**Measuring Social Loneliness: Psychometric and Conceptual properties**

Tom Yishay (12203424), Jeroen van der Meer (12658944), Britt Gerritsen (13292013), Konrad Mikalauskas (12185728), Laura Groot (12675733) & Jil Zoe Fuhrmann (12059935)

Department of Psychology, University of Amsterdam

7202BM04XY: Latent Variable Models

Dylan Molenaar

March 21, 2021

**Measuring Social Loneliness: Psychometric and Conceptual properties**

Two of the vastly vastly used measurment theories of psychological constructs are Latent Variables Models (LVM) and Classical Test Theory (CTT) (Borsboom, 2005). LVM conceptualize psychological constructs as unobservable (latent) variables measured by sets of observable manifestations- namely tests’ items. Latent variables explain all associations between the test’s items and provide unique information regarding both the test and its individual items. CTT unfolds constructs using tests’ sum scores consisting of the true (one’s true quantity of a construct) and error scores. Compared to CTT, LVM offers essential advantages for psychological measurement. First, LVM explicitly analyze the psychological constructs measured by tests rather than simply assessing its sum score. Such models assess the number of underlying test dimensions which are then interpreted as psychological constructs. Second, LVM offer improved falsifiability compared to CTT. Attempting to fit LVM using poorly specified models or highly variant data will result in clear errors (variance close to zero) while CTT’s scores are always applicable and thus hardly falsifiable. Third, unlike CTT’s dependecny on tests’ and smaples’ proeprties (mean, varince, and items’ paramters) LVM are invariant across tests and samples. Test properties are captured by items’ parameters and sample properties are captured by the latent variable. Finally, LVM account for item-specific error, highlighting individual differences and items’ individual contribution\deterioration of the test. CTT only allows the inspection of an aggregated sum score across items. Critics of LVM are the statistically more complex operations and the need for large samples. Nonetheless, given the high availability of processing speed and data from digital tools, such problems are mostly solvable.

Loneliness is a significant concern given the strategies used to regulate the spread of COVID-19 in current times. Social isolation and distancing are globally used to diminsih viral spread and have been linked to increased loneliness and poorer health (Sutin et al., 2020). Perlman et al. (1984) have defined social loneliness as one’s perceived discrepancy between desired and existing social relationships. Measuring social loneliness is thus crucial for optimizing people’s health. Cognitive, behavioral, and affective components of loneliness can be used to measure social loneliness systematically (Heinrich, & Gullone, 2006). Lonely people were found to have low self-esteem, experience a host of negative emotions (most commonly desperation and depression), and be socially inhibited. Nonetheless, a questionnaire developed to measure social loneliness based on these psychological facets is absent in the literature. Therefore, the present study investigates a self-reported survey of social loneliness among young adults.

The internal structure of loneliness, the number of its underlying dimensions (dimensionality), has been debated in loneliness literature (Heinrich, & Gullone, 2006). Russell (1980) found support for loneliness as a unidimensional construct measured by a self-reported survey (UCLA; Russell, 1980). However, later studies measuring loneliness have found considerable support for a bi-dimensional structure consisting of emotional and social loneliness (Heinrich, & Gullone, 2006; DiTommaso et al., 1997). For example, one study has found support for this bi-dimensional structure when analyzing 1,526 students (DiTommaso et al., 2004).

The literature enables the assessment of social loneliness using a self-reported survey among young adults. Prior research supports the validity of self-reported surveys in the context of loneliness and among young adults (DiTommaso et al., 2004; Russell, 1980). Secondly, measuring social loneliness as a unidimensional component of loneliness is supported by some past investigations (DiTommaso et al., 1977, 2004). To this end, the present study hypothesizes social loneliness to be a unidimensional construct.

The present study sheds light upon fundamental issues in the literature of loneliness. Investigating the internal structure of social loneliness can provide an initial resolution regarding loneliness’ dimensionality. When measuring social loneliness, the case of obtaining a unidimensional structure would most likely indicate that loneliness is at least bi-dimensional. Moreover, deriving a measurement tool from loneliness’s cognitive, behavioral, and affective elements constitutes a novel yet vital contribution to the literature. Assessing and relating the information provided by each element to the measurement of social loneliness, can refine loneliness’s theoretical and practical frameworks. Altogether, the internal structure and the psychological aspects of social loneliness are examined using a 15 items survey applied to a sample of young adults.

**Methods**

**Sample**

A convenience sample of 141 participants (*M* = 21.40, *SD* = 1.80) were administered the social loneliness questionnaire. Participants' age ranged from 17 to 63, consisting of mainly women (66.40%). Data of participants whose age falls outside of the range of 18-28 is excluded from data analysis to improve inference validity.

**Materials**

***Fitting the Item Response Theory and Factor Analysis models***

**R packages.** The “psych” R package consists of a wide array of general-purpose functions used in psychometric theory (Revelle, 2020). Polychoric correlations can be computed using the package, which are then used to test the dimensionality of a construct. Moreover, the package contains some factor analyses functionality. The “ltm” package is developed for various IRT analyses. Of relevance to this study are the graded response models to polytomous data. The “lavaan” package consists of several functions used for confirmatory factor analysis.

***Model fit assessment***

Model compression and fit statistics, elaborated in the analysis plan, will be evaluated by various methods. The RMSEA (Steiger & Lind, 1980) is a measure of how far off a hypothesized model is from a perfect model, and is therefore an absolute measure of model fit – values below 0.08 are deemed acceptable. The AIC (Akaike, 1974) and BIC (Schwarz, 1978), on the other hand, are relative measures, and are used for model comparison. The AIC estimates the distance between a model’s likelihood function and the true data-generating likelihood function. The BIC is a measure of the posterior probability of a fitted model given the observed data. Theoretical background aside, both measures perform similarly, with the BIC being more stringent regarding model complexity. In both measures, lower values indicate better relative model fit.

**Analysis Plan**

In this study, data will be analyzed using the “psych”, “ltm” and “lavaan” R packages (Revelle, 2020; Rizopoulos, 2020; Rosseel, 2012). The “psych” package allows for the construction of a polychoric correlation matrix whose eigenvalues will be computed to test for unidimensionality of the latent variable. The “ltm” package allows for the fitting of a graded response model to the gathered dataset. The estimated discrimination parameters of the model will then be tested for equality across items. Using base R, model fit statistics, such as the BIC and log-Likelihood, will be computed. Finally, using the “psych” and “lavaan” R packages, an exploratory factor analysis will be conducted to examine how well models with different numbers of factors fit to the dataset, from which the best fitting model will be selected using the RMSEA index. A confirmatory factor model will then be fitted, based on the factor structure of the model selected from the exploratory analysis, to measure goodness-of-fit, and observe whether the initial model could be recovered.

**Procedure**

The questionnaire aims to measure social loneliness. Items were based on the 3 facets of loneliness identified by previous research (affective, cognitive and behavioral) and were then attempted to be specified to social loneliness, for example by linking an emotion to a social situation. The questionnaire was created on Google Forms and consists of 4 items requesting demographic information (gender, age, occupation, country of residence), followed by 15 questions intended to measure the unidimensional construct of social loneliness. Five questions represented one of the three facets, affective (“Around my friends I often feel insecure”), cognitive (“My interests and ideas are shared by those around me”) and behavioral (“When I feel vulnerable, I withdraw from social situations”). The 15 construct items were scored on a Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”) and contained eigth indicative (“My relationships are superficial”) and seven contra-indicative (“Social interaction fulfills me”) items. Questions appeared in the same order for all participants and could be seen simultaneously on one screen. Participants were not rewarded for participation. Participants received no specific instructions regarding the procedure and no information about the intention of the study, but were informed about confidentiality regarding their data at the beginning of the questionnaire.

**Results**

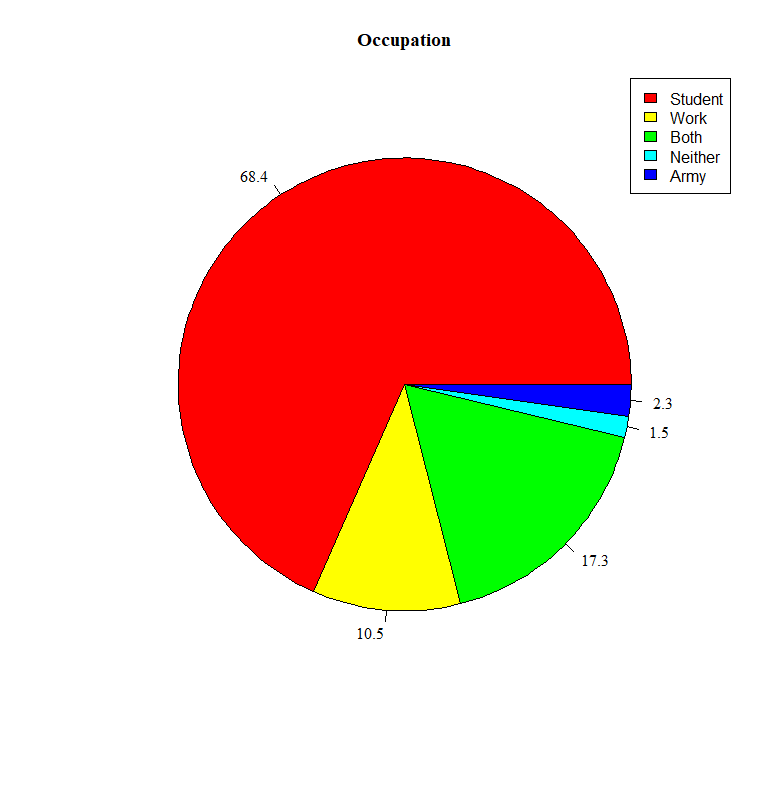
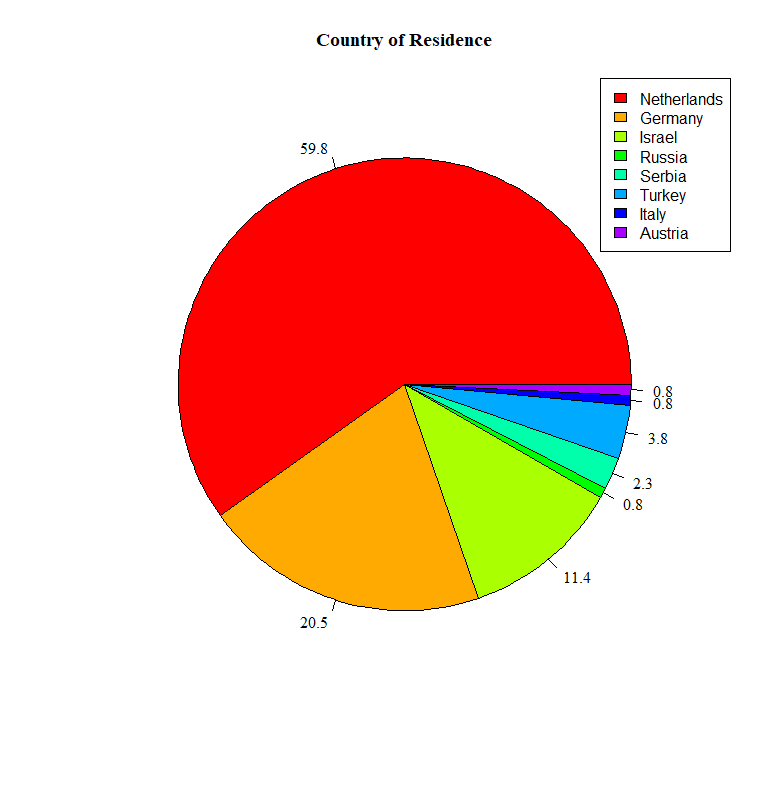
The descriptive statistics of all participates (N = 133) are shown in Table 1. Figure 1 visualizes participants’ occupations and current countries of residence.

**Table 1**

*Number of Participants per Condition, Gender Count and Mean Age. Values in Parenthesizes are Standard Deviations*

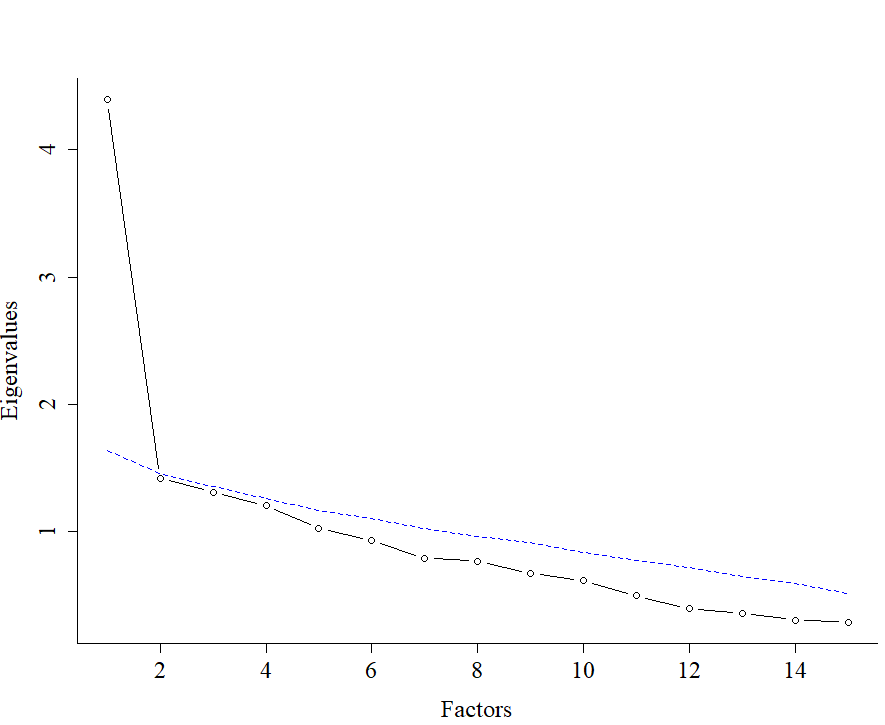
|  |  |
| --- | --- |
| Trait | Participants |
| Number of participants  Age  Men/Women | 133  21.39 (1.76)  39/91 |

**Figure 1**

*Participants’ Countries of Residency and Occupations. Numbers are percentage*

An item response analysis was carried to determine the survey’s (1) dimensionality, (2) parametric discrimination, and (3) information. As for the first, a polychoric correlation matrix was calculated to compute corresponding Eigenvalues of the data. The scree plot exhibited a considerable decrease from one to two factors, suggesting a singular factor (see Figure 2). Likewise, a parallel test analysis was conducted. As seen in Figure 2, the scree plot intersects with the parallel analysis between the first and second factor (blue line in Figure 2). In line with the hypothesis, the scree plot as well as the parallel test indicate a unidimensional internal structure. Secondly, a log-likelihood ratio test between the general (inequal parametric discrimination; *L* = -2533.32) and restricted (equal parametric discrimination; *L* = -2484.07) models was significant (See Table 2; *p* < .001). This indicates that equating the discrimination parameter across items deteriorated the model’s fit. Thus, discrimination was estimated per question as computed by the general model. Discrimination estimates (*M* = 1.16, *SD* = 0.70) were lowest for items number two (*a* = 0.30) and ten (*a* = 0.166) and highest for items eleven (*a* = 2.47) and four-teen (*a* = 2.40). Thirdly, information provided by the test was largest for theta levels between zero and three (see Figure 3). The test is thus maximally informative for higher than average levels of theta. Information per item was sorted according to theta’s three facets of this survey (cognitive, behavioral, and affective), as depicted by Figure 4. Most information was provided by affective items, followed by cogntivie and behavioral ones. Questions two and ten yielded the least information while questions eleven and four-teen clearly provided most information.

**Figure 2**

*Scree Plot with Principle Component Parallel Analysis (blue line)*

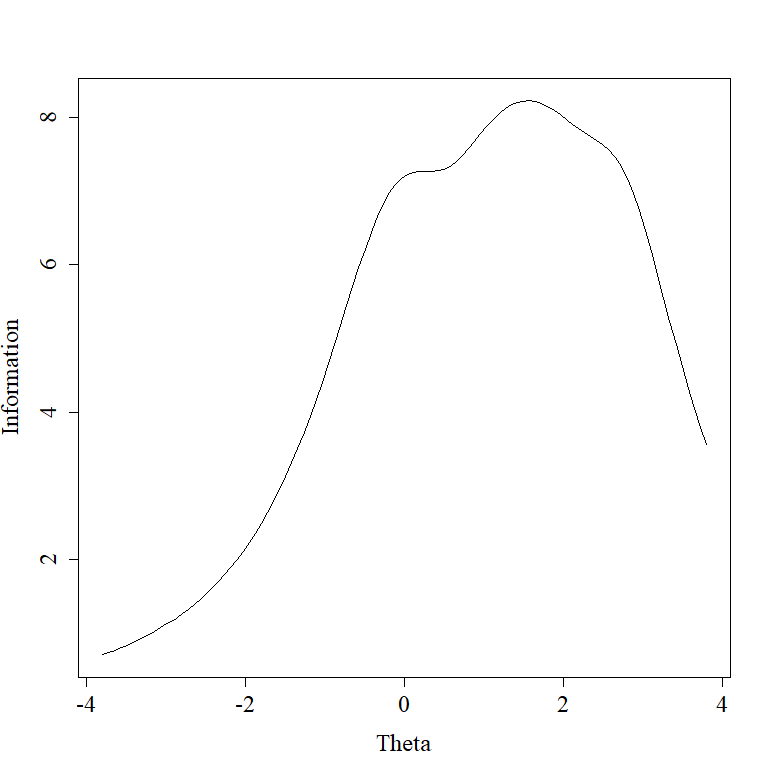
**Table 2**

*Fit Statistics for the Graded Response Models*

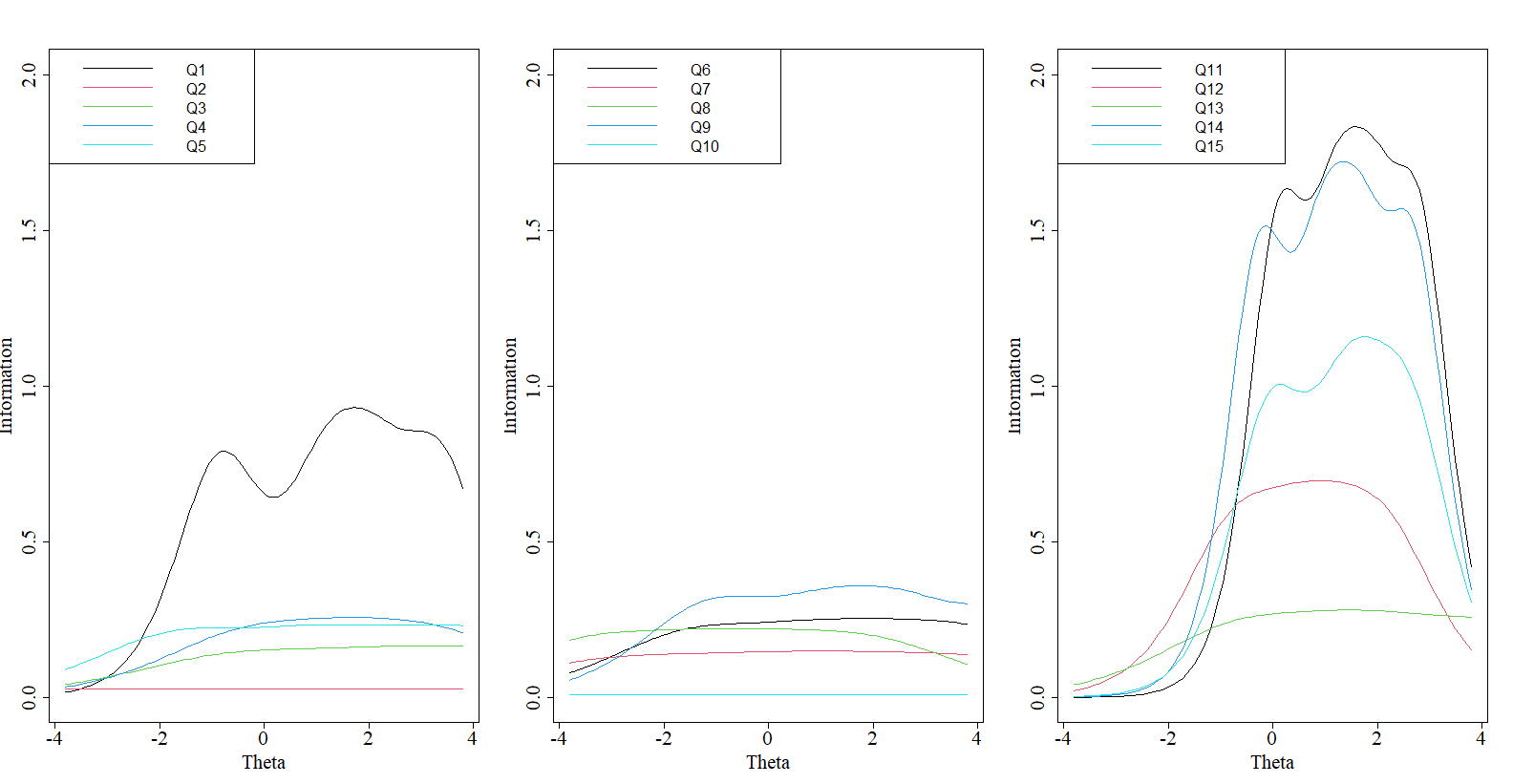
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AIC | BIC | Log-likelihood | | Ratio | df | P-value |
| General  Restricted | 5188.63  5118.15 | 5364.94  5334.92 | | -2533.32  -2484.07 | 98.48 | 14 | < .001 |

**Figure 3**

*The Survey’s Information for Different Levels of Theta. Values on the x-axis are z-scores*



**Figure 4**

*Information per Item, Sorted by Cogntivie (first panel from the left), Behavioral (middle panel), and Affective (third panel from the left) Facets*

A factor analysis study was conducted to investigate the factor structure of the test. A parallel analysis revealed that, at most, four factors could be retained in the model (see Figure 5). Exploratory factor analyses were then carried out, investigating the fit of models with 1-4 factors, the results of which are presented in Table 3.

**Figure 5**

*Scree Plot with Parallel Factor Analysis*

Chart, line chart

Description automatically generated

**Table 3**

*Fit Statistics for the Exploratory Factor Analyses*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # of factors | RMSEA | BIC | Chi-square | df | P-value |
| 1 | 0.082 | –269.24 | 170.90 | 90 | < .001 |
| 2 | 0.070 | –245.64 | 126.03 | 76 | < .001 |
| 3 | 0.051 | –223.14 | 84.95 | 63 | .034 |
| 4 | 0.043 | –185.76 | 63.65 | 51 | .110 |

According to the RMSEA index, only a 4-factor model displayed good fit, contrary to the hypothesis that a unidimensional factor structure is expected. The model’s standardized factor loadings were thus examined (see Table 4). Factors 3 and 4 each had only one item loading, indicating that these factors measured single traits, and not the latent variable of interest. Items 2, 3, 4, 7 and 10 failed to load on any factor as based on a common cutoff of 0.4, indicating that these items didnt capture sufficient factor variance (Brown, 2015).

A confirmatory factor analysis was then conducted with the following factor structure: Model 1: F1 =~ Q1 + Q5 + Q6 + Q9 + Q11 + Q12 + Q14 + Q15; F2 =~ Q11 + Q12 + Q13. The results showed that the new confirmatory model fit poorly (see Table 5), indicating that the factor structure might not be multidimensional after all. A 1-factor confirmatory model was fit, retaining only the items with sufficiently high factor loadings, i.e., greater than 0.6 (Model 2: F1 =~ Q1 + Q11 + Q12 + Q14 + Q15). The model displayed good fit (see Table 5), indicating that the previous five items should be retained.

**Table 4**

*4-factor Model Standardized Loadings, Cutoff of 0.4*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | F1 | F2 | F3 | F4 |
| Q1 | 0.575 |  |  |  |
| Q2 |  |  |  |  |
| Q3 |  |  |  |  |
| Q4 |  |  |  |  |
| Q5 | 0.401 |  |  |  |
| Q6 | 0.527 |  | 0.515 |  |
| Q7 |  |  |  |  |
| Q8 |  |  |  | 0.605 |
| Q9 | 0.457 |  |  |  |
| Q10 |  |  |  |  |
| Q11 | 0.616 | 0.436 |  |  |
| Q12 | 0.449 | 0.426 |  |  |
| Q13 |  | 0.997 |  |  |
| Q14 | 0.748 |  |  |  |
| Q15 | 0.675 |  |  |  |

**Table 5**

*Fit Statistics for the Confirmatory Factor Models*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | # of factors | RMSEA | CFI | Chi-square | df | *p*-value |
| Model 1 | 2 | 0.103 | 0.901 | 382.004 | 24 | < .001 |
| Model 2 | 1 | 0.029 | 0.997 | 5.542 | 5 | .353 |

**Discussion**

The present study has explored a newly constructed questionnaire for measuring social loneliness. Data was gathered and analyzed using Latent Variable Models (LVM). Both IRT and factor analyses found support for the unidimensional structure of social loneliness. This conclusion is in line with the present study’s hypothesis, previous research into loneliness, and the bi-dimensional conceptualization of loneliness (DiTommaso et al., 2004; Heinrich, & Gullone, 2006; DiTommaso et al., 1997). The results also indicated that the presented survey provides most information regarding above-average levels of social loneliness. Unexpectedly, affective rather than cognitive or behavioral items provided the utmost information about social loneliness. Possible explanations and applications for future research are discussed.

The study supported the advantages of Latent Variable Models compared to CTT. The insight regarding the unidimensional structure of social loneliness could not have been gained by CTT. Both IRT and factor analyses explictly assessed the number of dimensions underlying the relationships between the survy’s items using scree-plots and statistical assessmetns (e.g. paralell analysis). Furthermore, IRT analysis led to significant findings regarding the survey’s and item-specific information which are impossible to obtain by employing CTT. The survey has been shown to be most informative for above-average socially lonely people with affective items constituting the largest contribution to such information. These properties are beneficial since clinical diagnosis requires discriminating those suffering from high levels of loneliness from average levels. The above findings are not only useful for the construction of tests but manifest important implications to the theoretical conceptualization of social loneliness.

Contrary to the present study’s expectations, affective items undoubtedly yielded the most information on social loneliness. Whilst there was one cognitive item that also generated much information (“My friends don’t know who I really am”), none of the behavioral items did. This illustrated that the behavioral items were not highly effective measurements of social loneliness. This could be explained by the consistency bias which is a prominent phenomenon in self-reports (Leising, 2011). According to this phenomenon, people tend to base their assessment of their own interpersonal behavior on their general self-image even if that contradicts their actual behavior. This bias can lead to an inaccurate description of one’s own behavior which is also indicated by this research. Thus, in future research, more attention should be paid to the validity of behavioral items in self-reports. For example, such complications can be investigated by comparing objective and subjective reports of behaviors (e.g. exercise) linked to loneliness.

Items conveying the most and least amounts of information entail interesting limitations to the present study’s survey. The content of the maximally informative items (1, 11, 12, 14, & 15) regarded mostly friendships (items 1,11, & 12). This is an interesting observation that might be related to our choice of sample. Namely that the sample is limited to young adults between 18 and 28 years old of which the majority are students (68.4%). The social life of this group revolves more around friendships compared to older people, with younger adults spending more time with their friends (Nicolaisen et., 2017). This could be explained by the fact that most studying young adults aren't in the family formation phase. In this phase, young adults start spending less time with their friends and more time working on their careers and caring for their families (Pierret, 2006). Especially men report a decline in beyond-family activities with growing age (Field, 1999). Hence, older adults may agree to item one’s statement (“my friends don’t know who I really am”) without feeling lonely because of their close family ties. Therefore, the central role of friendships as social networks in the investigated sample might explain the positive association between highly informative items and content regarding friendships. As a result, the information of friendship-related items may be overetsimated comapred to the general population. Future studies could investigate this hypothesis by exploring the moderating role of friendships' significance in the relationship between social loneliness and age.

Taken together, the present study found considerable evidence for the unidimensional structure of social loneliness. It became apparent that the affective component of loneliness is most informative, whereas the behavioral component barely offered any information. The study brought forward a comparison of social loneliness’s components. This contribution to the test’s structure and facet-specific information may enhance the management of public health issues, specifically related to COVID-19 and loneliness. Measurement tools targeting the affective facet of social loneliness may be most appropriate in quantifying the prevalence of social loneliness in the wider population. Lastly, the advantages of LVM over CTT were clearly illustrated in the present study and might motivate future research to utilize LVM as a measurement model of psychological constructs.

**References**

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control, 19*(6), 716- 723. <https://doi.org/10.1007/978-1-4612-1694-0_16>

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

Borsboom, D. (2005).Measuring the Mind: Conceptual Issues in Contemporary Psychometrics.Cambridge University Press. <https://doi.org/10.1017/CBO9780511490026>

DiTommaso, E., Brannen, C., & Best, L. A. (2004). Measurement and validity characteristics of the short version of the social and emotional loneliness scale for adults. *Educational and Psychological Measurement*, *64*(1), 99-119. [https://doi.org/10.1177/0013164403258450](https://psycnet.apa.org/doi/10.1177/0013164403258450)

DiTommaso, E., & Spinner, B. (1997). Social and emotional loneliness: A re-examination of Weiss' typology of loneliness. *Personality and individual differences, 22*(3), 417-427. <https://doi.org/10.1016/S0191-8869(96)00204-8>

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>

Molenaar, D. (2021). Latent Variable Models, Lecture 1: Classical Test Theory. [Powerpoint Presentation]. <https://canvas.uva.nl/courses/21045/pages/lecture-1-classical-test-theory?module_item_id=832679>.

Norbury, R. (2021). Loneliness in the time of COVID. Chronobiology International, 1-3. <https://doi.org/10.1080/07420528.2021.1895201>

Perlman, D., Peplau, L. A., & Goldston, S. E. (1984). Loneliness research: A survey of empirical findings. *Preventing the harmful consequences of severe and persistent loneliness*, 13-46.

Revelle, W. (2020). psych: Procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois.

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, *17*, 1-25. <http://www.jstatsoft.org/v17/i05/>

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. <https://www.jstatsoft.org/v48/i02/>

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. <https://doi.org/10.1037/0022-3514.39.3.472>

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*,461–464. doi:10.1214/aos/1176344136

Sutin, A., Luchetti, M., & Terracciano, A. (2020). Has loneliness increased during COVID-19? Comment on “Loneliness: A signature mental health concern in the era of COVID-19.” *Psychiatry Research*, *291*, 113295–113295. <https://doi.org/10.1016/j.psychres.2020.113295>

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. Annual Meeting of the Psychometric Society, Iowa City, IA.

**Appendix**

data <- read.csv("cleaned\_data5.csv")

just.q <- data[,-c(1,2,3,4,5)] #create a data object of just the construct questions

N <- nrow(just.q)

windowsFonts(A = windowsFont("Times New Roman"))

library(lavaan)

library(psych)

library(ltm)

library(GPArotation)

################### UNI-DIMENSTIONALITY, EIGENVALUES AND POLYCHORIC CORELATION MATRIX ###################

pcor <- polychoric(just.q)$rho # Polychoric correlation matrix

E <- eigen(pcor)$values # Eigenvalues of poly matrix

plot(E,type ="b")

abline(h = 1, col = 2)

p <- fa.parallel(pcor,n.obs= nrow(just.q), show.legend = T ) #Parallel Analysis

plot(p$pc.values, family = "A", type ="b", lwd = 1.5, xlab = "Factors", ylab = "Eigenvalues", bty = "L",cex.lab= 1.5, cex.axis= 1.5, cex.main = 2)

lines(p$pc.sim, col = "blue", lwd = 1.5, lty = 2)

################### EQUALITY OF DISCRIMINATION ###################

nit= ncol(just.q) # number of items

sum.scores <- apply(just.q,1,sum) # sum score

item.test <- numeric() # item-test correlations

for(i in 1:nit){

item.test[i] <- cor(just.q[,i],sum.scores)

}

item.rest <- numeric() # item-rest correlations

for(i in 1:nit){

item.rest[i] <- cor(just.q[,i],sum.scores-just.q[,i])

}

Q <- c('Q1','Q2','Q3','Q4','Q5','Q6','Q7','Q8','Q9','Q10','Q11','Q12','Q13','Q14','Q15')

names(item.test) <- Q ; names(item.rest) <- Q #Assign question number to the correlations

barplot(item.test, xlab = 'item-test correlation')

barplot(item.rest, xlab = 'item-rest correlation')

#log likelihood test

general.m <- grm(just.q) #general model

restricted.m <- grm(just.q, constrained=TRUE) #restricted model (equal discrimination)

coef(general.m) ; coef(restricted.m)

anova(restricted.m, general.m) #compare the two models

################### ITEM & TEST INFORMATION ###################

# A plot of the item and test information functions.

res <- grm(just.q) #fit graded response model

layout(matrix(c(1:3),ncol=3,nrow=1)) #layout graphical device

Q <- c('Q1','Q2','Q3','Q4','Q5','Q6','Q7','Q8','Q9','Q10','Q11','Q12','Q13','Q14','Q15')

#Cognitive info

IIC <- plot(res,type="IIC", items=c(1:5),family = "A", lwd = 1.5,annot = F,xlab = "Theta", ylim = c(0,2),main = "",ylab = "Information", bty = "L",cex.lab= 1.5, cex.axis= 1.5, cex.main = 2)

legend("topleft", col =c(1:5), Q[1:5], lty = 1, cex = 1.5)

#Behavioral info

IIC <- plot(res,type="IIC", items=c(6:10),family = "A", lwd = 1.5,annot = F,

xlab = "Theta", ylim = c(0,2),main = "",ylab = "Information", bty = "L",cex.lab= 1.5, cex.axis= 1.5, cex.main = 2)

legend("topleft", col =c(1:5), Q[6:10], lty = 1, cex = 1.5)

#Affective info

IIC <- plot(res,type="IIC", items=c(11:15),family = "A", lwd = 1.5,annot = F,

xlab = "Theta", main = "",ylim = c(0,2),ylab = "Information", bty = "L",cex.lab= 1.5, cex.axis= 1.5, cex.main = 2)

legend("topleft", col =c(1:5), Q[11:15], lty = 1, cex = 1.5)

layout(1)

#Test information curves

TIC <- plot(res,type="IIC", items=0, family = "A",cex.lab= 1.5, cex.axis= 1.5,xlab = "Theta",main = "")

################### PIE CHARTS ###################

# Country of Origin

x <- c(sum(data$country == "Netherlands"), sum(data$country == "Germany"), sum(data$country == "Israel"), sum(data$country == "Russia")

, sum(data$country == "Serbia"), sum(data$country == "Turkey"), sum(data$country == "Italy"), sum(data$country == "Italy"))

pie(x, family = "A" ,

labels = round(100\*x/sum(x), 1), main = "Country of Origin", col = rainbow(length(2:10)))

legend("topright", c(unique(data$country)), cex = 1,

fill = rainbow(length(2:10)))

# Occupation

data$occupation = ifelse(data$occupation == "neither", "Neither",data$occupation) #Capitalise Neither

y <- c(sum(data$occupation == "Student"), sum(data$occupation == "Work"), sum(data$occupation == "Both"), sum(data$occupation == "Neither")

, sum(data$occupation == "Army"))

pie(y,

labels = round(100\*y/sum(y), 1), main = "Occupation", col = rainbow(length(2:7)), family = "A")

legend("topright", c(unique(data$occupation)), cex = 1,

fill = rainbow(length(2:7)))

################### FACTOR ANALYSIS ###################

### Factor extraction & selection ###

cor\_mat <- cor(just.q) #create correlation matrix

fa.parallel(cor\_mat, fa = "fa",n.obs=133)$fa.values #run parallel analysis

### EFA FACTORS 1:5 ###

efaList <- list()

RMSEAs <- numeric(5) ; BICs <- numeric(5) ; CHISQ <- numeric(5) ; DF <- numeric(5) ; PVAL <- numeric(5)

for(i in 1:5){

efaList[[i]] <- fa(r= cor\_mat, nfactors = i, n.obs = N, fm = "ml", rotate = 'none')

RMSEAs[i] <- efaList[[i]]$RMSEA[1]

BICs[i] <- efaList[[i]]$BIC

CHISQ[i] <- efaList[[i]]$STATISTIC

DF[i] <- efaList[[i]]$dof

PVAL[i] <- efaList[[i]]$PVAL

}

data.frame(RMSEAs,BICs,CHISQ,DF,PVAL) #Fit Statistics

### ONE FACTOR MODEL ###

model <-'f1=~Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10+Q11+Q12+Q13+Q14+Q15' #fit one factor model

fit\_factor <- cfa(model=model, data= just.q, meanstructure=TRUE, std.lv=FALSE)

summary(fit\_factor,header=T)

fitmeasures(fit\_factor)["rmsea"]

fitmeasures(fit\_factor)["srmr"]

### FOUR FACTOR MODEL LOADINGS ###

print(efaList[[4]]$loadings, cutoff = 0.4) #Check which questions load upon which factor

### TWO FACTOR SELECTIVE MODEL ###

just.select <- just.q[,-c(2,3,4,7,8,10)]

model.2 <-'

f1=~Q1+Q5+Q6+Q9+Q11+Q12+Q14+Q15

f2=~Q11+Q12+Q13

'

fit.2 <- cfa(model=model.2, data= just.select, meanstructure=TRUE, std.lv=FALSE)

summary(fit.2, fit.measures = T)

### ONE FACTOR SELECTIVE MODEL ###

model.1 <-'f1=~Q1+Q11+Q12+Q14+Q15'

data.1 <- just.q[,c(1,11,12,14,15)]

fit.1 <- cfa(model = model.1, data = data.1, meanstructure = T, std.lv = FALSE)

summary(fit.1, fit.measures = T)

## Descriptives ##

nrow(data) # Participants

mean(data$age) ; sd(data$age) # Age

length(which(data$gender == 'Woman')) # Women

length(which(data$gender == 'Man')) # Men