

# Measuring Child-Rearing Values. A Research Note

Bogdan Voicu<sup>1 2 3 \*</sup>

<sup>1</sup> CEPS/INSTEAD, 3 Avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg

<sup>2</sup> Romanian Academy, Research Institute for the Quality of Life, 13 Calea 13 Septembrie, 050711 București, Romania

<sup>3</sup> Lucian Blaga University of Sibiu, Department of Sociology and Social Work, 2-4 Lucian Blaga, 550169 Sibiu, Romania

---

## KEYWORDS

Child-rearing/parental values  
Value surveys  
Measurement issues  
Rating vs. ranking

## ABSTRACT

Large scale comparative studies, such as the value surveys (EVS and WVS) or the Eurobarometer, include measurements for parental/child-rearing values. This reflects a persistent interest for the topic, which produced salient studies starting with the first half of the twentieth century (Lynd and Lund 1929; Duvall 1946). Various scholars report data on parental values which use versions of the Q-sort methodology (Kohn 1977), ranking variables (Alwin 1990; Lenski 1961), scale indicators (Tulviste *et al.* 2007). Q-sort methodology remains the most widely employed. One of its versions is included in the value surveys as well. However, it fails to produce comparable indicators in different countries (Rabušić 2011; Xiao 2001) or at different moments in time (Wright and Wright 1976). This paper uses original data, provided by a Romanian convenience sample, to check if using various versions of the EVS/WVS items may lead to better ways to produce synthetic indicators for parental-values at individual level. SEM models are used to show that the best analytic solution would be to use individual items instead of producing summative indexes.

---

## Introduction

Defining social values with a short phrase become almost impossible due to the large variety of meanings, contexts and levels to which they may apply (Arts 2011; Jagodzinski 2004; van Deth and Scarbrough 1994). They may

---

\* Contact address: bogdan@iccv.ro (B. Voicu)

refer individual or societal level, can be defined as normative (patterns that should be followed, which are desirable for everybody) or in an evaluative way (what a specific individual or group of individuals usually think and do), can be defined through attitudes and behaviours that they trigger, or through their properties. I use the term 'to designate individual orientations towards desirable, which are not directly observable, imply cognitive, affective and evaluative elements, and manifest through attitudes and behaviours' (Voicu 2010b).

Parental values are defined by Kohn (1977: 18) as those characteristics that parents 'consider most desirable to inculcate in their children'. They are traits of the individuals, but are socially determined. They shape childrearing behaviours, children-parents relations, attitudes towards schooling, and children's values. Parental values is the term that was employed more often in the past, but the recent decades led to a smooth switch towards preferring the term child-rearing values, probably reflecting a tendency to centre the discourse rather around the child than on parents. No matter the labelling, most parental values scholars agree that they should include orientations towards authority and conformity, on one hand, and orientations towards independence and autonomy, on the other hand (Alwin 1988, 1990; Kohn 1977; Tufiş 2008; Xiao 2001).

European Values Study (EVS) and World Values Survey (WVS) include a set of items measuring child-rearing values. The two large-scale surveys are well-known and intensively used in social sciences. All together, they have the unique advantage to provide data for over 100 societies<sup>1</sup>. The measurement of parental values is quite simple. The interviewer presents a list of characteristics that one might consider important for children to learn, and the respondent may choose maximum five of them as being salient in his/her view. This is a simplified variant of the Kohn Q-sort methodology (Kohn 1977), which implied supplementary rankings among the items in the list. Unfortunately, if comparing different societies, it does not produce similar covariances among the items in the list (Rabušić 2011; Xiao 2000a, 2000b, 2001). This denies possibility to compute summative indicators.

This paper tries to see if using the same set of items in different ways

---

<sup>1</sup> Details about the two studies are available from their websites: [www.europeanvaluesstudy.eu](http://www.europeanvaluesstudy.eu) and [www.worldvaluessurvey.org](http://www.worldvaluessurvey.org).

leads to better results. I use data provided by a convenience sample of Romanian respondents, which answered to a short questionnaire in the summer of 2012. The purpose of the data collecting was to test the country-specific variables to be selected for the Romanian 2012 WVS wave. Among the items we have included the original WVS child-rearing values battery, one that involves partially ranking the qualities in the list, since only five out of eleven might be chosen. Two other versions of the respective set were added. In the first setting, respondents were asked to fully-rank all the eleven qualities that one may consider useful for children to learn. In the second one, the importance of each characteristic had to be evaluated on a 10-point scale. In the following, I test if any of the three solutions may lead to a reliable summative indicator. For this I check consistency at aggregate level, internal consistency in different analysis scenarios, and external validity in those scenarios that prove to be internally consistent.

Alwin and Krosnick (1985) proposed a similar analysis. They used a slightly different set of items, also tapping for parental values, and compared the outcome of ranking and rating scales. My goal is to go further by comparing three types of measurement, including partial-rankings to the two already analyzed by Alwin and Krosnick. I also consider confirmatory factor analysis instead of exploratory factor analysis, different models derived from the existing theory, and I apply this to the battery employed by the value surveys.

The paper begins with a short review of the literature on measuring parental values. It also considers the differences between rating and ranking variables. Then I introduce the methodology of the study. The findings section includes presenting some data from previous EVS and WVS waves to support the main argument. Then I focus on the results of the confirmatory factor analyses that I propose. In the final part I discuss the implications for further research.

## **Conceptual background**

In psychology, education science, and sociology of child and family child-rearing values are important factors to determine child development (Alwin 1986; Hitlin 2006; Kohn 1977; Tudge *et al.* 2000; Tulviste *et al.* 2007; Xiao

2001). Social values scholars (Inglehart 1997; Hageaars, Halman and Moors 2003; Tufiş 2008) treat the parental values or child-rearing values as part of a mix of value orientations which contrast two sets of preferences regarding what children should learn.

The classic measurement is the Kohn's Q-sort methodology: a list of qualities, typically 13, is presented to the respondent. They are said to be attributes that children may learn. The respondent is asked to choose the ones that she or he considers to be the most important three, respectively the least important three (Kohn 1977; Kohn and Slomczyński 1990; Tudge *et al.* 2000). Then, he/she is required to indicate the most important and the least important quality. This leads to a partial-ranking of the list of qualities. Some of the items in the list are clearly ranked as being the best of the least preferable options. Others are indicated of being preferable or not, but their rank is tied. Finally, for half of qualities in the list the only information provided is the fact that the respective trait is not present among the top 3 or the bottom 3 likeable characteristics. The work of Kohn was initially published in the 1960s and, in the line with previous works of Lynd and Lynd (1929) or Lenski (1961), it employed data collected from parents who were asked about their children or kids similar to their children. Latter studies showed that the same methodology can be applied to any adult and without the lead-in, asking about children in general (Wright and Wright 1976; Xiao 2001).

**Table 1.** *The WVS 2012 measurement of child-rearing values (partial ranking)*

**Here is a list of qualities that children can be encouraged to learn at home. Which, if any, do you consider to be especially important? Please choose up to five!**

		Mentioned	Not mentioned
V12.	<u>Independence</u>	1	0
V13.	<b><i>Hard work</i></b>	1	0
V14.	<u>Feeling of responsibility</u>	1	0
V15.	<u>Imagination</u>	1	0
V16.	Tolerance and respect for other people	1	0
V17.	<b><i>Thrift, saving money and things</i></b>	1	0
V18.	<u>Determination, perseverance</u>	1	0
V19.	<b><i>Religious faith</i></b>	1	0
V20.	Unselfishness*	1	0
V21.	<b><i>Obedience</i></b>	1	0
V22.	Self-expression	1	0

Table 1 depicts the variant of the Q-sort methodology that is employed by the European Values Study, the World Values Survey, and some Eurobarometers (EB28, 28.1, 37.2, 39.0, 44.0, 44.1, 63.1). The respondent is required to choose 5 out of 11 (sometimes 10 or 12) qualities that people may consider important to be learnt at home. In the following, I label this technique as *partial rankings*. Alternative measures include *rating* each attribute, or (fully) *ranking* them from 1 to 11. The rating approach is common to the Parental Beliefs Survey (Luster 1985, cited in Tudge et al., 2000), which includes a long list of 59 qualities to be rated by respondents. The fully ranking approach is typically employed for shorter lists of qualities (see Alwin 1990 for an example).

Some scholars use the items to describe a continuum opposing orientations towards authority and autonomy as polar dimensions (Kohn and Slomczyński 1990; Hagenaars, Halman and Moors 2003). Others opt for a multi-dimensional approach, usually describing a bi-dimensional space, shaped by two factors: authority or conformism, respectively autonomy or self-control (Alwin 1986; Luster, Rhoades and Haas 1989; Tulviste *et al.* 2007; Tufiş 2008; Xiao 2001). Another approach is to use the items as separate indicators, without any attempt to find latent concepts that explain their variations (Alwin 1996; Tudge *et al.* 2000).

The single-factor approach has the advantage to provide an easy to interpret measure, which may be used either as independent variable, or, in longitudinal perspective, to assess social change. Hagenaars, Halman and Moors (2003) employ the EVS/WVS 1999 battery (similar to the ones in Table 1, but not including ‘self-expression’ among the qualities) to compute a summative score of ‘preference of authority above autonomy’. There are four qualities tapping for authority: ‘Hard Work’, ‘Thrift, saving money and things’, ‘Religious Faith’, ‘Obedience’. For each respondent, the score adds as many points as many of these four qualities are mentioned. Then, for each of four qualities designing preference for autonomy (‘Independence’, ‘Feeling of Responsibility’, ‘Imagination’, ‘Determination, Perseverance’) which is mentioned, a point is subtracted from the score. The different colouring in Table 1 indicates the autonomy-authority opposition (underlined fonts vs. bolded-italic ones). There is an explicit idea that the four items are manifestations of a latent authority-autonomy value

orientation.

The multi-dimensional approach is typically used to provide a factor for self-control and another for authority or conformity<sup>2</sup>, following the classic classification proposed by Kohn (1969). Longer scales may produce more dimensions as well. For instance, Tulviste *et al.* (2007) use a list consisting of 20 qualities and identify a third factor, along with conformity and self-maximization ('development of one's self-potential and individuality, like self-confidence, independence, curiosity, and autonomy'). It taps for orientations towards power, and manifest through valuing qualities like being a leader, belief in own abilities, being a respected person, and being an influential person. Lin and Fu (1990) used a list of 28 qualities, each being evaluated by respondents on a 5-point scale. They extracted four factors, explaining parental control, independence, affection, and emphasis on achievement.

Replicating results in other context or at different moments in time proved to be difficult. Wright and Wright (1976) employed data from the 1973 General Social Survey to repeat Kohn's findings and to assess change. Using exploratory factor analysis, in the same manner as Kohn did, they failed to obtain identical or very close loadings. This might have been caused by considering all adults instead of parents, and renouncing to the lead-in. However, since the general sense of the findings was not affected, it may also indicate the instability of the initial solution.

Recent testing for validity of the single-factorial structure shows out that the eight items cannot be properly explained by a single latent variable (Rabušić 2011). Even when employing a two factors model, Tufiş (2008) found out that urban and rural Romania differ with respect to the way in which seven out of the ten child-rearing items cluster together (she considered independence, thrift, determination, obedience, hard work, imagination, responsibility). Exploratory factor analysis led Xiao (2000a, 2000b, 2001) to propose slightly different indicators for conformism in China and the US, although using the same WVS/EVS 1999-2000 data.

From here the idea to look for a better measurement, that may produce synthetic indexes closer to the conceptual expectations. Comparing partial

---

<sup>2</sup> The label is probably more meaningful in those collectivities/societies where orientation towards authority is the norm to conform to.

ranking, to rating and to fully ranking scales is the method for which I have opted, search for either a single-factor, either two factors to describe the space of parental values.

The seminal paper by Alwin and Krosnick (1985) reviews the advantages of ranking and rating scales. The first are said to be more precise, but sometimes may become too complicated to obtain reliable answers. In contrast, rating scales are easier to understand and to present to the respondent, also being less time consuming. Response styles or sets are also likely to be influential when using ratings. Alwin and Krosnick (1985) present data that supports the idea that, at aggregate level, ranking and rating scales produce similar hierarchies. Working on very small samples of Finnish, Swedish and Estonian mothers, and using a (different) 20-items long list of child-rearing qualities, Tulviste *et al.* (2007) include both ratings on a 4-point scale, and partial-rankings: the respondents were asked to order the most important 3 qualities. The authors present data showing no substantial difference when considering the covariances between the partial-ranked and the rated qualities, but the aggregate level hierarchies are different.

All these previous analyses led to three important questions to be answered, particularly when considering the child-rearing battery in the EVS/WVS questionnaire. First, one may search for an appropriate method to derive summative indicators which are useful in comparative research. This would mean to find an appropriate factorial structure that would remain invariant over nations and in time. Second, there is a related question, somehow implied by the first one: how many and which items from the battery should be used for computing the summative indicators. Third, if no positive solutions derive from the first two challenges, one might wonder if there is another way to produce the battery, using roughly the same items, which would lead to better results.

This paper tries to answer the three above objectives. For the first one, existing data from EVS and WVS may easily be employed. For selecting the items to be considered, I drawn on previous literature, particularly on the works of Hagenaars, Halman and Moors (2003), Tufiş (2008), Voicu (2010a), and Xiao (2000a, 2000b, 2001). For the third challenge, a specific research was designed, as I will explain in the following section. It aimed to see if using

the same items, but differently assessing them would lead to more effective information.

## **Data and Method**

Two types of data are used in this paper. Firstly, I consider EVS 2008 and WVS 2005 to depict variation of factorial structures across various European and non-European countries. The two studies are extensively described on their already mentioned websites. They provide data drawn from probabilistic representative samples in more than 100 societies.

Second, original data allows testing for the usefulness of various measurements for parental values. The interviews were carried out in Romania, by Metro Media Transilvania pollster, using a convenience sample. The aim was to test a 15-minutes questionnaire including items proposed for the country-specific part of the WVS 2012 Romanian study. 40 respondents were located in rural areas, 98 are women, and the average age is 45.22, with a maximum of 87 and a minimum of 17. A quarter of the respondents are under 28, and another quarter is older than 61. 76 graduated tertiary education (among which 61 graduated university), and 29 do not attended more than lower secondary.

Partial rankings, ratings (1-10 scales) and full rankings were designed. Partial rankings are the classic WVS/EVS measurement. Full rankings involved ordering all the 11 items depending on their perceived importance. The respondents were not allowed to provide tied-ranks. Finally, rating was done on a 10 point scale, each respondent indicating how important he/she considers each quality on the list.

Testing was done in three ways. First I have checked if any of the scales produces more missing data than others. Second, I was observing the aggregate level consistency: the question here is if the three measurements produce similar hierarchies of the items. Third, there was a check of the internal consistency. I have employed using confirmatory factor analysis/structural equations modelling (CFA/SEM) to test for internal consistency of various designs. I started with the single-factor solution using the reduced 4-item autonomy vs. authority (conformity) scale. Then I have considered the single-factor solution using the same 8 items as Hageenaars et

al (2003). Finally, I tested two-factor solutions, similar to the ones proposed by Xiao (2000a, 2000b, 2001), Tufiş (2007), and Tulaviste *et al.* (2007), but not necessary using the same qualities.

When testing for internal consistency, maximum likelihood estimates were used, except for the partial ratings model, in which the interest variables are dichotomous: 1 = the respective quality was mentioned by the subject, 0 = not mentioned. For the models based on the dichotomous variables WLSMV was employed. All models were run in MPlus6.

### **Cross-country lack of invariance of the one-factor models based on partial ratings**

Before presenting the Romanian results, it is useful to see some data resulting from previous WVS and EVS waves. The countries included in the presentations of results that follows were chosen such as to cover the variety of welfare and child-care regimes across Europe, to add comparison from outside the continent, and to reflect data collecting difficulties for the child-rearing battery. In fact, in some of the countries the maximum five choices limitation was not followed. The percentage of those who indicated more than 5 important qualities peaks 46% in the 2008 Romanian wave. However, Romania is not an exception. In the same wave other countries also had respondents to mention more choices. They were 12% in Luxembourg, 8% in Great Britain, 18% in Sweden, 28% in Hungary, etc. In the following analyses I have included surveys that followed the limitation as well as countries in which the constraint was not respected.

I have started with testing for invariance of the authority vs. autonomy one factor model, including only four items as indicators, as indicated in Voicu (2010a). According to theory, 'independence' and 'responsibility' should have been positively correlated to the factor, while 'religious faith' and 'obedience' should negatively covariate to the latent explanatory variable. Using the WVS 2005 and the EVS 2008 datasets, I have reached puzzling results, some of them being depicted in Table 2. The expected pattern is found in some of the Western European countries, including Western Germany and Portugal, but also in former communist societies, like Poland. In societies like Romania and Hungary, the factorial structure looks

similar but ‘responsibility’ is not significantly related to the factor. There are also countries, particularly USA, Belarus and Turkey, where the expected patterns are strongly violated.

The same happens when using the extended version of the index, which follows Hageñaars, Halman and Moors (2003). The design includes responsibility, independence, determination, and imagination as positive covariates for the latent orientation towards autonomy, while obedience, hard work, thrift and religiosity would stand as negative covariates. I found no society in which such model would fit the data. In most countries, at least one or two indicators had covariates with the factor which are contrary to theoretical expectations.

**Table 2.** *Estimates for the standardized regression weights (factor loadings) for the one-factor model with four indicators in various societies*

	Responsibility	Independence*	Religious Faith	Obedience
West Germany 2008	<b>0.05</b>	2.21	<b>-0.06</b>	<b>-0.13</b>
Portugal 2008	<b>0.02</b>	2.03	<b>-0.11</b>	<b>-0.12</b>
The Netherlands 2008	<u>-0.02</u>	2.01	<b>-0.10</b>	<b>-0.16</b>
Great Britain 2008	<u>-0.01</u>	2.00	<b>-0.04</b>	<b>-0.07</b>
Sweden 2008	0.01	2.10	-0.02	<b>-0.11</b>
Luxembourg 2008	0.00	2.00	0.00	-0.01
Greece 2008	<u>-0.00</u>	2.03	<b>-0.09</b>	<b>-0.10</b>
Poland 2008	<b>0.02</b>	2.03	<b>-0.11</b>	<b>-0.12</b>
Romania 2008	0.01	2.21	<b>-0.11</b>	<b>-0.07</b>
Belarus 2008	<u>-0.03</u>	2.30	<b>-0.07</b>	<b>-0.12</b>
Hungary 2008	<b>0.05</b>	2.35	-0.02	-0.01
Turkey 2008	<b>1.00</b>	0.76	<u>1.00</u>	<u>1.00</u>
Canada 2005	<b>-0.68</b>	0.88	0.02	-0.02
USA 2005	<u>-0.03</u>	2.01	<b>-0.11</b>	<b>-0.08</b>
Brazil 2005	<b>0.13</b>	0.82	0.02	-0.01
China 2005	<b>0.11</b>	2.15	-0.01	<b>-0.05</b>

\*the parameter was fixed to 1.

Notes: Bolded figures indicate regression weights which are significant at  $p \leq 0.05$ . Underlined figures indicate relations that have signs which are reversed as compared to the expected ones.

Another model to test was the one including the eight items, but defining two factors which negatively covariate. In this scenario only a few countries, including Sweden, Germany, China, Luxembourg, and Hungary, displayed the expected relations, but even in their case, not all regression weights were significant. More, in many other societies, at least one indicator was connected to the factors contrary to expectations. Most of the time this was the case of 'religious faith' or/and 'responsibility'. When testing models which included two factors and seven or six items, as suggested by the works of Tufiş (2008) or Xiao (2001), the results were also unsatisfying.

In all the above cases, I have not performed tests of invariance, since it was obvious from inspecting the factor loadings that none of the tested models kept unchanged factor structure in all, or at least in a majority of the considered countries. This confirms that one may need to look for alternative measurement options in order to enable computing summative indexes that could be used in comparative analyses.

## **The Romanian 2012 data**

### *Missing answers*

Expectations related to missingness were mainly derived from Alwin and Kosnick (1985) paper. Ranking scales, despite being more precise, were said to be more complicated. Therefore one may expect a higher volume of missing information as compared to ratings.

In the case of the Romanian 2012 dataset that I employ in this paper, three types of measurement were employed. I have labelled them as partial-rankings, full-rankings, and ratings. Applied for the 200 respondents, they did not produce different percentages of missing information. In the case of partial rankings, there are only two respondents who refused answering. They count for 1% out of the total sample. In the case of ratings, the missing answers reported in Table 2 include both refuse to response (NA) and indecision (DK – don't know). The share of cases missing information raise as high as 5% for altruism, but in most cases is under 2%. The same happens with the full rankings. Overall, the partial rankings seem to produce the

fewest cases with missing data, but the difference to the other two methods is negligible.

**Table 3.** *Percentages of missing answers for the three types of measurement*

		Partial rankings	Ratings	Full rankings
v12.	<b>Independence</b>	1% (2 cases out of 200)	1.5%	0.5%
v13.	<b>Hard work</b>		0.5%	0.5%
v14.	<b>Responsibility</b>		0.5%	0.5%
v15.	<b>Imagination</b>		1.5%	0.5%
v16.	<b>Tolerance and respect for others</b>		0.0%	1.0%
v17.	<b>Thrift</b>		0.5%	0.5%
v18.	<b>Perseverance</b>		0.0%	1.5%
v19.	<b>Religious faith</b>		2.0%	0.5%
v20.	<b>Altruism</b>		5.0%	2.0%
v21.	<b>Obedience</b>		2.0%	1.0%
v22.	<b>Self-expressing</b>		0.5%	2.0%

For all measurements, the respondents that provided missing answers were slightly better educated and younger than average, while no relation was seeable with gender. The sample was also better educated as compared to Romanian population. If drawing a probabilistic sample, I would expect that missingness would be not an issue.

***Hierarchies at the aggregate level***

Considering the average results for the whole sample, there is a strong correlation between the hierarchies resulting from the three types of measurement. When partial rankings used, responsibility, hard work and tolerance were the most often mentioned. These are the only qualities to be considered important by the majority (Table 4). When using ratings, these three items also received the higher scores: 9.0 for hard work, 8.9 for responsibility, 8.6 for tolerance. Full rankings provide a similar hierarchy. The three items are the ones to be mentioned most of the time among the priorities, with responsibility receiving an average rank of 3.8, hard work – 4.0, and tolerance 5.3.

At the bottom of the hierarchy, altruism and obedience are the last, no matter the measure. Self-expressing is the third less preferred one in the

partial and full rankings. When considering ratings, it comes on the fourth place, but this may be due to an acquiescence effect: self-expressing is the quality from the list that has the least familiar meaning.

**Table 4.** *Aggregate scores*

		<u>Partial rankings:</u> % mentioning the item as important	<u>Ratings:</u> average ratings on the 1-10 scale	<u>Full rankings:</u> average rank (1=highest, 11=lowest)
v12.	<b>Independence</b>	37%	7.8	5.6
v13.	<b>Hard work</b>	77%	9.0	4.0
v14.	<b>Responsibility</b>	79%	8.9	3.8
v15.	<b>Imagination</b>	35%	7.8	6.4
v16.	<b>Tolerance and respect for others</b>	72%	8.6	5.3
v17.	<b>Thrift</b>	43%	7.9	6.0
v18.	<b>Perseverance</b>	34%	8.3	6.2
v19.	<b>Religious faith</b>	49%	8.0	6.3
v20.	<b>Altruism</b>	13%	7.5	7.1
v21.	<b>Obedience</b>	9%	6.4	8.2
v22.	<b>Self-expressing</b>	33%	8.3	7.0

The rating strategy produced distributions that are skewed towards positive values (Table 4). This is consistent with Alwin's and Krosnick's (1985) findings and may reflect a tendency to positively answer to opinion questions in surveys (Jacoby 2011). In order to control the eventual effects of such response sets, I have computed a rough acquiescence score. For each respondent I counted the number of items measuring attitudes (not behaviours or evaluations) at which the answer was positive. 'Positive' meant 8, 9 or 10 on the 1-10 scales (how justified is 'To migrate for a while in another country', and 'To migrate for good in another country', as well as the rating child-rearing items), 3 or 4 on the 4-point scales coded in ascendant order (agreement with 'People in Romania should stay in their country', 'Everyone is due to contribute to the wealth of the country they were born in', 'Romania loses through the migration of its citizens to other countries', 'The ones that left abroad contribute to Romania's development', 'The families of the ones that left are suffering', 'Children suffer if their mother works', and 'Preschool children suffer if their mother works'), respectively 1 or 2 on the 5-point scales coded in reversed order (how worried the respondent was with the living conditions of specific groups:

family; neighbours; people in own region; people in Romania; people in Europe; people in the entire world; aged people in Romania; unemployed in Romania; immigrants from Romania; severely ill or disabled in Romania).

Then I have regressed each rating item on the acquiescence score, and I have ordered the average residuals (column 3 in Table 4). It turns out that the new hierarchy differs much from the ones based on rankings. Five out eleven items (Hard Work, Religious Faith, Altruism, Obedience, and Self-Expressing) substantially change their positions as compared to the initial order based on the ratings. The new hierarchy is quite far from the one derived from partial or full rankings. On its turns, the rating-based hierarchy which ignored the potential acquiescence effects is closer to the hierarchies derived from the ranking variables. This makes it more likely to be reliable. Since the results of controlling for acquiescence are inconclusive, in the subsequent analysis I choose to ignore the response set hypothesis.

**Table 5.** *Aggregate scores: the resulting hierarchies*

		Partial rankings	Ratings	Ratings: residuals (after controlling for acquiescence)	Full rankings
v12.	<b>Independence</b>	6	8	8	4
v13.	<b>Hard work</b>	2	1	9	2
v14.	<b>Responsibility</b>	1	2	2	1
v15.	<b>Imagination</b>	7	8	10	8
v16.	<b>Tolerance and respect for others</b>	3	3	4	3
v17.	<b>Thrift</b>	5	7	3	5
v18.	<b>Perseverance</b>	8	4	5	6
v19.	<b>Religious faith</b>	4	6	1	7
v20.	<b>Altruism</b>	10	10	7	10
v21.	<b>Obedience</b>	11	11	6	11
v22.	<b>Self-expressing</b>	9	4	11	9

In Table 5, with the mentioned exception of self-expressing, all the other items maintain almost unchanged their position in the hierarchies. The Pearson-correlations between the three resulting series are quite high: partial rankings vs. ratings – 0.87; partial rankings vs. full rankings – 0.93; ratings vs. full rankings – 0.86. This suggests that, if interested in working with aggregate level values, the three measures produces almost the same type of results, and can be used with approximately the same reliability.

*Individual level reliability: relations between measures for the same quality*

At individual level, the correlations between the three measurements of each concept are apparently weak (Table 6). However, a CFA analysis for each quality reveals excellent goodness of fit indexes for models explaining each of the three measurements through a common latent variable. This means that partial rankings, full rankings and ratings for each of the v12-v22 variables are closely interrelated and measure the same concepts.

**Table 6.** *Spearman correlations between measurements*

		Partial rankings - ratings	Partial rankings - Full rankings	Ratings - Full rankings
v12.	<b>Independence</b>	0.282	-0.242	-0.194
v13.	<b>Hard work</b>	0.291	-0.145	-0.249
v14.	<b>Responsibility</b>	0.053	-0.073	-0.242
v15.	<b>Imagination</b>	0.222	-0.248	-0.189
v16.	<b>Tolerance and respect for others</b>	0.325	-0.172	-0.185
v17.	<b>Thrift</b>	0.329	-0.256	-0.166
v18.	<b>Perseverance</b>	0.097	-0.180	-0.196
v19.	<b>Religious faith</b>	0.374	-0.369	-0.355
v20.	<b>Altruism</b>	0.315	-0.301	-0.202
v21.	<b>Obedience</b>	0.142	-0.186	-0.197
v22.	<b>Self-expressing</b>	0.112	-0.191	-0.193

*Internal consistency: the single factor solution, the four-item version*

In the existing literature there are at least two versions to produce the single factor model (Hagenaars, Halman and Moors 2003; Voicu 2010a). Using only four items (Voicu 2010a) is the first that I have tested, considering only religiosity, obedience, responsibility and independence. I have designed a simple model, in which a latent variable (AUTONOMY) explained the variance of the four indicators (Figure 1). The model was fitted using Mplus 6.12, first for the partial ranking scale, than for the ratings, and finally for the full rankings model.

**Figure 1.** *The one-factor model with four items considered*

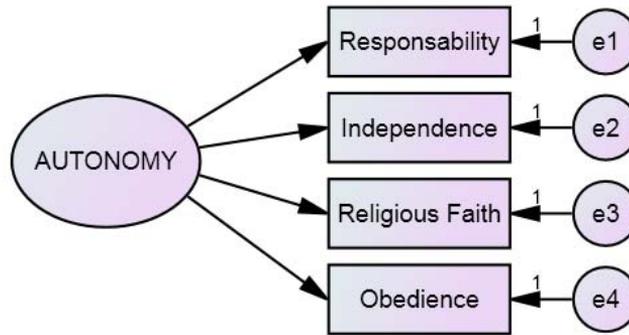


Table 7 displays results for the three models. None of them properly fits the data. When considering the sign of the indicators’ weights, the partial rankings and the full rankings measurements are consistent with the theory. This does not hold true when using ratings for measurement. Even when controlling for acquiescence effects, Religious Faith and Obedience positively load on the factor described by independence and responsibility.

**Table 7.** *Internal consistency of the three competing simplified one-factor measurement models*

		Partial rankings	Ratings	Full rankings
Fit indexes				
	<b>RMSEA</b>	0.112	0.189	0.253
	<b>CFI</b>	0.896	0.850	0.776
	<b>TLI</b>	0.687	0.551	0.328
	<b>WRMR/SRMR</b>	0.689	0.013	0.073
Unstandardized weights (standard errors in brackets)				
	<b>Independence</b>	1 (0)	1 (0)	1 (0)
	<b>Responsibility</b>	0,425 (0,205)	0.932 (0.193)	1.302 (0.329)
	<b>Religious faith</b>	-0,862 (0,324)	0.918 (0.209)	-0.890 (0.291)
	<b>Obedience</b>	-0,863 (0,306)	0.509 (0.196)	-1.860 (0.487)
Standardized weights				
	<b>Independence</b>	0.742	0.605	0.382
	<b>Responsibility</b>	0.315	0.785	0.583
	<b>Religious faith</b>	-0.640	0.471	-0.340
	<b>Obedience</b>	-0.641	0.246	-0.836
Number of cases				
	<b>N</b>	198	190	189

In the partial ranking model, responsibility is poorly predicted ( $R^2 \approx 5\%$ ). In the ratings model, Responsibility is almost unexplained by the latent variable ( $R^2 = 4\%$ ), and Independence is also weakly connected to the other variables ( $R^2 = 16\%$ ). In the full ranking model, Independence and Religious Faith have low R squares (12%, respectively, 15%). For all other indicators, in all models, the explained variance exceeds 40%.

***Internal consistency: the single factor solution, the eight-item version***

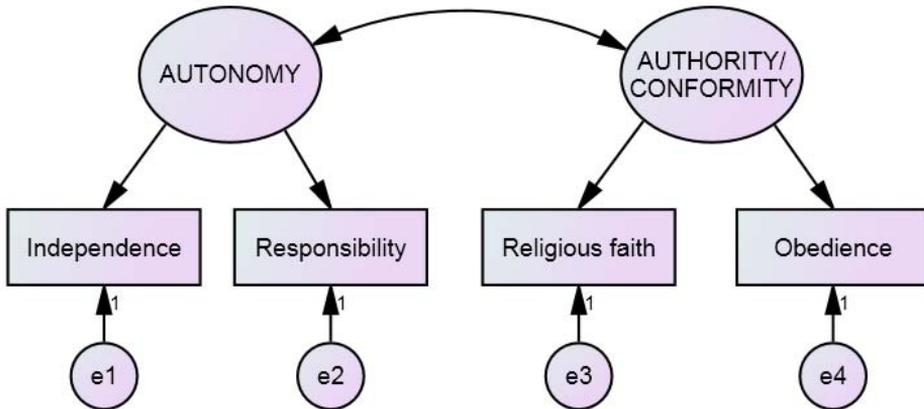
Using the eight items employed by Hageñaars, Halman and Moors (2003) is the second attempt to fit a one-dimensional model. For the partial ranking model, I had to drop perseverance in order to achieve convergence. The fit is not acceptable when considering TLI or CFI, but RMSEA has a better value (0.07). However, the Hard Work act opposed to theory, being positively associated with the latent factor, joining Independence, Imagination and Responsibility, while Religious Faith, Thrift, and Obedience have negative weights. In the model based on ratings, the fit is poor, and all observed variables positively covariate to the factor. The full rankings model poorly fits the data, and has the same problem with Hard Work which covariates with the factor contrary to the theoretical assumptions.

If excluding Hard Work from these models, all relations go in the expected way, but the goodness of fit indexes continue to indicate significant differences as compared to the data.

***Internal consistency: the two-factor solutions***

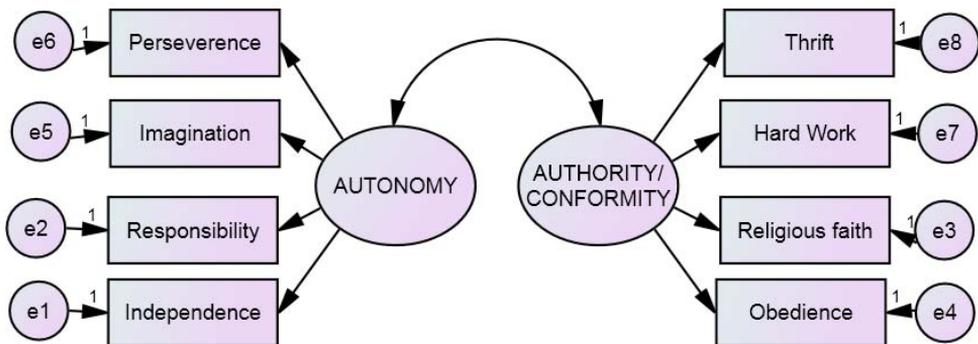
The reduced version of the two-factor solution is described in Figure 2. Independence and Responsibility are supposed to be explained by the latent orientation towards autonomy and self-development, while Religious Faith and Obedience are determined by the latent orientation towards authority and conformity patterns. The model is designed in the same way no matter the measurement.

**Figure 2.** *The two-factor model with four items*



The results are not encouraging. In the case of both partial ranking and full ranking models the two factors are negatively related and positively correlate with their indicators, as expected, but the goodness of fit indexes point out that the model is far from fitting the data:  $RMSEA > 0.1$ ,  $TLI < 0.9$ , even if  $CFI > 0.95$  and  $SMSR = 0.033$  in the case of the full rankings measurement. In the case of the ratings measurement, the fit looks good, but the two factors are positively related.

**Figure 3.** *The two-factor model with eight items*



The same happens with the two-factor solution depicted by Figure 3. In the ratings-based model, the two factors positively covariates, while the fit is not acceptable. The full rankings model does not converge unless perseverance is removed from the model, and even so, hard work is not connected to the authority factor as expected. The partial rankings

measurement leads to a model in which the relations between variables have the expected direction. However, the explained variance for the observed indicators is poor, exceeding 10% only for Obedience and Religious Faith. The goodness of fit indexes for the model based on partial rankings are almost acceptable: RMSEA=0.055; CFI=0.891; TLI=0.839; WRMR=0.916. Adding a supplementary covariance between error terms would increase the fit, but the explained variance for each indicator remains very low.

**Table 8.** *Internal consistency of the three competing measurement models applied to the modified Xiao US model*

		Partial rankings	Ratings	Full rankings
Fit indexes				
	<b>RMSEA</b>	0.050	0.106	NO CONVERGENCE
	<b>CFI</b>	0.959	0.924	
	<b>TLI</b>	0.923	0.858	
	<b>WRMR/SRMR</b>	0.738	0.051	
Explained variance (R <sup>2</sup> )				
	<b>Independence</b>	0.216	0.458	NO CONVERGENCE
	<b>Responsibility</b>	0.035	0.517	
	<b>Imagination</b>	0.158	0.464	
	<b>Thrift</b>	0.093	0.541	
	<b>Religious faith</b>	0.270	0.161	
	<b>Obedience</b>	0.181	0.163	
Covariance between the latent variables				
	<b>standardized</b>	-2.357	0.813	
Number of cases				
	<b>N</b>	198	190	189

Dropping perseverance as I did in the case of the full rankings two-factors and seven items model is consistent with Tufiş (2008). I have repeated the exercise for the case of partial rankings, respectively for the measurement through ratings. For both type of measurement, all goodness of fit indexes become acceptable. The partial rankings measurement, however, provide lower than 5% explained variances for Hard Working and Responsibility. On the other hand, the ratings-based model, although explains at least 15% of each observed variable, preserves the positive association between the two factors, finding which contradicts theoretical

expectations.

Another option would be to reproduce Xiao's models (2000a, 2000b). In both papers, she includes Independence, Perseverance, and Imagination as indicators for the autonomy dimension. Obedience, 'Good manners' and Religious Faith are considered as indicators for conformity in the US, while Thrift and Hard Work join Obedience in China. Since 'good manners' is not included in the questionnaire that I use, I tested only for the 'Chinese' version of the conformity factor. Then I have tested a modified version of the US version, in which I have replaced 'good manners' with Thrift.

In the case of Xiao's (2000b) China version, the partial rankings model does not converge to solution, the ratings one provides the same positive correlation between autonomy and authority/conformity, while for the full rankings the goodness of indexes indicate important differences between the theoretical model and the empirical data. Table 8 describes the results for the modified Xiao's US model. Full rankings and ratings do not produce acceptable models. In the case of partial ratings, the model fits the data, but it fails to explain much of the observed variables. However, this is the best one-dimension or two-dimension model that I managed to fit given the data.

### *Comparing models? External validity*

Table 9 summarizes the findings described in the previous sections. Until now, the analyses show that, at least for the case of this convenience sample, partial rankings provide better measurement, while the two-dimension modelling better fits the data. However, even these two-factor models do not explain much of the variance of the observed variables.

I use these two models for testing the external validity. The dataset is not very rich in independent variables, but it provides information on age, education, and perceived personal situation (better or worse as compared to the past – 5-point scale). The three of them should be positively correlated to the autonomy latent orientation and should negatively determine the authority/conformism factor (Tufiş 2008; Xiao 2001). I have also modelled supplementary impacts of age and education on the perceived current situation. In both models, the relations were the expected ones, further indicated that the partial ranking solution remains the most effective one.

**Table 9.** *A summary of the findings regarding internal consistency*

Measurement Model	Partial rankings	Ratings	Full rankings
4-item, one dimension	Fit: <del>bad</del> , relations: <b>good</b>	Fit: <del>bad</del> , relations: <del>bad</del>	Fit: <del>bad</del> , relations: <b>good</b>
8-item, one dimension	Fit*: <del>bad</del> , relations: <del>bad</del>	Fit: <del>bad</del> , relations: <del>bad</del>	Fit: <del>bad</del> , relations: <del>bad</del>
4-item, two dimensions	Fit: <del>bad</del> , relations: <b>good</b>	Fit: <b>good</b> , relations: <del>bad</del>	Fit: <del>bad</del> , relations: <b>good</b>
8-item, two dimensions	Fit: <del>bad</del> , relations: <b>good</b>	Fit: <del>bad</del> , relations: <del>bad</del>	Fit*: <del>bad</del> , relations: <del>bad</del>
7-Item, 2 dimensions (Tufiş)	Fit: <b>good</b> , relations: <b>good</b>	Fit: <b>good</b> , relations: <del>bad</del>	
6-item, 2 dimensions (Xiao-China)	No convergence.	Fit: <b>good</b> , relations: <del>bad</del>	Fit: <del>bad</del> , relations: <b>good</b>
6-item, 2 dimensions (Xiao-modified US)	Fit: <b>good</b> , relations: <b>good</b>	Fit: <b>good</b> , relations: <del>bad</del>	No convergence.

\*After dropping perseverance. Otherwise, there is no convergence.

## Discussion and implications

In the 1980s, the Alwin's and Krosnick's (1985) study on measuring parental values shed some light on the difficulties to produce a stable latent structure that would explain adults' preferences with regard to child rearing. A single factor solution was clearly rejected at the time, while rating and ranking scales were shown to generate different covariances among items. The findings of this paper extend their conclusions. I showed that neither partial-rankings, nor full-rankings, nor ratings generate single factor solutions, even if, unlike Alwin and Krosnick, I have worked with a selected number of valuable qualities. In this way I followed what actually the theory implied, and I have also overcome the difficulty given by rankings, which become linear combinations when all items are employed (van Leeuwen and Mandabach 2002). For the case of Romania, two factor solutions converged and fitted the data when the designed structure followed Tufiş's (2008) or Xiao's (2000a, 2001) models. However, I have also showed that there is no way to obtain measurement invariance when considering more culture. This cannot be achieved with the tested models even considering more homogenous regions, such as Europe, not to mention world-wide analyses. The finding is consistent with Rabušić (2011) analyses on partial-ranking

measures provided by EVS 2008, which this paper extends to sets of items based on full-rankings or on ratings.

The implication is not discouraging. Even if summative indexes are not legitimate to use, single items still produce dummy variables that may be successfully used in comparative analyses. Also, the partial ranking solution that is employed in surveys of values such as EVS and WVS is as reliable as using ratings or full-rankings. The list is short enough to avoid inducing more missing answers than ratings, is less complicated than full-rankings, and generates variables that are not as skewed as the ones resulting from ratings.

However, it remains the challenge of not following the limitation to choose maximum 5 items from the list. This is not affecting the structural equation models that I have previously produced when comparing different countries, since these models were based on covariance matrices and I was mostly interested in the relations among variables not in their levels or in the intercepts. If working with individual items, such as, for instance, 'Religious Faith' or 'Independence' things would be different. An individual who is allowed to choose more than 5 items would have a higher probability to choose any particular quality as compared to the situation when the choice is limited to five. If the variable would be employed as dependent or independent in a cross-national model, than one should carefully control for the effects of not respecting the limitation, otherwise the comparison risks to be futile. The same might be valid when slightly changing the list of qualities.

### **Acknowledgements**

Working on this paper was supported by CNCS grant PN-II-ID-PCE-2011-3-0210.

### **References:**

- Alwin, Duane. 1996. From Childbearing to Childrearing: The Link Between Declines in Fertility and Changes in the Socialization of Children. *Population and Development Review* 22: 176-196.
- Alwin, Duane. 1990. Cohort replacement and changes in qualities valued in children. *Journal of Marriage and the Family* 52 (2): 347-360.
- Alwin, Duane. 1988. From obedience to autonomy: Changes in traits desired

- in children. *Public Opinion Quarterly* 52 (1): 33-52.
- Alwin, Duane F. 1986. Religion and parental child-rearing orientations: Evidence of a Catholic-Protestant convergence. *American Journal of Sociology* 92 (2): 412-440.
- Alwin, Duane F. and Jon A. Krosnick. 1985. The Measurement of values in Surveys: A Comparison of Ratings and Rankings. *Public Opinion Quarterly* 49 (4): 535-552.
- Arts, Wil. 2011. Explaining European Value Patterns: Problems and Solutions. *Studia UBB Sociologia* LVI (1): 7-31.
- Duvall, Evelyn Millis. 1946. Conceptions of Parenthood. *American Journal of Sociology* 52 (3): 193-203.
- Hagenaars, Jacques, Loek Halman, and Guy Moors. 2003. Exploring Europe's basic values map. In *The Cultural Diversity of European Unity. Findings, Explanations and Reflections from the European Values Study*, eds. Wil Arts, Jacques Hagenaars, Loek Halman, Wim Van De Donk and Ton Van Schaik, 23-66. Leiden: Brill.
- Hitlin, Steven. 2006. Parental Influences on Children's Values and Aspirations: Bridging Two Theories of Social Class and Socialization. *Sociological Perspectives* 49 (1): 25-46.
- Inglehart, Ronald. 1997. *Modernization and Post-Modernization. Cultural, Economic and Political Change in 43 Societies*. Princeton: Princeton University Press.
- Jacoby, William G. 2011. Measuring Value Choices: Are Rank Orders Valid Indicators? Paper presented at the Annual Meeting of the Midwest Political Science Association, March 31, 2011, Chicago, IL, USA.
- Jagodzinski, Wolfgang. 2004. Methodological problems of value research, in 2004. In *Comparing Cultures: Dimensions of Culture in a Comparative Perspective*, eds. Henk Vinken, Joseph Soeters, and Peter Ester, 97-121. Leiden: Brill.
- Kohn, Melvin L. 1977. *Class and Conformity: A Study in Values*. Chicago, IL: University of Chicago Press.
- Kohn, Melvin L. and Kazimierz M. Slomczyński. 1990. *Social Structure and Self-direction: A Comparative Analysis of the United States and Poland*. Oxford: Basil Blackwell.
- Lenski, Gerhart Emmanuel. 1961. *The religious factor: a sociological study of religion's impact on politics, economics, and family life*. Garden City, N.Y.: Doubleday.
- Lin, Chin-Yau Cindy and Victoria R. Fu. 1990. A comparison of child-rearing practices among Chinese, immigrant Chinese, and Caucasian-American parents. *Child Development* 61 (2): 429-433.
- Luster, Tom. 1985. Parental Beliefs Survey. PhD thesis, Cornell University.
- Luster, Tom, Kelly Rhoades and Bruce Haas. 1989. The Relation between Parental Values and Parenting Behavior: A Test of the Kohn Hypothesis. *Journal of Marriage and the Family* 51 (1): 139-147.

- Lynd, Robert Staughton and Hellen Merrel Lynd. 1929. *Middletown: A Study in American Culture*. New York: Harcourt, Brace and Co.
- Rabušić, Ladislav. 2011. The values for children. Paper presented at the EVS Conference. The Value(s) of Europe, November 23-25, 2011, Tilburg, Netherlands.
- Tudge Jonathan R.H., Diane M. Hogan, Irina A. Snezhkovac, Natalya N. Kulakovac and Kathleen E. Etz. 2000. Parents' Child-rearing Values and Beliefs in the United States and Russia: The Impact of Culture and Social Class. *Infant and Child Development* 9 (2): 105–121.
- Tufiş, Paula. 2008. Social status and child-rearing values. In *The Values of the Romanians: 1993-2006. A sociological perspective*, eds. Bogdan Voicu and Mălina Voicu, 193-213. Iaşi: Institutul European.
- Tulviste, Tiia, Luule Mizera, Boel de Geer and Marja-Terttu Tryggvason. 2007. Child-rearing goals of Estonian, Finnish, and Swedish mothers. *Scandinavian Journal of Psychology* 48 (6): 487-497.
- van Deth, Jan and Elinor Scarbrough. 1994. The Concept of Value. In *The Impact of Values*, eds. Jan van Deth and Elinor Scarbrough, 21-47. Oxford: Oxford University Press.
- van Leeuwen, Dawn M. and Keith H. Mandabach. 2002. A Note on Reliability of Ranked Items. *Sociological Methods & Research* 31 (1): 87-105.
- Voicu, Bogdan. 2010. Impactul valorilor sociale asupra rezultatelor educaţionale. Implicaţii pentru politicile educaţionale [The Impact of social values on education achievements. Implications for education policy]. In *Politici educaţionale. Relevanţă şi eficienţă [Education policy. Relevance and efficiency]*, eds. Emilian M. Dobrescu and Niculaie Iancu, 386-406. Bucureşti: EuroLobby.
- Voicu, Bogdan. 2010b. Valorile şi sociologia valorilor [Values and the sociology of values]. In *Sociologie [Sociology]*, ed. Lazăr Vlăsceanu, 249-294. Iaşi: Polirom.
- Xiao, Hong. 2001. *Childrearing values in the United States and China: a comparison of belief systems and social structure*. Westport: Praeger.
- Xiao, Hong. 2000a. Class, Gender, and Parental Values in the 1990s. *Gender and Society* 14 (6): 785-803.
- Xiao, Hong. 2000b. Structure of Child-Rearing Values in Urban China. *Sociological Perspectives* 43 (3): 457-471.