



Mining concepts of health responsibility using text mining and exploratory graph analysis

Sofia Kjellström & Hudson Golino

To cite this article: Sofia Kjellström & Hudson Golino (2019) Mining concepts of health responsibility using text mining and exploratory graph analysis, *Scandinavian Journal of Occupational Therapy*, 26:6, 395-410, DOI: [10.1080/11038128.2018.1455896](https://doi.org/10.1080/11038128.2018.1455896)

To link to this article: <https://doi.org/10.1080/11038128.2018.1455896>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 24 May 2018.



Submit your article to this journal [↗](#)



Article views: 2179



View related articles [↗](#)





View Crossmark data [↗](#)



Citing articles: 7 View citing articles [↗](#)

Mining concepts of health responsibility using text mining and exploratory graph analysis

Sofia Kjellström^a  and Hudson Golino^b 

^aThe Jönköping Academy for Improvement of Health and Welfare, School of Health and Welfare, Jönköping University, Jönköping, Sweden; ^bDepartment of Psychology, University of Virginia, Charlottesville, VA, USA

ABSTRACT

Background: Occupational therapists need to know about people's beliefs about personal responsibility for health to help them pursue everyday activities.

Aim: The study aims to employ state-of-the-art quantitative approaches to understand people's health and responsibility at different ages.

Methods: A mixed method approach was adopted, using text mining to extract information from 233 interviews with participants aged 5 to 96 years, and then exploratory graph analysis to estimate the number of latent variables. The fit of the structure estimated via the exploratory graph analysis was verified using confirmatory factor analysis.

Results: Exploratory graph analysis estimated three dimensions of health responsibility: (1) creating good health habits and feeling good; (2) thinking about one's own health and wanting to improve it; and 3) adopting explicitly normative attitudes to take care of one's health. The comparison between the three dimensions among age groups showed, in general, that children and adolescents, as well as the old elderly (>73 years old) expressed ideas about personal responsibility for health less than young adults, adults and young elderly.

Conclusions: Occupational therapists' knowledge of the concepts of health responsibility is of value when working with a patient's health, but an identified challenge is how to engage children and older persons.

ARTICLE HISTORY

Received 25 June 2017

Revised 20 February 2018

Accepted 18 March 2018

KEYWORDS

Responsibility for health; text mining; exploratory graph analysis; occupational therapy; health



Introduction

Awareness of health notions is important for occupational therapists [1]. There is an increasing realization among health professionals that patients' health is a result of a co-production process among health professionals and patients [2]. This requires an improved partnership between professionals and patients, but also growing attention to self-care and self-management. However, if health is co-created, then it is of utmost importance to know what occupational therapists and people's views on health are and how they consider issues related to personal health responsibility. A challenge for the co-creation of health lies in the issues of power and responsibility, that should be shared [2]. Consequently, it is worthwhile to investigate how people reason about responsibility for health because they are expected to both recognize and fulfill that responsibility. Occupational therapists work with children, adults and older persons, so it is also of value to know

how the ideas of health responsibility vary across the lifespan.

Occupational therapists' views of health

There are some studies on occupational therapists' concepts of health, but the number of studies investigating their views on responsibility for health is scant, at best. A qualitative study of occupational therapists' notions of health identified three main definitions [3]. The first defines health as feeling fine, and defines health as a subjective notion based upon experiences and as feeling well, strong or contented. The second defines health as being able to act, which means able to choose for oneself, exert control over one's life, and to participate in activities. Health is also described as a factual presence of certain bodily or mental conditions, such as arthritis or diabetes. The results from the qualitative study were tested in a quantitative study which showed that there was very strong

CONTACT Sofia Kjellström  sofia.kjellstrom@ju.se  The Jönköping Academy for Improvement of Health and Welfare, School of Health and Welfare, Jönköping University, P.O. Box 1026, SE-551 11 Jönköping, Sweden

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

support for a holistic view of health with a focus on health as feeling well [1].

Peoples' views of health and responsibility across the lifespan

There is a research area focused on lay beliefs about health [4,5] which emphasizes the importance of studying persons' beliefs or views on what health is, how it is maintained, and its place in daily life, as well as its impact on actual health behaviors [4]. Some authors propose that lay health beliefs and reasoning should be explored and used to influence general societal health [4]. Most research on people's views and beliefs about health has focused on the concepts of health, and only a small part focused on people's views and beliefs about responsibility for health. The ideas about personal responsibility for health appeared in the 1970's in medical and political discourse [22].

Studies on children's' views of health show a developmental pattern, since the notions of health become more varied, complex, thoughtful and elaborated with age [6–8]. The two most important components of health found were to feel good, and being able to participate in desired activities [6]. In short, the ideas about health mature with age. Children decrease the use of reporting concrete practices (eating vegetables, keeping clean), and grow into the ability to conceptualize health as being in bodily good shape and feeling good, and, lastly, adding ideas of mental health [6]. A study comparing mothers and children found, once again, that the concepts became more complex with age [8]. Large differences within each age group were also found [8].

Adults, on the other hand, describe issues of health and illness in terms of responsibility for health [9–13]. A study conducted with elderly Canadians found that they see health responsibility as part of a wider responsibility for later life, and that to live responsibly is a civic duty and a way to obtain government support [14]. Responsible living included daily monitoring of all parts of life. Older Finnish, on the other hand, expressed the need of being both physically and socially active in order to have good health, as well as taking care of the body by a healthy diet, exercising and having a positive attitude towards life. Taking care of oneself and being active were seen as a moral responsibility [15]. These studies indicate that older persons assume personal responsibility for their own health, but they also raise several factors that can limit individual influence [16]. Consequently, health is a multifaceted subject for them, involving not only the

concept of 'responsibility for health' but also activities, social pressures and moral issues, access to social services, and broader domains of concern.

Development of qualitatively different notions of responsibility for health

Adults also experience and interpret concepts qualitatively different in the domain of health and responsibility. Studies show that people reason about health issues at different degrees of complexity [17–21]. There are at least three developmental sequences in reasoning about responsibility issues [22]. Firstly, some maturity is required to naturally start talking about personal responsibility in different domains [19,23–25]. Secondly, the ability to comprehend and take responsibility grows from childhood, adolescence and continues through adulthood [26–29]. People expand their ability to take on responsibility for things (e.g. thoughts, roles, emotions, and actions) [27]. Thirdly, the term 'responsibility' is interpreted differently depending on one's stage of development [23,30].

A study tying together the notions of health and responsibility showed that adults reason in four qualitatively different ways about responsibility for health [22]. The first way shows some difficulty in addressing responsibility per se, which was evidenced by presenting concrete actions without explaining their connection to health or responsibility for health (e.g. walk, not eat specific things like fat on pork steak). Secondly, adults may generalize about what people should do, but without explaining the connection of the actions to health or responsibility. Thirdly, they describe responsibility for health by emphasizing knowledge and explaining things by making causal connections to responsibility in specific health contexts. A few reflected on one's own thinking. All these three ways treated 'health' and 'responsibility' as if they were 'things.' The fourth way shows an elaboration of the learning acquired from living with a health condition to give the requested example of responsibility. To sum up, many people struggled to define and explain what responsibility for health is [22].

As pointed before, studies in the domain of health responsibility are scarce. New studies, aiming to verify how people describe and conceptualize responsibility for health, are required to understand various notions and how they differ between age groups. The current paper aims to adopt state-of-the-art quantitative approaches to understand people's views of health and responsibility at different ages.

Method

This mixed method study is based on structured interviews, conducted as part of a study designed to explore the concept of responsibility for health. The COREQ 32-item checklist for qualitative studies informed this study to assure standards of quality were met [31].

The research was designed and initiated by a senior researcher (SK), who contacted (HG) for the analysis process. All interviews were conducted by students at School of Health and Welfare, Jönköping University 2011-2014. The students were prepared by listening to three lectures on previous research on responsibility for health [22,32], performing qualitative developmental interviews, and research ethics. They were also provided written material about the informed consent process and a flow chart of the process. The majority of students were women and each student contacted one person of their own choice, which meant that they often had some prior relationship with the interviewee. When a person was contacted, he or she was provided with an information booklet providing details of the study design, research ethics, and the requirements in terms of time for the participant. The students were instructed to provide time for the person to reflect on the decision of participation. The students performed the interviews face-to face at a place chosen by the participant. They audio recorded and then transcribed the interviews verbatim.

A structured interview guide was developed and used in the interviews. It was divided into two sections (health and responsibility for health). Each section had 5 questions: What is health/responsibility for health? Why do you choose this way of describing health/responsibility for health? Why is your description a good way of describing health/responsibility for health? If you wanted to teach someone else about health/responsibility for health, how would you do that? You have explained how you would teach someone about health/responsibility for health - why is this a good explanation? Probing by asking to provide more examples was recommended. The questions were constructed to create reflection on the answers and to provide reasons.

The participants were also given the opportunity to answer a questionnaire covering sociodemographic data such as age, education, and occupation. The students provided a printed questionnaire that was filled in after the interview, sealed in an envelope with stamps on and sent it directly to the researcher. The questionnaire included questions from *The national*

public health survey in Sweden on health, lifestyle, and living conditions.

There were 233 participants, ranging from 5 to 96 years of age (Mean = 40.11, SD = 26.02), with 67% being women. In total, 21% of the sample answered only the interview and 79% answered both the interview and the questionnaire.

Permission to carry out the study was approved by the Research Ethics Committee at the Jönköping University. An informed consent process was initiated and consent was obtained by the participants, including assurances of confidentiality.

Text mining

A number of health-related research projects interview people in order to understand, for example, their habits, their opinion on one or more subjects, as well as various concepts and attitudes. These interviews are, in general, transcribed into text for qualitative analysis. However, we can use the text generated from the interviews as data frames of individual words, in order to manipulate, summarize, visualize the characteristics of text, and search for patterns [33]. Converting units of text into analyzable data requires a procedure termed text mining [34], that has been applied in a range of areas, from bibliometrics [35] to systematic reviews [34,36], content analysis [37], biomedical research [38], biology and medicine [36], and health issues [39], among other areas [40].

Text mining is a data driven, exploratory method utilized to find patterns and trends in large data sets [41]. Text mining helps to transform unorganized information into succinct knowledge [36]. This technique makes it possible to collect, maintain, interpret and discover information relevant to research in a systematic and efficient way. Researchers have pointed out that text mining is epistemologically compatible with content analysis [42,43] and grounded theory [43] for three reasons. They share an open mindedness to the data, the categories emerge from the data and are redefined in an iterative process, and sound text mining fulfills the quality criteria of qualitative researchers [43].

Thomas, McNaught and Ananiadou [34] point to three main areas in which the text mining technique can be employed: 1) retrieval of text information; 2) term extraction; and 3) data mining. The retrieval of text information concerns the identification and extraction of pieces of text related to a certain area, task, or focus of the research. The term extraction automatically identifies key words or phrases that are relevant. Data mining, on the other hand, is based on

identifying and extracting parts of the text, sentence fragments, words or any other information present in the texts investigated to assemble an analyzable database, to which statistical techniques can be applied.

In order to apply the text mining in the interviews, the *tm package* [44] and the *SnowballC package* [45] of the R environment for statistical computing [46] were used. The first package allows the application of sequences of programming that will allow the identification and extraction of elements of the texts analyzed. The second package allows the cleaning of textual elements in several languages, including Swedish. The mining process is applied in seven steps: 1) creation of an analyzable vector of text data; 2) conversion of all textual elements to the low-case format; 3) removal of all punctuation; 4) removal of so-called stop words (connectives, articles, prepositions, etc.); 5) stemming, which is the identification of sets of textual elements that may differ from one another only in the modification of the structural root of the words; 6) conversion of the textual units into a frequency database, where each textual unit is allocated to a column, and the lines represent the observations (or participants) and the values of the cells represent the number of times that textual element pops up in the interview; 7) removal of sparse terms, i.e. terms whose incidence is very low. All these procedures usually lead to a very high number of textual elements (i.e. terms, words or word roots). For this reason, it is necessary to use strategies to decrease the level of sparsity.

Analysis

After creating a document term matrix for each construct (health and health responsibility), a new technique to assess the number of latent factors in multivariate data termed Exploratory Graph Analysis (EGA) [47,48] was used. This is a new technique to estimate the number of latent factors using a network psychometrics approach [49]. Firstly, EGA estimates the correlation matrix of the observable variables, then proceeds to use the LASSO estimation to obtain a regularized partial correlation matrix, with the regularization parameter defined via the extended Bayesian Information Criteria (EBIC) over 100 different values. In the last step, the walktrap algorithm [50] is used to find the number of clusters in the network. A fundamental rule of network psychometrics is that the number of clusters in networks equals the number of latent variables [51]. It is important to note that EGA not only estimates the number of dimensions underlying the data, but also clearly shows which items belong to which dimensions. Golino and

Epskamp [51] studied the accuracy of six techniques to estimate the number of dimensions (very simple structure with complexity one, minimum average partial procedure, BIC and EBIC, parallel analysis, Kaiser-Guttman eigenvalue greater than one rule), and compared their accuracy with the accuracy of EGA on 32,000 simulated data sets with known factor structures (varying the number of factors, the number of dichotomous items, sample size and correlation between factors).

Golino and Epskamp [51] showed that EGA performed as well as parallel analysis, in general. However, when the correlation between factors was high (0.70), the number of items per factor was five, and the number of factors was four, only EGA could correctly estimate the number of latent factors, achieving a mean accuracy of 100% when the sample size was large. They also verified how the controlled conditions and their high-order interactions affected the mean accuracy of each method, via ANOVA. The results showed that EGA was the only technique attaining a high partial eta squared effect size in only one condition (sample size). The other techniques attained high effect sizes from three to nine conditions or their high-order interactions.

Exploratory Graph Analysis was applied using the *EGA package* [48] of the R environment for statistical computing [46]. The structure estimated via EGA was submitted to a confirmatory factor analysis, using the *EGA package* [48], with the diagonally weighted least squares (DWLS) estimator. The DWLS approach uses the weighted least squares estimator with polychoric correlations as input to create the asymptotic covariance matrix, being suited for ordinal data. The fit of the models was verified using the root mean-square error of approximation (RMSEA), the comparative fit index (CFI) [52], the normed fit index (NFI) and the nonnormed fit index (NNFI) [53]. A good fit is indicated by a RMSEA < 0.06 [54], a CFI > 0.95 [55], a NFI and NNFI > 0.90 [53].

After verifying the fit of the health responsibility model, the factor scores were compared in different groups of participants, split by age and by the categorical items of the Sweden's National Public Health Survey. There were five age groups: 1) from 5 to 17 years ($n = 36$); 2) from 18 to 26 years ($n = 36$); 3) from 27 to 48 years ($n = 36$); 4) from 49 to 73 years ($n = 36$); 5) from 74 to 96 years ($n = 35$). Tukey's honest significant difference test [56] was used in order to compare the means of factor scores in each group of participants. This approach was preferred over other, more robust, approaches (e.g. multiple

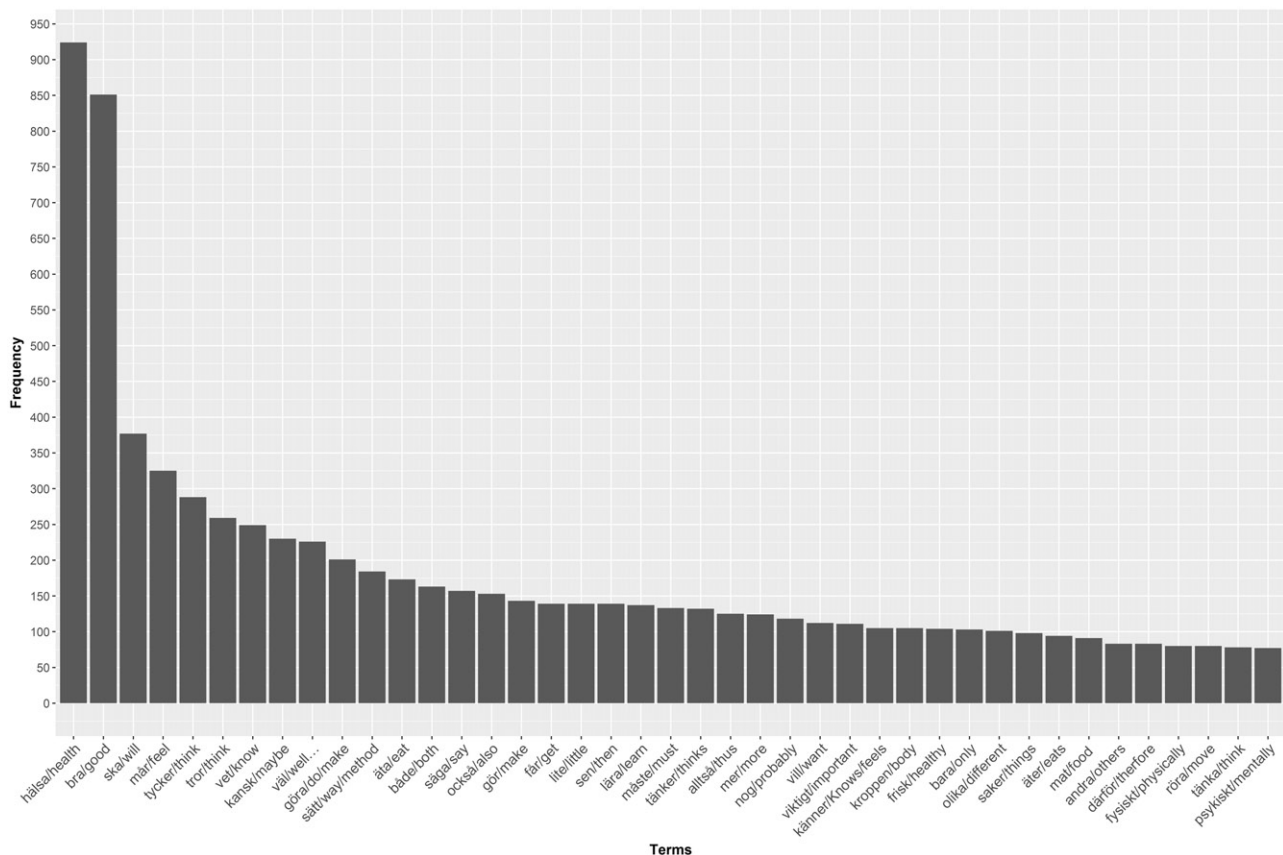


Figure 1. Frequency plot of the terms extracted after the text mining process.

group confirmatory factor analysis) because the sample size for the age groups were very small.

To further check the impact of age on the factors identified via exploratory graph analysis, two new models were investigated where age was regressed on each one of the factors. In the first model, age was constrained to have a regression coefficient of zero, while in the second model the age regression parameter was freely estimated. The Satorra-Bentler scaled chi-square difference test [57] was used to check if the two measurement models were significantly different.

Results

Concepts of health

In the first round of text mining, 4,011 terms were extracted, with a sparsity of 99%. Sparsity can be defined as the degree of zero-entries in a dataset, so a dataset with a sparsity of 10% means that 90% of the cells have non-zero entries. In the first step of the text mining, the terms appeared in at least 1% of the interviews. To decrease the level of sparsity, terms that did not appear in at least 20% of the interviews were deleted, resulting in a final document term

matrix with 41 terms (or words). The relative frequency of the words can be viewed in Figure 1. The distribution of the terms or words frequency shows a pattern that is very common in any given corpus of natural language, called Zipf's law. This law states that the frequency that a word appears is inversely proportional to its rank [33].

The exploratory graph analysis showed a structure with six factors. However, one of the factors had only one term ('both' or 'både'). Golino and Demetriou [47] described a rule to use EGA to explore the dimensionality of instruments or multivariate data: if one item or variable forms a single factor, it should be removed and the EGA procedure should be applied again. This is an iterative process, that ends when each estimated factor has at least two items or variables. After removing the term 'both', EGA pointed to a five-factor structure (see Figure 2).

The five-factor structure estimated via EGA was submitted to a confirmatory factor analysis. The results show a very good fit to the data [$\chi^2(730) = 544.23$; $p = 1.00$; CFI = 0.97; RMSEA = 0.02; NFI = 0.97; NNFI = 0.97]. Figure 3 shows the standardized factor loadings and the correlation between factors.

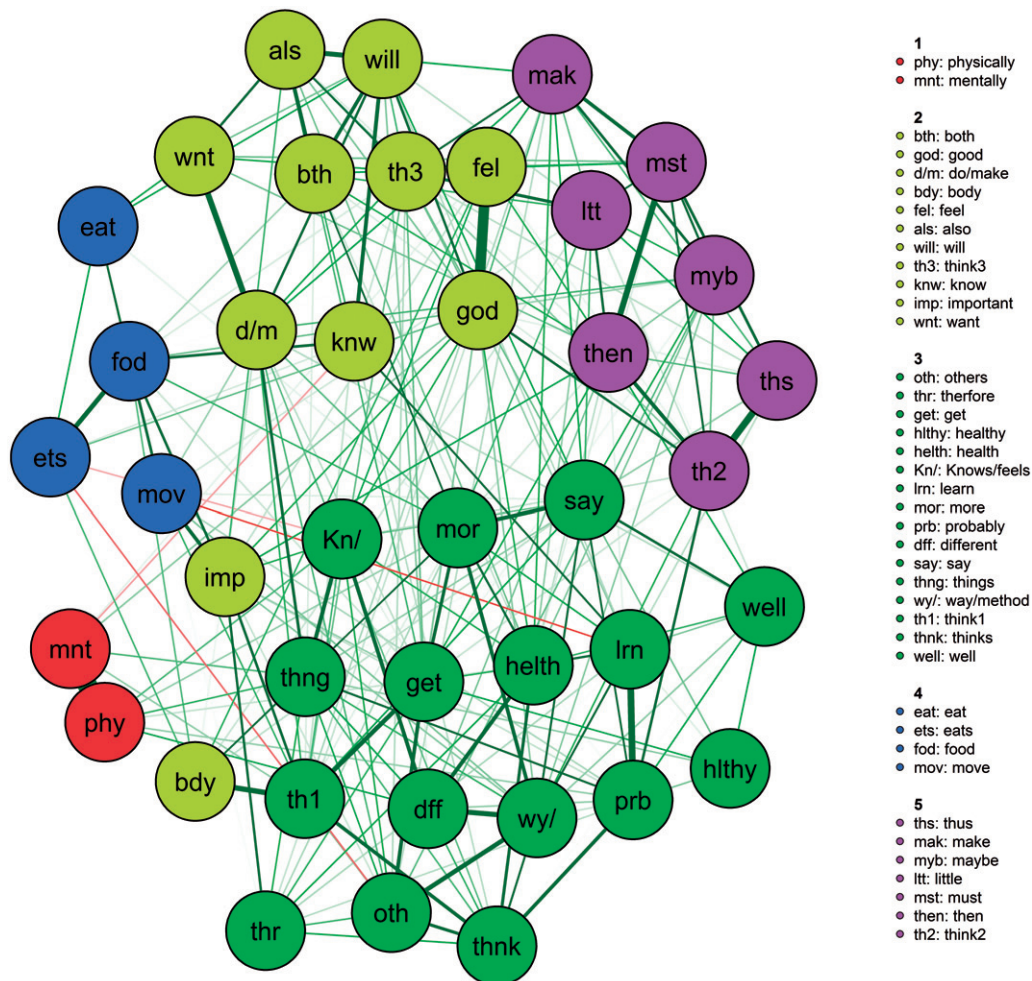


Figure 2. Structure of the document term matrix estimated via EGA.

Health factor one: physical and mental health

The first factor has only two terms: physical and mental. In the interviews, it was common to find participants referring to health as having physical and mental facets. Defining health as meaning to ‘feel good physically and mentally’ was also common. However, these terms were also used separately by the interviewees. As can be seen in Figure 3, the standardized loadings of both terms were high (0.99 and 0.79), however the correlation between factor one and the other factors is small (< 0.30).

Health factor two: feeling good and doing as an imperative

The second factor included the following terms: only, good, to do, body, feel, also, will/shall, thinks, know, important and want. The terms good, feeling and body were combined as ‘feeling good’ and ‘feeling good in the body’ by the interviewees. People mentioned health, in general, as tied to the body. For

example, some interviewees described health as ‘feeling harmony in the body’, while others as ‘listening to the body’ or ‘building up the body’. There was a related aspect referring to activity through the words ‘do/make’. Health means to be able to do things or to do the right things. The verb ‘doing/making’ was also combined with the word ‘want’ as expressed in being able ‘to do what I want’. There was a normative and performative aspect in this factor, evidenced by the terms ‘important’ and ‘will/shall’. The participants talked about some aspects as being important things that a person should do, like ‘you should keep in shape’. The term ‘only’ is particularly used as ‘not only’ and the term ‘also’ is used to combine aspects which shows that health was conceived as a complex term involving more than singular aspects. The most important terms in the second factor were good (standardized factor loading of 0.64) and will/shall (standardized factor loading of 0.63). The term with the lowest factor loading was body (0.20).

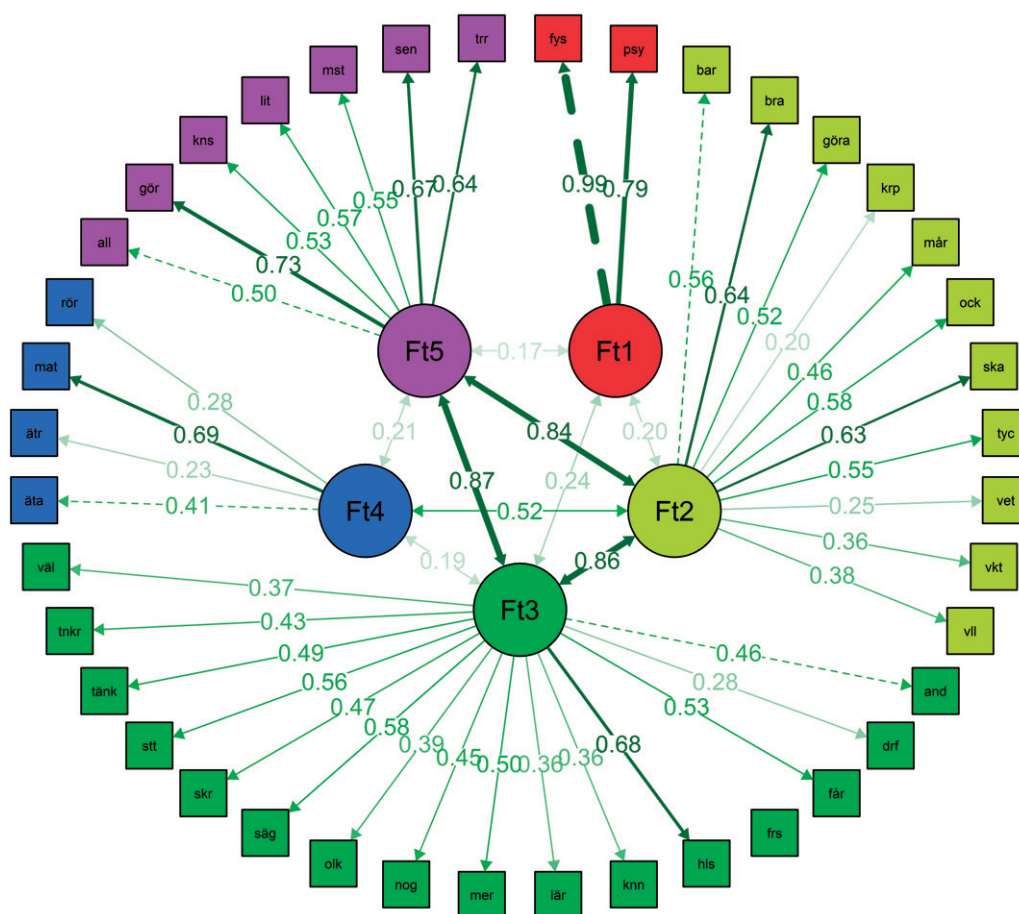


Figure 3. Standardized factor loadings and correlation between factors.

Health factor three: being healthy

The third factor includes most terms that are common in ordinary language, and also refer to how you might explain the topic to other people but are not specific to the topic of health. A core aspect was that health is to be healthy, as in having no diseases. Learning was another aspect which, on one hand, referred to teaching others about health, but also was the concept of learning to know oneself and learning to live with one's physical condition. There are different 'ways', and 'good' ways that health can be achieved. It is also framed in terms of what a person should 'consider' or think about regarding health. The most important terms in the third factor were 'healthy' (standardized factor loading of 0.68) and 'say' (standardized factor loading of 0.58). The term with the lowest factor loading was 'because' (0.28).

Health factor four: eating and exercising

The fourth factor referred to eating food (eat, eating, food) and physical exercise (move), which can be interpreted as two main activities that were expressed as being essential for health. There was also a

normative element evidenced in the terms of 'have to' do. The terms with the highest standardized factor loadings were food (0.69) and eat (0.48).

Health factor five: actions and beliefs

The fifth factor had the following terms: thus, do, maybe, little, have to, thus and believe/think. This factor combined actions and beliefs as well as several terms that are used in everyday language. The term 'does' evidences health as an achievement. The terms with the highest factor loading were 'do' (0.73), 'thus' (0.67) and 'believe' (0.64).

Concepts of health responsibility

In the first round of text mining, 4031 terms were extracted, with a sparsity of 99%. To decrease the level of sparsity, terms that did not appear in at least 20% of the interviews were deleted, resulting in a final document term matrix with 41 terms (or words). The relative frequency of the words about health responsibility can be viewed in Figure 4. The distribution of

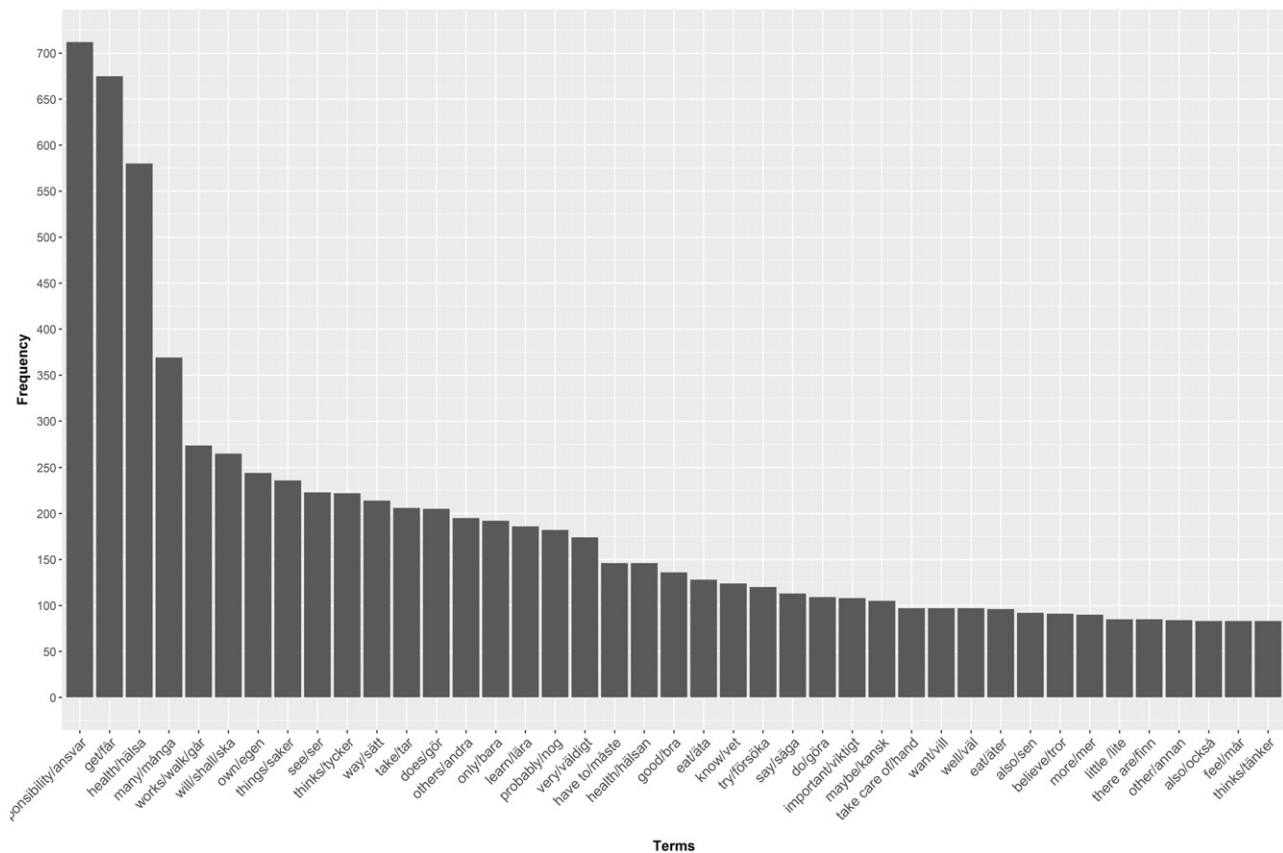


Figure 4. Frequency plot of the terms extracted after the text mining process – concepts of health responsibility.

the terms or words frequency also showed a pattern similar to Zipf's law.

The exploratory graph analysis showed a structure with nine factors. However, three of the factors had only one term ('other' or 'annan'; (to) 'eat' or 'äta'; 'eat'(eating) or 'äta'). After removing these terms, EGA pointed to a three-factor structure (see Figure 5).

The three-factor structure estimated via EGA was submitted to a confirmatory factor analysis. The result showed a very good fit to the data [$\chi^2(662) = 498.21$; $p = 1.00$; CFI = 0.98; RMSEA = 0.03; NFI = 0.98; NNFI = 0.97]. Figure 6 shows the standardized factor loadings and the correlation between factors.

Health responsibility factor one: creating good health and feeling good

The first factor had the following terms: others, responsibility, good, health, maybe, learn, little, feel, and also. Responsibility for health was conceptualized as referring to creating good health, but also good habits and feeling well. Another aspect was the use of the term 'others' as in taking care of others, but the term was also used in a more neutral way to explain that there are other things involved in health

responsibility. The interviewees talked about learning in two senses: firstly, as something that you need to learn yourself and secondly, as teaching others about what responsibility for health implies. The terms with the highest standardized factor loadings were responsibility (0.77), health (0.76), maybe (0.68) and good (0.63). In sum, this factor can be interpreted as follows: responsibility for health is creating good health and feeling good.

Health responsibility factor two: importance of taking care of one's own health

The second factor included the terms: only, own, get, find, try, works/walk, health, have to, more, probably, things, also, thinks, very, important, want. While in the first factor the focus is on others, in the second factor the focus seemed to be on the self, switching to each person's own responsibility for health, expressed with the term 'own'. Health responsibility had a normative dimension demonstrated by the word 'have to', implying that a person has to take care of his/her health. The term 'have to' was also used as verbalizing that some things needs to be done, but primarily that the individual have to take care of its health. The verb 'want' was also used as in wanting to improve the

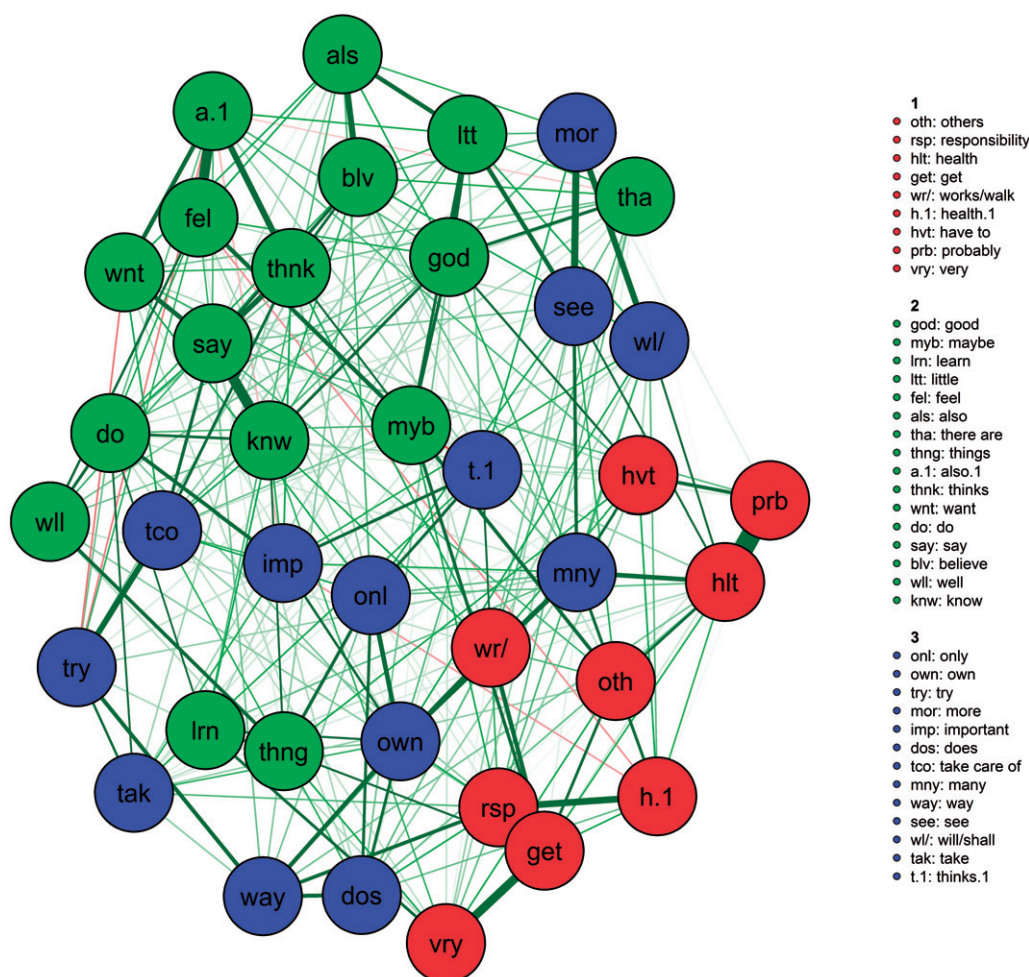


Figure 5. Structure of the document term matrix estimated via EGA.

health. The terms 'important' and 'very' were also used to emphasize that health is a significant matter. The term 'try' was used with two meanings, partly as trying to improve your health, and partly as trying to explain for other persons the importance of being healthy. Terms with the highest loadings were: thinks (0.55), also (0.54), find (0.54) and want (0.53). An interpretation of this factor is: self-responsibility involves thinking about one's own health and wanting to improve it, by doing things that has to be done to achieve this goal.

Health responsibility factor three: taking care about yourself

The third factor is based upon the following terms: does, to do, care/hand, many, say, way, see, will/shall, take, believe, thinks, well, know. There was an activity aspect evidenced by the terms 'does' and 'to do'. The interviewees talked about doing things for your health, but a few also mentioned doing the correct things. The term 'take' was used as in the expression of

taking responsibility for your health. Common expressions that were used are 'take care about yourself', but also 'to take care about your body or health'. There was also a normative aspect in this factor as articulated with the use of the term 'will/shall'. Terms with the highest factor loadings were: do (0.72) and will/shall (0.67). The third factor involved more explicitly a normative aspect of health responsibility.

Responsibility for health at different age groups

A multiple comparison, via Tukey's honest significant difference test [56], was applied in order to compare the means of factor scores by age group. The result (Table 1) showed a statistically significant difference in factor one between the group of children and adolescents (age group one: 5 to 17, $p < 0.05$) and young adults (age group two: 18 to 26, $p < 0.05$), adults (age group three: 27 to 48, $p < 0.05$) and older adults plus the young elderly (age group four: 49 to 73, $p < 0.05$). It also shows a significant difference between the old

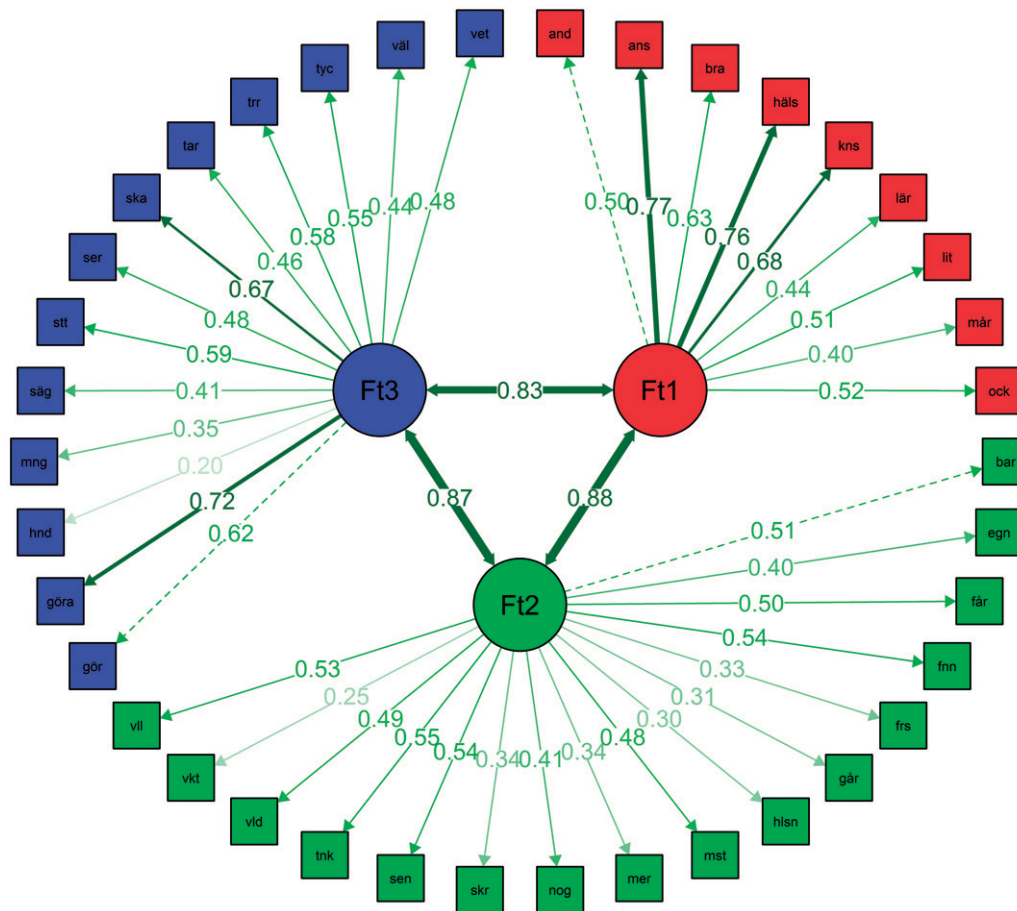


Figure 6. Standardized factor loadings and correlation between factors – health responsibility.

Table 1. Tukey's honest significant difference test.

Factors	Age groups	Difference	Lower	Upper	<i>p</i> adjusted	
Factor 1	2-1	0.43	0.02	0.83	0.03*	
	3-1	0.78	0.37	1.19	0.00*	
	4-1	0.43	0.02	0.84	0.03*	
	5-1	0.20	-0.24	0.65	0.70	
	3-2	0.36	-0.05	0.76	0.12	
	4-2	0.00	-0.40	0.41	1.00	
	5-2	-0.22	-0.66	0.22	0.64	
	4-3	-0.35	-0.76	0.06	0.13	
	5-3	-0.58	-1.02	-0.13	0.00*	
	5-4	-0.22	-0.67	0.22	0.63	
	Factor 2	2-1	0.29	-0.01	0.59	0.06
		3-1	0.53	0.23	0.84	0.00*
4-1		0.35	0.05	0.65	0.01*	
5-1		0.26	-0.06	0.59	0.17	
3-2		0.24	-0.06	0.54	0.19	
4-2		0.06	-0.24	0.36	0.98	
5-2		-0.03	-0.35	0.30	1.00	
4-3		-0.18	-0.48	0.12	0.47	
5-3		-0.27	-0.60	0.06	0.17	
5-4		-0.09	-0.41	0.24	0.95	
Factor 3		2-1	0.57	0.06	1.07	0.02*
		3-1	0.74	0.24	1.25	0.00*
	4-1	0.55	0.05	1.05	0.02*	
	5-1	0.46	-0.08	1.00	0.14	
	3-2	0.17	-0.33	0.68	0.88	
	4-2	-0.01	-0.52	0.49	1.00	
	5-2	-0.11	-0.65	0.44	0.98	
	4-3	-0.19	-0.69	0.32	0.84	
	5-3	-0.28	-0.83	0.27	0.62	
	5-4	-0.09	-0.64	0.45	0.99	

*Significant < 0.05.

elderly (age group five: 74 to 96) and adults ($p < 0.05$).

Comparing the means of the second factor, a statistically significant difference was found between age group three (adults) and age group one (children and adolescents, $p < 0.05$), and between age group four (young elderly) and age group one ($p < 0.05$). Finally, a statistically significant difference was found between age group two (young adults) and age group one (children and adolescents, $p < 0.05$), between age group three (adults) and age group one ($p < 0.05$), and between age group four (young elderly) and age group one ($p < 0.05$). Figures 7-9 show the boxplots of the factor scores by age group.

In terms of the items from the Sweden's National Public Health Survey, there was no significant difference in the factor scores, for all factors, when comparing the categories presented in each item.

Two new models (Figure 10) of health responsibility were investigated, where age was regressed on each one of the factors identified via exploratory graph analysis. In the first model, age was constrained to have a regression weight of zero [$\chi^2(700) = 574.97$;

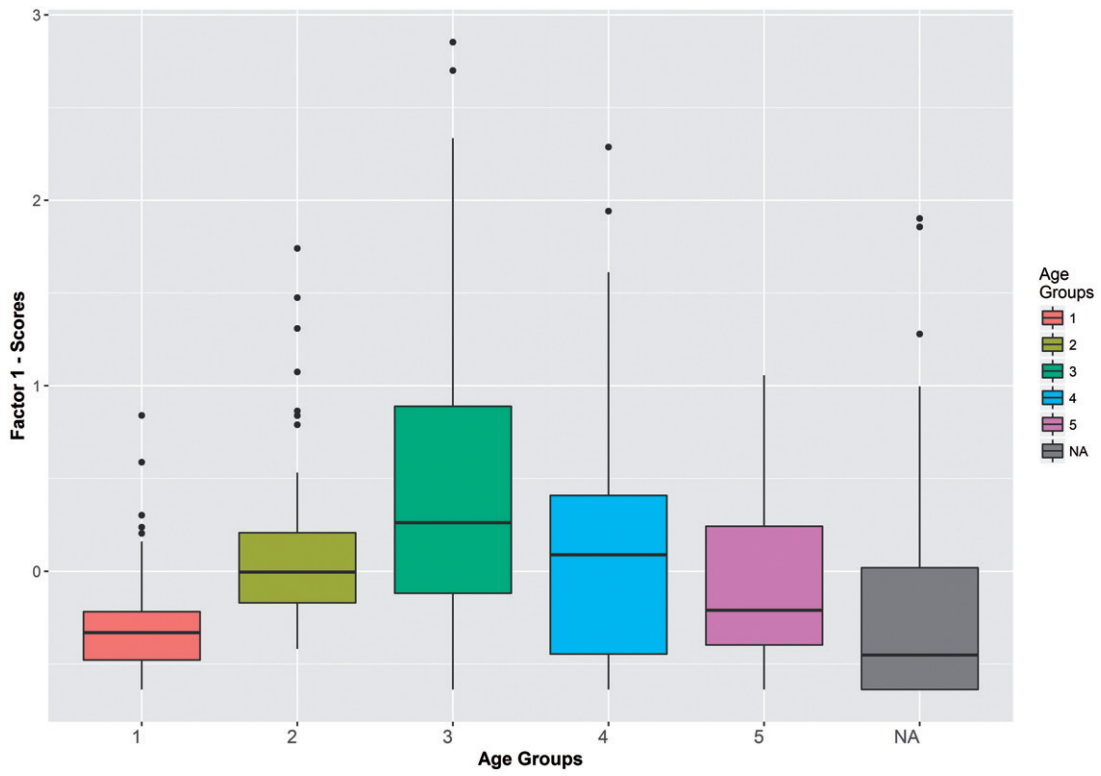


Figure 7. Factor One scores by age group.

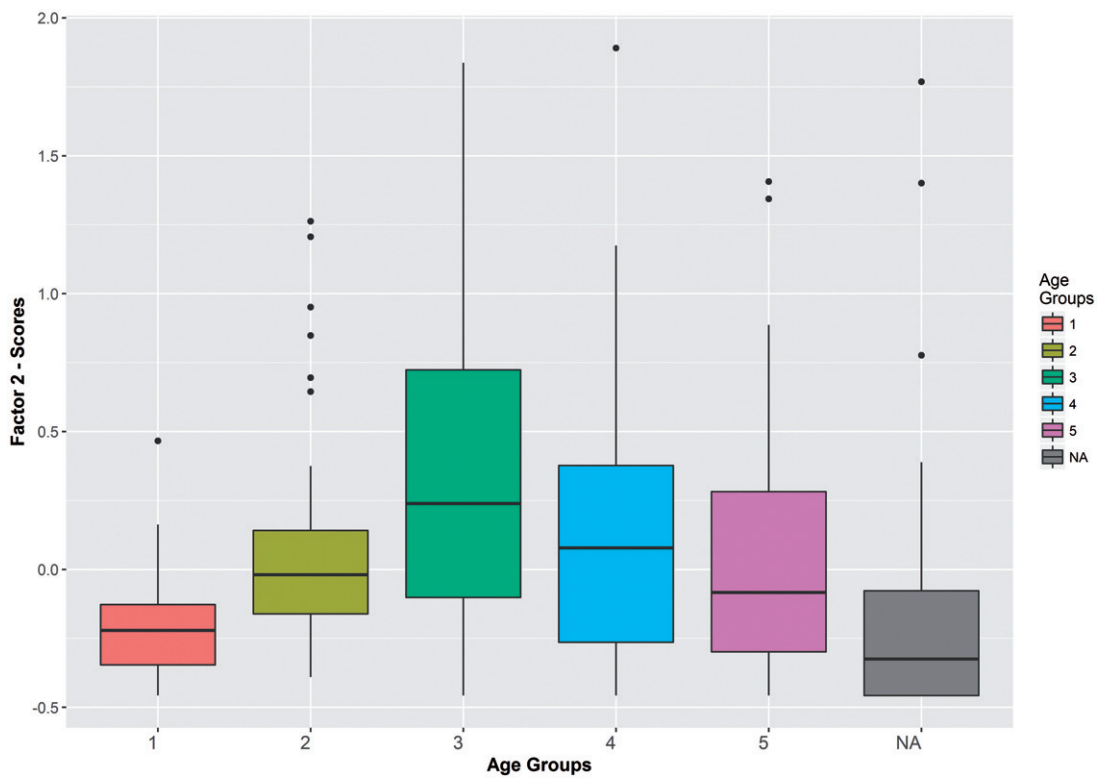


Figure 8. Factor Two scores by age group.

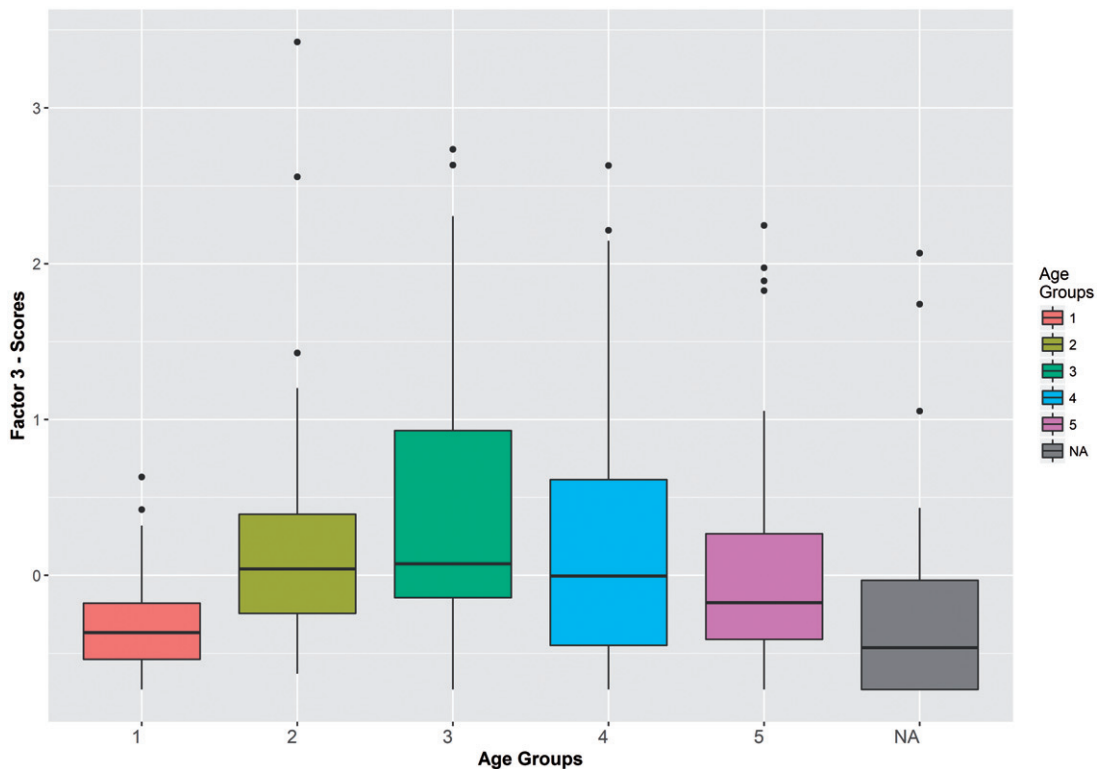


Figure 9. Factor Three scores by age group.

$p = 1.00$; CFI = 0.96; RMSEA = 0.03], while in the second model the age regression parameter was freely estimated [$\chi^2(697) = 551.05$; $p = 1.00$; CFI = 0.96; RMSEA = 0.03]. Age presented a regression coefficient of -0.03 with factor one (creating good health and feeling good) with a p -value of 0.41, a coefficient of 0.14 with factor two (importance of taking care of one's own health) with a p -value of 0.001, and a coefficient of 0.13 with factor three (taking care about yourself) also with a p -value of 0.001. The Satorra-Bentler scaled chi-square difference test [57] showed no statistically significant difference between the two models (Table 2).

Discussion

Literature shows that adults working in the field of occupational therapy define health as feeling well, strong and connected, it being a subjective notion based on their experiences [3]. It is also defined in terms of the capacity to act and to exert control in one's own life and activities, besides the factual presence of physical and mental conditions [3]. Children, however, focus their definition of health as feeling good and being able to participate in activities they like [6]. From childhood to adulthood, there is a shift in the way people think about health, from concrete thinking towards more complex abstract definitions

of health. The notion that people have personal responsibility for their own health goes in the same direction [25–28]. Among older persons, the concept of responsibility for health became a little bit different from adults, since the former define it as one aspect of broader responsibility in later life. However, in general, for older persons health responsibility is being active, having a healthy diet and a positive attitude toward life [22].

In the current research, five factors related to the concepts of health were found, with a very good fit to the data processed using text mining in the interviews. In general, people described and conceptualized health as feeling good physically and mentally, as a general well-being (or feeling good in general) with a normative and performative aspect, as being healthy or having no diseases, knowing what to eat and a combination between action (what you do) and beliefs. The correlations between the factors showed that three of them are highly related: feeling good in general (factor two), being healthy (factor three) and the combination of actions and beliefs (factor five). These results have some connection to the definitions presented by occupational therapist, especially that health is a subjective notion, based on experiences, and strongly related to feeling well both physically and mentally.

In terms of the health responsibility, three factors were found, also with a very good fit to the data.

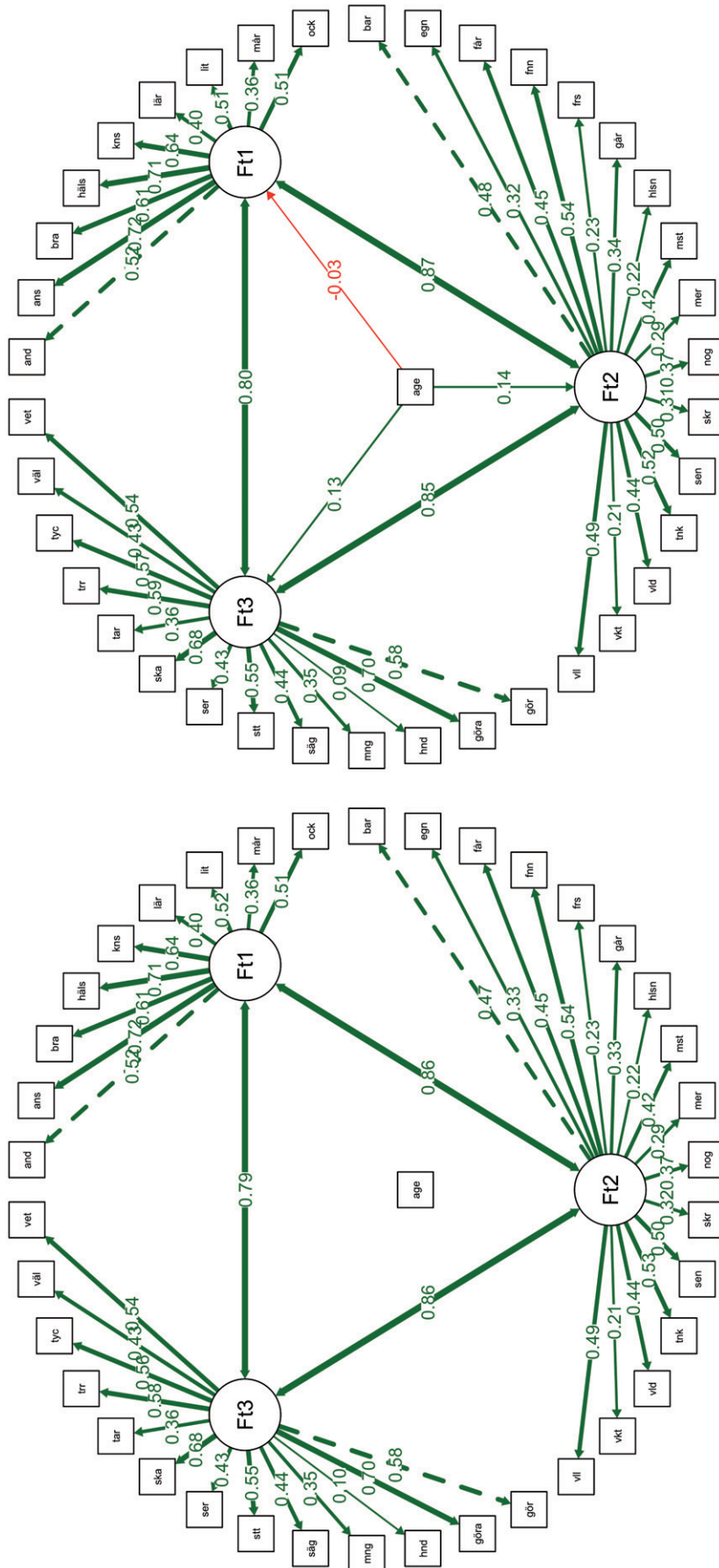


Figure 10. Models of health responsibility and age.

Table 2. Satorra and bentler (2001) scaled chi-square difference test.

Models	DF	Chi-Square	Chi-Square difference	DF difference	<i>p</i>
Model 2 - Age regression parameter freely estimated	697	551.06			
Model 1 - Age parameter constrained to zero	700	574.98	5.8242	3	0.1205

Health responsibility was defined as creating good health habits and feeling good, thinking about one's own health and wanting to improve it, and adopting explicitly normative attitudes in order to take care of yourself, your body and your health. These three general dimensions were investigated in different age groups, to verify if there was any difference between children and adolescents, young adults, adults, adults and young elderly and old elderly people.

Adults, aged between 27 and 46 years old, are the ones with the higher factor scores in factor one, i.e. responsibility for health is creating good health habits and feeling good, which was significantly different from children and adolescents and old elderly. Young adults (age between 18 and 26 years old), as well as adults and young elderly (aged between 49 and 73 years), also presented significantly higher factor scores, compared to children and adolescents. The old elderly (from 74 to 96 years) was the only group that did not present a significantly higher factor score in factor one compared to children and adolescents. When the factor scores of the second factor (health responsibility involves thinking about one's own health and wanting to improve it), were compared between age groups, children and adolescents showed significantly lower factor scores, compared to all age groups except the old elderly group. The same scenario was repeated in the third factor (health responsibility is adopting explicitly normative attitudes to take care of yourself, your body and your health), where children and adolescents' factor scores were significantly lower than all other age groups, except the old elderly group. These results can be interpreted in at least two ways. They could indicate some kind of a functional decline in the old elderly, who may be less inclined to create healthy habits, and then less likely to address this topic in their interviews, or may be indicative of a generational difference between the old elderly group and the other groups.

In general, adults presented the highest factor scores on all three factors. It is important to point out that previous studies also showed that the concepts of health develop dramatically from childhood and adolescent to adulthood [6].

When the three-factor model identified using exploratory graph analysis was modified to see the impact of age as a continuous variable on each of the factors, a slightly different scenario appeared. Age did

not impact people's conceptions that responsibility for health is creating good health habits and feeling good. On the other hand, age demonstrated a small but positive and significant relationship with the two other factors (health responsibility involves thinking about one's own health and wanting to improve it, and health responsibility is adopting explicitly normative attitudes to take care of yourself, your body and your health). This result indicates that both factors are slightly stronger as people gets more mature, presenting a possible developmental pattern. Future studies could investigate how both concepts of health responsibility changes over time, adopting a longitudinal design in order to understand if this interpretation can be supported by evidences.

Occupational therapists working with patients need to be aware of the notions of health responsibility and how they change throughout life. Particularly challenging is to engage children and older persons in health responsibility discussions and behaviors since this study showed that they do express these ideas as much as adults in general. An important question is if health conversations could become even more individualized and contextualized based upon these findings, which need to be further investigated. This is essential for occupational therapists to improve people's abilities to live active and healthy lives.

There are several kinds of methodological considerations in this study. Researchers claim that text mining and qualitative research are epistemological compatible [43], which means that methodological quality criteria apply. Text data mining improves consistency and replicability since it is not subjected to human factors like fatigue, boredom or carelessness [43]. However, there is also an interpretive element in naming and describing the factors identified via exploratory graph analysis.

A strength of the study is the use of a new innovative approach, termed exploratory graph analysis, to identify the number of factors underlying the dataset constructed after the text mining process. The combination of the two strategies can identify structures in people's responses, that can then be used to guide the construction of new questionnaires. This study demonstrates the promise of this procedure. A minor limitation is that the categories seemed to include some stop-words (e.g. connectives like 'thus') which means that not all stop-words had been excluded.

This might be due to the language used in the current research (Swedish), which may have less elaborate lists of stop-words. This makes the interpretation of the factors a bit harder, but we do not think this has severely affected the results. Future research should collect data in other languages and examine if the concept of health responsibility includes similar features and factors. Another limitation of the current study is that the confirmatory factor model was fitted in the same data used in the Exploratory Graph Analysis. Future research should investigate if a similar structure, with a similar fit, can be identified using new datasets.

In the current research, the questions were constructed with the intent to discover how people describe and define health and health responsibility. After text mining the interviews, it was possible to obtain a small set of terms or words that grouped together in reasonable way. It was possible to identify three general factors of health responsibility, highly correlated, that seem to increase with age, from childhood to adulthood. Future studies should increase the sample size to obtain more information about how people think, describe and conceptualize health and health responsibility. The current findings can provide support for the construction of new instruments to assess concepts of health and health responsibility.



Acknowledgements

We would like to thank for funding from Stationmaster Lennart Israelsson Foundation for the Promotion of Elderly Activity and Health (Aktiestinsens Lennart Israelssons stiftelse för främjande av äldres aktivitet och hälsa).

Disclosure statement

The authors report no conflicts of interest.

ORCID

Sofia Kjellström  <http://orcid.org/0000-0001-8952-8773>
Hudson Golino  <http://orcid.org/0000-0002-1601-1447>

References

- [1] Björklund A, Svensson T, Read S. Holistic and biomedical concepts of health: a study of health notions among Swedish occupational therapists and a suggestion for developing an instrument for comparative studies. *Scand J Occup Ther.* 2006;13:141–150.
- [2] Batalden M, Batalden P, Margolis P, et al. Coproduction of healthcare service. *BMJ Qual Safety.* 2016;25:509–517.
- [3] Svensson T, Björklund B. Health, the body and occupational therapy. *Scand J Occup Ther.* 2000; 7:26–32.
- [4] Hughner RS, Kleine SS. Views of health in the lay sector: a compilation and review of how individuals think about health. *Health (London).* 2004;8: 395–422.
- [5] Lawton J. Lay experience of health and illness: past research and future agendas. *Sociol Health Illn.* 2003;25:23–40.
- [6] Natapoff J. Children's views of health: a developmental Study. *Am J Public Health.* 1978;68: 995–1000.
- [7] Natapoff JN. A developmental analysis of children's ideas of health. *Health Educ Q.* 1982;9:34–45.
- [8] Schmidt LR, Fröhling H. Lay conceptions of health and illness from a developmental perspective. *Psychol Health.* 2000;15:229–238.
- [9] Donnelly TT, McKellin W. Keeping healthy! Whose responsibility is it anyway? Vietnamese Canadian women and their healthcare providers' perspectives. *Nurs Inq.* 2007;14:2–12.
- [10] Lawton J, Ahmad N, Peel E, et al. Contextualising accounts of illness: notions of responsibility and blame in white and South Asian respondents accounts of diabetes causation. *Sociol Health Illn.* 2007;29:891–906.
- [11] Pill R, Nigel CHS. Choice or change: further evidence on ideas of illness and responsibility for health. *Soc Sci Med.* 1985;20:981–991.
- [12] Pill R, Stott NCH. Concepts of illness causation and responsibility: some preliminary data from a sample of working class mothers. *Soc Sci Med.* 1982;16: 43–52.
- [13] Shye D, Javetz R, Shoval JT. Lay self-care in health: the views and perspectives of Israeli laypeople. *Soc Sci Med.* 1991;33:297–308.
- [14] Kemp CL, Denton M. The allocation of responsibility for later life: Canadian reflections on the roles of individuals, government, employers and families. *Aging Soc.* 2003;23:737–760.
- [15] Jolanki O. Moral argumentation in talk about health and old age. *Health (London).* 2004;8:483–503.
- [16] Jolanki O. Discussing responsibility and ways of influencing health. *IJAL.* 2008;3:45–76.
- [17] Gast HL. The Relationship Between Stages of Ego Development and Developmental Stages of Health Self Care Operations: Texas Woman's University; 1983.
- [18] Kajanne A. Structure and content: The relationship between reflective judgment and laypeople's viewpoints. *J Adult Develop.* 2003;10:173–188.
- [19] Kajanne A, Pirttila-Backman A-M. Laypeople's viewpoints about the reasons for expert controversy regarding food additives. *Public Underst Sci.* 1999;8:303–315.
- [20] Stilley CS, Lawrence K, Bender A, et al. Maturity and adherence in adolescent and young adult heart recipients. *Pediatr Transpl.* 2006;10:323–330.
- [21] Lawrence K, Stilley CS, Olshansky E, et al. Further exploration: maturity and adherence in adolescent

- and young adult heart recipients. *Prog Transpl*. 2008;1:50–54.
- [22] Kjellström S, Ross SN. Older persons reasoning about responsibility for health: variations and predictions. *Int J Aging Hum Dev*. 2011;73:99–124.
- [23] Dawson TL, Gabrielian S. Developing conceptions of authority and contract across the lifespan: two perspectives. *Dev. Rev* 2003;23:162–218.
- [24] Gilligan C. In a different voice: psychological theory and women's development. Cambridge, Mass.: Harvard University Press; 1982.
- [25] Robbins JG, Greenwald R. Environmental attitudes conceptualized through developmental theory: a qualitative analysis. *J. Social Issues*. 1994;50:29–47.
- [26] Kegan R. The evolving self: problem and process in human development. Cambridge, London: Harvard University Press; 1982.
- [27] Kegan R. In over Our Heads: The mental demands of modern life. Cambridge, Mass: Harvard University Press; 1994.
- [28] Loevinger J, Blasi A. Ego development: conceptions and theories. 1st ed. San Francisco: Jossey-Bass; 1976. (Jossey-Bass behavioral science series).
- [29] Rybash JM, Roodin PA, Making decisions about health-care problems: a comparison of formal and postformal modes of competence. In: Commons ML, Sinnott JD, Richards FA, et al., editors. Adult development: comparisons and applications of developmental models. Vol. 1. New York: Praeger; 1989. p. 217–235.
- [30] Dawson TL, Xie Y, Wilson M. Domain-general and domain-specific developmental assessments: do they measure the same thing? *Cogn Devel*. 2003;18:61–78.
- [31] Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19:349–357.
- [32] Kjellström S. Ansvar, Hälsa Och Människa: En Studie Av Idéer Om Individens Ansvar För Sin Hälsa [Responsibility, health and the individual: a study of ideas on personal responsibility for health]. Linköping: Linköping University; 2005.
- [33] Silge J, Robinson D. Text mining with R: a tidy approach. Sebastopol, CA: O'Reilly Media; 2017.
- [34] Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. *Res Syn Meth*. 2011;2:1–14.
- [35] Kostoff RN, del Rio JA, Humenik JA, et al. Citation mining: Integrating text mining and bibliometrics for research user profiling. *J Am Soc Inf Sci*. 2001;52:1148–1156.
- [36] Ananiadou S, McNaught J. Text mining for biology and biomedicine. Boston and London: Artech House London; 2006.
- [37] Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15:1277–1288.
- [38] Cohen AM, Hersh WR. A survey of current work in biomedical text mining. *Brief Bioinform*. 2005;6: 57–71.
- [39] Chen W-T, Barbour R. Life priorities in the HIV-positive Asians: a text-mining analysis in young vs. old generation. *AIDS Care* 2017;29:507–510.
- [40] Feldman R, Sanger J. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge, UK: Cambridge University Press; 2007.
- [41] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2016.
- [42] Singh N, Hu C, Roehl WS. Text mining a decade of progress in hospitality human resource management research: identifying emerging thematic development. *Int J Hosp Manage*. 2007;26:131–147.
- [43] Yu CH, Jannasch-Pennell A, DiGangi S. Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. *Qual Rep*. 2011;16:730.
- [44] Feinerer I, Hornik K. tm: Text Mining Package, 2014. URL <http://CRAN.R-project.org/package=tm> R package version 05-83[p 188]. 2011.
- [45] Bouchet-Valat M. SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library. R package version 05. 2014.
- [46] Team RC. A language and environment for statistical computing. R Foundation for statistical computing, 2015; Vienna, Austria, 2016.
- [47] Golino HF, Demetriou A. estimating the dimensionality of intelligence like data using exploratory graph analysis. *Intelligence*. 2017;62:54–70.
- [48] Golino HF. EGA package. 2016. URL <https://github.com/hfgolino/EGA>
- [49] Epskamp S, Rhemtulla M, Borsboom D. Generalized network psychometrics: Combining network and latent variable models. arXiv preprint arXiv: 160509288. 2016. doi: 10.1007/s11336-017-9557-x
- [50] Pons P, Latapy M. Computing communities in large networks using random walks. *JGAA*. 2006;10: 191–218.
- [51] Golino HF, Epskamp S. Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One*. 2017;12:e0174035.
- [52] Bentler PM. Comparative fit indexes in structural models. *Psychol Bull*. 1990;107:238.
- [53] Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol Bull*. 1980;88:588.
- [54] Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing structural equation models. Newbury Park: Sage; 1993. p. 136–162.
- [55] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6:1–55.
- [56] Abdi H, Williams LJ. Tukey's honestly significant difference (HSD) test. *Encyclopedia of research design*. Thousand Oaks, CA: Sage. 2010:1–5.
- [57] Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001; 66: 507–514.