Erasmus+ Project Redefining nursing skills for ai and robotization in health car

Development of the Autonomous Technologies in Nursing Practice Scale: Preliminary Results

Olga LEZHNINA[1], Johanna MINK[2], Gábor KISMIHÓK[3], Adrienn Siket UJVÁRINÉ[4], Ádám KESZLER[5], Miklós ZRÍNYI[6], Stefan T. MOL[7], Sofija PAJIC [8], Martina HASSELER[9]

Funding: This work was supported by the EU programme Erasmus+ [grant agreement number: 2018-1-DE02-KA202-005101]

Conflict of Interest Statement: No conflict of interest has been declared by the authors.

Acknowledgements: We are grateful to all individuals who participated in the study.

The corresponding author: Olga LEZHNINA, Leibniz Information Centre for Science and Technology University Library, Hannover, Germany, Olga.Lezhnina@tib.eu

---

[1]TIB Leibniz Information Centre for Science and Technology University Library, Hannover, Germany

[2]University Hospital Heidelberg, Department of General Practice and Health Services Research, Heidelberg, Germany

[3]TIB Leibniz Information Centre for Science and Technology University Library, Hannover, Germany

[4]University of Debrecen, Faculty of Health, Institute of Health Studies, Department of Nursing Sciences, Nyíregyháza, Hungary

[5]University of Debrecen, Faculty of Economics and Business, Debrecen, Hungary

[6]University of Debrecen, Faculty of Health, Institute of Health Studies, Department of Nursing Sciences, Nyíregyháza, Hungary

[7]Amsterdam Business School, University of Amsterdam, Amsterdam, The Netherlands

[8]Amsterdam Business School, University of Amsterdam, Amsterdam, The Netherlands

[9]Ostfalia University of Applied Sciences, Faculty of Health Sciences, Wolfsburg, Germany

ABSTRACT

Aims: The aim of the study was to develop and psychometrically test the Autonomous Technologies in Nursing Practice scale for measuring self-assessed ability of nurses to deal with autonomous technologies in nursing practice.

Design: Cross-sectional survey design was used for data collection.

Methods: Item generation and selection was conducted based on cultural decentring principle by nursing experts in Germany, Hungary, and the Netherlands. The data was collected in Germany (n = 104) in June 2020 and Hungary (n = 700) in November 2019 - January 2020, the participants were nurses and nursing undergraduates. The best functioning items were selected in the process of item analysis and differential item functioning analysis, and psychometric properties of the resulting scale were evaluated.

Results: The resulting scale is a 16-item unidimensional instrument demonstrating good item functioning, homogeneity and internal consistency reliability.

Conclusion: The current version of the scale, which showed good psychometric properties, can be used by nursing professionals to evaluate their self-assessed ability to deal with autonomous technologies in nursing practice, or can be further developed by researchers.

Impact: This study addresses the problem of lack of psychometric instruments evaluating self-assessed ability of nurses to deal with autonomous technologies. A unidimensional scale, the Autonomous Technologies in Nursing Practice scale, was developed and psychometrically tested by the authors. The scale can be useful for nursing professionals, while researchers might benefit from our data analytic procedures outlined in the text and in our freely available R script.

Key words: nurses, autonomous technologies, instrument development, psychometric testing

INTRODUCTION

In 2016, the World Health Organization (WHO) stated that that nursing education and practice "are taking place in an era of progressive technological advancement, and its promotion is an important element for the future" (WHO, 2016, p. 11). In times of COVID-19, benefits and challenges brought about by autonomous technologies, such as robots and artificial intelligence, have become the topic of research and debate (Feizi et al., 2021). To engage with these technologies, nurses need competencies comprising knowledge, skills, values, attitudes, and performance (Cowan et al., 2007). In order to evaluate ability of health professionals to deal with autonomous technologies in nursing practice, valid and reliable instruments are required. As autonomous technologies are still rare in healthcare systems of many countries, objective measures of performance in interaction with technology would be less generally applicable than self-assessed measures of various competencies. According to contemporary guidelines in nursing research (Streiner & Kottner, 2014), reliability and validity are not immutable properties of a scale but depend on the sample and the circumstances, and thus, the development of the scale should be conducted as an iterative process. In this study, a few important steps are undertaken towards the development of an instrument measuring self-assessed ability to deal with autonomous technologies in nursing practice, the Autonomous Technologies in Nursing Practice (ATNP) scale. International significance of this work is determined by rapidly changing working environment, which introduces autonomous technologies to nursing practice all over the world. In addition, psychometric procedures that we describe in the text and implement in our R script can be helpful for nursing researchers developing a psychometric instrument and dealing with problems of measurement invariance typical for international research.

Background

Autonomous technologies are broadly defined as technologies which interact with the world without human help, such as robots or artificial intelligence (International Committee of the Red Cross, 2019). These technologies are rapidly changing nursing practice (Locsin & Ito, 2018), as they can be used in monitoring, for mobility aids, for reducing physical workload of nursing staff (Maalouf et al., 2018), or for assisting in decision-making processes (Liao et al., 2015). Although nurses, according to previous studies, appreciated technological support in their daily practice, especially in terms of monitoring and physical tasks (Lee et al., 2017), they also expressed fear that they would be replaced by such technologies as artificial intelligence (Abdullah & Fakieh, 2020). Nursing professionals should be increasingly involved into making decisions related to technological change in their working environment to be able to accept the new technology and interact with it effectively (Pepito & Locsin, 2018), as it was shown that psychological empowerment of nurses is important for their job satisfaction (Li at al., 2018). Valid and reliable psychometric instruments are required to measure ability of nurses to deal with new technology in terms of psychological attitudes, which are important for motivation, such as self-assessed competence (Deci & Ryan, 2008). However, existing instruments in nursing field measure nursing competencies in more general terms without focusing on technology (Nilsson et al., 2014), or address knowledge and skills related to nursing informatics (Phillips et al., 2017). Psychological aspects of human-technology interaction were elaborated by general psychological frameworks, such as the Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh et al., 2003). The UTAUT was applied to research in healthcare, and the findings suggested that psychological factors related to medical professionals' attitudes to technology, as operationalized in self-assessed measures, are important predictors of intention to use the technology (Kim et al., 2016; Maillet et al., 2015; Sharifian et al., 2014; Zhou et al., 2016). Psychometric scales capturing nurses' self-assessed ability to interact with autonomous technologies were not reported in previous studies, to the best knowledge of the authors.

THE STUDY

Aims

The aim of the study was to develop and psychometrically test the ATNP scale for measuring self-assessed ability of nurses to deal with autonomous technologies in nursing practice. Self-assessed ability is understood as one's belief that he or she is able to accomplish tasks specified in the questions.

Methodology

The instrument was developed in accordance with methodological recommendations for psychometric practices (Boateng et al., 2018; Streiner & Kottner, 2014; DeVellis, 2003). Item generation was conducted based on theoretical considerations related to existing competency frameworks in the field of nursing informatics (Mantas & Hasman, 2017; Nagle et al., 2017), information literacy of nurses (Technology Informatics Guiding Education Reform, n.d.), and the nursing process (Yura & Walsh, 1988). In literature, the nursing process was conceptualized in terms of assessing, planning, implementing, and evaluating (Yura & Walsh, 1988), while information literacy of nurses (Technology Informatics Guiding Education Reform, n.d.), as well as technological literacy in general, was described as "the ability to use, manage, assess, and understand technology" (International Technology Education Association, 2007, p. 9). Therefore, our framework included assessment, planning, implementing, and evaluating as areas of competencies required for interaction with autonomous technologies. These areas were covered by the items, which were generated and selected by nursing experts.

The authors applied the principle of cultural decentring (He & van de Vijver, 2012) to prevent a potential construct bias and aim for the cross cultural equivalence (Vandenberg & Lance, 2000; van de Vijver & Tanzer, 2004), and experts from local cultures were involved in preventing the method bias at the stage of project implementation. At the stage of item reduction, differential item functioning (DIF) analysis was applied to secure multicultural invariance of the resulting instrument.

Cognitive interviews with the target population (nursing professionals) were conducted, two in each country, by the nursing experts who are native speakers of the language. The interviews were structured in accordance with methodological recommendations to combine "think aloud" and "verbal probing" techniques (Willis, 2005). Face validity of 60 items generated at the previous stage was assessed. The results of the interviews were analysed by the nursing experts, which led to deletion of redundant and unclearly formulated items. When the cognitive interviews were processed by the consortium, 26 items were selected as the preliminary version of the ATNP scale to be administered in all participating countries (see Appendix A).

Instrument

The instrument used for the data collection was the 26-item ATNP scale (see Appendix A). The participants were asked to assess their ability to deal with specific tasks related to autonomous technologies. A brief definition of autonomous technology was given. Self-assessed ability to deal with autonomous technologies was measured with the 5-point Likert scale as follows: *to a very low degree* = 1, *to a low degree* = 2, *to a moderate degree* = 3, *to a high degree* = 4, *to a very high degree* = 5. A sample item (ATNP_12): "How able are you to share tasks with an autonomous technology in nursing practice?" The scale was administered in local languages (German, Hungarian, and Dutch). In this paper, the English translation is presented.

Ethical considerations

Ethical approval for the study was obtained in participating countries in accordance with their respective ethical regulations. For data collection in Germany, the study received ethical approval from the ethics committee of the medical faculty of the University Hospital Heidelberg. For data collection in Hungary, the study received ethical approval from the Dean of University of Debrecen Faculty of Health (DEEK) and from director and nursing directors of three participating hospitals: University of Debrecen Clinical Center, Debrecen; Szabolcs-Szatmár Bereg County hospital, Nyíregyháza; and Felsőszabolcsi Hospital, Kisvárda. For data collection in the Netherlands, the study received ethical approval from the Economics and Business Ethical Committee (EBEC) of the Faculty of Economics and Business of the University of Amsterdam.

Data collection

In Germany, paper-based data collection started in February 2020 at the University Hospital Heidelberg but was interrupted due to COVID-19 restrictions, and online data collection started in June 2020. The link to the online questionnaire was sent to the nursing directors in each department of the University Hospital Heidelberg and the nursing school Heidelberg to be distributed among nurses and nursing undergraduates. In addition, the participants were invited via social media. For online data collection, LimeSurvey software (Limesurvey, n.d.) was used. In Hungary, paper-based data collection was conducted in November 2019 - January 2020 in three hospitals: University of Debrecen Clinical Center, Debrecen; Szabolcs-Szatmár Bereg County hospital, Nyíregyháza; and Felsőszabolcsi Hospital, Kisvárda. For processing the questionnaires, EvaSys software (EvaSys Survey Automation Suite, 2019) was used. In the Netherlands, the COVD-19 situation thwarted data collection. Of the four hospitals with whom access was negotiated, two ended up declining participation at the last moment, and the two hospitals where the survey was distributed yielded an insufficient response (n = 14 and n = 2) to warrant any meaningful data analysis. Other organizations that were contacted (including a nursing journal and two professional nursing organizations) also declined participation either by not responding to our request for participation or citing Corona.

Participants

The Hungarian sample (n = 700) and the German sample (n = 104) were used for the analysis, with the total sample of N = 804. Demographic statistics of the samples are given in Table 1.

Table 1: Demographic Statistics of the Samples

Sample	Total	Gender			Status		Age
		Female	Male	Unspecified	Nurses	Undergraduates	
German	104	82	20	2	79	16	38(13)
Hungarian	700	616	68	16	482	218	37(11)

Note. In the German sample, nine participants did not specify whether they were nurses or undergraduates. The age is given as the mean value with the standard deviation in brackets.

Data analysis

The data analysis was conducted with R, version 4.0.2 (R Core Team, 2020). The R script is available in Supplemental Materials. The items that did not meet the criteria specified below were removed from the scale, and the resulting instrument was explored. In terms of item reduction, statistical power considerations were taken into account, so that the results on the German sample (n = 104) informed our decisions less substantially than the results obtained with the total sample (N = 804) or the Hungarian sample (n = 700). Domain knowledge considerations influenced the final decisions on item reduction: The nursing experts suggested keeping at least one item for each of the intended areas (assessment, planning, implementation, and evaluation) in the resulting scale.

Pre-processing of the data included exploring and visualising missing data with the package VIM (Kowarik & Templ, 2016). For imputation, the random forest algorithm, which was shown to outperform other commonly used methods (Golino & Gomes, 2016; Waljee et al., 2013), was used. The implementation of the algorithm from the package missForest (Stekhoven, 2013) was applied, which is more robust than other implementations (Tang & Ishwaran, 2017).

Validation of the scale was conducted by means of the Classical Test Theory (CTT) and the Item Response Theory (IRT) in accordance with methodological guidelines (Boateng et al., 2018; Dima, 2018; Streiner & Kottner, 2014; Terwee et al., 2007). In the frame of CTT, the items were checked for (i) excessively low (< .30) or high (>.90) correlations, or negative correlations; (ii) excessively low (< .05) or high (> .95) frequency of endorsement, and (iii) floor and ceiling effects with the cut-off of .15, as recommended by Boateng et al., (2018) and Terwee et al. (2007).

DIF was analysed in accordance with the guidelines suggested by Fischer and Karl (2019). We used three procedures to compare their results and remove the items that showed the differential item bias: (i) Multi-Group Confirmatory Factor Analysis (MGCFA) with the special attention to differences in mean and covariance structures (dMACs), (ii) Multi-Group Factor Analysis Alignment (hereinafter alignment, see Asparouhov & Muthén, 2014), and (iii) DIF using ordinal regression in the frame of IRT (hereinafter lordif). With MGCFA, the following hypotheses were tested: The set of items evokes the same conceptual framework in defining the latent construct in each comparison group; the regression slopes linking the manifest measures to the underlying construct are invariant across groups; the regression intercepts linking the manifest measures to the underlying construct are invariant across groups; the CFA model holds equivalently and assumes a common form across groups; unique variances for like manifest measures are invariant across groups; and variances and covariances among the latent variables are invariant across groups (Vandenberg & Lance, 2000). As the classical approach to MGCFA does not estimate the effect size of item bias, dMACs as an effect size measure were calculated with the package ccpsyc (Jo-Karl, 2020). Alignment was conducted with the package sirt (Robitzsch, 2020), and the cut-offs for invariance tolerance in loadings and intercepts were used as suggested by Fischer and Karl (2019). For lordif, we chose the package lordif (Choi et al., 2011), which conducts ordinal logistic regression analysis and applies the graded response model for IRT trait estimates. Monte-Carlo approach, which is incorporated in the package, was used to determine empirical thresholds for DIF detection. Visualisations provided by the package lordif (Choi et al., 2011), which include trait distribution plots and Item Characteristic Curves (ICCs), were explored. After DIF analysis and item removal, the further procedures were conducted on the whole sample. Item discrimination was explored in the frame of CTT and IRT as recommended by Boateng et al. (2018), and information plots were built.

Psychometric properties of the resulting scale were evaluated. Dimensionality was checked with exploratory factor analysis (EFA), Very Simple Structure Analysis (VSS) and Item Cluster Analysis (ICLUST). EFA was conducted in accordance with recommendations for the best practices in EFA by Howard (2016). Assumptions for EFA were checked with Barlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. Principal axis factoring was employed as a factor analytic method. For factor retention, we combined scree plot analysis, parallel analysis, and Velicer's minimum average partial (MAP) test. With VSS, the fit of increasingly complex models was assessed. With ICLUST, hierarchical clustering was applied to the items, and the cluster diagram was plotted. In the frame of non-parametric IRT, Mokken Scaling Analysis (MSA) was used to explore dimensionality of the scale, as it is an effective unbiased method which does not require multivariate normality assumption (Dima, 2018). Assumptions of homogeneity, monotonicity, local independence, and invariant item ordering were checked. Automated Item Selection Procedure (AISP) with increasing thresholds of homogeneity and the cut-off .30 for homogeneity was used as suggested by Dima (2018). To explore the internal consistency of the scale, Cronbach's alpha with the confidence interval and McDonalds omega with the confidence interval were used (Crutzen & Peters, 2017).

In the frame of the nomological network evaluation, factors Internet Affect and Internet Exhilaration of the General Internet Attitude Scale (INT; Joyce & Kirakowski, 2015) and the shortened 10-item version of the Big Five Scale (BFI; Rammstedt et al., 2013) were used to ensure that the self-assessed ability in ATNP forms a separate construct distinct from general technology (Internet) acceptance and personality dimensions. Agglomerative hierarchical clustering of variables was applied for this purpose (Çokluk et al., 2010; Farelly et al., 2017), which starts with each variable forming a separate cluster, then the number of clusters is reduced based on a similarity criterion, until all variables are agglomerated in a single cluster. This method allows uniting variables that are close to each other in terms of containing similar information (Chavent et al., 2017), and thus, the nomological network can be clearly presented.

RESULTS

In the German sample, there was no missing data in the scales (ATNP, INT, and BFI). In the Hungarian sample, there was missing data (.012 of the scales dataset, .005 of the ATNP subset), which was explored and visualised. The aggregation plot for the Hungarian sample (see Figure 1) shows no clear patterns in missingness, thus indicating that the data was missing at random. The missing data was imputed with the random forest algorithm, and in further analysis, the sample obtained by the imputation was used.

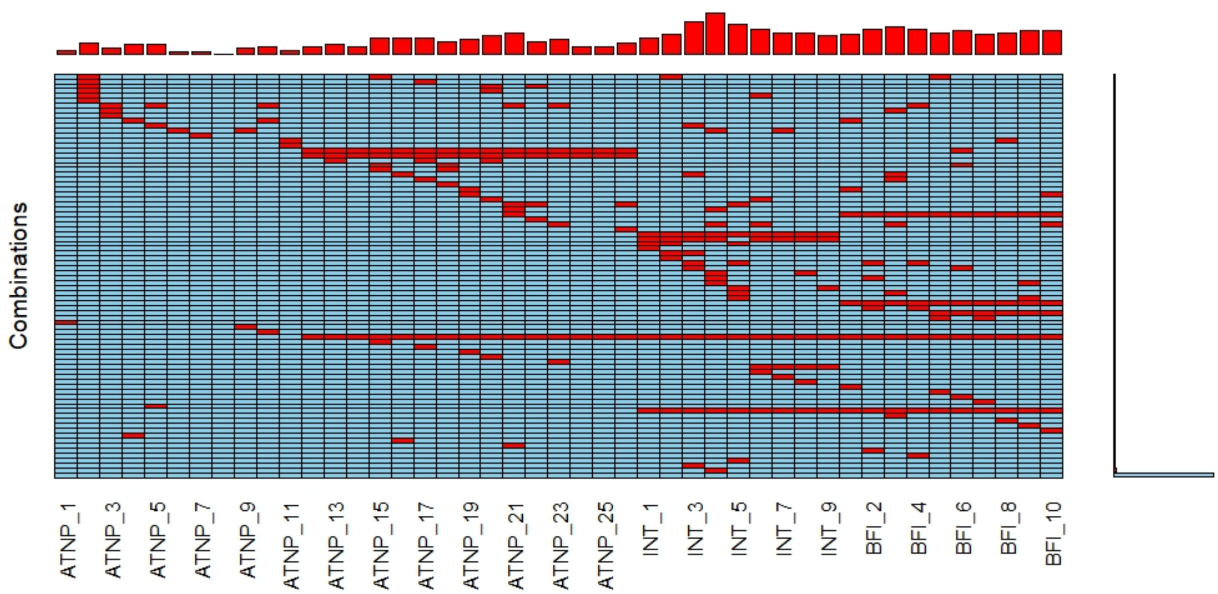

Figure 1: Aggregation Plot for Missing Data in the Hungarian Sample

Note. Red fields = missing data, blue fields = observed data. Three scales (ATNP, BFI, and INT) are included. Each row in the plot presents a certain combination of missing and present data. The barplot on the right shows the number of cases for a specific combination of missing and present data. The rows are sorted based on the number of cases.

In both samples, negative correlations between the items were not detected. Correlations were in the range from .37 to .86 in the German sample and from .31 to .73 in the Hungarian sample, with the exception of the correlation of .27 between items ATNP_5 and ATNP_15 in the Hungarian sample. Excessively high (more than .95) frequency of endorsement in any of response categories was not detected in either of the samples. Excessively low (less than .05) frequency of endorsement in any of response categories was detected in the following items: The German sample: ATNP_3, ATNP_5, ATNP_6, and ATNP_7. The Hungarian sample: ATNP_5 and ATNP_25. For floor and ceiling effects, the cut-off of .15 was taken. The floor effect was detected in the following items: The German sample: ATNP_1 and ATNP_ 23. The Hungarian sample: ATNP_15. The ceiling effect was detected in the following items: The German sample: ATNP_5 and ATNP_25. The Hungarian sample: ATNP_2, ATNP_3, ATNP_5, ATNP_6, ATNP_7, ATNP_18, and ATNP_25. After this stage, items ATNP_2, ATNP_3, ATNP_5, ATNP_6, ATNP_7, ATNP_15, ATNP_18, and ATNP_25, which showed insufficient performance on the Hungarian sample, were removed from further analysis.

DIF for the Hungarian and the German samples was analysed. First, the CFA model was fit to both groups, and dMACs were calculated. Two items had the largest dMACs: for item 2 (ATNP_4) it was .089, and for item 1 (ATNP_1) it was .070, while for the rest of items it was in the range from .007 to .048. Alignment, which was conducted with the sirt package and the cut-offs suggested by Fischer and Karl (2019), did not show DIF in the groups. DIF with the package lordif was conducted with alpha level .05, with 200 replications in the Monte-Carlo simulation. According to the trait distributions plot (Figure 2), there was a broad overlap in the distributions of the trait as shown by smoothed histograms of ATNP scores.

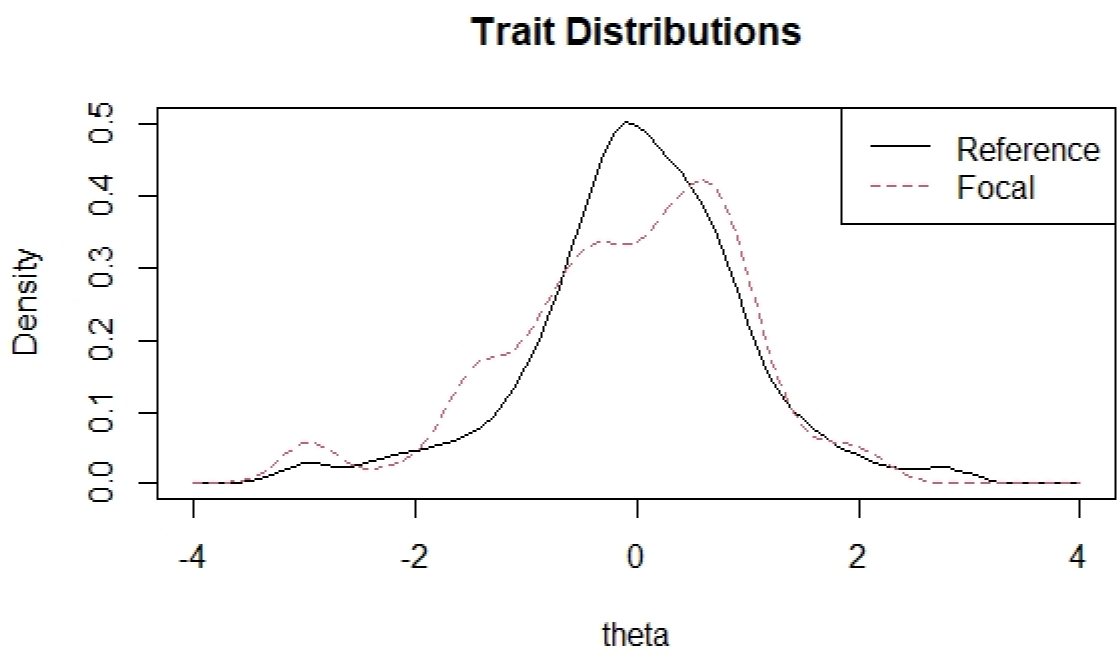

Figure 2: Trait Distributions Plot for the Hungarian and the German Samples

Note. The reference group is the Hungarian sample, and the focal group is the German sample.

Two items showed differential functioning in the Hungarian and the German samples: item 2 (ATNP 4) and item 15 (ATNP_22). Sample DIF plot for item ATNP_4 illustrates the functioning of the item (see Figure 3), and the plot for item ATNP_22 reveals the same pattern.

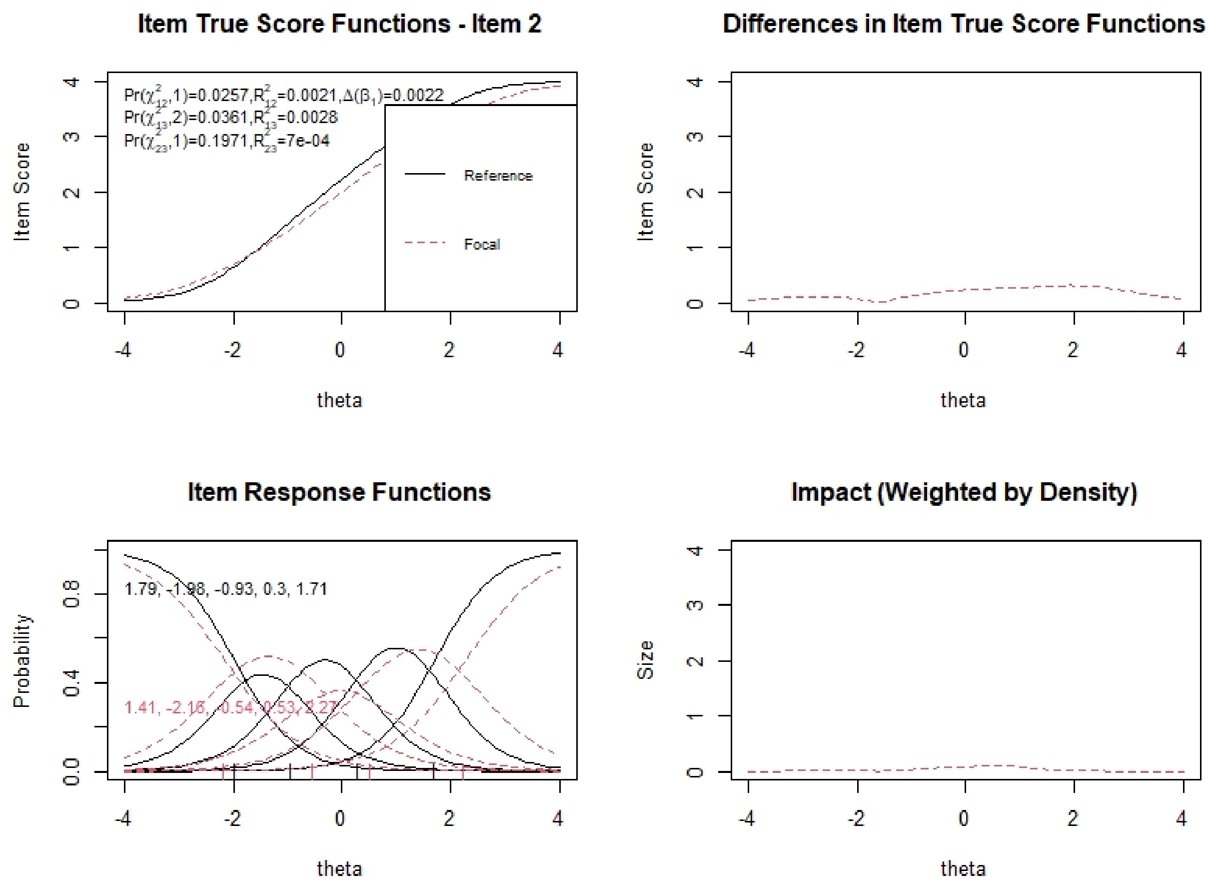

Figure 3: DIF Plots for ATNP_4 in the Hungarian and the German Samples

Note. The reference group is the Hungarian sample, and the focal group is the German sample.

The DIF plot includes the ICCs for the two samples (the upper-left graph), the absolute difference between the ICCs (the upper-right graph), this difference weighted by the score distribution for the focal group (the lower-right graph), and item response functions for the two groups based on threshold values by the group (the lower-left graph). We can see that the item showed the non-uniform DIF, and the difference is increasing with the increase in ATNP scores, but the impact of the difference is still very low. Test characteristic curves (see Figure 4) show the expected total scores at each level of ATNP (theta), which on the left graph is shown for all items, and on the right graph for items with DIF.

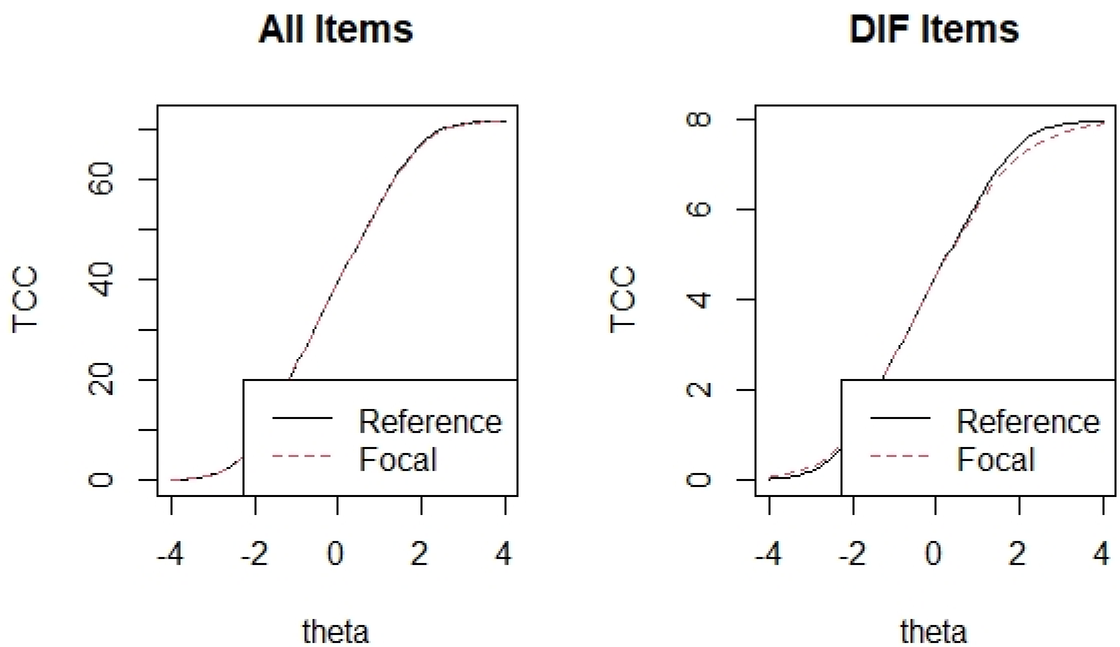

Figure 4: Impact of DIF Items on Test Characteristic Curves

Note. The reference group is the Hungarian sample, and the focal group is the German sample.

We can conclude that there was minimal difference in the total expected score at any level for the Hungarian and the German samples. This is confirmed by the individual-level DIF impact plot (see Figure 5): The box plot on the left shows that the interquartile range, which represents the middle 50% of the differences between scores ignoring DIF and scores accounting for DIF, has the median of approximately zero.

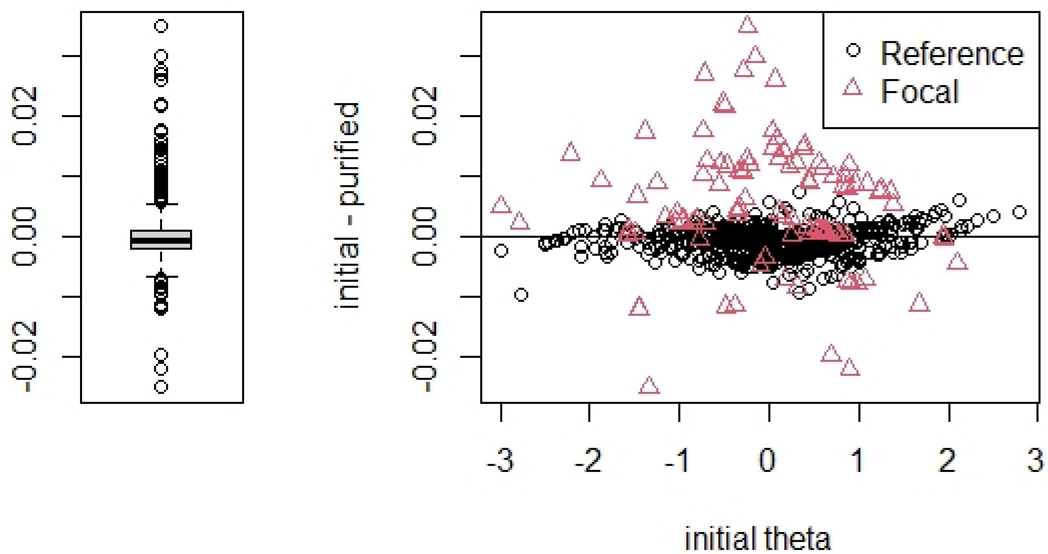

Figure 5: Individual-Level DIF Impact

Note. The reference group is the Hungarian sample, and the focal group is the German sample.

In the graph on the right, the differences are plotted against scores ignoring DIF (initial theta). It can be seen that for the reference group (the Hungarian sample), accounting for DIF did not change the scores, while for the focal group (the German sample) accounting for DIF led to either lower or higher scores in comparison to ignoring DIF. Overall, the measurement invariance for the two countries was established by the procedures. Based on the results of DIF analysis, we removed item ATNP_4, as it was flagged by lordif and had the largest dMACs.

Further analysis was conducted on the whole sample, the Hungarian and the German together (N = 804). Item discrimination was explored. In the frame of CTT, item ATNP_1 had the lowest discrimination value (.62) and was followed by items ATNP_9 (.67) and ATNP_10 (.70). The same result was obtained with the IRT analysis, item ATNP_1 had the lowest discrimination value (.90) and was followed by items ATNP_9 (1.03) and ATNP_10 (1.10).

Information plots for the items showed the same pattern (see Figure 6): items ATNP_1, ATNP_9 and ATNP_10 performed worse than others. Due to theoretical considerations (items from all four areas covered by the framework should be presented in the instrument), the nursing experts suggested keeping items ATNP_1 and ATNP_9, and item ATNP_10 was removed from the scale.

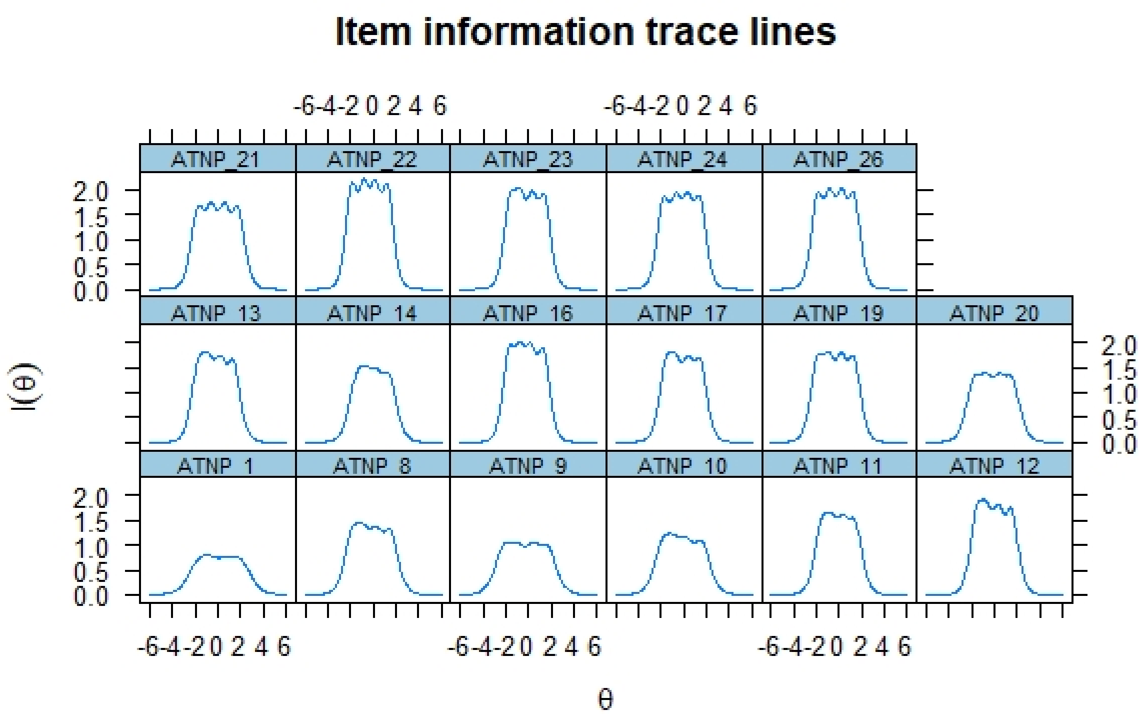

Figure 6: Information Trace Lines for ATNP Items

To summarise, 10 items were removed from the scale in the process of item reduction. Items ATNP_2, ATNP_3, ATNP_5, ATNP_6, ATNP_7, ATNP_15, ATNP_18, and ATNP_25 showed insufficient performance in terms of frequency of endorsement, ceiling effect, or floor effect; item ATNP_4 was flagged by DIF analysis; and item ATNP_10 was removed in the result of discrimination analysis. The remaining 16 items were kept in the scale (see Limitations section).

The resulting instrument, hereinafter called ATNP, a 16-item scale (see Appendix B), was explored further. EFA was conducted as described in the previous section. The assumptions for the EFA were met: The KMO of the scale was .97, and KMO values for items above .96, which indicated the sampling adequacy; and Bartlett's test of sphericity, χ^2 (120) = 8891.36, was significant with p < .001. Scree plot analysis, parallel analysis, and Velicer's MAP test indicated the one-factor solution (see Figure 7).

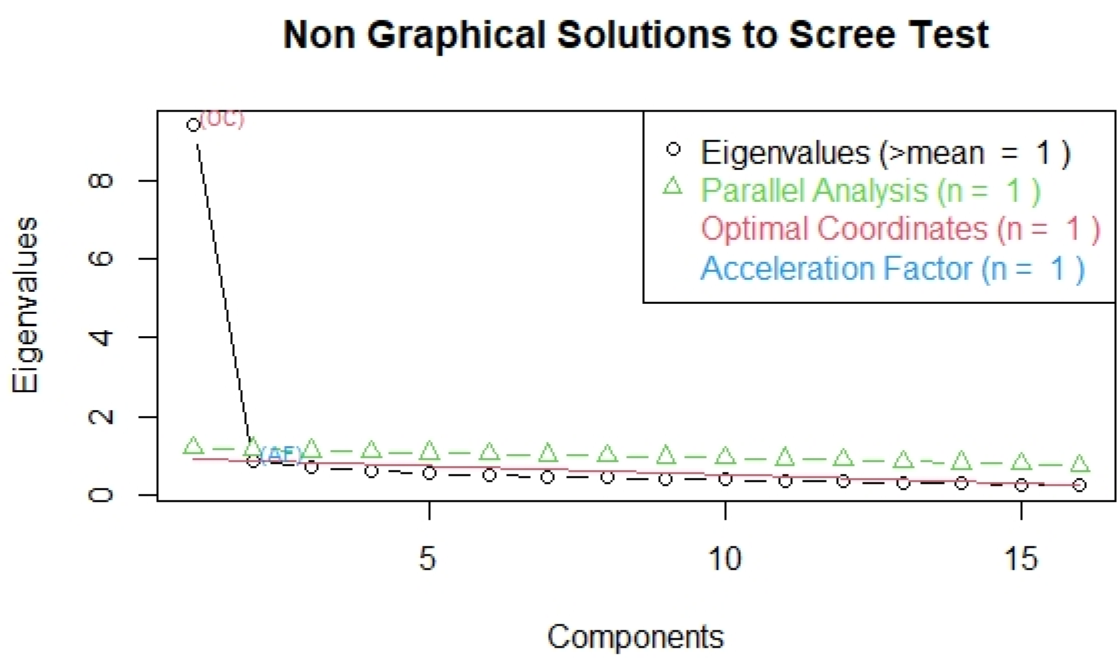

Figure 7: EFA Plot for ATNP Scale

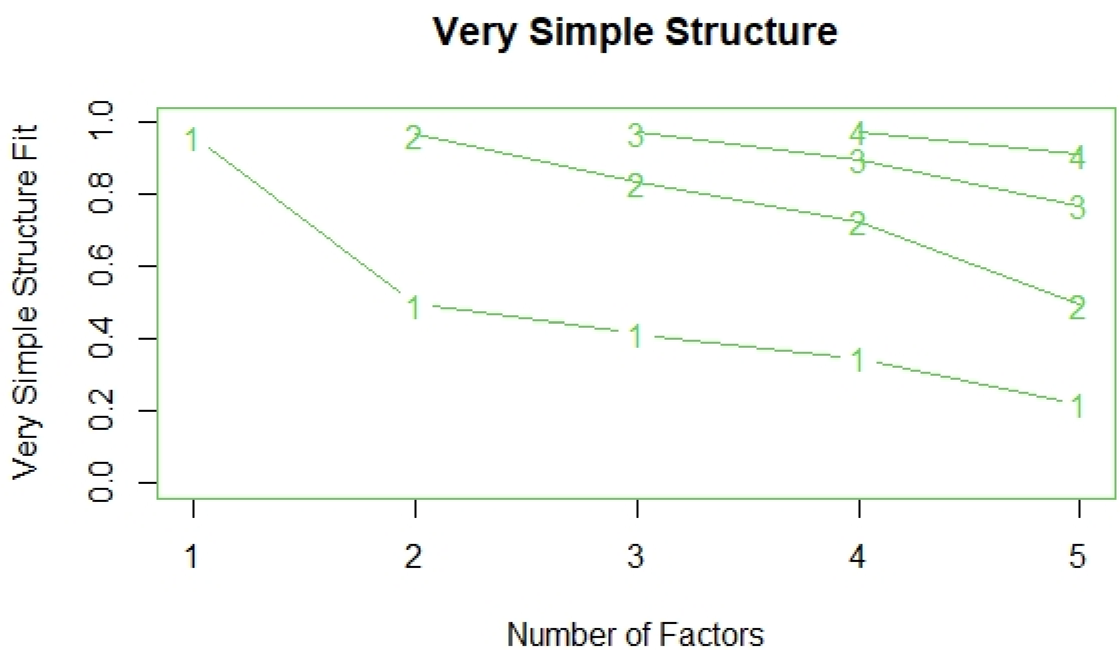

Figure 8: Very Simple Structure Plot for ATNP Scale

VSS complexity 1 reached the maximum of .96 with one factor (see Figure 8). The same conclusions were obtained with ICLUST (see Figure 9).

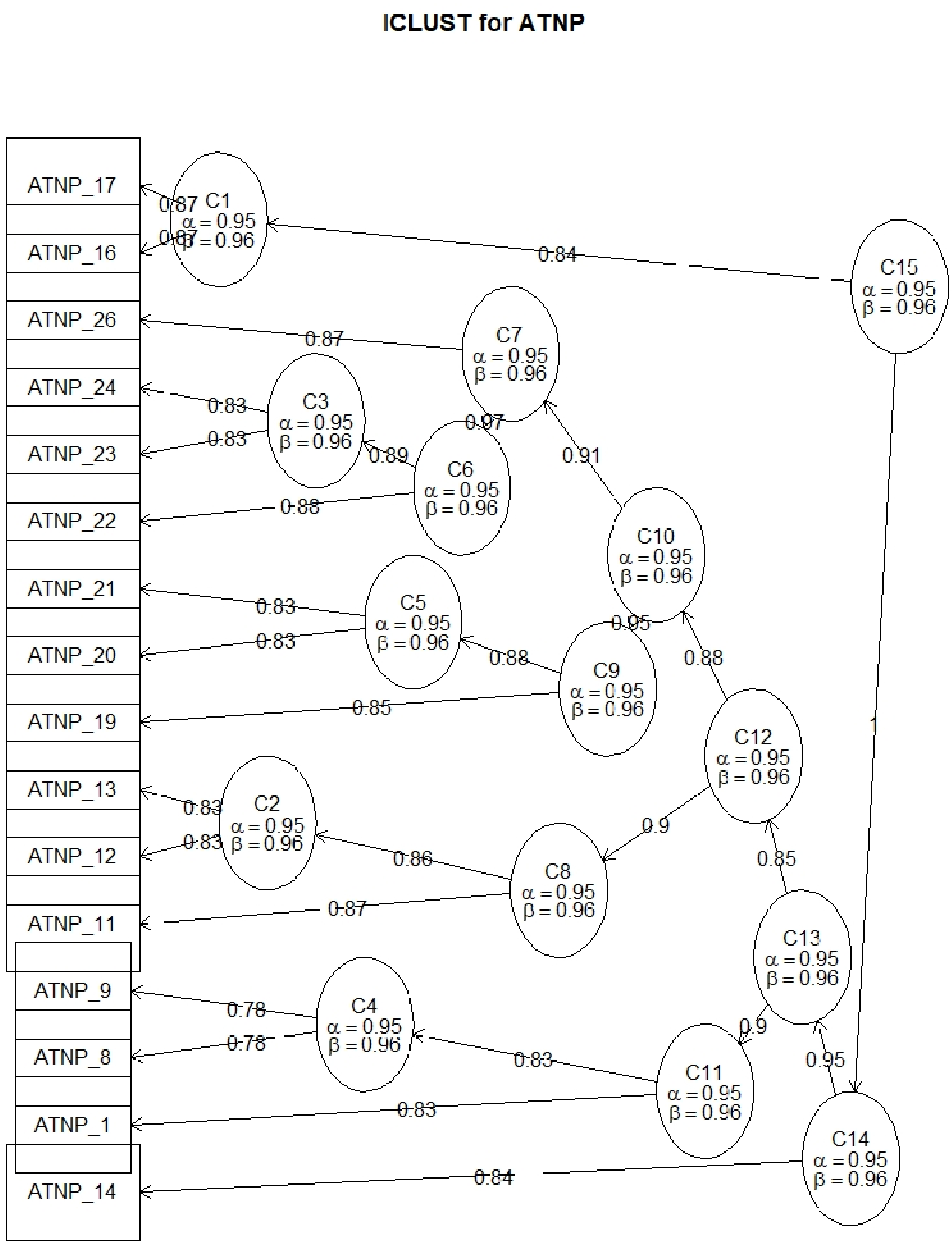

Figure 9: ICLUST Plot for ATNP Scale

Thus, the ATNP scale could be considered a unidimensional instrument. Cronbach's alpha of the scale was .95 [.95, .96], and McDonald's omega was .94 [.93, .94].

In the frame of non-parametric IRT, the ATNP scale was unidimensional: AISP showed good scalability of the items at the threshold level of homogeneity as high as .50, while the minimal level recommended by Dima (2018) is .30. The scale had good homogeneity H = .59 (SE = .017), and values of homogeneity for items were in the range from .51 to .62. Monotonicity test with the default minisize of n = 80 gave criterion values that were not equal to zero in seven items, but they were lower than the threshold of .40, and the assumption was met. The local independence assumption was not met by the most items, and the assumption of non-intersecting item response functions (invariant item ordering) was violated for six items of the scale. Thus, in terms of local independence and invariant item ordering, the instrument is still to be improved. ICCs of the items were peaked and dispersed across all levels of the latent trait (see Figure 10).

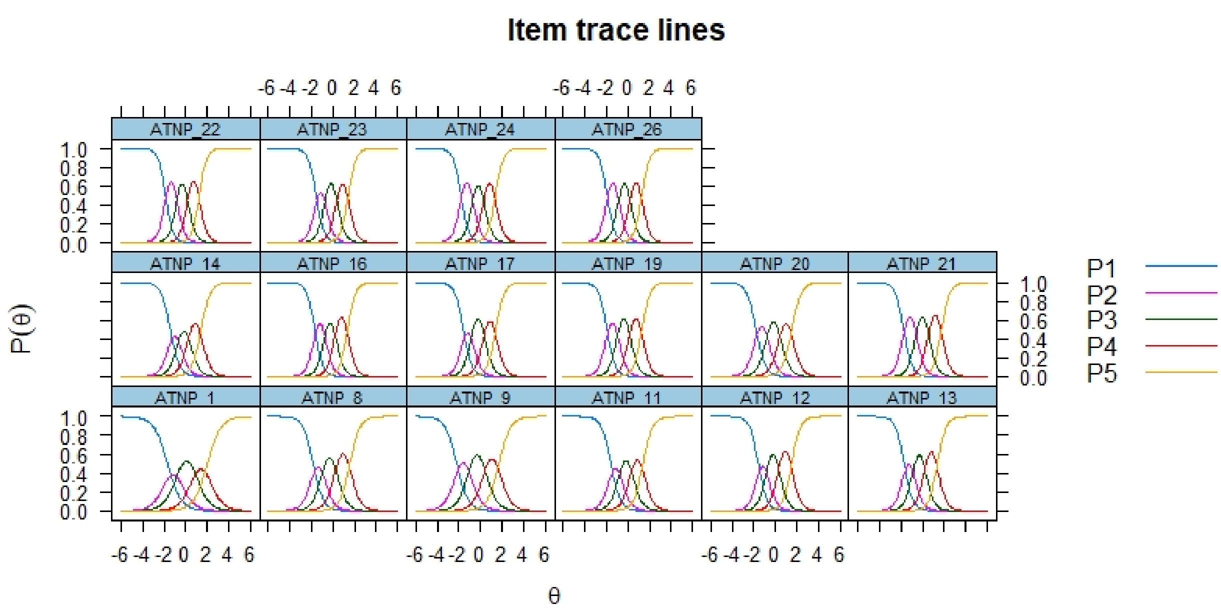

Figure 10: Item Trace Lines for ATNP Items

Descriptive statistics for the ATNP items are given in Table 2. It can be seen that the values of skew and kurtosis are acceptable.

Table 2: ATNP Items With Means, Standard Deviations, Standard Errors, Skew, and Kurtosis

Items	M	SD	Skew	Kurtosis	SE
ATNP_1	2.93	1.10	- 0.03	- 0.51	0.04
ATNP_8	3.22	1.07	- 0.32	- 0.38	0.04
ATNP_9	3.21	1.00	- 0.20	- 0.21	0.04
ATNP_11	3.15	1.15	- 0.22	- 0.64	0.04
ATNP_12	3.14	1.09	- 0.30	- 0.47	0.04
ATNP_13	3.25	1.09	- 0.35	- 0.43	0.04
ATNP_14	3.11	1.18	- 0.22	- 0.78	0.04
ATNP_16	3.22	1.11	- 0.29	- 0.57	0.04
ATNP_17	3.15	1.11	- 0.26	- 0.48	0.04
ATNP_19	3.31	1.06	- 0.30	- 0.39	0.04
ATNP_20	3.13	1.08	- 0.16	- 0.49	0.04
ATNP_21	3.01	1.01	- 0.09	- 0.40	0.04
ATNP_22	3.29	1.06	- 0.28	- 0.46	0.04
ATNP_23	3.14	1.09	- 0.20	- 0.51	0.04
ATNP_24	3.18	1.07	- 0.18	- 0.52	0.04
ATNP_26	3.29	1.05	- 0.27	- 0.42	0.04

Barplots for ATNP items visualising frequencies of endorsement show that all response options are represented in the data (see Figure 11).

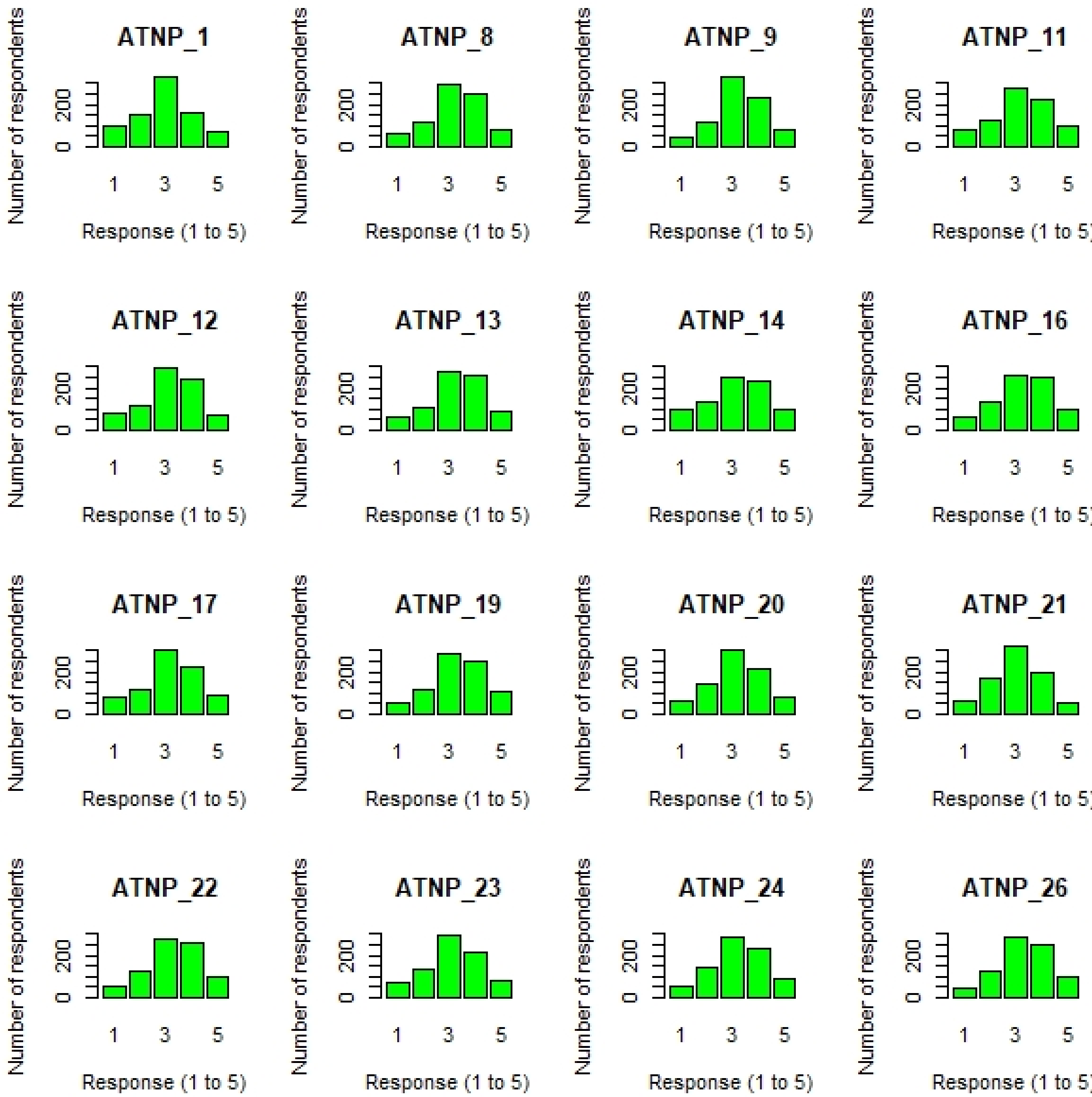

Figure 11: Barplots for ATNP Items

The dendrogram for all items from ATNP, BFI, and INT scales shows that the ATNP items form a separate cluster from BFI and INT items (see Figure 12). Stability of partitions was checked with the default of 100 bootstrap samples (see Figure 13). Most stable partitions with three clusters (see Figure 14) and eight clusters (see Figure 15) were explored. The gain in cohesion was 21.06% for the three-cluster partition and 43.13% for the eight-cluster partition. ATNP items in both partitions formed a single cluster separate from BFI and INT items.

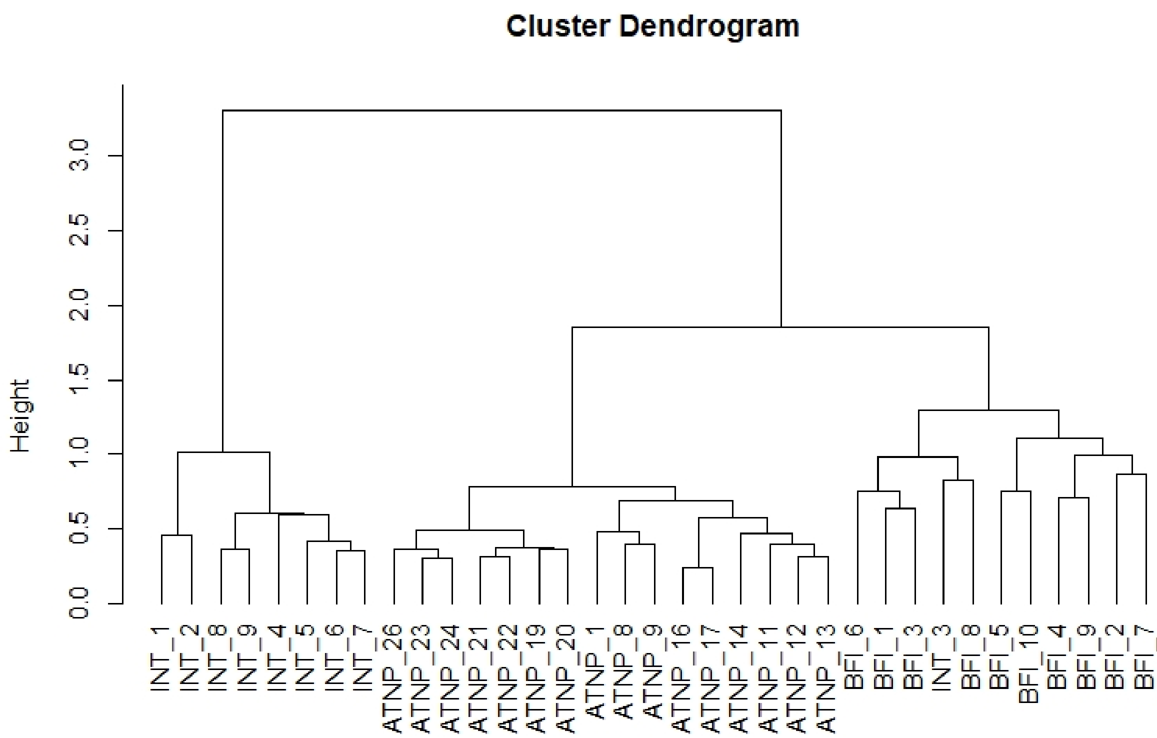

Figure 12: Cluster Dendrogram for ATNP, BFI and INT Items

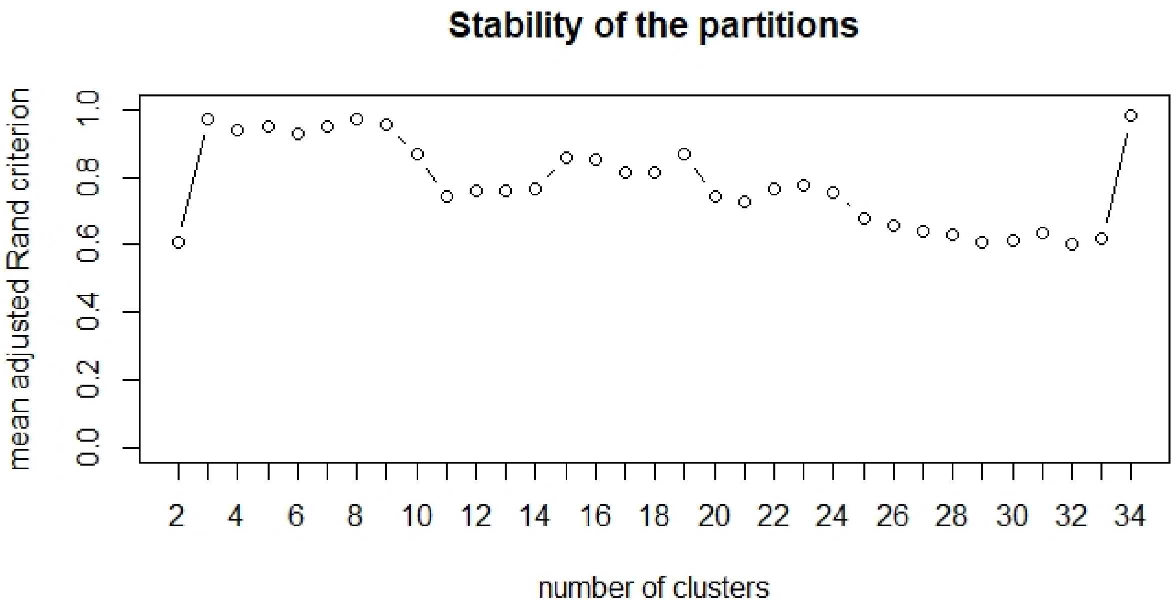

Figure 13: Stability of Partitions for Hierarchical Clustering of ATNP, BFI and INT Items

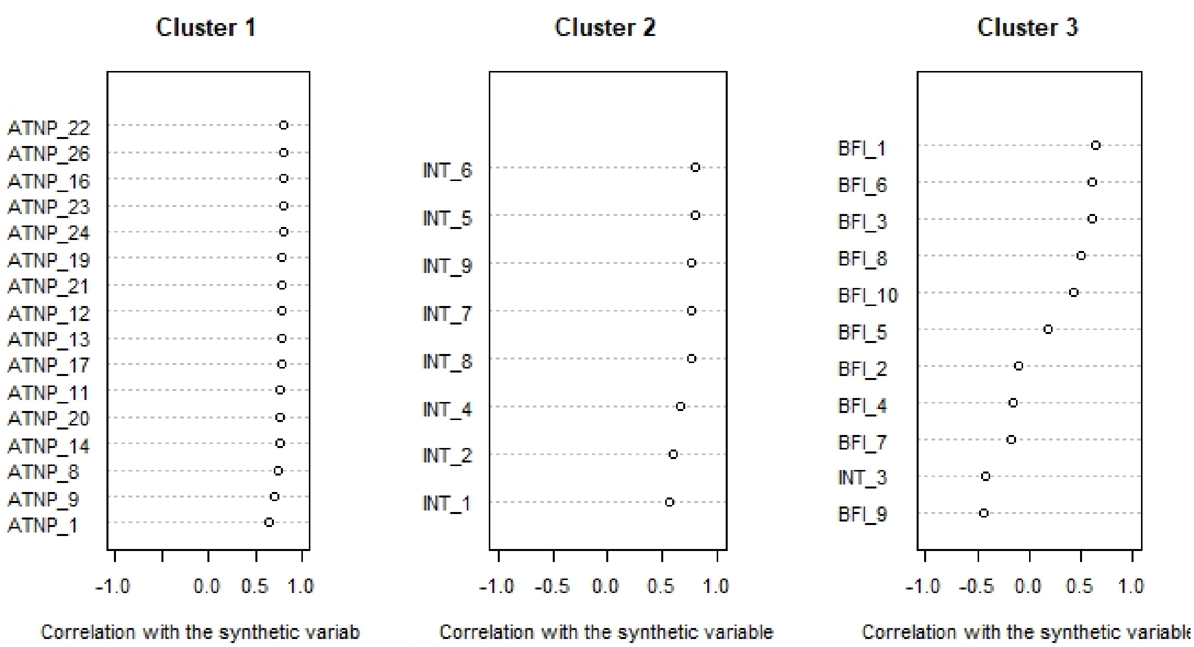

Figure 14: Three-Cluster Partition of ATNP, BFI and INT Items

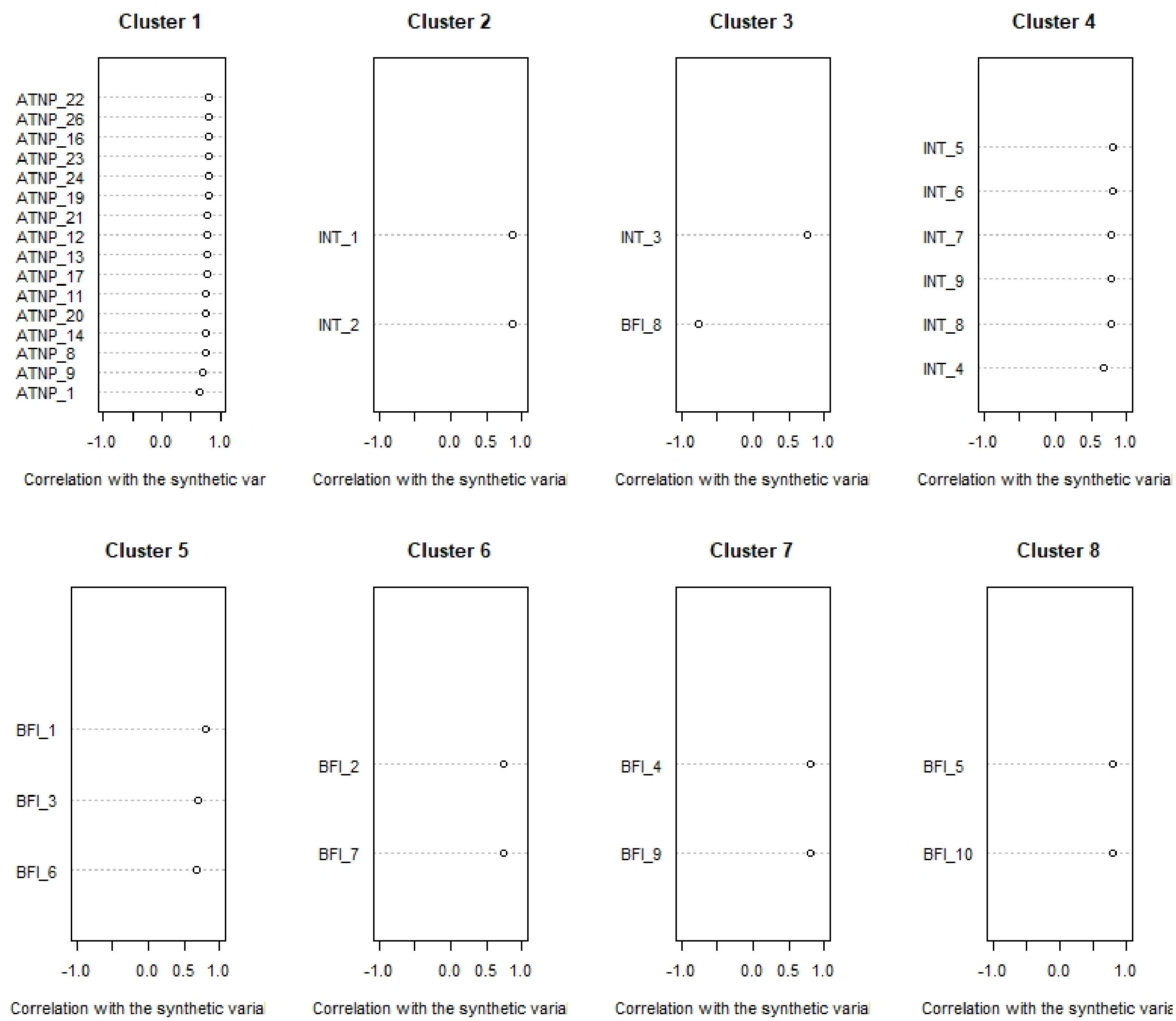

Figure 15: Eight-Cluster Partition of ATNP, BFI and INT Items

DISCUSSION

This study addressed the problem of lack of psychometric instruments evaluating self-assessed ability of nurses and nursing undergraduates to deal with autonomous technologies. We developed the ATNP scale measuring self-assessed ability of nurses to deal with autonomous technologies in nursing practice. ATNP is a 16-item unidimensional instrument, which showed good item functioning, homogeneity H = .59 (SE = .017), and internal consistency reliability (alpha = .95 [.95, .96], omega = .94 [.93, .94]). Despite some limitations outlined in the next section, the current version of ATNP can be used as a measure of self-assessed ability to deal with autonomous technologies in nursing practice, as it showed good reliability and validity in the frame of this study. In addition, psychometricians will benefit from our data analytic procedures, which are outlined in the text and in our freely available R script and include the state-of-the-art missing data imputation with the random forest algorithm, psychometric analysis by means of CTT and IRT, DIF analysis, and informative visualizations. Thus, we believe that our findings will be useful for nursing science and for psychometric research.

Limitations

Limitations of the study are related to sampling procedures, the use of the self-report measure as our instrument, and characteristics of the data. In the absence of randomization, with participants volunteering for the survey, sampling bias is always to be taken into account, and social desirability is always to be considered an issue when self-report measures are used. In addition, due to the COVID-related restrictions, the data collection resulted in the insufficient German sample, while the data in the Netherlands is yet to be collected. Thus, our findings are preliminary, and further research, including research in countries beyond Germany, Hungary and the Netherlands, is required to finalise the instrument.

Another limitation of the study is determined by inevitable subjectivity of analytical decisions (Silberzahn et al., 2018). For instance, item ATNP_1 showed floor effect in the German sample, the second largest dMACs, and the lowest discrimination value, but the experts decided to keep it for theoretical reasons (items from all four areas covered by the framework should be presented in the instrument). As another example, a researcher might set different cut-off values for any of analytical procedures, e.g., DIF analysis, and either more or fewer items would be removed from the scale. In this paper, we reported our analytical decisions thoroughly and transparently, as transparency is crucial for replicability of scientific findings and for further development of psychometric instruments (Streiner & Kottner, 2014).

CONCLUSION

In our age of digitalisation, new technologies are increasingly used in nursing practice, and psychometric instruments to measure ability of nursing professionals to interact with autonomous technologies should be developed. The current version of ATNP, which showed good item functioning, homogeneity and internal consistency reliability, can be used by nurses and nursing undergraduates to evaluate their self-assessed ability to deal with autonomous technologies in nursing practice, and can be further developed by researchers.

Conflict of Interest Statement

No conflict of interest has been declared by the authors.

References

Abdullah, R., Fakieh, B. (2020). Health care employees' perceptions of the use of artificial intelligence applications: Survey study. *Journal of Medical Internet Research*, *22*(5), Article e17620. https://doi.org/10.2196/17620

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495–508. https://doi.org/10.1080/10705511.2014.919210

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Chavent, M., Kuentz, V., Liquet, B., & Saracco, J. (2017). ClustOfVar: Clustering of variables (R Package Version 1.1) [Computer software]. Comprehensive R Archive Network. https://CRAN.R-project.org/package=ClustOfVar

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An *R* package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, *39*(8). https://doi.org/10.18637/jss.v039.i08

Çokluk, Ö., Büyüköztürk, S., & Kayri, M. (2010). Analyzing factor structure of the scales by hierarchical clustering analysis: An alternative approach. *Australian Journal of Basic and Applied Sciences*, *4*(12), 6397–6403. https://bit.ly/3ssDdzc

Cowan, D. T., Norman, I., & Coopamah, V. P. (2007). Competence in nursing practice: A controversial concept – a focused review of literature. *Accident and Emergency Nursing*, *15*(1), 20–26. https://doi.org/10.1016/j.aaen.2006.11.002

Crutzen, R., & Peters, G.-J. Y. (2017). Scale quality: Alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review*, *11*(3), 242–247. https://doi.org/10.1080/17437199.2015.1124240

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne*, *49*(3), 182–185. https://doi.org/10.1037/a0012801

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Sage Publications.

Dima, A. L. (2018**).** Scale validation in applied health research: Tutorial for a 6-step R-based psychometrics protocol. *Health Psychology and Behavioral Medicine*, *6*(1), 136–161, https://doi.org/10.1080/21642850.2018.1472602

EvaSys Survey Automation Suite. (2019). *EvaSys* [Computer Software]. Electric Paper Evaluationssysteme. https://evasys.de/

Farrelly, C. M., Schwartz, S. J., Lisa Amodeo, A., Feaster, D. J., Steinley, D. L., Meca, A., & Picariello, S. (2017). The analysis of bridging constructs with hierarchical clustering methods: An application to identity. *Journal of Research in Personality*, *70*, 93–106. https://doi.org/10.1016/j.jrp.2017.06.005

Feizi, N., Tavakoli, M., Patel, R., & Farokh Atashzar, S. (2021). Robotics and AI for teleoperation, tele-assessment, and tele-training for surgery in the era of COVID-19: Existing challenges, and future vision. *Frontiers in Robotics and AI*. https://www.frontiersin.org/articles/10.3389/frobt.2021.610677/abstract

Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, *10*, Article 1507. https://doi.org/10.3389/fpsyg.2019.01507

Golino, H. F., & Gomes, C. M. A. (2016). Random forest as an imputation method for education and psychology research: Its impact on item fit and difficulty of the Rasch model. *International Journal of Research & Method in Education*, *39*(4), 401–421. https://doi.org/10.1080/1743727X.2016.1168798

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, *2*(2). https://doi.org/10.9707/2307-0919.1111

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, *32*(1), 51–62. https://doi.org/10.1080/10447318.2015.1087664

International Committee of the Red Cross. (2019). *Autonomy, artificial intelligence and robotics: Technical aspects of human control*. https://bit.ly/3aKlYRR

International Technology Education Association (Ed.). (2007). *Standards for technological literacy: Content for the study of technology* (3rd ed.). https://www.iteea.org/42511.aspx

Jo-Karl [@Jo-Karl]. (2020). Jo-Karl/ccpsyc: Package for cross-cultural analyses (R Package Version 0.2.3) [Computer software]. R Package Documentation. https://rdrr.io/github/Jo-Karl/ccpsyc/

Joyce, M., & Kirakowski, J. (2015) Measuring attitudes towards the internet: The General Internet Attitude Scale. *International Journal of Human–Computer Interaction*, *31*(8), 506–517. https://doi.org/10.1080/10447318.2015.1064657

Kim, S., Lee, K.-H., Hwang, H., & Yoo, S. (2015). Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the Unified Theory of Acceptance and Use of Technology (UTAUT) in a tertiary hospital. *BMC Medical Informatics and Decision Making*, *16*, Article 12. https://doi.org/10.1186/s12911-016-0249-8

Kowarik, A., & Templ, M. (2016). Imputation with the R package VIM. *Journal of Statistical Software*, *74*(7). https://doi.org/10.18637/jss.v074.i07

Lee, J. Y., Song, Y. A., Jung, J. Y., Kim, H. J., Kim, B. R., Do, H. K., & Lim, J. Y. (2018). Nurses' needs for care robots in integrated nursing care services. *Journal of Advanced Nursing*, *74*(9), 2094–2105. https://doi.org/10.1111/jan.13711

Li, H., Shi, Y., Li, Y., Xing, Z., Wang, S., Ying, J., Zhang, M., & Sun, J. (2018). Relationship between nurse psychological empowerment and job satisfaction: A systematic review and meta-analysis. *Journal of Advanced Nursing*, *74*(6), 1264–1277. https://doi.org/10.1111/jan.13549

Liao, P. H., Hsu, P. T., Chu, W., Chu, W. C. (2015). Applying artificial intelligence technology to support decision-making in nursing: A case study in Taiwan. *Health Informatics Journal*, *21*(2), 137–148. https://doi.org/10.1177/1460458213509806

Limesurvey. (n.d.). *LimeSurvey: An open source survey tool* (Version 3.25.10+210128) [Computer software]. http://www.limesurvey.org

Locsin, R. C., & Ito, H. (2018). Can humanoid nurse robots replace human nurses. *Journal of Nursing*, *5*(1), 1–6. https://doi.org/10.7243/2056-9157-5-1

Nilsson, J., Johansson, E., Egmar, A.-C., Florin, J., Leksell, J., Lepp, M., Lindholm, C., Nordström, G., Theander, K., Wilde-Larsson, B., Carlsson, M., & Gardulf, A. (2014). Development and validation of a new tool measuring nurses self-reported professional competence—the Nurse Professional Competence (NPC) Scale. *Nurse Education Today*, *34*(4), 574–580. https://doi.org/10.1016/j.nedt.2013.07.016

Maalouf, N., Sidaoui, A., Elhajj, I. H., & Asmar, D. (2018). Robotics in nursing: A scoping review. *Journal of Nursing Scholarship*, *50*(6), 590–600. https://doi.org/10.1111/jnu.12424

Maillet, É., Mathieu, L., & Sicotte, C. (2015). Modeling factors explaining the acceptance, actual use and satisfaction of nurses using an electronic patient record in acute care settings: An extension of the UTAUT. *International Journal of Medical Informatics*, *84*(1), 36–47. https://doi.org/10.1016/j.ijmedinf.2014.09.004

Mantas, J., & Hasman, A. (2017). IMIA educational recommendations and nursing informatics. *Studies in Health Technology & Informatics*, *232*, 20–30. https://doi.org/10.3233/978-1-61499-738-2-20

Nagle, L. M., Sermeus, W., & Junger, A. (2017). Evolving role of the nursing informatics specialist. In J. Murphy, W. Goossen, & P. Weber (Eds.), *Forecasting informatics competencies for nurses in the future of connected health: Proceedings of the Nursing Informatics Post Conference 2016* (Vol. 232, pp. 212–221). IOS Press. https://doi.org/10.3233/978-1-61499-738-2-212

Pepito, J. A., & Locsin, R. (2018). Can nurses remain relevant in a technologically advanced future? *International Journal of Nursing Sciences*, *6*(1), 106–110. https://doi.org/10.1016/j.ijnss.2018.09.013

Phillips, A., Yen, P.-Y., Kennedy, M., & Collins, S. (2017). Opportunity and approach for implementation of a self-assessment tool: Nursing Informatics Competencies for Nurse Leaders (NICA-NL). In J. Murphy, W. Goossen, & P. Weber (Eds.), *Forecasting informatics competencies for nurses in the future of connected health: Proceedings of the Nursing Informatics Post Conference 2016* (Vol. 232, pp. 207–211). IOS Press. https://doi.org/10.3233/978-1-61499-738-2-207

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A short scale for assessing the Big Five dimensions of personality: 10 Item Big Five Inventory (BFI-10). *methoden, daten, analysen*, *7*(2), 233–249. https://doi.org/10.12758/mda.2013.013

R Core Team. (2020). *R: A language and environment for statistical computing* (R Version 4.0.2) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Robitzsch, A. (2020). *sirt: Supplementary item response theory models* (R Package Version 3.9-4) [Computer software]. Comprehensive R Archive Network. https://cran.r-project.org/web/packages/sirt/index.html

Sharifian, R., Askarian, F., & Nematolahi, M. (2014). Factors influencing nurses' acceptance of hospital information systems in Iran: Application of the Unified Theory of Acceptance and Use of Technology. *Health Information Management Journal*, *43*(3), 23–28. https://doi.org/10.12826/18333575.2014.0001.Sharifian

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., . . . Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Stekhoven, D. J. (2013). *missForest: Nonparametric missing value imputation using random forest* (R Package Version 1.4) [Computer software]. Comprehensive R Archive Network. https://cran.r-project.org/web/packages/missForest/index.html

Streiner, D. L., & Kottner, J. (2014). Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing*, *70*(9), 1970–1979. https://doi.org/10.1111/jan.12402

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *10*(6), 363–377. https://doi.org/10.1002/sam.11348

Terwee, C. B., Bot, S. D. M., & de Boer, M. R. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Technology Informatics Guiding Education Reform. (n.d.). The TIGER initiative: Informatics competencies for every practicing nurse: Recommendations from the TIGER Collaborative.

http://s3.amazonaws.com/rdcms-himss/files/production/public/FileDownloads/tiger-report-informatics-competencies.pdf

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Venkatesh, V., Morris, M.G., Davis, G.B., & Davis, F.D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., & Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, *3*(8), Article e002847. https://doi.org/10.1136/bmjopen-2013-002847

Willis, G. B. (2005). *Cognitive interviewing*. Sage Publications. https://www.doi.org/10.4135/9781412983655

World Health Organization. (2016). Global strategic directions for strengthening nursing and midwifery 2016–2020. https://www.who.int/hrh/nursing_midwifery/global-strategic-midwifery2016-2020.pdf?ua=1

Yura, H., & Walsh, M. B. (1988). *The nursing process. Assessing, planning, implementing, evaluating* (5th ed.). Appleton & Lange.

Zhou, M., Herselman, M., & Coleman, A. (2016, April 24–28). *Investigating factors determining the use of the clinical care module by nurses through the UTAUT model* [Conference paper].

eTELEMED 2016: The Eighth International Conference on eHealth, Telemedicine, and Social Medicine, Venice, Italy. https://bit.ly/3tB1BiQ