Exploratory Likert Scaling as an Alternative to Exploratory Factor Analysis: Methodological Foundation and a Comparative Example Using an Innovative Scaling Procedure

Müller-Schneider, Thomas

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

methods, data, analyses | Vol. 16(1), 2022, pp. 51-76 DOI: 10.12758/mda.2021.12

Exploratory Likert Scaling as an Alternative to Exploratory Factor Analysis. Methodological Foundation and a Comparative Example Using an Innovative Scaling Procedure

Thomas Müller-Schneider

Universität Koblenz-Landau, Campus Landau

Abstract

Identifying the dimensional structure of a set of items (e.g., when studying attitudes) is an important and intricate task in empirical social research. In research practice, exploratory factor analysis is usually employed for this purpose. Factor analysis, however, has known problems that may lead to distorted results. One of its central methodological challenges is to select an adequate multidimensional factor space. Purely statistical decision heuristics to determine the number of factors to be extracted are of only limited value. As I will illustrate using an example from lifestyle research, there is a considerable risk of fragmenting a complex unidimensional construct by extracting too many factors (overextraction) and splitting it across several factors. As an alternative to exploratory factor analysis, this paper presents an innovative scaling procedure called *exploratory Likert scaling*. This methodologically based technique is designed to identify multiple unidimensional scales. It reliably finds even extensive latent dimensions without fragmenting them. To demonstrate this benefit, this paper takes up an example from lifestyle research and analyzes it using a novel R package for exploratory Likert scaling. The unidimensional scales are constructed sequentially by means of bottom-up item selection. Exploratory Likert scaling owes its high analytical potential to the principle of multiple scaling, which is adopted from Mokken scale analysis and transferred to classical test theory.

Keywords: dimensional analysis, classical test theory, multiple scaling, exploratory factor analysis, exploratory Likert scaling

A fundamental task and activity of empirical social research involves measuring latent dimensions and assessing their content by means of related indicators. Attitudes, action patterns, preferences, motives, and abilities are typical areas of such dimensions—also referred to as latent constructs or dispositions. Empirically, latent dimensions are inferred from item response patterns by employing specific statistical techniques. Major questions and issues in data analysis and methodology concern the dimensionality of a given domain (or universe of items) such as political attitudes or lifestyle preferences: Is the phenomenon in question structured by only one dimension or by several, and if so, how many dimensions are to be meaningfully distinguished in a certain theoretical context? How might one determine the dimensional structure of a set of items in number and content, and how might one then construct scales for measuring the identified dimensions? Understanding the latent dimensional structure of the data in question is essential to achieving conceptual clarity (Rose, 2014, pp. 21–45).

Among practitioners of social science research, a two-step approach of dimensional analysis predominates, which can also be found in relevant textbooks on data analysis. This approach begins by exploring and determining the dimensional structure of a set of items by means of factor analysis. The items of each of the extracted dimensions are then subjected to an item analysis in order to construct Likert scales according to classical test theory (e.g., Fromm, 2012; Kopp & Lois, 2014). The following article deals with the first step, that is, the exploration of dimensional structures. As for exploratory factor analysis (EFA), it is well known that determining the number of factors to be extracted may be a "knotty issue," as DeVellis (2012, p. 127) puts it. Finding an adequate multidimensional solution is still a crucial methodological challenge of EFA. Underextraction on the one hand and overextraction on the other may lead to substantial misinterpretation. This contribution presents an innovative scaling procedure that can serve as a useful alternative to EFA. I refer to this procedure as exploratory Likert scaling (ELS).

Since exploratory Likert scaling, unlike exploratory factor analysis, is based on the concept of multiple unidimensionality, this article begins with a methodological foundation of exploratory dimensionality analysis in the social sciences. Subsequently, the problem of EFA that pertains here is highlighted and illustrated with an example from lifestyle research. As it turns out, there is an imminent risk of splitting complex latent dimensions across the multidimensional factor space. This renders a gainful and appropriate application of EFA technically complicated and demanding in research practice. Against this backdrop, exploratory Likert scaling is outlined. It is shown to be a straightforward technique of multiple unidimen-

Direct correspondence to
Thomas Müller-Schneider, Universität Koblenz-Landau, Campus Landau
E-mail: tms@uni-landau.de

sional scaling based on classical test theory. Each scale is constructed by employing a "bottom-up" clustering procedure using item discrimination as fundamental criterion. The same example of lifestyle research is then used again to illustrate the analytical potential of ELS for the identification of multiple unidimensional constructs.

The Objective: Identifying Multiple Unidimensional Scales

From a methodological perspective, exploring the dimensional structure of a complex data set is anything but trivial. The starting point for further considerations is a unidimensional item response pattern. In general, unidimensionality (aka homogeneity) means that the components of a test or scale measure the same underlying property (latent dimension). In the context of classical test theory (CTT), unidimensionality implies a high degree of interrelatedness among items (Green et al., 1977; Heidenreich, 1984, p. 370; Nunnally, 1978, p. 274). The concept of internal consistency as a measure of reliability is essentially based on inter-item correlations. In addition, highly associated items are each correlated with the total score of all other items. This corrected item–total correlation (or item discrimination) is considered to be an indicator of the relationship between an item and the true score of the latent dimension in question (DeVellis, 2006, p. 52). Likert scaling, which uses Cronbach's alpha as a measure of internal consistency and the corrected item–total correlation as the criterion for item analysis, is not only the most common application of CTT but also by far the dominant scaling method in the social sciences. It is in this specific sense (i.e., with reference to CTT) that the term *Likert scaling* is used here.

The correlational approach to building unidimensional scales can also be found in item response theory. In Mokken scale analysis[1] for dichotomous items, measures of item discrimination and overall scale homogeneity are derived from the coefficient H_{ij}, which is equivalent to the corrected Phi (Phi/Phi_{max}) in 2×2 tables (Stokman & van Schuur, 1980, p. 23). In contrast to the correlational approach, the Rasch model uses the principle of local independence to define unidimensionality. In practical application, however, the model is not convincing. In the social sciences and especially in sociology it is rarely even used as a scaling method (international educational assessment studies such as PISA are something of an exception). Furthermore, and more fundamentally, it has been found to be unsuitable for assessing unidimensionality (Hattie, 1985; Stelzl, 1979). The option to relax the

1 Mokken scale analysis can be regarded as a nonparametric probabilistic version of Guttman scaling (van Schuur, 2003).

assumption of local independence, which was implemented in the context of model improvements (e.g., TenVergert et al., 1993), does not diminish the severe criticism of a problematic concept of unidimensionality.

The characterization of a unidimensional data structure as a definable group of highly correlated items can be readily extended to complex data structures. An item set with two or more underlying dimensions contains a corresponding number of item groups (or clusters) with specific properties: Within these groups the respective items are highly correlated; between groups the items are only poorly correlated or are not correlated at all. In the case of small and well-ordered correlation matrices, simple data inspection is sufficient to identify homogeneous item clusters. Upon expanding the matrices and increasing their complexity, the limits of this visual method are reached very quickly, so that specific multivariate techniques are required for the exploration, determination, and interpretation of dimensional structures. Basically, any technique capable of identifying homogeneous groups (clusters) of items in correlation matrices, such as factor or cluster analysis, is applicable. In this context, it should be noted that factor analysis can be described as a multidimensional extension of CTT (Fischer, 1974, p. 77).

The identification of suitable multivariate techniques of exploratory dimensionality analysis also requires clarifying the appropriate concept of dimensionality. One has to distinguish between multidimensionality and multiple unidimensionality (Jacoby, 1991, p. 35). Multidimensionality does not simply mean that an area of interest cannot be adequately captured by a single dimension. Additionally, multidimensionality entails the concept of locating objects (variables, individuals) simultaneously within an *n*-dimensional space. In factor analysis, for example, each item is defined multidimensionally by its loadings (correlations) on every factor of the selected solution. Multiple unidimensionality, in contrast, signifies that a complex data structure is represented by two or more unidimensional scales. Each item can be characterized by its relation to the respective set of scale items. To emphasize the difference, complex data structures with more than one underlying dimension can be described by multiple unidimensional scales (latent constructs) as opposed to a single multidimensional solution (Jacoby, 1991, p. 36).

So, then, what is the objective of exploratory dimensionality analysis? Is it to find a single multidimensional solution or multiple unidimensional scales? Research practice yields mixed messages that differ with respect to two typical steps of dimensional analysis. In the first step, an exploratory factor analysis (a multidimensional procedure) is conducted to determine the number of dimensions along with their associated items. In the second, the obtained multidimensional information is used to develop unidimensional scales (i.e., Likert scales according to CTT criteria) (e.g., Kopp & Lois, 2014). This common two-step practice with its switch from multidimensionality to multiple unidimensionality reflects the prominent methodological status of unidimensional constructs. Following McIver and Carmines (1981), "social scientists should strive to develop and use unidimensional concepts because they are more susceptible to theory-relevant research" (p. 14). They subsequently state:

> Multidimensional concepts, on the other hand, typically hamper such research because they are too ambiguous in terms of their meaning, too difficult to measure in a clear and precise manner, and too theoretically oriented themselves. Their complexity and ambiguity renders them less optimal for the development and assessment of social-science theories. In other words, using unidimensional scaling models to measure unidimensional concepts puts the measurement strategy on the same analytical level (p. 14).

Matters get more complicated by the fact that strict unidimensionality at the level of single items is only of ideal-typical nature. For example, a response to the item "reading a book" may be influenced by different latent dimensions, such as an inclination to enjoy high culture as well as a domestic leisure orientation. The fact that a single item may underlie more than one latent dimension does not, however, imply that the measurement concept of unidimensionality is invalid. This is because unidimensionality refers to the level of the overall scale and thus to the common core of meaning of all scale items. Single items are useful in building a unidimensional scale only to the extent that they tap into this common core (Nunnally, 1978, p. 274). If too strongly affected by one or more "interfering dimensions", an item has to be removed from the scale in question.

Against the methodological background presented and in accordance with common research practice, the objective in exploratory dimensionality analysis is to identify multiple unidimensional scales—or, with respect to the most common scaling model in the social sciences, to identify multiple Likert scales.

The concept of multiple unidimensionality does not disqualify exploratory factor analysis as a helpful first-step tool for the exploration of dimensional structures.[2] However, the multidimensional model of EFA is prone to certain problems that may result in distorted results. To avoid these problems, this paper suggests an alternative that is directly connected to the objective of identifying multiple Likert scales. Before turning to this alternative, exploratory Likert scaling, I will highlight one of the key challenges for exploratory dimensionality analysis that arises from the multidimensional model of factor analysis.

2 Due to model extensions, Rasch scaling can also now be used to define multidimensional structures (Cheng et al., 2009). Unlike EFA, however, it is not even a helpful first-step tool, as it is inherently unsuitable for the *exploration* of dimensional structures.

A Key Challenge for Exploratory Factor Analysis: Selecting an Adequate Factor Space

The term *exploratory factor analysis* is not used consistently in the literature, so its definition should be briefly clarified. First of all, strictly speaking, a distinction has to be made between two analytical models, namely, principal component analysis (PCA) and factor analysis “proper” (FA), the latter of which is based on the model of common factors (Fabrigar et al., 1999, p. 275). In addition to its confirmatory variant (CFA), which is not relevant here and therefore not considered, factor analysis (FA) can also be used for the dimensional exploration of complex correlation matrices. This type of procedure is called exploratory factor analysis. In social science research (especially in German-speaking countries), however, the term exploratory factor analysis is also used when principal component analysis (PCA) is employed as a statistical method for exploring dimensional structures. It is in this latter sense that I will speak of exploratory factor analysis hereafter. It should be noted that the common factor model and PCA differ significantly in their analytical basis but usually lead to equivalent results—even with regard to the number of factors or components considered (Wolff & Bacher, 2010, p. 349; Velicer & Jackson, 1990). It should also be stated that the problems encountered in the context of EFA do not arise for confirmatory factor analysis, as the number of factors in CFA is predetermined by theoretical considerations. However, the social science applications of factor analysis are mostly exploratory in nature because they are typically not focused on hypothesis testing but on the dimensional interpretation of complex item batteries.

EFA is undoubtedly a useful tool to discover multiple dimensions in terms of homogenous item clusters: “Variables that are correlated with one another but largely independent of other subsets of variables are combined into factors” (Tabachnick & Fidell, 2007, p. 607). In PCA, factors (used here synonymously with principal components) are extracted stepwise from the correlation matrix in such a way that they explain the maximum of the (remaining) variance among all items. Therefore, the first factor accounts for the most variance, and each successive factor accounts for a decreasing portion of the item variance. All factors are uncorrelated with each other. This iterative extraction procedure is used to construct an n-dimensional space of orthogonal factors, whereby the maximum number of extracted factors corresponds to the number of items entered in the analysis (aka the full component model). As the objective of exploratory dimensionality analysis is to identify clusters of highly intercorrelated items, the full component model is obviously worthless. A major task of EFA, therefore, is to determine a more parsimonious solution with an adequate number of factors being extracted. To this end, a variety of procedures (or stopping rules for further extraction) have been suggested (e.g., Peres-Neto et al., 2005; Hoyle & Duvall, 2004). The most relevant in research practice are based on the eigenvalue (Wolff & Bacher, 2010). This measure represents the amount of variance explained by each factor and equals the number of items in the full component model. The best-known and most commonly applied stopping rule for factor extraction, at least in the social sciences, is the so-called Kaiser rule or eigenvalue-greater-than-one rule (the default option in SPSS). As the name of the rule indicates, it states that a factor is to be extracted (retained) as long as its eigenvalue is greater than one. Another criterion is derived from the visual inspection of a scree plot, which represents factors with respect to their eigenvalue in a downward curve. According to the scree test (Cattell 1966), the point before the curve levels off (the "elbow") denotes the number of factors to be retained as significant. A less well-known procedure is the parallel test, which compares randomly generated eigenvalues with empirical ones. It needs to be stressed that the rules mentioned here typically do not lead to the same results; moreover, they (especially the scree test) may at times be ambiguous and therefore open to interpretation (Ledesma et al., 2015; Tabachnick & Fidell, 2007, p. 644–646; Wolff & Bacher, 2010, p. 343). To put it more generally, determining the number of factors is a fundamental issue in factor analysis and remains a major challenge that has come under considerable debate (Peres-Neto et al., 2005; Ledesma et al., 2015; Hoyle & Duval, 2004). Problems with (orthogonal or oblique) factor rotation[3] are also substantial (Sakaluk & Short, 2017) but not directly relevant to the present topic and can thus be neglected in the context of this article. Regardless, the selection of the number of factors is at any rate more critical than the selection of a rotation method (Tabachnick & Fidell, 2007, p. 644).

With respect to an appropriate number of factors to extract, factor analysis is susceptible to misspecifications, and this might lead to biased results in terms of identifying relevant latent dimensions. Such misspecifications can take the form of underextraction (i.e., extracting too few factors) or overextraction (i.e., extracting too many factors). Whereas it is widely agreed that underextraction leads in many cases to more severe distortions than overextraction (de Winter & Dodou, 2016; Fava & Velicer, 1996, p. 908; Wood et al., 1996), the extracting of surplus factors is presumably the more common problem. This is due to the fact that overextraction usually occurs in cases where the popular Kaiser criterion is employed, especially in combination with a large number of variables (Zwick & Velicer, 1986; Fava & Velicer, 1992, p. 388). It is assumed that a small number of extra factors may do little harm, but substantial overextraction results in the severe problem of factor fission (Cattel, 1978, p. 168; Fava & Velicer, 1992, p. 389; Wood et al., 1996). Factor fission (or factor splitting) denotes the phenomenon that items belonging to a common latent dimension are dispersed across different factors. An older study

3 In orthogonal rotation, extracted factors remain statistically independent (uncorrelated). In oblique rotation, the factors are allowed to correlate.

with real data sets even documented a massive change in the factor structure as the number of extra factors extracted was increased (Levonian & Comrey, 1966).

I will illustrate the problem of factor splitting by an example from lifestyle research. It will be picked up again in the following section for comparison with exploratory Likert scaling. The focus of the example is on three well-known schemes of everyday aesthetics identified and theorized by Schulze (1992), which are the high-culture scheme, the trivial scheme, and the tension scheme. Along with age and education, these three aesthetic patterns of everyday life are among the constituent characteristics of five social milieus. These aesthetic schemes are complex cross-situational response tendencies that are considered here as theoretically relevant dimensions (unidimensional constructs). The schemes were obtained by means of exploratory analysis from a very broad range of individual preferences and action tendencies within a total of 110 relevant items (Schulze, 1992, pp. 595–598). All items were measured on a five-point response scale. For the following analyses, the original data of the study with a total of 1,024 interviews were used. First, a unidimensional secondary analysis of each of the three schemes was performed to obtain rather homogeneous scales with items having discriminatory power of at least 0.45. The results are presented in Table 1. As can be seen, all three scales, as measured by Cronbach's alpha, have a very high internal consistency.

These three dimensions (unidimensional scales) are now contrasted with the results of an exploratory factor analysis (PCA), which was computed for the same data with the usual specifications. Using the eigenvalue-greater-than-one rule, 25 factors were extracted (see appendix, Table A1) and then subjected to varimax rotation. What matters here is not the individual results of the factor analysis but the mapping of the three unidimensional scales in the multidimensional factor space. As Table 1 shows, both the trivial scheme and the tension scheme are fully captured by a single factor each (factor 1 and 2). Almost all scale-specific items show high factor loadings. Only pub attendance (tension scheme) stands out with a lower loading on factor 2 and a simultaneous loading on factor 8. This, however, is completely unproblematic, provided that the item would be included in a subsequent unidimensional scale analysis.

A substantially different picture emerges for the high-culture scheme. First, it should be noted that factor 3 reflects significant manifestations of this aesthetic orientation. A number of relevant items (e.g., classical music and literature, visiting exhibitions and museums) show high factor loadings. Nevertheless, the problem of factor fission is evident. Fragments of the latent dimension and single items are scattered across the multidimensional factor space. For example, the fragment that primarily addresses literature about the inner life (self-awareness and psychological problems) is found on factor 7. Only the item that captures the preference for poetry also loads on factor 3. Further, the marker items of factor 9, which revolve around private educational inclination, have no discernible connection to factor 3. Reading

Der Spiegel loads exclusively on factor 14, and reading *Die Zeit* shows no substantial loading on any of the 25 factors at all. Engagement with literature has also broken out of the high-culture scheme and is located on factor 12 (with classical and modern literature building a "bridge" to factor 3). Overall, it must be noted that the high-culture scheme in the totality of its meanings is not evident in the exploratory factor analysis presented here.

Table 1 Everyday aesthetic schemes: Results from unidimensional scaling and exploratory factor analysis

Dimension	Items (preference for, interest in, inclination to…)	Corrected item–total correlation	Factor and loadings	
			FA 1	
Trivial scheme	*Heimat* films[1]	0.70	0.76	
	Shows/quizzes (TV)	0.64	0.62	
	Popular theater (TV)	0.73	0.75	
	Local broadcasts	0.51	0.47	
	Nature broadcasts	0.46	0.41	
	Light music	0.57	0.62	
	German hits	0.67	0.71	
	German folk songs	0.79	0.76	
	Bavarian folk music	0.79	0.79	
	Brass music	0.78	0.78	
		$\alpha = 0.91$		
			FA2	FA8
Tension scheme	Pop and rock music (TV)	0.73	0.70	
	Rock music	0.77	0.72	
	Oldies (e.g., The Beatles)	0.56	0.61	
	Reggae music	0.68	0.72	
	Soul music	0.69	0.77	
	Pop music	0.81	0.78	
	Folk music	0.60	0.68	
	Blues music	0.58	0.69	
	Attending concerts (rock, pop, jazz)	0.63	0.55	
	Going to the movies	0.62	0.51	
	Going to a pub	0.48	0.36	0.40
	Going to a discotheque	0.54	0.44	
		$\alpha = 0.91$		

Table 1 continued

Dimension	Items (preference for, interest in, inclination to…)	Corrected item–total correlation	Factor and loadings				
			FA3	FA7	FA9	FA12	FA14
High-culture scheme	Classical music	0.62	0.76				
	Contemporary classical music	0.51	0.64				
	Classical concerts	0.56	0.68				
	Theater (TV)	0.49	0.70				
	Newspaper: culture section	0.50	0.48				
	Visiting exhibitions/ galleries	0.55	0.46				
	Poems	0.58	0.45	0.45			
	Self-awareness literature	0.55		0.80			
	Psychological problem literature	0.59		0.76			
	Writing (e.g., diary)	0.46		0.38			
	Classical literature	0.75	0.56	0.31		0.33	
	Modern literature	0.70	0.39			0.38	
	Books on social/ political issues	0.63		0.38		0.37	
	Book reading	0.50				0.57	
	Courses, education	0.50			0.72		
	Language learning	0.47			0.71		
	Professional training (at home)	0.54			0.67		
	Reading *Der Spiegel*	0.51					0.51
	Reading *Die Zeit*	0.52					
		$\alpha = 0.91$					

[1] Sentimental films in an idealized rural setting

Note: For clarity, only factor loadings greater than 0.3 are reported (for "Reading *Die Zeit*", there was no factor loading with an absolute value greater than 0.3).

However, factor analysis does not fundamentally fail to represent fragments and individual items of the high-culture scheme within one single factor. One only has to reduce the number of extracted factors in such a way that the relevant construct (i.e., the high-culture scheme) is not split up and becomes visible in its entirety. The Kaiser criterion was not used for this purpose; rather a series of analyses with a gradually decreasing and predetermined number of factors was computed. Reducing the number of factors from eight to seven yielded the expected switch in factor structure, which is to say, the entire dimension of the high-cultural orientation was mapped onto a single (the first) factor. The second factor represents the tension scheme and the third the trivial scheme. The remaining four factors reveal further aspects of everyday preferences, which have to do with sports, shopping, maintaining social contacts, and domestic activities. This solution with seven factors is roughly in the range considered by the scree plot (five or six factors, depending on its interpretation). The solutions with three to six factors also represent each of the three relevant aesthetic patterns in a single factor. In social science research, it is quite common not to adhere too strictly to potentially problematic or ambiguous statistical criteria in the search for an appropriate *n*-dimensional factor space but rather to consider a range of conceivable solutions. This procedure is not only completely in line with the above methodological considerations for exploratory dimensionality analysis (identifying multiple unidimensional constructs) but also explicitly advised (Wolff & Bacher 2010, p. 343), and for good reason. Ultimately, it comes down to the interpretability and scientific usability of factors. As Tabachnick and Fidell (2007) concisely put it, "A good PCA … 'makes sense'; a bad one does not" (p. 608).

At this point, the question arises as to whether there is a different and possibly more suitable method to identify multiple unidimensional constructs than to try out a range of conceivable *n*-dimensional factor solutions. Exploratory Likert scaling offers a useful alternative to this strategy.

Exploratory Likert Scaling

Exploratory Likert scaling is an innovative method for discovering clusters of internally well and externally poorly correlated items within a given data set. It is based on a generalizable scaling procedure that works according to the crystallization principle and was originally proposed by Mokken (1971) for a step-by-step scale construction. The nucleus of crystallization is the maximally homogeneous "two-item scale" of a data set, which is then gradually extended by "bottom-up item selection" to a scale that meets the conditions of the monotone homogeneity model (Hemker et al., 1995, p. 342; Sijtsma et al., 1990, pp. 181–183). By taking the corrected item–total correlation (item discrimination according to CTT) as a coefficient of scalability, the crystallization principle can be applied immediately to the construction of a Likert scale. I suggest the following algorithm on the basis of bottom-up Mokken scale analysis:

1. Find the two items with the highest positive correlation.[4] Consider this pair of items as the potential crystallization nucleus of a Likert scale and calculate their total score.
2. From the remaining items, select the one that correlates most highly with the total score (i.e., has the highest item discrimination). Expand the scale nucleus by this item and recalculate the total score (with *n*+1 items).

4 The algorithm can also account for negative correlations (reversed items), although this is not relevant in the present context.

3. Repeat the process of step 2 until a predefined lower bound (minimum item–total correlation) is reached. The bottom-up item selection is then completed for this scale.

The use of the (corrected) item–total correlation as a criterion for scale extension not only aims at finding items of a latent dimension that are as discriminative as possible but serves at the same time to establish the internal consistency of the emerging scale in an optimal way. Higher item–total correlations of the selected items will result in a higher average inter-item correlation and thus also in a higher value for Cronbach's alpha (Lord & Novick, 1968, pp. 330–331). After the construction of the first scale is completed with step three of the algorithm, the search for additional scales begins. This is accomplished by means of "multiple scaling" (Mokken, 1971, pp. 194–195; Sijtsma et al., 1990, p. 185), which means that the entire scaling process is iterated. The algorithm therefore has to be extended by a fourth step:

4. Try to create a further scale from the remaining item pool by repeating steps 1 to 3. Then start again with step 4 and continue the process until no new scale nucleus can be found (specified by the minimum item–total correlation).

Multiple scaling according to the crystallization principle enables a sequential identification of groups of internally highly correlated items and thus of multiple unidimensional scales. Multiple scaling is also appropriately seen as "sequential clustering" of items (van Abswoude et al., 2004). Especially with regard to the main objective here, the exploration of the dimensional structure of an item pool using this procedure may, however, lead to problematic results. Depending on the value for the specified item–total correlation, this can be expected to obscure the dimensional structure of the data. The corresponding problem is already familiar from multiple Mokken scaling (Sijtsma & Molenaar, 2002, p. 80). A value close to zero would merge (almost) all items into a single scale, even if two or even more dimensions clearly underlie the data. In the context of EFA, one would use the term *under-extraction*. If, on the other hand, one chooses a rather high value for the minimum item discrimination, the above algorithm would split unidimensional scales into a number of fragments. This is equivalent to the problem of factor fission. However, the problem can be easily solved, and above all in a way that optimizes the exploratory potential of multiple scaling. The process of multiple scaling is divided into two steps wherein a high value for the minimum item discrimination is deliberately set in the first step in order to search for very homogeneous kernels of potential scales (search procedure). These kernels then serve as starting sets for the second scaling step (extension procedure), in which the minimum item–total correlation is significantly lowered and overlapping scaling is allowed. Overlapping scale construction means that each item can be assigned not only to the first but also to all subsequent potential scale kernels. Each of them has the opportunity, so to speak, to collect all scalable items. Now the exploratory potential of the entire scaling procedure, or exploratory Likert scaling, becomes visible: Each latent dimension can be determined reliably and completely even if only one scale fragment of the respective dimension was identified in the first step of the scaling procedure. What remain are only single non-scalable items or item groups that cannot be interpreted in a meaningful way or have no scientific use in the given context.

I will now contrast the factor analysis discussed above with an exploratory Likert scaling based on the same items. For this purpose, the R package "elisr" (Bißantz, 2021) was used. This package was developed on the author's initiative specifically for exploratory Likert scaling. All 110 items of the everyday aesthetic preferences were included in the search procedure. For the construction of potential scale kernels, an item–total correlation of 0.60 was set as the lower bound (in principle, it is reasonable to start several runs with varying lower bounds to ensure that kernels of all relevant dimensions are found). Table 2 presents the potential kernels in the order of their construction.

The software reports the average inter-item correlation and Cronbach's alpha as descriptive measures of internal consistency as well as the corrected item–total correlation that is essential for scale construction. To be precise, one should speak of a marginal item–total correlation, since this value reflects the item discrimination that is found at the moment when the scale is extended by the item in question (with subsequent scale expansion, this value may change). As can be seen in Table 2, a total of nine potential kernels is found at a minimum item–total correlation of 0.60, with some of them consisting of only a two-item scale. The first line of the respective item lists shows the two items that were fused first. Due to the specification of the search procedure, all kernels show a very high internal consistency (measured by the average inter-item correlation or Cronbach's alpha). The three relevant dimensions (the high-culture, trivial, and tension scheme) are all represented by scale fragments. The high-culture scheme even appears in five fragments with different contents (scales 3, 5, 6, 7, and 8), whereby the two items signaling an interest in opera (scale 3) are curiously not included in the overall scale presented in Table 1 above. The contents of the remaining two kernels (scales 2 and 9), which indicate a preference for information about sports and for domestic pursuits, were also identified in the EFA.

In the second scaling step (the extension procedure), the minimum item–total correlation was substantially decreased in order to allow each scale kernel to be extended with relevant items. The value of the minimum item–total correlation is now no longer oriented towards the search for very homogeneous scale kernels but rather towards the still acceptable item discrimination with respect to an overall scale. In this context, the value of 0.3 is often mentioned, but content-related aspects should also be taken into account. So as not to generate scales that were too extensive for reasons of clarity, a lower bound of 0.40 was selected in the present

Table 2 Results of the search procedure

Scales and steps	Items	r_{itm}	$\bar{r}$	α
Scale 1				
1	Brass music ‖ Bavarian folk music	0.87	0.87	0.93
2	German folk songs	0.78	0.79	0.92
3	Popular theater (TV)	0.66	0.70	0.91
4	*Heimat* films[1]	0.69	0.67	0.91
5	German hits	0.64	0.63	0.91
6	Shows/quizzes (TV)	0.61	0.59	0.91
Scale 2				
1	Sports (newspaper) ‖ Sports (TV)	0.80	0.80	0.89
2	Sports magazines	0.62	0.65	0.85
Scale 3				
1	Opera (music) ‖ Opera (TV)	0.78	0.78	0.88
Scale 4				
1	Pop music ‖ Rock and pop (TV)	0.76	0.76	0.86
2	Rock music	0.78	0.74	0.90
3	Soul music	0.63	0.66	0.89
4	Reggae music	0.66	0.62	0.89
Scale 5				
1	Self-awareness ‖ Psychological problem literature	0.76	0.76	0.86
Scale 6				
1	Classical literature ‖ Modern literature	0.69	0.69	0.82
2	Books on social/political issues	0.61	0.60	0.82
Scale 7				
1	Courses/education ‖ Professional training (at home)	0.66	0.66	0.79
Scale 8				
1	Classical concerts ‖ Classical music (preference)	0.62	0.62	0.76
Scale 9				
1	Cleaning up ‖ Tidying	0.62	0.62	0.76

[1] See footnote 1, Table 1.

Note: Minimal item–total correlation for the search procedure = 0.6; r_{itm} = marginal item–total correlation; $\bar{r}$ = average inter-item correlation; α = Cronbach's alpha.

exploratory analysis. Of the nine scales extended in the second scaling step, four are documented in this paper (scales 3 and 5 below in the text and scales 1 and 4 in the appendix, Table A2). The trivial scheme (extended scale 1) is now represented by 12 items. Drawing a line below the item “nature broadcasts” yields the precise set of ten items listed in Table 1. The two remaining items (preference for comedy movies and light fiction) also belong to the extended scale because the minimum item–total correlation chosen for the extension procedure (0.40) is lower than that for the scales compiled in Table 1 (which is 0.45). The same can be said for the tension scheme. In the expansion process of scale kernel 4, three additional items (visiting a night club, meeting in the city, interest in sci-fi/fantasy on TV) were included in addition to those shown in Table 1.

The scale extensions that affect the high-culture scheme are of particular interest and warrant closer scrutiny. The most important result can be seen in the fact that the scheme crystallizes completely at all its scale fragments found in the first step. Contrary to the EFA, no splitting of the latent dimension occurs. This is exactly what is ensured by the extension procedure in the second scaling step. If we look at extended scale 5 (Table 3), for example, we can see that the scale kernel, which is about self-awareness and dealing with psychological problems, is first expanded to include indicative topics (e.g., classical music and literature) and then educationally relevant content. Again, if one were to draw a boundary line at a marginal discriminatory power of 0.45, one would find all items of the high-culture scheme from Table 1. The same holds for the extended scales 6 and 7 (not documented in the appendix), whereby scale 6 starts with indicative high-culture topics and scale 7 with education-specific content. The extended scale 3, with its crystallization nucleus of the two preferences for opera (music, TV), also collects all relevant items, but the education-specific content is now included only below a minimum item discrimination of 0.45.[5] This demonstrates that the lower bound for exploratory purposes should be set rather lower than higher. Items that are borderline in terms of content or statistics can be excluded again for the final scales at a later stage. A final item selection is warranted in any case, given that, as already mentioned, the item–total correlation of an item can change its value in the course of the expansion procedure. Thus, the (marginal) item–total correlation of the two opera items, which is very high when they are merged to form a scale nucleus (0.78, identical to the bivariate correlation), falls below 0.45 in the further extension process. This is the reason why the two items were not included in the high-culture scale reported in Table 1. Extended scale 8 (not documented in the appendix) also contains all items of the high-culture dimension.

5 The items “Courses, education” and “Language learning” both reach the threshold of 0.45, but only after the item “Professional training” is included, which has a marginal discrimination power below this value (0.44).

As for the remaining two scale kernels from the search procedure (scales 2 and 9), these were enlarged by two and four items, respectively (not documented in the appendix). Although these scales are easy to interpret, they remain fragmentary (at least in the analyzed data set and for the specified minimum item–total correlation of 0.40). The extended scale 2 with a total of five items still focuses on sports, and scale 9, with its six items, revolves all around topics stereotyped as female (domestic chores, fashion, cosmetics). It is important to note that these two scales (fragments) neither influence nor even interfere with the bottom-up and sequential construction of the three relevant dimensions. Exploratory Likert scaling is not affected by irrelevant items or scale fragments. This, however, does not apply to the same extent to EFA. First of all, the irrelevant fragments "build" factors with an eigenvalue greater than one and are thus involved in determining the *n*-dimensional

Table 3 Results of the extension procedure: extended scales 3 and 5

Scales and steps	Items	r_{itm}	$\bar{r}$	α
Scale 3				
1	Opera (music) ǁ Opera (TV)	0.78	0.78	0.88
2	Classical music	0.60	0.63	0.84
3	Concerts with classical music	0.59	0.57	0.84
4	Theater (TV)	0.58	0.54	0.85
5	Classical literature	0.56	0.51	0.86
6	Contemporary classical music	0.56	0.49	0.87
7	Poems	0.51	0.46	0.87
8	Modern literature	0.53	0.45	0.88
9	Newspaper: culture section	0.52	0.43	0.88
10	Visiting exhibitions/galleries	0.53	0.42	0.89
11	Books on social/political issues	0.49	0.41	0.89
12	Psychological problem literature	0.47	0.40	0.89
13	Self-awareness literature	0.48	0.38	0.90
14	Book reading	0.47	0.38	0.90
15	Reading *Die Zeit*	0.45	0.36	0.90
16	Reading *Der Spiegel*	0.45	0.36	0.90
17	Professional training (at home)	0.44	0.35	0.91
18	Courses, education	0.45	0.34	0.91
19	Language learning	0.45	0.33	0.91
20	Writing (e.g., diary)	0.43	0.33	0.91
21	Documentaries (TV)	0.43	0.32	0.91
22	Newspaper: politics section	0.43	0.31	0.91
23	Jazz music	0.42	0.31	0.91

Table 3 continued

Scales and steps	Items	r_{itm}	$\bar{r}$	α
Scale 5				
1	Self-awareness ‖ Psychological problem literature	0.76	0.76	0.86
2	Books on social/political issues	0.48	0.55	0.79
3	Modern literature	0.57	0.52	0.81
4	Classical literature	0.66	0.52	0.84
5	Poems	0.57	0.50	0.86
6	Classical music	0.52	0.47	0.86
7	Classical concerts	0.53	0.45	0.87
8	Visiting exhibitions/galleries	0.52	0.43	0.87
9	Contemporary classical music	0.52	0.42	0.88
10	Newspaper: culture section	0.53	0.41	0.88
11	Theater (TV)	0.54	0.40	0.89
12	Book reading	0.49	0.39	0.89
13	Reading *Die Zeit*	0.47	0.38	0.90
14	Reading *Der Spiegel*	0.48	0.37	0.90
15	Professional training (at home)	0.47	0.36	0.90
16	Courses, education	0.47	0.36	0.90
17	Language learning	0.46	0.35	0.91
18	Writing (e.g., diary)	0.46	0.34	0.91
19	Jazz music	0.43	0.33	0.91
20	Documentaries (TV)	0.41	0.33	0.91
21	Newspaper: politics section	0.42	0.32	0.91
22	Opera (music)	0.41	0.31	0.91
23	Opera (TV)	0.43	0.31	0.91

Note: Minimal item–total correlation for the extension procedure = 0.4; r_{itm} = marginal item–total correlation; $\bar{r}$ = average inter-item correlation; α = Cronbach's alpha.

factor space (at least according to the greater-than-one rule). In any case, the fragments must be represented in the selected *n*-dimensional space. This cannot leave relevant factors completely unaffected because they have to be mapped in the same multidimensional factor space too. Whether insignificant scales (or fragments) lead to distortions in exploratory factor analysis is difficult to answer in general, if only because the results also involve substantial subjective decisions by the researcher. It must be added here, however, that the very existence of irrelevant item clusters makes it difficult in principle to speak of a "true" dimensionality or a "true" number of factors, contrary to what is sometimes found in the literature (e.g., Fava & Velicer, 1996, p. 908; Wood et al., 1996).

The explanatory scaling process to identify relevant latent constructs is followed by a second step of dimensional analysis: item analysis to construct the final scales. This second step is “business as usual” and outside the focus of this contribution (see Introduction). Nevertheless, some procedural remarks might be helpful. The final item analysis is based on all items of an extended scale that was selected for representing a latent dimension. Items found to have too little discriminatory power (corrected item–total correlation) must be removed from the scale in question. Usually, a respondent’s scale value is then computed as the summated score of all items included in the scale. But how does one deal with overlapping items? Following the concept of multiple unidimensionality, each item should be assigned to one scale only (according to statistical or content criteria). This should also be done in order not to overestimate the correlation between the final scales on grounds of multiply allocated items.

Conclusion and Discussion

This contribution has focused on exploratory dimensionality analysis in the social sciences. It began with methodological considerations on dimensional structures in complex data sets and their empirical identification. It was noted that, with reference to CTT, structures with more than one latent dimension are empirically reflected in a corresponding number of clusters with internally well and externally poorly correlated items. This contribution then further elaborated that the main objective of exploratory dimensional analysis is to find multiple unidimensional constructs as opposed to a single multidimensional solution. With reference to the common research practice of unidimensional scaling within the framework of CTT, this means identifying multiple Likert scales.

Since exploratory factor analysis is a genuinely multidimensional procedure, it is not designed to identify multiple unidimensional structures. Instead, the technique searches for a single n-dimensional (orthogonal) factor space to adequately represent multiple item clusters. One of the main methodological challenges of EFA is to determine the number of factors that span this n-dimensional space. If the most frequently used statistical criterion, the eigenvalue greater-than-one rule, is applied, it is widely acknowledged that one has to reckon with overextraction and factor fission. As was illustrated by the example provided from lifestyle research, this risks capturing extensive latent dimensions only in the way of disconnected fragments and to completely overlook single items scattered in the overextracted n-dimensional space. The best practice of an exploratory dimensionality analysis by means of factor analysis, at least in the social sciences, is therefore not to rely primarily on ambiguous statistical criteria but (as is often done anyway) to check a range of conceivable solutions and then decide on the interpretability and scientific significance of the factors found. The respective items of each factor can then be subjected to item analysis in order to construct unidimensional Likert scales.

Exploratory Likert scaling is a useful alternative for analyzing dimensional structures. This novel method belongs to the multiple unidimensional scaling approach, as it has already been implemented in Mokken scale analysis. Compared to the statistically complex EFA, exploratory Likert scaling (ELS) is a straightforward and completely different technique. It does not require a predefinition of a multidimensional (orthogonal) space and is better suited for identifying multiple unidimensional constructs than EFA owing to its bottom-up item selection and sequential clustering. A two-step multiple scaling strategy that combines an initial search for homogeneous scale kernels with their subsequent expansion not only avoids a methodological problem with sequential clustering but also optimizes the exploratory potential of the scaling procedure. Starting with any fragment of a unidimensional construct, the procedure interlinks all relevant contents of the construct. No splitting will occur, even if the unidimensional construct is complex in terms of the underlying empirical association structure. Also, large numbers of items do not pose any difficulties for the multiple unidimensional scaling approach. Especially in exploratively demanding data situations—large numbers of items, and high degrees of complexity but with unidimensional associations between items nevertheless—ELS is superior to EFA, the latter of which may quickly become confusing or even misleading in the case of substantial overextraction.

In the literature, there have been proposals on how to optimize factor analysis in order to make better decisions on the number of factors to retain. Lawrence and Hancock (1999), for example, state that "[t]he implementation of more precise factor extraction decision heuristics is essential" (p. 569). Referring to Zwick and Velicer (1986), they point to the minimum average partial procedure and parallel analysis as "extremely promising alternatives" (p. 569) to conventional practice. Ledesma et al. (2015) suggest enhancements of the scree test in the hope of providing better tools to determine the number of factors to retain. However, for identifying multiple unidimensional constructs, the approach of optimizing statistical (formal) criteria to define the number of factors is only of limited value. One reason for this is that the number of factors cannot be totally objectified on the basis of statistics alone. Apart from simulation purposes, there is, as mentioned above, no absolutely "true" dimensionality of a set of items, at least in the field of social sciences. Above all, the attempt at statistical optimization proceeds in the *wrong direction*. From the methodological point of view of multiple unidimensional scaling, the main problem of EFA is that an n-dimensional space has to be defined at all. The multidimensional approach creates unnecessary statistical complexity in exploratory dimensionality analysis, which in turn may lead to misspecifications and inappropriate results.

Multiple scaling can in principle also be performed using hierarchical clustering methods, as has already been suggested for Mokken scaling (van Abswoude et al., 2004). The dendrogram visualizes the fusion process and can be interpreted similarly to exploratory Likert scaling in terms of bottom-up scaling. With an appropriate fusion algorithm, a hierarchical agglomerative cluster analysis of items, as was demonstrated in a case study for Mokken scaling (Müller-Schneider, 2001), leads to substantially the same results as a two-step sequential scale construction (i.e., a search and extension procedure). Nevertheless, there are reasons to prefer exploratory Likert scaling. As the number of items increases, the dendrogram becomes less clear, which considerably impairs the visual analysis of bottom-up item selection and the dimensionality of the data. In addition, and more importantly, exploratory Likert scaling with its characteristic coefficients is, unlike cluster analysis, directly integrated into the analytical framework of dimensional analysis. Item–total correlation determines the constitution as well as the extension of a scale kernel, and at each step, the internal consistency of the resulting scale can be precisely traced by the average item correlations and Cronbach's alpha.

Besides the reliable identification of multiple unidimensional constructs, there is another noteworthy advantage of ELS. In order to interpret the determined scales appropriately, there is no need for such a thing as factor rotation. This being the case, ELS avoids unnecessary model complications and all the specific issues involved therein. Consequently, there is also no need for an always somewhat arbitrary oblique rotation to map any given correlations between latent dimensions. Since the statistical identification of multiple dimensions using ELS does not demand a predefined space of orthogonal dimensions, the constructs can correlate with each other (or not) in a natural way from the outset.

References

Bißantz, S. (2021). elisr: Exploratory Likert Scaling. R package version 0.1.1.

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. doi:10.1207/s15327906mbr0102_10

Cattell, R.B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences.* New York: Plenum Press.

Cheng, Y.-Y., Wang W.-C., & Ho, Y.-H. (2009). Multidimensional Rasch Analysis of a Psychological Test With Multiple Subtests: A Statistical Solution for the Bandwidth-Fidelity Dilemma. *Educational and Psychological Measurement, 69*(3), 369–388. doi:10.1177/0013164408323241

DeVellis, R. F. (2006). Classical Test Theory. *Medical Care*, *44*(11), 50–59. doi:10.1097/01.mlr.0000245426.10853.30

DeVellis, R. F. (2012). *Scale Development: Theory and Applications* (3rd ed.). Thousand Oaks, CA: Sage Publications.

De Winter, J. C. F., & Dodou, D. (2016). Common Factor Analysis versus Principal Component Analysis: A Comparison of Loadings by Means of Simulations.

Communications in Statistics – Simulation and Computation, *45*(1), 299–321. doi:10.1080/03610918.2013.862274

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, *4*(3), 272–299. doi:10.1037/1082-989X.4.3.272

Fava, J. L., & Velicer, W. F. (1992). The Effects of Overextraction on Factor and Component Analysis. *Multivariate Behavioral Research*, *27*(3), 387–415. doi:10.1207/s15327906mbr2703_5

Fava, J. L., & Velicer, W. F. (1996). The Effects of Underextraction in Factor and Component Analyses. *Educational and Psychological Measurement*, *56*(6), 907–929. doi:10.1177/0013164496056006001

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern, Stuttgart, Vienna: Huber.

Fromm, S. (2012). *Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten* (2. Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of Coefficient Alpha as an Index of Test Unidimensionality. *Educational and Psychological Measurement*, *37*(4), 827–838. doi:10.1177/001316447703700403

Hattie, J. (1985). Methodology Review: Assessing Unidimensionality of Tests and Items. *Applied Psychological Measurement*, *9*(2), 139–164. doi:10.1177/014662168500900204

Heidenreich, K. (1984). Grundbegriffe der Meß- und Testtheorie. In E. Roth (Eds.), *Sozialwissenschaftliche Methoden: Lehr- und Handbuch für Forschung und Praxis* (pp. 352–384). Munich, Vienna: Oldenbourg.

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of Unidimensional Scales From a Multidimensional Item Bank in the Polytomous Mokken IRT Model. *Applied Psychological Measurement*, *19*(4), 337–352. doi:10.1177/014662169501900404

Hoyle, R., & Duvall, J. (2004). Determining the Number of Factors in Exploratory and Confirmatory Factor Analysis. In D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 302–317). Thousand Oaks, CA: Sage Publications.

Jacoby, W. G. (1991). *Data Theory and Dimensional Analysis*. Newbury Park, London, New Delhi: Sage Publications.

Kaplan, D. (Ed.) (2004). *The Sage Handbook of Quantitative Methodology for the Social Sciences*. Thousand Oaks, CA: Sage Publications.

Kopp, J., & Lois, D. (2014). *Sozialwissenschaftliche Datenanalyse: Eine Einführung* (2. überarbeitete und aktualisierte Auflage). Wiesbaden: Springer VS.

Lawrence, F. R., & Hancock, G. R. (1999). Conditions Affecting Integrity of a Factor Solution under Varying Degrees of Overextraction. *Educational and Psychological Measurement*, *59*(4), 549–579. doi:10.1177/00131649921970026

Ledesma, R. D., Valero-Mora, P., & Macbeth, G. (2015). The Scree Test and the Number of Factors: A Dynamic Graphics Approach. *The Spanish Journal of Psychology*, *18*. doi:10.1017/sjp.2015.13

Levonian, E., & Comrey, A. L. (1966). Factorial Stability as a Function of the Number of Orthogonally-Rotated Factors. *Behavioral Science*, *11*(5), 400–404. doi:10.1002/bs.3830110511

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

McIver, J. P., & Carmines, E. G. (1981). *Unidimensional Scaling.* Beverly Hills: Sage Publications.

Mokken, R. J. (1971). *A Theory and Procedure of Scale Analysis: With Applications in Political Research*. The Hague: Mouton.

Müller-Schneider, T. (2001). Multiple Skalierung nach dem Kristallisationsprinzip: Eine Alternative zur explorativen Faktorenanalyse. *Zeitschrift für Soziologie, 30*(4), 305–315. doi:10.1515/zfsoz-2001-0404

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited. *Computational Statistics & Data Analysis*, *49*(4), 974–997. doi:10.1016/j.csda.2004.06.015

Rose, J. (2014). *The Public Understanding of Political Integrity: The Case for Probity Perceptions*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

Sakaluk, J. K., & Short, S. D. (2017). A Methodological Review of Exploratory Factor Analysis in Sexuality Research: Used Practices, Best Practices, and Data Analysis Resources. *Journal of Sex Research*, *54*(1), 1–9. doi:10.1080/00224499.2015.1137538

Schulze, G. (1992). *Die Erlebnisgesellschaft: Kultursoziologie der Gegenwart.* Frankfurt am Main: Campus-Verlag. [(2005). *The Experience Society*. London: Sage Publications.]

Sijtsma, K., Debets, P., & Molenaar, I. W. (1990). Mokken Scale Analysis for Polychotomous Items: Theory, a Computer Program and an Empirical Application. *Quality and Quantity*, *24*(2), 173-188 doi:10.1007/BF00209550

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to Nonparametric Item Response Theory* (Vol. 5). Thousand Oaks, CA: Sage Publications.

Stelzl, I. (1979): Ist der Modelltest des Rasch-Modells geeignet, Homogenitätshypothesen zu überprüfen? Ein Bericht über Simulationsstudien mit inhomogenen Daten. *Zeitschrift für experimentelle und angewandte Psychologie, 26*(4)*,* 652–672.

Stokman, F., & van Schuur, W. (1980). Basic Scaling. *Quality and Quantity*, *14*(1), 5–30. doi:10.1007/BF00154792

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston, MA: Pearson.

TenVergert, E., Gillespie, M., & Kingma, J. (1993). Testing the Assumptions and Interpreting the Results of the Rasch Model Using Log-Linear Procedures in SPSS. *Behavior Research Methods, Instruments, & Computers, 25*(3), 350–359. doi:10.3758/BF03204525

Van Abswoude, A. A. H., Vermunt, J. K., Hemker, B. T., & van der Ark, L. A. (2004). Mokken Scale Analysis Using Hierarchical Clustering Procedures. *Applied Psychological Measurement*, *28*(5), 332–354. doi:10.1177/0146621604265510

Van Schuur, W. H. (2003). Mokken Scale Analysis: Between the Guttman Scale and Parametric Item Response Theory. *Political Analysis*, *11*(2), 139–163. doi.org/10.1093/pan/mpg002

Velicer, W. F., & Jackson, D. N. (1990). Component Analysis versus Common Factor Analysis: Some Further Observations. *Multivariate Behavioral Research*, *25*(1), 97–114. doi:10.1207/s15327906mbr2501_12

Wolff, H.-G., & Bacher, J. (2010). Hauptkomponentenanalyse und explorative Faktorenanalyse. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Da-*

tenanalyse (pp. 333–365). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-92038-2_15

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of Under- and Overextraction on Principal Axis Factor Analysis with Varimax Rotation. *Psychological Methods, 1*(4), 354–365. doi:10.1037/1082-989X.1.4.354

Zwick, W. R., & Velicer, W. F. (1986). Comparison of Five Rules for Determining the Number of Components to Retain. *Psychological Bulletin*, *99*(3), 432–442. doi:10.1037//0033-2909.99.3.432

Appendix

Table A1 Eigenvalues of extracted components (greater than 1)

Component	Eigenvalue	Component	Eigenvalue	Component	Eigenvalue
1	15.18	11	1.77	21	1.11
2	7.88	12	1.65	22	1.08
3	7.11	13	1.50	23	1.06
4	5.12	14	1.43	24	1.05
5	3.08	15	1.35	25	1.02
6	2.52	16	1.32		
7	2.21	17	1.27		
8	2.08	18	1.22		
9	2.01	19	1.19		
10	1.88	20	1.16		

Table A2 Results of the extension procedure: extended scales 1 and 4

Scales and steps	Items	r_{itm}	$\bar{r}$	α
Scale 1				
1	Brass music ‖ Bavarian folk music	0.87	0.87	0.93
2	German folk songs	0.78	0.79	0.92
3	Popular theater (TV)	0.66	0.70	0.91
4	*Heimat* films[1]	0.69	0.67	0.91
5	German hits	0.64	0.63	0.91
6	Shows/quizzes (TV)	0.61	0.59	0.91
7	Light music	0.56	0.56	0.91
8	Local broadcasts	0.48	0.52	0.91
9	Nature broadcasts	0.46	0.49	0.90
10	Comedy movies	0.44	0.46	0.90
11	Light fiction	0.41	0.43	0.90
Scale 4				
1	Pop music ‖ Rock and pop (TV)	0.76	0.76	0.86
2	Rock music	0.78	0.74	0.90
3	Soul music	0.63	0.66	0.89
4	Reggae music	0.66	0.62	0.89
5	Going to the movies	0.58	0.58	0.89
6	Attending concerts (rock, pop, jazz)	0.60	0.55	0.90
7	Going to a discotheque	0.59	0.53	0.90
8	Folk music	0.55	0.50	0.90
9	Blues music	0.57	0.49	0.91
10	Oldies (e.g., The Beatles)	0.55	0.47	0.91
11	Going to a pub	0.48	0.45	0.91
12	Visiting a night club	0.43	0.43	0.91
13	Meeting in the city	0.43	0.41	0.91
14	Science Fiction, fantasy (TV)	0.40	0.39	0.91

[1] See footnote 1, Table 1.
Note: Minimal item–total correlation for the extension procedure = 0.4; r_{itm} = marginal item–total correlation; $\bar{r}$ = average inter-item correlation; α = Cronbach's alpha.