



A solution for factorial validity testing of three-item scales: An example of tau-equivalent strict measurement invariance of three-item loneliness scale

Stanisław K. Czerwiński¹ · Paweł Andrzej Atroszko¹

Accepted: 26 February 2021
© The Author(s) 2021

Abstract

Ultra-short scales are increasingly popular in surveys. Congeneric model fit of a three-item scale cannot be tested with Confirmatory Factor Analysis (CFA) without additional assumptions because the number of degrees of freedom is equal to zero. A more rigorous tau-equivalent model, assuming equality of factor loadings can be tested instead. The objective of this study was to demonstrate this approach with an example of the psychometric study of the Polish version of the Three-Item Loneliness Scale (TILS), and to discuss the arising problems and possible solutions. There seems to be a high need for such analysis because currently, some properties of CFA make it an approach still predominant over Item Response Theory (IRT) models in the quality of life research. A sample of 3510 students completed TILS together with the questionnaires measuring a variety of indicators of well-being. The results provided evidence for a good fit of a tau-equivalent model. Furthermore, multi-group CFAs provided support for strict measurement invariance of this model. To the Authors' knowledge, it is the first practical application of a tau-equivalent model to testing the factorial validity of an ultra-short scale and probably the first empirical case of tau-equivalent measurement invariance in psychological literature in general. TILS showed good criterion validity and satisfactory reliability. Unidimensionality of three-item scales can be examined with a tau-equivalent model that has some favorable psychometric properties. However, it might be exceedingly restrictive in certain practical cases. When developing a new short scale, it is recommended to maintain at least four items.

Keywords Loneliness · Measurement invariance · Quality of life · Scale · Tau-equivalence · Validity

Introduction

Loneliness

Loneliness is defined by de Jong Gierveld as 'a situation experienced by a participant as one where there is an unpleasant or unacceptable lack of (quality of) certain social relationships. The extent to which the situation is experienced as serious depends upon the participant's perception of his or her ability to realize new relationships, or to improve existing ones.' (de Jong Gierveld, 1989). It is important to emphasize

that loneliness is a subjective feeling and to distinguish it from social isolation, as they are two separate constructs (Coyle & Dugan, 2012; Perissinotto & Covinsky, 2014), for one can be in the company of others and still feel alone or live in seclusion and have no negative feelings about it whatsoever.

A large number of studies concerning loneliness focus on the elderly population (Holt-Lunstad, Smith, Baker, Harris, & Stephenson, 2015) whilst neglecting younger samples, even though the relationship between age and loneliness has been found to be U-shaped (Pinquart & Sorensen, 2001). Loneliness is not a problem specific only to the elderly. Studies show that younger generations are showing less concern for other people and gravitate away from civic orientation (Twenge, Campbell, & Freeman, 2012), which in turn could explain their increase in loneliness (Cigna U.S. Loneliness Index, 2018; Twenge, Spitzberg, & Campbell, 2019). With social support being one of the fundamental factors in human well-being (Cohen & Wills, 1985), it is not surprising that each successive generation shows more and more signs of mental problems (Twenge et al., 2010). Poland could be exceptionally vulnerable to this effect. For example, it is one of

✉ Paweł Andrzej Atroszko
p.atroszko@ug.edu.pl

Stanisław K. Czerwiński
stanislaw.czerwinski96@gmail.com

¹ Institute of Psychology, University of Gdańsk, Bażyńskiego 4,
80-309 Gdańsk, Poland

the few countries in Europe to have increasing suicide rates (Höfer, Rockett, Vämik, Etzersdorfer, & Kapusta, 2012). The rapid social change in post-communist countries is among the postulated causes for a decrease in well-being (Domagała-Krecioch & Majerek, 2014; Höfer et al., 2012). Poland as a country that transitioned to the market economy and is still in the process of adapting the Western “way of life” is a compelling case for the process of increasing alienation among youth (Zoutewelle-Terovan & Liefbroer, 2017). The existing data seems to support the notion that dynamic economic growth comes with a price. For example, compulsive overworking prevalence evaluated with the same measure and cut-off score seems to be 2 to 3 times higher in Poland than in countries with established and stable economies such as Denmark and Norway (Andreassen, Nielsen, Pallesen, & Gjerstad, 2019; Atroszko, Pallesen, Griffiths, & Andreassen, 2017; Lichtenstein, Malkenes, Sibbersen, & Hinze, 2019), and this trend is already visible among undergraduate students (Lawendowski, Bereznowski, Wróbel, Kierzkowski, & Atroszko, 2019). Having that in mind, further investigation of loneliness in young adults, especially in rapidly economically developing countries, is of great importance. Our study focused on the undergraduate students because they comprise almost half of the population age 19–24 in Poland (Główny Urząd Statystyczny, 2019) and previous studies showed that student populations are highly vulnerable to loneliness (American College Health Association, 2016). Polish studies showed high prevalence of depression and considerable hopelessness among university students (Czerwiński, Mackiewicz, Mytlewska, & Atroszko, 2020; Koryczan, Piotrowski, Roj, Czerwiński, & Atroszko, 2020).

The previous studies showed significant relationships of loneliness with a wide range of well-being indicators, such as health (Swami et al., 2007; Tobiasz-Adamczyk et al., 2017), sleep quality (Matthews et al., 2017; Yu, Steptoe, Niu, Ku, & Chen, 2017), self-esteem (Heinrich & Gullone, 2006; Vanhalst, Luyckx, Scholte, Engels, & Goossens, 2013), satisfaction with personal relationships (Mellor, Stokes, Firth, Hayashi, & Cummins, 2008), social support (Segrin & Passalacqua, 2010; Tobiasz-Adamczyk et al., 2017), general quality of life (Fanakidou et al., 2017), satisfaction with life (Buelga, Musitu, Murgui, & Pons, 2008; Huo & Kong, 2014; Liu & Guo, 2008), stress (Hughes, Waite, Hawkey, & Cacioppo, 2004; Lee & Goldstein, 2015), anxiety (Fanakidou et al., 2017; Moore & Schultz, 1983), and depression (Cacioppo, Hughes, Waite, Hawkey, & Thisted, 2006; Erzen & Çikrikci, 2018).

Issues in Factorial Validity Testing of Three-Item Scales

Ultra-short scales are becoming increasingly popular in educational and psychological research due to the convenience of

application, often satisfactory psychometric properties and reduction of bias introduced by the excessive burden on participants with long questionnaires. They are often used, for example, in the quality of life research (Cheung & Lucas, 2014), health psychology, including clinical settings (Krebs et al., 2009), and epidemiology (Beutel et al., 2017) due to the convenience of their use. Since loneliness is a variable strictly related to the well-being (Erzen & Çikrikci, 2018; Rico-Uribe et al., 2016; Valtorta, Kanaan, Gilbody, Ronzi, & Hanratty, 2016; Vanderweele, Hawkey, & Cacioppo, 2012), we will discuss the psychometric issues related to the ultra-brief scales within the context of quality of life research and health psychology. Even though in the case of individual diagnosis a smaller pool of items comes with a significant loss in precision, when conducting a large scale surveys focused on the relationships between numerous variables and controlling for a wide range of covariates, more concise instruments usually perform almost as good as their lengthier alternatives (Gogol et al., 2014; Kemper, Trapp, Kathmann, Samuel, & Ziegler, 2018; Rammstedt & Beierlein, 2014), including having good predictive validity (Bergkvist & Rossiter, 2007). In such case, short, valid and reliable tools are invaluable. In consequence, a proper analytical approach to substantiate their adequate psychometric properties is crucial and poses specific challenges. For example, the reliability of single-item measures cannot be substantiated with internal consistency coefficients, and may require test-retest coefficients such as Intraclass Correlation Coefficient (ICC). It should be emphasized that the focus of this paper is on the specific circumstances in which the already existing ultra-short scale is being validated in a new sample, such as in different culture or different demographic. The measure is not used for precise diagnosis of individuals but to investigate the relationships with other variables or as a means to control for a confounding variable, for example, as in Health and Retirement Study (Chen & Feeley, 2014). Furthermore, such a measure can be particularly valuable if it was used extensively in previous studies and showed good psychometric properties, such as in the case of the Three-Item Loneliness Scale (TILS). This is of the highest importance from the perspective of the direct comparisons of results and emphasis laid on the reproducibility and replicability of effects (Open Science Collaboration, 2015; Pashler & Wagenmakers, 2012; Patil, Peng, & Leek, 2016; Plesser, 2018). In this context, a specific class of ultra-short measures is three-item scales which are not infrequently used in the quality of life research (Cuijpers, Smits, Donker, Ten Have, & de Graaf, 2009; Jenkins, Stanton, Niemczyk, & Rose, 1988; Kelly, 2004; Krebs et al., 2009; Leon, Olfson, Portera, Farber, & Sheehan, 1997).

The standard psychometric approach is the latent variable model (Borsboom, Mellenbergh, & Van Heerden, 2003; Borsboom, Molenaar, & Wright, 2015; Muthén, 1984; Smith, 2000). It is a statistical model that relates a set of observable/

manifest variables to one or more latent variables. In such a model, loneliness is assumed to be a latent continuous variable that is measured by either continuous or categorical manifest variables. One of the first steps in the validation process aimed at substantiating construct validity is demonstrating that the items measure the same construct. This is most often performed with factor analysis or methods more suitable for nonlinear data often used for Rasch models and Item Response Theory (IRT) models, such as nonlinear factor analysis or multidimensional scaling (De Ayala & Hertzog, 1991), or item fit approach (Tay & Drasgow, 2012), and other methods with specific approach depending on the manifest variable measurement scale and linearity/nonlinearity of the relationship between item score and latent trait (Smith, 2000; Stochl, Jones, & Croudace, 2012). An overview of the proposed procedures for dimensionality testing of a set of item responses can be found elsewhere (De Champlain & Gessaroli, 1998). In some cases, including ordered categorical data, IRT models are equivalent to factor analysis (Bartholomew, 1987; Birnbaum, 1968; Samejima, 1970; Takane & De Leeuw, 1987), and both models allow for measurement invariance testing (Kim & Yoon, 2011; Muthén & Asparouhov, 2014). The choice of a specific psychometric approach is determined by multiple factors (Pettilo, Cano, McLeod, & Coon, 2015), and there are still no explicit guidelines of which solution is preferable in all possible circumstances (Coulacoglou & Saklofske, 2017). The detailed discussion of the technicalities behind particular approaches as well as philosophical underpinnings behind the latent variable model exceeds the scope of this paper and can be found elsewhere (Borsboom et al., 2003; Coulacoglou & Saklofske, 2017). While acknowledging the ontological doubts regarding the status of latent variables, this paper focuses on the solutions related to confirmatory factor analytic (CFA) approach (sometimes called “restricted factor analysis” (McDonald, 2013)) which is one of the most widely used statistical approaches in the quality of life research, and in social and behavioral sciences in general (Depaoli, Tiemensma, & Felt, 2018).

The advantages of this approach include its congruence with modern scientific standards based on hypothesis testing and confirmatory approach (Gorsuch, 1983; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012), relative ease of practical application for quality of life researchers non-specializing in advanced statistics, and suitability for particular practical aims such as using a measure to enable investigating the relationships between variables and controlling for covariates in large scale surveys. Studies show that for specific purposes such as measurement invariance testing, CFA results are to a large extent similar to Rasch measurement and can be treated as a good approximation of a more comprehensive approach (Salzberger & Sinkovics, 2006). These properties of CFA make it an approach overwhelmingly predominant over IRT models in the quality of life research (Depaoli et al., 2018). Possible explanations for why IRT approaches

are rarely used in health-related studies include them being perceived as challenging to implement and interpret or requiring large sample sizes that are oftentimes difficult or impossible to obtain, especially in clinical populations (Salzberger & Sinkovics, 2006).

Within the CFA approach, there are different models used for testing the structure of a measure depending on the assumptions underlying the measurement. The most restrictive *parallel model* requires that all items must measure the same latent variable, on the same scale, with the same degree of precision, and with the same amount of error (Raykov, 1997a, 1997b). The *tau-equivalent model* allows individual item error variances to differ from one another. The *essentially tau-equivalent model* further allows for different degrees of precision of measurement. The least restrictive and the most general *congeneric model* assumes that each item measures the same latent variable, with possibly different scales, different degrees of precision, and with different amounts of error. Even though commonly used Cronbach’s alpha coefficient requires at least a tau equivalent model, most of the studies use the congeneric model due to its least restrictive nature, and consequently report biased estimates of internal consistency with Cronbach’s alpha. For more up to date and approachable description of the models, please see for example paper by Graham (2006).

In the case of single-factor three-item scales, the congeneric measurement model is just-identified, has zero degrees of freedom, and thus its fit to the data cannot be meaningfully tested. This problem can be solved by assuming a more rigorous tau-equivalent model which constrains all loadings to be held equal (Graham, 2006). Thus, it yields a non-zero number of degrees of freedom and makes it possible to test the model’s goodness of fit. With three items, high and equal loadings provide favorable psychometric properties in terms of factorial validity and reliability, and therefore improve other types of validity such as criterion validity (especially when sums of items instead of latent factors are used in the subsequent analyses). To the Authors’ knowledge, tau-equivalence has never been tested before in a practical setting due to strict and often empirically unfeasible assumptions that it takes. Furthermore, testing measurement invariance of a tau-equivalent model in most cases would yield unsatisfactory results due to its extremely rigorous assumptions. Therefore, while it seems empirically feasible to develop ultra-short tau-equivalent scales, the subsequent invariance testing of such scales across a variety of groups could bring more problems than benefits.

There are few other possible solutions to the problem of the factorial validity of three-item scales, such as using Rasch or IRT model, analyzing a model with more items and assuming unidimensionality of the shorter version if it was confirmed for the longer version, and analyzing scale together with other scales in a multifactor model. However, when taking into account specific circumstances of interest, i.e., validating

already existing three-item measure used for investigating associations with other variables, these solutions have some disadvantages in comparison to testing a tau-equivalent model.

The first alternative could be a polytomous Rasch model or a more flexible graded response model (Baker, Rounds, & Zevon, 2000; Samejima, 1970; Tuerlinckx & Wang, 2004). These approaches are commonly used to analyze the performance of particular items with multiple category responses without guessing. As it was already mentioned, these approaches were shown to be equivalent to CFA; however, they provide more sophisticated information (Jöreskog & Moustaki, 2001; Kamata & Bauer, 2008; Takane & De Leeuw, 1987). While arguably they could be used to investigate the psychometric performance of a three-item scale, they require more expertise with psychometrics, larger sample sizes, and preferably more items (De Champlain & Gessaroli, 1998). These analyses represent a feasible, challenging, perhaps fascinating but conceivably disproportionately overcomplicated approach in the discussed context. Studies show that for specific purposes such as measurement invariance testing (Kim & Yoon, 2011; Meade & Lautenschlager, 2004; Muthén & Asparouhov, 2014; Samejima, 1970; Tay, Meade, & Cao, 2015), CFA results are to a large extent similar to Rasch measurement and can be treated as a good approximation of a more comprehensive approach (Salzberger & Sinkovics, 2006).

A different possible solution to the problem of three-item scales having zero degrees of freedom could be testing them alongside another measure in a multi-factor solution. This, however, brings a few problems. Firstly, the other scale(s) used in such a model should be very carefully chosen. It should be preferably more than three-item so that its fit could be independently examined. Secondly, it should measure a somewhat dissimilar construct in order to minimize any overlapping, which could affect the fit of the multifactor model. However, had such a model failed to show satisfactory fit, sometimes it could be challenging to identify the exact source of problems as in CFA models, problems with fit tend to propagate throughout the model. For example, residuals of items pertaining to one scale could have substantial covariance with residuals of items of another scale. This would require allowing for correlated error terms in order to obtain acceptable fit, which equals to imposing arbitrary ad hoc assumptions on the model. Additionally, subsequent studies would have to use the same scales to allow for comparing the results, and sampling variability could influence the results on any of the measures making it virtually impossible to conduct meaningfully such comparisons.

Another alternative could be the analysis of a congeneric model with more items, e.g., a full 20-item Revised UCLA Loneliness Scale (R-UCLA; Russell, Peplau, & Cutrona, 1980). If the unidimensionality of the full scale showed a good fit to the data, then it could be assumed that the shorter version

of the scale also shows a good fit for a single factor solution. However, this approach still does not provide an exact measure of fit for the short three-item version of the scale. Furthermore, if only the three-item version is used in the subsequent studies, the fit of the data in a new sample would be unknown. This approach is related to another issue, i.e., potential multidimensionality of loneliness. One could argue that if a longer version of the scale is multidimensional and each item of the shorter version is taken from different factors, then the shorter measure is not unidimensional. However, there are two issues that need to be taken into account in this situation.

Firstly, even multidimensional measures of loneliness need to assume one general construct of loneliness. Otherwise, there would be three different but interrelated constructs. That would mean that there is no such phenomenon as loneliness but multiplicity of different types of loneliness, which poses practical and ontological problems. Therefore, a multidimensional measure of loneliness would have to assume a single general factor with several first-order factors measuring specific components of loneliness (Salzberger & Sinkovics, 2006). The correlated factors of loneliness most often stem from the inadequate analytical approach such as exploratory factor analysis (Mahon & Yarcheski, 1990; Mahon, Yarcheski, & Yarcheski, 1995; Wilson, Cutts, Lees, Mapungwana, & Maunganidze, 1992) or principal component analysis (Joiner, Catanzaro, Rudd, & Rajab, 1999) without cross-validation with CFA or CFA without testing model with higher-order factor (Goossens et al., 2009; Goossens & Beyers, 2002; Maes, Klimstra, Noortgate, & Goossens, 2014; Penning, Liu, & Chou, 2013). This is incongruent with the assumptions of hypothesis testing underlying the modern scientific method (Wagenmakers et al., 2012). Very few studies on loneliness investigate hierarchical structures, and when they do, it still does not conform to the strict requirements of confirmatory approach (Joiner et al., 1999). What follows from the assumption of one multidimensional construct of loneliness is that even three items, each measuring a different component of loneliness, still measure one general construct of loneliness. However, whether these items would show unidimensionality or not is another fascinating and complex psychometric problem which is yet to be systematically approached in simulation studies.

Secondly, a more technical problem with studies yielding multidimensional models of loneliness is that, to some extent, the multidimensionality seems to be an artifact of the item wording. TILS is a shorter version of R-UCLA (Hughes et al., 2004), one of the most commonly used measures of loneliness. R-UCLA originally consisted of 20 items and was assumed to have a single factor solution (Russell et al., 1980). Later, some studies provided evidence that two or three-factor solutions have a better fit for the instrument, with

the most popular one being a three-factor model of “intimate others”, “social others” and “belonging and affiliation” (Austin, 1983; Kwiatkowska, Rogoza, & Kwiatkowska, 2017; McWhirter, 1990). However, it was argued that the multifactor solutions are, to some extent, an artifact of reversed item wording, and with the factors being very strongly correlated, the scale is measuring a singular construct of loneliness (Hartshorne, 1993; Russell, 1996). It is worth noting that in the case of another widely used measure of loneliness, De Jong Gierveld Loneliness Scale (de Jong Gierveld & Kamphuis, 1985), commonly used as a two-dimensional instrument with one dimension being comprised of only positively worded items and another with only negatively worded items, a bifactor analyses provided evidence that all items were better represented by one general factor of loneliness (Grygiel, Humenny, & Rebisz, 2016; Grygiel, Humenny, Rebisz, Świtaj, & Sikorska, 2013). The factorial artifacts produced by reverse coding of items seem to be a common problem in the quality of life research, and more attention should be devoted to making researchers more aware of it since many of the most widely used tools face this problem (Røysamb & Strype, 2002; Salerno, Ingoglia, & Coco, 2017). More information on analytical approaches to bifactor models can be found elsewhere (Chen, West, & Sousa, 2006; Tay & Drasgow, 2012; Wang, Chen, & Jin, 2014). Regardless of dimensionality controversies surrounding R-UCLA, the three items forming TILS are all negatively worded and derive from the same factor.

In order to meaningfully compare results obtained in different groups, measurement invariance of a scale should be demonstrated. The most commonly used method to test factorial invariance of a measure is based in CFA (Marsh, 1987), which is to a large extent, equivalent to the methods used in Rasch and IRT models (Salzberger & Sinkovics, 2006). This allows testing the equivalence of a scale across multiple groups. A more detailed description of the meaning and statistics behind testing invariance can be found elsewhere (Putnick & Bornstein, 2016).

Present Study

The present study examined the psychometric properties of the Polish version of TILS in the student sample, including tau-equivalent measurement invariance between genders. The scale showed both concurrent and discriminant validity in previous studies, and satisfactory reliability, however, its factorial validity has not been effectively tested before, as the researchers mostly focused on criterion validity, tested measurement invariance without testing the model fit on its own or used exploratory factor analysis (Hawkey, Duvoisin, Ackva, Murdoch, & Luhmann, 2015; Hughes et al., 2004; Matthews-Ewald & Zullig, 2013).

Materials and Methods

Participants

The sample consisted of 3510 students from nine Polish universities, of which 1970 (56.1%) were female, and 1506 (42.9%) were male (34 respondents did not specify their gender). Participants' mean age was 20.92 years ($SD = 2.65$). Thirty eight participants took part in test-retest procedure with three-week interval between measurements, 31 females and 7 males, with mean age of 20.14 years ($SD = 1.31$). Convenience sampling was used; however, in principle, it was aimed at assuring diversity of students to some extent representing the population of undergraduate students in Poland. Therefore, the sample included most of the types of universities (e.g., technological, business schools, humanistic, sport academy), both public and private, variety of faculties and courses of study from each university, both full-time and part-time students, and students of all years of study.

Measures

The measures used were chosen due to their brevity, convenience of application, good psychometric properties and a need to reduce bias due to long survey fulfillment time typical for lengthier tools. All these measures are or are based on a widely recognized and commonly applied questionnaires.

Three-Item Loneliness Scale (TILS) consists of three items: “How often do you feel that you lack companionship?”, “How often do you feel left out?” and “How often do you feel isolated from others?”. The response options for each item were “hardly ever,” “some of the time” or “often.” The previous studies suggest that it has good validity and reliability, and measurement invariance in different countries, however, inadequate psychometric models applied in those studies need to be taken into account (Hawkey et al., 2015; Matthews-Ewald & Zullig, 2013). In the present sample, the Cronbach's alpha reliability coefficient was .80.

General health, sleep quality, global self-esteem, satisfaction with personal relationships, satisfaction with support received from friends, general quality of life, satisfaction with life and meaning in life were measured with single-item measures with nine-point response scales developed on the basis of items from WHOQOL-BREF scale (Skevington, Lotfy, & O'Connell, 2004). The scales showed good validity and reliability in previous research (Atroszko, Bagińska, Mokosińska, Sawicki, & Atroszko, 2015; Atroszko, Pianka, Raczyńska, & Atroszko, 2015; Atroszko, Sawicki, Mąkinia, & Atroszko, 2017; Atroszko, Sawicki, Sendal, & Atroszko, 2017). In the previous studies, the reported ICC for test-retest reliability were .72 for general health, .81 for sleep quality, .79 for global self-esteem, .80 for satisfaction with personal relationships, .64 for support received from friends, .86 for

general quality of life, .88 for satisfaction with life and .86 for meaning in life.

Perceived stress was measured with the Perceived Stress Scale-4 (PSS-4; Cohen, Kamarck, & Mermelstein, 1983). The PSS-4 consists of 4 items, a 5-point Likert response format scale, rated from 0 – “Never” to 4 – “Very often.” The Polish version of the scale showed good validity and reliability in previous research (Atroszko, 2015). In the present sample, the Cronbach’s alpha reliability coefficient was .72.

Anxiety and depression were measured by the Hospital Anxiety and Depression Scale (HADS), which includes 14 items with a 4-point response format (Zigmond & Snaith, 1983). Seven items measure anxiety, and seven measure depression. The Polish version of the scale showed good validity and reliability in previous research (Czerwiński et al., 2020). In the present sample, the Cronbach’s alpha reliability coefficient was .74 for depression, and .85 for anxiety.

Procedure

Convenience sampling was used. Those willing to participate (more than 95%) filled in ‘paper and pencil’ anonymous questionnaires during regular university classes. There were no specific inclusion and exclusion criteria. No monetary reward was provided as an incentive to complete the survey. Attaining formal and written informed consent was not regarded as necessary as voluntary completion of the questionnaires was regarded as providing consent. Data were gathered between 2013 and 2017 as a part of subsequent research projects concerning survey studies on behavioral addictions among undergraduate students in Poland. The invitation to the study stated that it concerns psychosocial functioning of students, including wide range of variables such as personality and well-being. The order of presentation of the measures was changed a few times during the course of the projects in order to minimize the potential bias related to questions’ order. Data was gathered both by the principal investigator and research assistants.

The scale was translated from English to Polish in a multi-step translation process conforming to the commonly used standards of psychometric instruments translation. The process included the following procedures: i) translation from English into Polish separately by one bilingual person and one psychologist fluent in English, ii) developing an agreement on the initial Polish version within a panel consisting of both translators and a psychometrician, iii) back translation by two different translators: a bilingual person and a psychologist fluent in English, iv) comparing back translation with the original version and with initial Polish translation within a panel consisting of all four translators and a psychometrician, and choosing item wording for the final Polish version, v) pre testing among a group of individuals ($n = 15$) for any problems with understanding the items and their intended meaning, and introducing any necessary corrections to items’ wording.

Statistical Analyses

Missing data (less than 1.5%) were imputed using expectation maximization (EM) algorithm, which provides unbiased estimates of parameters and improves statistical power of analyses (Enders, 2001; Scheffer, 2002). Confirmatory factor analyses were performed using Mplus 6.11. Due to the strictly ordinal character of the response scale, the CFA models were tested using Weighted Least Square Mean and Variance Adjusted (WLSMV) estimator. Due to the number of items, the standard congeneric model had zero degrees of freedom and could not yield meaningful results on the model fit to data. A tau-equivalent model was tested instead, as it increases the number of degrees of freedom to two. Tau-equivalence means that all items in the model load onto one factor equally. The basic model was initially tested separately in two groups of females and males. The following measures were used to evaluate the fit of the model: χ^2 divided by degrees of freedom (χ^2/df), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Squared Error of Approximation (RMSEA). CFI and TLI values above .95 are indicative of good fit (Hu & Bentler, 1999), however values above .90 are considered as adequate (Kline, 2004). RMSEA scores of .08 and lower are acceptable (Brown & Cudeck, 1993). Measurement invariance between genders was assessed using multiple-group procedures in which sets of parameters were freed sequentially in a series of hierarchically nested models. These models were also tau-equivalent. Configural invariance means testing the equivalence of factor structure between groups in the first step. Metric and scalar invariance could not be tested due to the structure of the response scale (three items with a three-point response format). Metric invariance assumes the same factor loadings between the groups, and scalar invariance assumes the same thresholds for items between groups. However, for ordered ternary data, any set of parameters with configural invariance can be transformed to satisfy the invariance of thresholds, of loadings and intercepts, or of intercepts and unique variances. As a result, there is no degree of freedom to test these invariance conditions (Wu & Estabrook, 2016). In consequence, metric and scalar invariance were not tested as, in this case, they are equivalent to configural invariance. As a result, only residual invariance can be meaningfully compared to the previous step. Strict invariance is the most rigorous model which tests whether the construct, item loadings, item thresholds, and residual variances are all the same in both groups. ICC along with the 95% confidence interval (CI) was used as a measure of test-retest reliability. Means, standard deviations, percentages, skewness and kurtosis of scales, Cronbach’s α for reliability, ICC and Pearson correlation coefficients were calculated using IBM SPSS 25.

Results

Descriptive data on TILS is presented in Table 1. An ICC for TILS was .84 (95% CI = .67–.92, $p < .001$). It indicates good test-retest reliability.

In the female group a tau-equivalent one-factor model showed following fit indices: $\chi^2/df = 10.44$, CFI = .997, TLI = .995, RMSEA = .070 (90% CI = .045–.099), with each item loading of .86. In the male group, model fit indices were: $\chi^2/df = 8.61$, CFI = .996, TLI = .993, RMSEA = .071 (90% CI = .043–.104), and with each item loading of .84. Table 2 shows the model fit for each successively stringent test of invariance. The indicators of a model fit showed acceptable strict measurement invariance of a tau-equivalent model.

Means, standard deviations, percentages, skewness, kurtosis and correlations between loneliness and other measured variables are presented in Table 3. All correlations between TILS and criterion variables were as expected. The scales did not show extreme skewness nor kurtosis.

Discussion

The aim of the study was to address the problem of testing model fit in scales that consist of three items and to validate the Polish version of TILS in a student sample using the suggested psychometric approach. Factorial validity was supported by a good fit to the data of the tau-equivalent factorial model of TILS, and strict measurement invariance between genders using a very stringent tau-equivalent model. Cronbach’s alpha provided evidence for good reliability of the tool. The scale showed good concurrent validity. The results confirmed the expected correlations with the indicators of subjective well-being, perceived stress, anxiety, and depression, consistent with previous data showing worse psychosocial functioning of lonely individuals (Buelga et al., 2008; Cacioppo et al., 2006; Fanakidou et al., 2017; Heinrich & Gullone, 2006; Hughes et al., 2004; Liu & Guo, 2008; Matthews et al., 2017; Mellor et al., 2008; Moore & Schultz, 1983; Segrin & Passalacqua, 2010; Swami, et al., 2006; Tobiasz-Adamczyk et al., 2017; Yu et al., 2017). Lack of association with

gender suggests adequate divergent validity. These, together with very good factorial structure and measurement invariance, support the construct validity of the scale. Therefore, the scale presents significant advantages for large scale epidemiological studies, especially in the population of university students. Alongside other advancements in the measurement of loneliness (Auné, Abal, & Attorresi, 2020), including among the undergraduate students (Caballer et al., 2020), this study constitutes a significant contribution to the field.

Loneliness is an increasing problem among young generations (Cigna U.S. Loneliness Index, 2018; Twenge et al., 2019). More than 10% of the respondents in the current study answered that they often lack companionship, feel left out, or feel isolated from others. More than 56% of the students answered that some of the time or often they lack companionship, more than 43% feel left out, and 40% feel isolated from others. The correlations between loneliness and depression, anxiety, stress, self-esteem, general quality of life, satisfaction with life, and meaning in life ranged from .37 to .46 (absolute values). This means that there may be around 40% more students with considerably decreased well-being among lonely individuals (Rosenthal, 2005). With about 1.5 million students in Poland, and 10% being lonely, it would mean 60,000 students with anxiety and depression related to loneliness. To some extent it might explain such phenomena in Poland as the so called “crisis of child and adolescent psychiatry”, increasing suicide rates among underaged, and emotional and behavioral problems among affluent youth (see Atroszko & Atroszko, 2020). While these are crude and preliminary estimates, more studies on representative samples should analyze this problem in-depth, and provide more precise estimates.

Strengths and Limitations

In terms of limitations, a convenience sample was used. Therefore, the results of the present study cannot be generalized to other populations without some reservations. The absence of a clinical evaluation of participants prevented any analyses and conclusions on the relationship between symptoms of disorders and results of the survey. Furthermore, all

Table 1 Response rates of each of the TILS items

	Entire sample			Females			Males		
	hardly ever (1)	some of the time (2)	often (3)	hardly ever (1)	some of the time (2)	often (3)	hardly ever (1)	some of the time (2)	often (3)
How often do you feel that you lack companionship?	43.6%	46.1%	10.1%	42.7%	47.4%	9.8%	44.6%	44.4%	11.0%
How often do you feel left out?	56.9%	31.5%	11.6%	55.3%	33.6%	11.2%	58.7%	29.0%	12.1%
How often do you feel isolated from others?	60.0%	29.1%	10.8%	60.4%	28.8%	10.7%	59.6%	29.5%	10.9%

Table 2 Measurement invariance between genders

Model	χ^2	<i>df</i>	CFI	Δ CFI	RMSEA	90% CI	Δ RMSEA
Tau-equivalent configural invariance	39.716	4	.996	–	.073	[.053, .094]	–
Tau-equivalent residual invariance	64.270	9	.994	–0.002	.060	[.047, .074]	–0.013

data were self-reported and are therefore prone to weaknesses of such data (such as social desirability bias, recall biases, common method bias, etc.). Metric and scalar invariance could not be tested due to the structure of the response scale (three-point response format). Instead, the configural model was compared directly to the strict model. This should be kept in mind if scales with more response options are tested. In such cases, metric and scalar invariance could be tested separately. Also, no other validated loneliness measures were used for testing the convergent validity of the instrument. A more detailed comparison of information from CFA and IRT/Rasch approaches has not been presented. Regarding the strengths of the present study, a large sample and valid psychometric tools were used. As far as the Authors are aware, it is the first practical application of a tau-equivalent model to testing factorial validity and measurement invariance of a three-item scale, and probably the first substantiated empirical case of tau-equivalent measurement invariance in psychological literature in general. This study has demonstrated how to overcome technical restrictions on psychometric testing in specific circumstances by applying the most rigorous assumptions,

which can significantly improve the quality of research with ultra-short measures.

The aim of this study was not to present a new solution or “the solution” to the problem of factorial validity of three item scales. Instead, it analyses the existing solutions, discussing their advantages and disadvantages, and points to the optimal available option showing a relevant example of TILS. Such an analysis is clearly needed since none of the papers on the validity of the three item scales that the authors found in the relevant literature showed adequate factorial validity testing. The paper is mostly directed to non-experts in the psychometrics field for whom it may be an accessible description of a solution to the problem they may face in their research. Therefore, this paper constitutes a valuable and practical contribution to a wide range of fields applying ultra-brief psychometric measures.

Conclusions and Future Studies Directions

This paper analysed the existing solutions to the problem of factorial validity of three item scales, discussing their advantages and disadvantages. Based on this analysis, an optimal available

Table 3 Mean scores, standard deviations (*SD*), percentages and correlation coefficients between loneliness and criterion variables

Variable	Mean (<i>SD</i>) / Percentages	Skewness	Kurtosis	Loneliness
Loneliness	4.72 (1.71)	.88	–.11	–
Gender ^a	56.1% females	–	–	–.01
Age	20.92 (2.65)	–	–	–.05**
General health	5.92 (2.10)	–.49	–.56	–.20**
Sleep quality	5.35 (2.15)	–.23	–.81	–.22**
Global self-esteem	5.97 (1.82)	–.65	.18	–.41**
Satisfaction with personal relationships	5.83 (2.40)	–.44	–.80	–.38**
Satisfaction with support received from friends	6.64 (1.87)	–.81	.40	–.37**
General quality of life	6.85 (1.40)	–.86	1.45	–.37**
Satisfaction with life	5.90 (1.88)	–.25	–.58	–.43**
Meaning in life	5.84 (2.08)	–.35	–.62	–.40**
Stress	10.75 (3.04)	.19	–.15	.41**
Anxiety	7.19 (4.55)	.63	–.08	.46**
Depression	4.64 (3.50)	.89	.38	.37**

^a Point-biserial correlation coefficient (0 = female, 1 = male)

* $p < .05$. ** $p < .01$

Score ranges: loneliness: 3–9, general health: 1–9, sleep quality: 1–9, global self-esteem: 1–9, satisfaction with personal relationships: 1–9, satisfaction with support received from friends 1–9, general quality of life: 1–9, satisfaction with life: 1–9, meaning in life: 1–9, stress: 4–20, anxiety: 0–21, depression: 0–21

option is suggested along with a relevant example of TILS. Within CFA approach, tau-equivalent models allow for favorable psychometric performance of a scale, including better estimates of concurrent validity when a simple sum score is used instead of the latent score. However, it should be noted that when developing an ultra-short scale, different approaches could be applied and compared. Future studies could investigate the relative usefulness of using Rasch models or IRT models in comparison to CFA tau-equivalent models when creating a short scale. Perhaps more detailed information on the performance of particular items could allow for psychometrically better performing ultra-short scales and higher measurement invariance. Graded response model, a less restrictive approach than the Rasch model, could allow for items that contribute to the latent construct to varying degrees as it does not assume that all items have the same discrimination parameter. Since the tau-equivalent model is very restrictive, it could be recommended to use at least four-item scales, which allow testing the factorial structure of the more general congeneric model. Furthermore, uni- vs. multi-dimensionality of loneliness scales awaits more in-depth investigation, specifically more simulation studies in relation to developing ultra-short measures.

TILS showed good factorial validity, measurement invariance between genders, adequate concurrent and divergent validity, as well as good reliability in terms of internal consistency and test-retest measures. These results support the construct validity of the scale. Since loneliness seems to be a growing problem among younger populations, it is necessary to investigate its sources and consequences in more detail. Having a shorter tool for measuring loneliness can prove to be particularly useful in large surveys, which require the participants to fill in a sizable number of questionnaires, and when reducing the burden on the respondents is of great importance. Future research should look into its predictive validity in relation to health and well-being, particularly among young populations. More studies on loneliness among undergraduate students are highly warranted, as previous research indicates that this is a vulnerable population. In recent years indices of well-being among undergraduate and graduate students across countries tend to decline. The academic pressures that interfere with healthy social life leading to increasing feelings of loneliness, as well as social isolation, are suggested as one of the reasons. This has even more significance in the advent of global COVID-19 pandemic.

Authors' Contribution SKC assisted with literature search, study design and concept, statistical analyses, data interpretation, generation of the initial draft of the manuscript, manuscript preparation and editing, and final editing; PAA assisted with obtaining funding, literature search, study design and concept, data collection, statistical analyses, data interpretation, manuscript preparation and editing, and final editing.

Funding On the basis of decision number DEC-2013/08/T/HS6/00403 the author (Paweł Andrzej Atroszko) received funds from National Science Centre Poland within doctoral scholarship for preparing PhD dissertation. The data was collected in the subsequent research projects financed from: Grant of the Faculty of Social Sciences of the University of Gdansk for the projects of young scientists and doctoral students in 2013 (538–7422-B121–13), 2015 (538–7422-B910–15), and 2016 (538–7422-B286–16).

Data Availability Statement The datasets generated during and/or analysed during the current study are not publicly available due to the fact that participants were not informed that the dataset will be openly available but are available from the corresponding author on reasonable request.

Declarations

Ethics Ethical principles were carried out in accordance with the Declaration of Helsinki. The project was approved by Research Ethics Committee at the University of Gdańsk. Participation in the study was voluntary.

Conflict of Interest All authors declare that they have no conflict of interest regarding this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- American College Health Association. (2016). *American College Health Association-National College Health Assessment II: Canadian reference group data report spring 2016*. Retrieved September 12, 2020 from <https://www.acha.org/documents/ncha/NCHA-II%20SPRING%202016%20CANADIAN%20REFERENCE%20GROUP%20DATA%20REPORT.pdf>.
- Andreassen, C. S., Nielsen, M. B., Pallesen, S., & Gjerstad, J. (2019). The relationship between psychosocial work variables and workaholism: Findings from a nationally representative survey. *International Journal of Stress Management*, 26(1), 1–10. <https://doi.org/10.1037/str0000073>.
- Atroszko, P. A. (2015). *The structure of study addiction: Selected risk factors and the relationship with stress, stress coping and psychosocial functioning* (unpublished doctoral thesis). University of Gdansk, Poland.
- Atroszko, P. A., & Atroszko, B. (2020). Integrating national and international research, and exploring country-specific factors contributing to work addiction commentary to "Workaholism – Psychological and social determinants of work addiction". *Current Problems of Psychiatry*, 21(4). <https://doi.org/10.2478/cpp-2020-0017>.

- Atroszko, P. A., Bagińska, P., Mokosińska, M., Sawicki, A., & Atroszko, B. (2015). Validity and reliability of single item self-report measures of general quality of life, general health and sleep quality. In M. McGreevy & R. Rita (Eds.), *Proceedings of the 4th biannual CER comparative European research conference* (pp. 207–211). London, UK: Science Publishing.
- Atroszko, P. A., Pallesen, S., Griffiths, M. D., & Andreassen, C. S. (2017). Work addiction in Poland: Adaptation of the Bergen work addiction scale and relationship with psychopathology. *Health Psychology Report, 4*, 345–355. <https://doi.org/10.5114/hpr.2017.68759>.
- Atroszko, P. A., Pianka, L., Raczynska, A., & Atroszko, B. (2015). Validity and reliability of single-item self-report measures of social support. In M. McGreevy & R. Rita (Eds.), *Proceedings of the 4th biannual CER comparative European research conference* (pp. 216–219). London, UK: Science Publishing.
- Atroszko, P. A., Sawicki, A., Mąkinia, A., & Atroszko, B. (2017). Further validation of single-item self-report measure of satisfaction with life. In M. McGreevy & R. Rita (Eds.), *Proceedings of the 7th biannual CER comparative European research conference* (pp. 107–110). London, UK: Science Publishing.
- Atroszko, P. A., Sawicki, A., Sendal, L., & Atroszko, B. (2017). Validity and reliability of single-item self-report measure of global self-esteem. In M. McGreevy & R. Rita (Eds.), *Proceedings of the 7th biannual CER comparative European research conference* (pp. 120–123). London, UK: Science Publishing.
- Auné, S. E., Abal, F. J., & Attorresi, H. F. (2020). Modeling of the UCLA loneliness scale according to the multidimensional item response theory. *Current Psychology, 19*(1), 1–10. <https://doi.org/10.1007/s12144-020-00646-y>.
- Austin, B. A. (1983). Factorial structure of the UCLA loneliness scale. *Psychological Reports, 53*(3), 883–889. <https://doi.org/10.2466/pr0.1983.53.3.883>.
- Baker, J. G., Rounds, J. B., & Zevon, M. A. (2000). A comparison of graded response and rasch partial credit models with subjective well-being. *Journal of Educational and Behavioral Statistics, 25*(3), 253–270. <https://doi.org/10.2307/1165205>.
- Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. New York: Oxford Univ. Pr.
- Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*(2), 175–184. <https://doi.org/10.1509/jmkr.44.2.175>.
- Beutel, M. E., Klein, E. M., Brähler, E., Reiner, I., Jünger, C., Michal, M., Wiltink, J., Wild, P. S., Münzel, T., Lackner, K. J., & Tibubos, A. N. (2017). Loneliness in the general population: Prevalence, determinants and relations to mental health. *BMC Psychiatry, 17*(1), 97. <https://doi.org/10.1186/s12888-017-1262-x>.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479). Reading: Addison-Wesley.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219. <https://doi.org/10.1037/0033-295X.110.2.203>.
- Borsboom, D., Molenaar, D., & Wright, J. D. (2015). Psychometrics. *International Encyclopedia of the Social & Behavioral Sciences, 19*(2), 418–422.
- Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (Vol. 21, pp. 136–162). Newbury Park, CA: Sage. <https://doi.org/10.1177/004912419201002005>.
- Buelga, S., Musitu, G., Murgui, S., & Pons, J. (2008). Reputation, loneliness, satisfaction with life and aggressive behavior in adolescence. *The Spanish Journal of Psychology, 11*(1), 192–200. <https://doi.org/10.1017/s1138741600004236>.
- Caballer, A., Belmonte, O., Castillo, A., Gasco, A., Sansano, E., & Montoliu, R. (2020). Equivalence of chatbot and paper-and-pencil versions of the De Jong Gierveld loneliness scale. *Current Psychology, 19*(1), 1–10. <https://doi.org/10.1007/s12144-020-01117-0>.
- Cacioppo, J. T., Hughes, M. E., Waite, L. J., Hawkley, L. C., & Thisted, R. A. (2006). Loneliness as a specific risk factor for depressive symptoms: Cross-sectional and longitudinal analyses. *Psychology and Aging, 21*(1), 140–151. <https://doi.org/10.1037/0882-7974.21.1.140>.
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5.
- Chen, Y., & Feeley, T. H. (2014). Social support, social strain, loneliness, and well-being among older adults: An analysis of the health and retirement study. *Journal of Social and Personal Relationships, 31*(2), 141–161. <https://doi.org/10.1177/0265407513488728>.
- Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research, 23*(10), 2809–2818. <https://doi.org/10.1007/s11136-014-0726-4>.
- Cigna, Ipsos. Cigna U.S. loneliness index: Survey of 20,000 Americans examining behaviors driving loneliness in the United States. 2018. Retrieved July 24, 2018 from: https://www.multivu.com/players/English/8294451-cigna-us-lonelinesssurvey/docs/IndexReport_1524069371598-173525450.pdf.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396. <https://doi.org/10.2307/2136404>.
- Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin, 98*(2), 310–357. <https://doi.org/10.1037/0033-2909.98.2.310>.
- Coulacoglou, C., & Saklofske, D. H. (2017). *Psychometrics and psychological assessment principles and applications*. Saint Louis: Elsevier Science.
- Coyle, C. E., & Dugan, E. (2012). Social isolation, loneliness and health among older adults. *Journal of Aging and Health, 24*, 1346–1363. <https://doi.org/10.1177/0898264312460275>.
- Cuijpers, P., Smits, N., Donker, T., Ten Have, M., & de Graaf, R. (2009). Screening for mood and anxiety disorders with the five-item, the three-item, and the two-item mental health inventory. *Psychiatry Research, 168*(3), 250–255. <https://doi.org/10.1016/j.psychres.2008.05.012>.
- Czerwiński, S. K., Mackiewicz, J. J., Mytlewska, W. M., & Atroszko, P. A. (2020). Factorial validity, measurement invariance and concurrent validity of hospital anxiety and depression scale in polish undergraduate student sample. *Psychiatria i Psychologia Kliniczna, 20*(1), 13–18. <https://doi.org/10.15557/pspik.2020.0002>.
- De Ayala, R. J., & Hertzog, M. A. (1991). The assessment of dimensionality for use in item response theory. *Multivariate Behavioral Research, 26*(4), 765–792. https://doi.org/10.1207/s15327906mbr2604_9.
- De Champlain, A., & Gessaroli, M. E. (1998). Assessing the dimensionality of item response matrices with small sample sizes and short test lengths. *Applied Measurement in Education, 11*(3), 231–253. https://doi.org/10.1207/s15324818ame1103_2.
- de Jong Gierveld, J. D. (1989). Personal relationships, social support, and loneliness. *Journal of Social and Personal Relationships, 6*(2), 197–221. <https://doi.org/10.1177/026540758900600204>.
- de Jong Gierveld, J. D., & Kamphuis, F. (1985). The development of a Rasch-type loneliness scale. *Applied Psychological Measurement, 9*(3), 289–299. <https://doi.org/10.1177/014662168500900307>.
- Depaoli, S., Tiemensma, J., & Felt, J. M. (2018). Assessment of health surveys: Fitting a multidimensional graded response model. *Psychology, Health & Medicine, 23*(Sup1), 13–31. <https://doi.org/10.1080/13548506.2018.1447136>.

- Domagała-Krecioch, A., & Majerek, B. (2014). The issue of loneliness in the period of “emerging adulthood”. *European Scientific Journal, ESJ*, 9(10). <https://doi.org/10.19044/esj.2013.v9n10p%25p>.
- Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(1), 128–141. https://doi.org/10.1207/S15328007SEM0801_7.
- Erzen, E., & Çikrikci, Ö. (2018). The effect of loneliness on depression: A meta-analysis. *International Journal of Social Psychiatry*, 64(5), 427–435. <https://doi.org/10.1177/0020764018776349>.
- Fanakidou, I., Zyga, S., Alikari, V., Tsironi, M., Stathoulis, J., & Theofilou, P. (2017). Mental health, loneliness, and illness perception outcomes in quality of life among young breast cancer patients after mastectomy: The role of breast reconstruction. *Quality of Life Research*, 27(2), 539–543. <https://doi.org/10.1007/s11136-017-1735-x>.
- Główny Urząd Statystyczny. (2019). *Szkoły wyższe i ich finanse w 2018 roku*. Retrieved September 12, 2020 from: https://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultaktualnosci/5488/2/15/1/szkoły_wyższe_i_ich_finance_w_2018.pdf.
- Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). “My questionnaire is too Long!” the assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, 39(3), 188–205. <https://doi.org/10.1016/j.cedpsych.2014.04.002>.
- Goossens, L., & Beyers, W. (2002). Comparing measures of childhood loneliness: Internal consistency and confirmatory factor analysis. *Journal of Clinical Child and Adolescent Psychology*, 31(2), 252–262. <https://doi.org/10.1207/153744202753604520>.
- Goossens, L., Lasgaard, M., Luyckx, K., Vanhalst, J., Mathias, S., & Masy, E. (2009). Loneliness and solitude in adolescence: A confirmatory factor analysis of alternative models. *Personality and Individual Differences*, 47(8), 890–894. <https://doi.org/10.1016/j.paid.2009.07.011>.
- Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement*, 66(6), 930–944. <https://doi.org/10.1177/0013164406288165>.
- Grygiel, P., Humenny, G., & Rebisz, S. (2016). Using the De Jong Gierveld loneliness scale with early adolescents: Factor structure, reliability, stability, and external validity. *Assessment*, 26(2), 151–165. <https://doi.org/10.1177/1073191116682298>.
- Grygiel, P., Humenny, G., Rebisz, S., Świtaj, P., & Sikorska, J. (2013). Validating the polish adaptation of the 11-item De Jong Gierveld loneliness scale. *European Journal of Psychological Assessment*, 29(2), 129–139. <https://doi.org/10.1027/1015-5759/a000130>.
- Hartshome, T. S. (1993). Psychometric properties and confirmatory factor analysis of the UCLA loneliness scale. *Journal of Personality Assessment*, 61(1), 182–195. https://doi.org/10.1207/s15327752jpa6101_14.
- Hawkey, L. C., Duvoisin, R., Ackva, J., Murdoch, J. C., & Luhmann, M. (2015). Loneliness in older adults in the USA and Germany: Measurement invariance and validation. Working paper series, NORC at the University of Chicago, paper 2015-002.
- Heinrich, L. M., & Gullone, E. (2006). The clinical significance of loneliness: A literature review. *Clinical Psychology Review*, 26(6), 695–718. <https://doi.org/10.1016/j.cpr.2006.04.002>.
- Höfer, P., Rockett, I. R., Värnik, P., Etzersdorfer, E., & Kapusta, N. D. (2012). Forty years of increasing suicide mortality in Poland: Undercounting amidst a hanging epidemic? *BMC Public Health*, 12(1). <https://doi.org/10.1186/1471-2458-12-644>.
- Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, 10(2), 227–237. <https://doi.org/10.1177/1745691614568352>.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys. *Research on Aging*, 26(6), 655–672. <https://doi.org/10.1177/0164027504268574>.
- Huo, Y., & Kong, F. (2014). Moderating effects of gender and loneliness on the relationship between self-esteem and life satisfaction in Chinese University students. *Social Indicators Research*, 118(1), 305–314. <https://doi.org/10.1007/s11205-013-0404-x>.
- Jenkins, C. D., Stanton, B. A., Niemcryk, S. J., & Rose, R. M. (1988). A scale for the estimation of sleep problems in clinical research. *Journal of Clinical Epidemiology*, 41(4), 313–321. [https://doi.org/10.1016/0895-4356\(88\)90138-2](https://doi.org/10.1016/0895-4356(88)90138-2).
- Joiner, T. E., Catanzaro, S. J., Rudd, M. D., & Rajab, M. H. (1999). The case for a hierarchical, oblique, and Bidimensional structure of loneliness. *Journal of Social and Clinical Psychology*, 18(1), 47–75. <https://doi.org/10.1521/jscp.1999.18.1.47>.
- Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36(3), 347–387. <https://doi.org/10.1207/s15327906347-387>.
- Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(1), 136–153. <https://doi.org/10.1080/10705510701758406>.
- Kelly, W. E. (2004). A brief measure of general worry: The three item worry index. *North American Journal of Psychology*, 6(2), 219–225.
- Kemper, C. J., Trapp, S., Kathmann, N., Samuel, D. B., & Ziegler, M. (2018). Short versus Long scales in clinical Assessment: Exploring the Trade-Off Between Resources Saved and Psychometric Quality Lost Using Two Measures of Obsessive–Compulsive Symptoms. *Assessment*. <https://doi.org/10.1177/1073191118810057>.
- Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 212–228. <https://doi.org/10.1080/10705511.2011.557337>.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioural research*. Washington: American Psychological Association. <https://doi.org/10.1037/10693-000>.
- Koryczan, P., Piotrowski, P., Roj, W., Czerwiński, S., & Atroszko, P. (2020). Validity and reliability of the polish version of the short hopelessness scale. *Health Psychology Report*, 8, 445–452. <https://doi.org/10.5114/hpr.2019.91428>.
- Krebs, E. E., Lorenz, K. A., Bair, M. J., Damush, T. M., Wu, J., Sutherland, J. M., Asch, S. M., & Kroenke, K. (2009). Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. *Journal of General Internal Medicine*, 24(6), 733–738. <https://doi.org/10.1007/s11606-009-0981-1>.
- Kwiatkowska, M., Rogoza, R., & Kwiatkowska, K. (2017). Analysis of the psychometric properties of the revised UCLA loneliness scale in a polish adolescent sample. *Current Issues in Personality Psychology*, 6(2), 164–170. <https://doi.org/10.5114/cipp.2017.69681>.
- Lawendowski, R., Bereznowski, P., Wróbel, W. K., Kierzkowski, M., & Atroszko, P. A. (2019). Study addiction among musicians: Measurement, and relationship with personality, social anxiety, performance, and psychosocial functioning. *Musicae Scientiae*, 24, 449–474. <https://doi.org/10.1177/1029864918822138>.

- Lee, C. S., & Goldstein, S. E. (2015). Loneliness, stress, and social support in young adulthood: Does the source of support matter? *Journal of Youth and Adolescence*, 45(3), 568–580. <https://doi.org/10.1007/s10964-015-0395-9>.
- Leon, A. C., Olfson, M., Portera, L., Farber, L., & Sheehan, D. V. (1997). Assessing psychiatric impairment in primary care with the Sheehan disability scale. *The International Journal of Psychiatry in Medicine*, 27(2), 93–105. <https://doi.org/10.2190/t8em-c8yh-373n-luwd>.
- Lichtenstein, M. B., Malkenes, M., Sibbersen, C., & Hinze, C. J. (2019). Work addiction is associated with increased stress and reduced quality of life: Validation of the Bergen work addiction scale in Danish. *Scandinavian Journal of Psychology*, 60(2), 145–151. <https://doi.org/10.1111/sjop.12506>.
- Liu, L., & Guo, Q. (2008). Life satisfaction in a sample of empty-nest elderly: A survey in the rural area of a mountainous county in China. *Quality of Life Research*, 17(6), 823–830. <https://doi.org/10.1007/s11136-008-9370-1>.
- Maes, M., Klimstra, T., Noortgate, W. V., & Goossens, L. (2014). Factor structure and measurement invariance of a multidimensional loneliness scale: Comparisons across gender and age. *Journal of Child and Family Studies*, 24(6), 1829–1837. <https://doi.org/10.1007/s10826-014-9986-4>.
- Mahon, N. E., & Yarcheski, A. (1990). The dimensionality of the UCLA loneliness scale in early adolescents. *Research in Nursing & Health*, 13(1), 45–52. <https://doi.org/10.1002/nur.4770130108>.
- Mahon, N. E., Yarcheski, T. J., & Yarcheski, A. (1995). Validation of the revised UCLA loneliness scale for adolescents. *Research in Nursing & Health*, 18(3), 263–270. <https://doi.org/10.1002/nur.4770180309>.
- Marsh, H. W. (1987). The factorial invariance of responses by males and females to a multidimensional self-concept instrument: Substantive and methodological issues. *Multivariate Behavioral Research*, 22(4), 457–480. https://doi.org/10.1207/s15327906mbr2204_5.
- Matthews, T., Danese, A., Gregory, A. M., Caspi, A., Moffitt, T. E., & Arseneault, L. (2017). Sleeping with one eye open: Loneliness and sleep quality in young adults. *Psychological Medicine*, 47(12), 2177–2186. <https://doi.org/10.1017/s0033291717000629>.
- Matthews-Ewald, M. R., & Zullig, K. J. (2013). Evaluating the performance of a short loneliness scale among college students. *Journal of College Student Development*, 54(1), 105–109. <https://doi.org/10.1353/csd.2013.0003>.
- McDonald, R. P. (2013). *Test theory a unified treatment*. London: Taylor and Francis.
- McWhirter, B. T. (1990). Factor analysis of the revised UCLA loneliness scale. *Current Psychology*, 9(1), 56–68. <https://doi.org/10.1007/bf02686768>.
- Meade, A. W., & Lautenschlager, G. J. (2004, April). Same question, different answers: CFA and two IRT approaches to measurement invariance. In: 19th Annual Conference of the Society for Industrial and Organizational Psychology (Vol. 1).
- Mellor, D., Stokes, M., Firth, L., Hayashi, Y., & Cummins, R. (2008). Need for belonging, relationship satisfaction, loneliness, and life satisfaction. *Personality and Individual Differences*, 45(3), 213–218. <https://doi.org/10.1016/j.paid.2008.03.020>.
- Moore, D., & Schultz, N. R. (1983). Loneliness at adolescence: Correlates, attributions, and coping. *Journal of Youth and Adolescence*, 12(2), 95–100. <https://doi.org/10.1007/bf02088307>.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132. <https://doi.org/10.1007/bf02294210>.
- Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. *Frontiers in Psychology*, 5, 978. <https://doi.org/10.3389/fpsyg.2014.00978>.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716. <https://doi.org/10.1126/science.aac4716>.
- Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on Replicability in psychological science. *Perspectives on Psychological Science*, 7(6), 528–530. <https://doi.org/10.1177/1745691612465253>.
- Patil, P., Peng, R. D., & Leek, J. (2016). A statistical definition for reproducibility and replicability. *BioRxiv*, 066803. <https://doi.org/10.1101/066803>.
- Penning, M. J., Liu, G., & Chou, P. H. (2013). Measuring loneliness among middle-aged and older adults: The UCLA and de Jong Gierveld loneliness scales. *Social Indicators Research*, 118(3), 1147–1166. <https://doi.org/10.1007/s11205-013-0461-1>.
- Perissinotto, C. M., & Covinsky, K. E. (2014). Living alone, socially isolated or lonely—What are we measuring? *Journal of General Internal Medicine*, 11, 1429–1431. <https://doi.org/10.1007/s11606-014-2977-8>.
- Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*, 18(1), 25–34. <https://doi.org/10.1016/j.jval.2014.10.005>.
- Pinquant, M., & Sorensen, S. (2001). Influences on loneliness in older adults: A meta-analysis. *Basic and Applied Social Psychology*, 23(4), 245–266. <https://doi.org/10.1207/153248301753225702>.
- Plesser, H. E. (2018). Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Frontiers in Neuroinformatics*, 11, 76. <https://doi.org/10.3389/fninf.2017.00076>.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>.
- Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and ways to overcome them. *Journal of Individual Differences*, 35, 212–220. <https://doi.org/10.1027/1614-0001/a000141>.
- Raykov, T. (1997a). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 173–184. <https://doi.org/10.1177/01466216970212006>.
- Raykov, T. (1997b). Scale reliability, Cronbachs coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, 32(4), 329–353. https://doi.org/10.1207/s15327906mbr3204_2.
- Rico-Urbe, L. A., Caballero, F. F., Martin-Maria, N., Cabello, M., Ayuso-Mateos, J. L., & Miret, M. (2016). Association of loneliness with all-cause mortality: A meta-analysis. *PLoS One*, 13, e0190033. <https://doi.org/10.1371/journal.pone.0190033>.
- Rosenthal, R. (2005). Binomial effect size display. *Encyclopedia of Statistics in Behavioral Science*. <https://doi.org/10.1002/0470013192.bsa050>.
- Røysamb, E., & Strype, J. (2002). Optimism and pessimism: Underlying structure and dimensionality. *Journal of Social and Clinical Psychology*, 21(1), 1–19. <https://doi.org/10.1521/jscp.21.1.1.22403>.
- Russell, D., Peplau, L. A., & Cutrona, C. E. (1980). The revised UCLA loneliness scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*, 39(3), 472–480. <https://doi.org/10.1037/0022-3514.39.3.472>.
- Russell, D. W. (1996). UCLA loneliness scale (version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, 66(1), 20–40. https://doi.org/10.1207/s15327752jpa6601_2.
- Salemo, L., Ingoglia, S., & Coco, G. L. (2017). Competing factor structures of the Rosenberg self-esteem scale (RSES) and its measurement invariance across clinical and non-clinical samples. *Personality and Individual Differences*, 113, 13–19. <https://doi.org/10.1016/j.paid.2017.02.063>.

- Salzberger, T., & Sinkovics, R. R. (2006). Reconsidering the problem of data equivalence in international marketing research: Contrasting approaches based on CFA and the Rasch model for measurement. *International Marketing Review*, 23(4), 390–417. <https://doi.org/10.1108/02651330610678976>.
- Samejima, F. (1970). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 35(1), 139–139.
- Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, 3, 153–160 Retrieved from http://equinetrust.org.nz/massey/fms/Colleges/College%20of%20Sciences/IIMS/RLIMS/Volume03/Dealing_with_Missing_Data.pdf.
- Segrin, C., & Passalacqua, S. A. (2010). Functions of loneliness, social support, health behaviors, and stress in association with poor health. *Health Communication*, 25(4), 312–322. <https://doi.org/10.1080/10410231003773334>.
- Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of Life Research*, 13(2), 299–310. <https://doi.org/10.1023/B:QURE.0000018486.91360.00>.
- Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, 2, 199–218.
- Stochl, J., Jones, P. B., & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A non-parametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology*, 12(1). <https://doi.org/10.1186/1471-2288-12-74>.
- Swami, V., Chamorro-Premuzic, T., Sinniah, D., Maniam, T., Kannan, K., Stanistreet, D., & Furnham, A. (2007). General health mediates the relationship between loneliness, life satisfaction and depression. *Social Psychiatry and Psychiatric Epidemiology*, 42(2), 161–166. <https://doi.org/10.1007/s00127-006-0140-5>.
- Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393–408. <https://doi.org/10.1007/bf02294363>.
- Tay, L., & Drasgow, F. (2012). Theoretical, statistical, and substantive issues in the assessment of construct dimensionality: Accounting for the item response process. *Organizational Research Methods*, 15(3), 363–384. <https://doi.org/10.1177/1094428112439709>.
- Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, 18(1), 3–46. <https://doi.org/10.1177/1094428114553062>.
- Tobiasz-Adamczyk, B., Galas, A., Zawisza, K., Chatterji, S., Haro, J. M., Ayuso-Mateos, J. L., Koskinen, S., & Leonardi, M. (2017). Gender-related differences in the multi-pathway effect of social determinants on quality of life in older age—The COURAGE in Europe project. *Quality of Life Research*, 26(7), 1865–1878. <https://doi.org/10.1007/s11136-017-1530-8>.
- Tuerlinckx, F., & Wang, W. C. (2004). Models for polytomous data. In: *Explanatory Item Response Models* (pp. 75–109). New York, NY: Springer.
- Twenge, J. M., Campbell, W. K., & Freeman, E. C. (2012). Generational differences in young adults life goals, concern for others, and civic orientation, 1966–2009. *Journal of Personality and Social Psychology*, 102(5), 1045–1062. <https://doi.org/10.1037/a0027408>.
- Twenge, J. M., Gentile, B., Dewall, C. N., Ma, D., Laceyfield, K., & Schurtz, D. R. (2010). Birth cohort increases in psychopathology among young Americans, 1938–2007: A cross-temporal meta-analysis of the MMPI. *Clinical Psychology Review*, 30(2), 145–154. <https://doi.org/10.1016/j.cpr.2009.10.005>.
- Twenge, J. M., Spitzberg, B. H., & Campbell, W. K. (2019). Less in-person social interaction with peers among U.S. adolescents in the 21st century and links to loneliness. *Journal of Social and Personal Relationships*, 36(6), 1892–1913. <https://doi.org/10.1177/0265407519836170>.
- Valtorta, N. K., Kanaan, M., Gilbody, S., Ronzi, S., & Hanratty, B. (2016). Loneliness and social isolation as risk factors for coronary heart disease and stroke: Systematic review and meta-analysis of longitudinal observational studies. *Heart*, 102(13), 1009–1016. <https://doi.org/10.1136/heartjnl-2015-308790>.
- Vanderweele, T. J., Hawkey, L. C., & Cacioppo, J. T. (2012). On the reciprocal association between loneliness and subjective well-being. *American Journal of Epidemiology*, 176(9), 777–784. <https://doi.org/10.1093/aje/kws173>.
- Vanhalst, J., Luyckx, K., Scholte, R. H., Engels, R. C., & Goossens, L. (2013). Low self-esteem as a risk factor for loneliness in adolescence: Perceived - but not actual - social acceptance as an underlying mechanism. *Journal of Abnormal Child Psychology*, 41(7), 1067–1081. <https://doi.org/10.1007/s10802-013-9751-y>.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632–638. <https://doi.org/10.1177/1745691612463078>.
- Wang, W., Chen, H., & Jin, K. (2014). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement*, 75(1), 157–178. <https://doi.org/10.1177/0013164414528209>.
- Wilson, D., Cutts, J., Lees, I., Mapungwana, S., & Maunganidze, L. (1992). Psychometric properties of the revised UCLA loneliness scale and two short-form measures of loneliness in Zimbabwe. *Journal of Personality Assessment*, 59(1), 72–81. https://doi.org/10.1207/s15327752jpa5901_7.
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014–1045. <https://doi.org/10.1007/s11336-016-9506-0>.
- Yu, B., Steptoe, A., Niu, K., Ku, P., & Chen, L. (2017). Prospective associations of social isolation and loneliness with poor sleep quality in older adults. *Quality of Life Research*, 27(3), 683–691. <https://doi.org/10.1007/s11136-017-1752-9>.
- Zigmond, A. S., & Snaith, R. E. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67(6), 361–370. <https://doi.org/10.1111/j.1600-0447.1983.tb09716.x>.
- Zoutewelle-Terovan, M., & Liefbroer, A. C. (2017). Swimming against the stream: Non-normative family transitions and loneliness in later life across 12 nations. *The Gerontologist*, 58(6), 1096–1108. <https://doi.org/10.1093/geront/gnx184>.