs and health of self while also exhibiting lower levels of worry about work and school, compared to the younger group (i.e., age 20–39 years). In addition, the younger age group had a higher reported rate of social worries (Correa and Brown, 2019). Future research should investigate the longitudinal invariance of the GAD-7 scale across wider age ranges over the lifespan, which may show higher variability. Finally, the findings on longitudinal measurement invariance provide complementary evidence that the GAD-7 scale responds well psychometrically to societal and population disturbances such as the COVID-19 pandemic.

Limitations of the study include that evidence on measurement invariance of the GAD-7 scale is restricted to young adults (between ages 30 and 35) and may not generalize to wider age ranges. NDIT participants were almost all born in Canada and reported higher education (63 % attended university) and employment (84 % were employed), than Canadian averages of census data on a similar age group (in 2021, 57.5 % of Canadians reported university-level education and 57.1 % were employed) (Statistics Canada, 2022). However, our sample included participants with a wide range of education and income and therefore the findings may be generalizable to a large segment of the Canadian population. In fact, one recent study on measurement invariance of the GAD-7 scale across ethnicity and sociodemographic groups showed promising evidence of psychometric stability across several population strata (Moreno et al., 2019).

5. Conclusion

This study found that the GAD-7 scale measuring anxiety symptom severity has a unidimensional latent structure, which is invariant over time and across sex and language (English and French). The observed stability in key parameters of the GAD-7 scale is important for epidemiological studies investigating differences in scores across subgroups and changes in scores over time, and for clinical practice.

CRediT authorship contribution statement

Teodora Riglea: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Robert J. Wellman:** Writing – review & editing, Methodology. **Marie-Pierre Sylvestre:** Writing – review & editing, Supervision, Funding acquisition. **Catherine Sabiston:** Writing – review & editing, Supervision. **Jennifer O'Loughlin:** Writing – review & editing, Supervision, Funding acquisition.

Role of the funding source

Teodora Riglea is funded by doctoral bursaries from the Fonds de recherche du Québec - Santé, Programme Ministère de l'Éducation et de l'Enseignement supérieur-Universités, and the Réseau de Recherche en Santé des Populations du Québec. Marie-Pierre Sylvestre is supported by an FRQS Junior 2 salary award. Jennifer O'Loughlin held a Canada Research Chair in the Early Determinants of Adult Chronic Disease 2004–2021. Funders were not involved in any step during the conduct of the research and/or preparation of the article.

Declaration of competing interest

None.

Acknowledgements

The authors would like to thank the NDIT participants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2025.01.117.

Data availability

NDIT data are available on request. Access to NDIT data is open to any university-appointed or affiliated investigator upon successful completion of the application process. Masters, doctoral and post-doctoral students may apply through their primary supervisor. To gain access, applicants must complete a data access form and return it to the principal investigator (jennifer.oloughlin@umontreal.ca). The procedure to obtain access to NDIT data is described in O'Loughlin, J., Dugas, E. N., Brunet, J., DiFranza, J., Engert, J. C., Gervais, A., Gray-Donald, K., Karp, I., Low, N. C., Sabiston, C., Sylvestre, M. P., Tyndale, R. F., Auger, N., Belanger, M., Barnett, T., Chaiton, M., Chenoweth, M. J., Constantin, E., Contreras, G., Kakinami, L., Labbe, A., Maximova, K., McMillan, E., O'Loughlin, E. K., Pabayo, R., Roy-Gagnon, M. H., Tremblay, M., Wellman, R. J., Hulst, A., Paradis, G., 2015. Cohort Profile: The Nicotine Dependence in Teens (NDIT) Study. Int J Epidemiol. 44(5), 1537-1546. doi: https://doi.org/10.1093/ije/dyu135.

References

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. American Psychiatric Association, Washington, DC.

Borgogna, N.C., Brenner, R.E., McDermott, R.C., 2021. Sexuality and gender invariance of the PHQ-9 and GAD-7: implications for 16 identity groups. J. Affect. Disord. 278, 122–130. https://doi.org/10.1016/j.jad.2020.09.069.

Brattmyr, M., Lindberg, M.S., Solem, S., Hjemdal, O., Havnen, A., 2022. Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder Scale-7 in a Norwegian

psychiatric outpatient sample. BMC Psychiatry 22, 461. https://doi.org/10.1186/s12888-022-04101-z.
Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct. Equ. Model. 9, 233–255.
Clark, D.A., Donnellan, M.B., 2021. Chapter 32 — What if apples become oranges? A primer on measurement invariance in repeated measures research. In: Rauthmann, J. F. (Ed.), The Handbook of Personality Dynamics and Processes. Academic Press, pp. 837–854.
Correa, J.K., Brown, T.A., 2019. Expression of generalized anxiety disorder across the lifespan. J. Psychopathol. Behav. Assess. 41, 53–59. https://doi.org/10.1007/s10862-018-9697-z.
De Man, J., Absetz, P., Sathish, T., Desloge, A., Haregu, T., Oldenburg, B., Johnson, L.C. M., Thankappan, K.R., Williams, E.D., 2021. Are the PHQ-9 and GAD-7 suitable for use in India? A psychometric analysis. Front. Psychol. 12. https://doi.org/10.3389/fpsyg.2021.676398.
Delamain, H., Buckman, J.E.J., Stott, J., John, A., Singh, S., Pilling, S., Saunders, R., 2024. Measurement invariance and differential item functioning of the PHQ-9 and GAD-7 between working age and older adults seeking treatment for common mental disorders. J. Affect. Disord. 347, 15–22. https://doi.org/10.1016/j.jad.2023.11.048.
Doi, S., Ito, M., Takebayashi, Y., Muramatsu, K., Horikoshi, M., 2018. Factorial validity and invariance of the 7-item Generalized Anxiety Disorder Scale (GAD-7) among populations with and without self-reported psychiatric diagnostic status. Front. Psychol. 9. https://doi.org/10.3389/fpsyg.2018.01741.
Eid, M., Geiser, C., Koch, T., Heene, M., 2017. Anomalous results in G-factor models: explanations and alternatives. Psychol. Methods 22, 541–562. https://doi.org/10.1037/met0000083.
Flora, D.B., 2020. Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. Adv. Methods Pract. Psychol. Sci. 3, 484–501. https://doi.org/10.1177/2515245920951747.
Flora, D.B., Curran, P.J., 2004. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol. Methods 9, 466–491. https://doi.org/10.1037/1082-989X.9.4.466.
Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res. 18, 39–50. https://doi.org/10.2307/3151312.
Hinz, A., Esser, P., Friedrich, M., Glaesmer, H., Mehnert-Theuerkauf, A., Schroeter, M.L., Petrowski, K., Toussaint, A., 2023. Changes in anxiety in the general population over a six-year period. PLoS One 18, e0291206. https://doi.org/10.1371/journal.pone.0291206.
Hu, L.T., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. Multidiscip. J. 6, 1–55. https://doi.org/10.1080/10705519909540118.
Kline, R.B., 2023. Principles and Practice of Structural Equation Modeling. Guilford Publications.
Liu, Y., Millsap, R., West, S., Tein, J.-Y., Tanaka, R., Grimm, K., 2017. Testing measurement invariance in longitudinal data with ordered-categorical measures. Psychol. Methods 22, 486–506. https://doi.org/10.1037/met0000075.
Mackinnon, S., Curtis, R., O'Connor, R., 2022. A tutorial in longitudinal measurement invariance and cross-lagged panel models using lavaan. Meta-Psychology. 6.
Micoulaud-Franchi, J.-A., Lagarde, S., Barkate, G., Dufournet, B., Besancon, C., Trébuchon-Da Fonseca, A., Gavaret, M., Bartolomei, F., Bonini, F., McGonigal, A., 2016. Rapid detection of generalized anxiety disorder and major depression in epilepsy: validation of the GAD-7 as a complementary tool to the NDDI-E in a French sample. Epilepsy Behav. 57, 211–216. https://doi.org/10.1016/j.yebeh.2016.02.015.
Mills, S.D., Fox, R.S., Malcarne, V.L., Roesch, S.C., Champagne, B.R., Sadler, G.R., 2014. The psychometric properties of the Generalized Anxiety Disorder-7 Scale in Hispanic Americans with English or Spanish language preference. Cult. Divers. Ethn. Minor. Psychol. 20, 463–468. https://doi.org/10.1037/a0036523.
Moreno, E., Muñoz-Navarro, R., Medrano, L.A., González-Blanch, C., Ruiz-Rodríguez, P., Limonero, J.T., Moretti, L.S., Cano-Vindel, A., Moriana, J.A., 2019. Factorial invariance of a computerized version of the GAD-7 across various demographic groups and over time in primary care patients. J. Affect. Disord. 252, 114–121. https://doi.org/10.1016/j.jad.2019.04.032.
O'Loughlin, J., Dugas, E.N., Brunet, J., DiFranza, J., Engert, J.C., Gervais, A., Gray-Donald, K., Karp, I., Low, N.C., Sabiston, C., Sylvestre, M.P., Tyndale, R.F., Auger, N., Auger, N., Mathieu, B., Tracie, B., Chaiton, M., Chenoweth, M.J., Constantin, E., Contreras, G., Kakinami, L., Labbe, A., Maximova, K., McMillan, E., O'Loughlin, E.K., Pabayo, R., Roy-Gagnon, M.H., Tremblay, M., Wellman, R.J., Hulst, A., Paradis, G., 2015. Cohort profile: the Nicotine Dependence In Teens (NDIT) study. Int. J. Epidemiol. 44, 1537–1546. https://doi.org/10.1093/ije/dyu135.
Ong, C.W., Pierce, B.G., Klein, K.P., Hudson, C.C., Beard, C., Björgvinsson, T., 2022. Longitudinal measurement invariance of the PHQ-9 and GAD-7. Assessment 29, 1901–1916. https://doi.org/10.1177/10731911211035833.
Perry, J.L., Nicholls, A.R., Clough, P.J., Crust, L., 2015. Assessing model fit: caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. Meas. Phys. Educ. Exerc. Sci. 19, 12–21. https://doi.org/10.1080/1091367X.2014.952370.
Romano, I., Ferro, M.A., Patte, K.A., Leatherdale, S.T., 2022. Measurement invariance of the GAD-7 and CESD-R-10 among adolescents in Canada. J. Pediatr. Psychol. 47, 585–594. https://doi.org/10.1093/jpepsy/jsab119.
Saunders, R., Moinian, D., Stott, J., Delamain, H., Naqvi, S.A., Singh, S., Wheatley, J., Pilling, S., Buckman, J.E.J., 2023. Measurement invariance of the PHQ-9 and GAD-7 across males and females seeking treatment for common mental health disorders. BMC Psychiatry 23, 298. https://doi.org/10.1186/s12888-023-04804-x.
Sawaya, H., Atoui, M., Hamadeh, A., Zeinoun, P., Nahas, Z., 2016. Adaptation and initial validation of the Patient Health Questionnaire — 9 (PHQ-9) and the Generalized Anxiety Disorder — 7 Questionnaire (GAD-7) in an Arabic speaking Lebanese psychiatric outpatient sample. Psychiatry Res. 239, 245–252. https://doi.org/10.1016/j.psychres.2016.03.030.
Shevlin, M., Butter, S., McBride, O., Murphy, J., Gibson-Miller, J., Hartman, T.K., Levita, L., Mason, L., Martinez, A.P., McKay, R., Stocks, T.V.A., Bennett, K.M., Hyland, P., Vallieres, F., Valiente, C., Vazquez, C., Contreras, A., Peinado, V., Trucharte, A., Bertamini, M., Panzeri, A., Bruno, G., Granziol, U., Mignemi, G., Spoto, A., Vidotto, G., Bentall, R.P., 2022. Measurement invariance of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) across four European countries during the COVID-19 pandemic. BMC Psychiatry 22, 154. https://doi.org/10.1186/s12888-022-03787-5.
Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092.
Sriken, J., Johnsen, S.T., Smith, H., Sherman, M.F., Erford, B.T., 2022. Testing the factorial validity and measurement invariance of college student scores on the Generalized Anxiety Disorder (GAD-7) scale across gender and race. Meas. Eval. Couns. Dev. 55, 1–16. https://doi.org/10.1080/07481756.2021.1902239.
Statistics Canada, 2022. Census profile. In: Canada, S. (Ed.), 2021 Census of Population.
Statistics Canada, 2023. Survey on COVID-19 and Mental Health, February to May 2023.
Stochl, J., Fried, E.I., Fritz, J., Croudace, T.J., Russo, D.A., Knight, C., Jones, P.B., Perez, J., 2022. On dimensionality, measurement invariance, and suitability of sum scores for the PHQ-9 and the GAD-7. Assessment 29, 355–366. https://doi.org/10.1177/1073191120976863.
Svetina, D., Rutkowski, L., Rutkowski, D., 2020. Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using M plus and the lavaan/semtools packages. Struct. Equ. Model. Multidiscip. J. 27, 111–130.
Sylvestre, M.-P., Tchouangue Dinkou, G.D., Armasu, A., Pelekanakis, A., Plourde, V., Bélanger, M., Maximova, K., Thombs, B.D., O'Loughlin, J., 2022. Symptoms of depression and anxiety increased marginally from before to during the COVID-19 pandemic among young adults in Canada. Sci. Rep. 12, 16033. https://doi.org/10.1038/s41598-022-20379-1.
Teymoori, A., Real, R., Gorbunova, A., Haghish, E.F., Andelic, N., Wilson, L., Asendorf, T., Menon, D., von Steinbüchel, N., 2020. Measurement invariance of assessments of depression (PHQ-9) and anxiety (GAD-7) across sex, strata and linguistic backgrounds in a European-wide sample of patients after Traumatic Brain Injury. J. Affect. Disord. 262, 278–285. https://doi.org/10.1016/j.jad.2019.10.035.
Watkins-Martin, K., Orri, M., Pennestri, M.-H., Castellanos-Ryan, N., Larose, S., Gouin, J.-P., Ouellet-Morin, I., Chadi, N., Philippe, F., Boivin, M., Tremblay, R.E., Côté, S., Geoffroy, M.-C., 2021. Depression and anxiety symptoms in young adults before and during the COVID-19 pandemic: evidence from a Canadian population-based cohort. Ann. General Psychiatry 20, 42. https://doi.org/10.1186/s12991-021-00362-2.
Widaman, K.F., Ferrer, E., Conger, R.D., 2010. Factorial invariance within longitudinal structural equation models: measuring the same construct across time. Child Dev. Perspect. 4, 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x.
Wu, H., Estabrook, R., 2016. Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. Psychometrika 81, 1014–1045. https://doi.org/10.1007/s11336-016-9506-0.
Xia, Y., Yang, Y., 2019. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. Behav. Res. Methods 51, 409–428. https://doi.org/10.3758/s13428-018-1055-2.