Chapter 14

Aberrant Response Patterns in Elderly Respondents: Latent Class Analysis of Respondent Scalability

G. van den Wittenboer, J.J. Hox and E.D. de Leeuw

Department of Education, University of Amsterdam
Department of Education, University of Amsterdam
Department of Social Research Methodology, Vrije Universiteit

1. Introduction

One well-known source of measurement error in surveys is the respondent. Errors made by respondents are difficult to control: respondents can be instructed in what is expected from them, and they can be motivated to do their best, but it is difficult to manipulate respondents to reduce their errors. Therefore, research has concentrated on identifying respondents who produce many errors.

Person fit analysis investigates whether a person's answering behavior deviates from that predicted by a measurement model, or from the answer patterns of the majority in the group to which that person belongs. For example, if a student answers eight out of ten questions correctly, one generally expects him/her to have missed the two most difficult questions. If the two easiest questions are answered incorrectly, the response pattern is completely unexpected. For persons detected as aberrant, the total test score does not adequately reflect the attribute that is being measured and this total score should not be used in substantial research or decision processes (e.g., educational research, school examinations). Further research is needed before firm conclusions can be drawn about their test performance. To investigate response patterns, data have to be gathered on a test or scale that consists of a number of questions about the same topic. Furthermore, the test or scale should have good psychometric properties; the reliability and the scalability must be high.

One promising approach is the use of „person fit“ indices to detect inconsistent respondents (cf. Meijer & De Leeuw, 1993; De Leeuw & Hox, 1994). Another approach, with fewer assumptions, is the application of latent class analysis to distinguish latent classes of respondents that are consistent according to the measurement model and additional latent classes of respondents that are inconsistent (i.e., aberrant; cf. Dayton & Macready, 1980). In both approaches, a psychometric scaling model (e.g., Guttman model) is adopted to serve as a benchmark for aberrancy. The aim of this paper is to show that the approach with latent class analysis allows a more detailed analysis of the erratic response patterns.

Most research on aberrant response patterns has concentrated on person fit indices, also known as „caution indices.“ In general, three types of person fit indices can be distinguished. The first type consists of indices that are based on the assumptions of parametric IRT-models, such as the Rasch model (cf., Tarnai & Rost, 1990, and Molenaar & Hoytink, 1990). The second type evaluates a response pattern given the assumptions of a nonparametric IRT model (e.g., the Mokken model; Sijtsma, 1988), and the third type evaluates an individual response pattern by means of statistics based on the group to which a person belongs (e.g., the proportion of correct items, cf., Harnisch & Linn, 1981; for an overview see Meijer, 1994.) Person fit indices have been applied, for example, by Levine and Rubin (1979) who discuss person fit indices for the detection of cheating on aptitude tests. Harnisch and Linn (1981) used person fit indices to differentiate schools with a special curriculum on math and reading. Van Tilburg and De Leeuw (1991) applied person fit indices to compare different data collection methods, and Meijer and De Leeuw (1993) used person fit indices to investigate aberrant respondents in a general survey.

Person fit indices have two important drawbacks. First, criteria for cutting scores are not clear. Sometimes rules of thumb exist (cf., Harnisch & Linn, 1981), sometimes a statistical test is used (Van der Flier, 1982). The statistical test gives rise to the well-known type I error; even if there are no aberrant respondents, a certain number will still be detected. Secondly, person fit indices divide respondents into the two classes „normal“ and „aberrant,“ and allow no further specification of different types of aberrant respondents.

Examples of desirable specifications include a random response process, or a systematic but erroneous response process. For instance, in educational research one may distinguish between „guessers“ and „cheaters“. When a student does not know the answer to a difficult question, s/he can guess (producing a random response) or cheat by looking up the answers or looking at a neighbor's answers (a systematic response process). Both strategies may result in an aberrant response pattern, but regular person fit indices do not distinguish between them. With measures on attitude scales, „guessing“ would refer to respondents who answer without much thinking about the precise content of the question, a phenomenon known as the „top-of-the-head“ response. „Cheating“ would refer to respondents that show unexpected response behavior on extreme questions, which can be the result of a social desirability bias that shows up only with extreme questions. Because „guessing“ and „cheating“ are mostly associated with testing data, we will use the more general terms „random aberrant“ and „systematic aberrant“ response patterns. In the next section we will show that latent class analysis, using Lazarsfeld's latent distance model, can be used to distinguish between random aberrant response patterns (e.g., guessing) and systematic aberrant response patterns (e.g., cheating). Just as in most person fit research, deviation from the perfect Guttman pattern is used as a criterion for aberrancy.

2. The Latent Distance Model

Latent class analysis has a long tradition in analyzing dichotomous responses with a Guttman-like structure for the underlying scale (Lazarsfeld, 1950; Lazarsfeld & Henry, 1968; Goodman, 1974; McCutcheon, 1987; Langeheine, 1988; Clogg, 1988). Other latent class models have been proposed for various IRT models, such as the Rasch model. In both approaches, serious attention has been paid to the problem of unscalable respondents (Goodman, 1975; Clogg & Sawyer, 1981; Rost, 1990), who in fact are respondents with aberrant response patterns according to some restricted latent class model.

We confine ourselves to the **LA**tent **D**istance model (LAD) for probabilistic Guttman-type data (see Clogg & Sawyer, 1981; McCutcheon, 1987; Langeheine, 1988, see also section 2.1 in chapter 1). This model was first developed by Lazarsfeld (1950). We apply it to dichotomous data, with category '1' denoting correct answers and category '2' denoting incorrect answers. If there are k dichotomous items, the LAD has k+1 classes. The items are in order of increasing difficulty. Likewise the classes are ordered, with class 1 containing respondents who master all items, class 2 containing respondents who master all items except the most difficult one, and finally class k+1 containing subjects who answer all items incorrectly. The main ideas of the LAD model are presented in Table 1. Given the class, Table 1 shows the conditional probabilities of a correct, or positive, answer to each of the dichotomous items. Cell (3,2), for example, contains the probability $'b_1'$ that someone of class 3 answers item 2 correctly. The probability of an incorrect answer is $'b_2' (= 1 - b_1)$ which is not in the table.

latent classes	*item 1*	*item 2*	*item 3*	*item 4*	*item 5*	*item 6*	*item 7*
class 1	a_1	b_1	d_1	f_1	h_1	j_1	m_1
class 2	a_1	b_1	d_1	f_1	h_1	j_1	$1-m_1$
class 3	a_1	b_1	d_1	f_1	h_1	k_1	$1-m_1$
class 4	a_1	b_1	d_1	f_1	i_1	k_1	$1-m_1$
class 5	a_1	b_1	d_1	g_1	i_1	k_1	$1-m_1$
class 6	a_1	b_1	e_1	g_1	i_1	k_1	$1-m_1$
class 7	a_1	c_1	e_1	g_1	i_1	k_1	$1-m_1$
class 8	$1-a_1$	c_1	e_1	g_1	i_1	k_1	$1-m_1$

Table 1: The latent distance model (the LAD model) as a restricted latent class model

The difference between these conditional probabilities and the corresponding conditional probabilities of 1 and 0 in the deterministic Guttman scale is called the „error rate“. The error rate of cell (1,4) is $f_2 = 1 - f_1$,or the false negative, because in the deterministic Guttman scale f_2 should be zero. Likewise, the error rate of cell (8,4), or the false positives, is g_1 since in the deterministic case g_1 should be zero. Characteristic for the LAD model is that the conditional probabilities follow the deterministic pattern. Exceptions are found at the end points of the scale; to avoid identification problems, the error rates of item 1 and those of item 7 are set equal in each class (Lazarsfeld & Henry, 1968). For items 2 to 6 there are two conditional probabilities that are equal across classes given the Guttman pattern.

The latent distance model can be extended to accommodate aberrant respondents. One possibility is to add additional classes that are completely unrestricted. More specific hypotheses can be investigated by adding classes that model a specific type of aberrant response behavior. For instance, guessing or randomly answering respondents can be modeled by adding an additional class that has equal probabilities of a correct answer for all items (if this probability is fixed at .5 we have total guessing; allowing this probability to be estimated freely models partial knowledge in the guessing process). Cheating can be modeled by systematic aberrants that have unrestricted probabilities at the five easiest items, and which answer correctly to the two most difficult items with probability one.

3. Application to a „Need for Affiliation“ questionnaire

In 1992 a large scale survey was held in the Netherlands about the living arrangements and social networks of older adults (Broese van Groenou et al., 1995). A stratified sample of older adults (aged 55-89) was taken from the population registers of eleven municipalities in the Netherlands. The questionnaire contained questions on demographics, living arrangements and household characteristics, health, family, network members, employment and residential history, loneliness, well-being, social skills and need for affiliation. The overall response rate was 61.7%. The average age of these respondents was 70; slightly more respondents (53%) were female. Most of them lived independently (i.e., not in an institution); Seventy-six (5%) lived in an institution of some sort, including nursing homes, old people's homes and psychiatric hospitals. The majority (64%) lived with a partner, 3% lived alone but had a partner (e.g., in nursing home) and 33% lived alone without a partner. Of this last category the majority (70%) was widowed.

The data were collected in the context of the NESTOR research program conducted by the Vrije Universiteit Amsterdam and the Netherlands Interdisciplinary Demographic Institute in the Hague, funded by the Ministry of Education and Sciences and the Ministry of Welfare, Health and Cultural Affairs.

The „need for affiliation“ scale (developed by Van Tilburg, 1988) contains nine dichotomously scored statements that deal with what one considers to be important in one's relations with others. In total, 1532 respondents completed this scale. For the total scale, the reliability (rho) was .82 and the scalability (Loevinger's H) was .47. Three item pairs had p-values very close together (e.g., .62 and .64). In cases in which the item difficulties are so close together it is hard to decide whether an inversion in the response pattern at that point is an indication of real aberrancy or a failure to ascertain the correct item order. We therefore decided to randomly choose one of each two items that were very close together. This resulted in a six item scale with rho=.75 and H=.46. The values for the scalability and reliability suggest that an analysis of this scale is likely to turn up several aberrant respondents. The results of a sequence of latent class models for these data are summarized in Table 2.

	model	*classes*	*chi- square*	*df*
1	basic LAD model	7	127	47
2	model 1 + one free class	8	64	41
3	model 1 + two free classes	9	47	34
4	model 1 + random & systematic error	9	71	44
5	model 4 + one free class	10	45	38

Table 2: Fit of (extended) LAD models for need for affiliation data

The first model for these data is the basic LAD model, which does not fit well. With two additional free error classes we obtain a satisfactory model (p = .07). When we restrict the error classes to model random and systematic aberrants as defined in section 3, the chi-square goes up again and the model is rejected (p = .01). Adding another free class produces a model that is again acceptable (p = .20). We prefer model (5) to model (3) because it has a slightly lower chi-square with more degrees of freedom, and because it gives more information about the different types of aberrant respondents. Model (5) estimates the total proportion of aberrant respondents as .078: .041 random aberrants, .006 systematic aberrants, and .031 in the free error class. The free error class consists of respondents who respond with virtual certainty „yes“ to question 4 (response probability .99) and „no“ to question 2 (response probability .93). Given the small estimated proportion of respondents in this class, we regard this class as consisting of respondents with a highly specific response set.

Usually, respondents are assigned to one of the classes according to the highest recruitment probability. Following that rule, we find 4% of the respondents assigned to aberrant classes, which is acceptable given the combined probability of .078. However, assigning respondents to one of the aberrancy classes implies that they do not receive a score on the need for affiliation scale. Therefore, we have used a more intricate coding scheme than assignment to the latent class with the highest recruitment probability. In this scheme, all respondents whose probability of being aberrant was less than 1 were assigned a LAD score based on the highest recruitment probability. Next, a variable was created that indicated the probability of being aberrant. Then, all respondents with a nonzero probability of being aberrant were assigned to an aberrancy class. As a result, respondents may have an assigned score on the substantive scale and also belong to an aberrancy class.

This procedure results in 13 respondents classified as systematic aberrant, 146 as random aberrant. The free error class is ignored in the following analyses; we concentrate on the systematic and the random aberrants. If we take a closer look at the 13 respondents that are classified as systematic aberrants, we find that only seven of them live independently (54%, as opposed to 68% for the other respondents), and that three of them receive some institutional care (23%, as opposed to 7% for the rest). We also find that one of them is not Dutch (8%, as opposed to 5% for the rest). This suggests that at least some of these aberrant response patterns may be linked to specific respondent problems. If we take a closer look at the 146 respondents classified as random aberrants, we find the intriguing fact that 17 aberrant respondents find the interview „tiring“ (12% against 18% for the rest). Thus, respondents who have few aberrant patterns, more often describe the interview as tiring. Perhaps these respondents „work“ harder during the interview, resulting in better (less aberrant) responses. This suggests that random aberrancy may be associated with not thinking too hard about the answers, a phenomenon that survey researchers know as „top-of-the-head responses.“

4. Discussion

The latent class analysis with the LAD model of the data estimates that only 8 percent of the respondents belong to a latent aberrancy class. This confirms the conclusion from the psychometric analyses that it is a fairly strong scale. Looking at the available background data, we could formulate a hypothesis about what went wrong in the response process for a few of the aberrant cases. The general impression is that in this case the aberrant responses are the consequence of highly idiosyncratic aspects of the response process. There is no distinct profile that characterizes the aberrant respondents. It should be noted that the computer assisted interview contained an automated check for signs of senile dementia. If these occurred, the program switched to a dementia test, and if this was positive, the interview was cut short. This eliminates all potentially extreme aberrant patterns from our data set.

It is possible that there were only a few aberrant respondents in our sample of elderly people. It is also conceivable that problems emerge from using latent class models for this kind of data. These problems concern the interpretation of the models and their estimation. The interpretation problem occurs because, even if the data contain many perfect Guttman patterns and the log-ratio chi square indicates a good fit, the estimated probabilities do not necessarily follow the expected Guttman pattern. Therefore, the degree to which the parameter estimates follow the ideal Guttman pattern is considered another criterion for the adequacy of the LAD model. In our example, the estimated probabilities of some models in Table 2 differ from the ideal Guttman pattern. For instance, both model 3, the model with two extra free classes (which fits fairly well with p = .07) and model 4, the model with one class of random and one class of systematic aberrants (which does not fit well), show more different probabilities in class 1, 4, and 8 than in all the other models. Some of these probabilities produce a reversion in the ideal Guttman pattern. Model 5, the one finally chosen, links properly with the remaining models, and the parameter estimates do not violate the Guttman pattern.

A technical problem encountered with all models is the sensitivity to starting values. Both in LCAG (Hagenaars & Luijkx, 1987) and Panmark (Van der Pol et al., 1991) the estimation procedure often leads to local maxima (a local maximum is suspected when different starting values lead to different results, or when in a sequence of models a more restricted model has a smaller chi-square than a less restricted model). The program Panmark has the option to generate a large set (by default 1000) of random starting values; to compare the results of the first iterations (by default eight); and to continue the estimation process with the best three starting sets. If the results of these converge, one may have some confidence that a global maximum is found. In our example, the default procedure did not solve the problem. We needed twenty, time consuming, initial iterations for all 1000 sets of starting values before the results became stable.

A related technical problem is that the computer programs show a strong tendency to converge on probabilities equal to zero or one. For example, Guttman classes with relatively low frequencies, which are nevertheless present in the data, often are set equal to zero instead of a small probability as one should expect. It is not obvious whether this is because of the data, the specific models, or the EM algorithm used in the analysis.

A problem that is related to the algorithm is that, once a latent class is estimated to have a probability of zero, this value does not change in subsequent iterations. In explorative analyses (not reported here) we noted that this effect became more striking with longer sets of items than the six items we have used in this example. To deal with perfect Guttman classes with „zero“ probability, Clogg and Sawyer (1981) suggest that the item which causes the problem can be eliminated under certain conditions. It remains unclear, however, which item this should be. Eliminating items is particularly problematic in a probabilistic environment where the errors also give information about the scalability. Therefore we decided to leave the scale intact, and to tolerate boundary values for the estimates.

It should be noted that there are probably more aberrant response patterns than the two specified above. The challenge is to formulate a plausible, but erroneous response process, and to formulate this as a set of restrictions on the basic model. This can pose formidable difficulties. For instance, our definition of „systematic aberrants“ allows free probabilities for some items, while the response probabilities for the most difficult items are fixed at one. A more restrictive definition would require that the free probabilities are all fixed at .5, which corresponds to pure guessing, while our definition allows for partial knowledge. Moreover, systematic aberrants of this type are scarce in our real life data. Relaxing the restrictions to a probability of one for errors only made at the most difficult item would interfere with the usual error probability that should be allowed in the case of probabilistic Guttman scales. Despite the problems, we are nevertheless convinced that error modeling by means of restrictions in latent class analysis leads to a better understanding of aberrant response behavior.

References

Broese van Groenou, M.I. T.G. van Tilburg, T.G., De Leeuw, E.D. & Liefbroer, A.C. (1995). Data Collection. In: C.P.M. Knipscheer, J. de Jong Gierveld, T.G. van Tilburg, & P.A. Dykstra (eds). *Living arrangements and social networks of older adult.* Amsterdam: VU University Press.

Clogg, C.C., (1988). Latent class models for measurement. In: R. Langeheine & J. Rost. *Latent trait and latent class models.* New York, Plenum Press.

Clogg, C.C., & Sawyer, D.O., (1981). A comparison of alternative models for analyzing the scalability of response patterns. In: S. Leinhardt (ed.). *Sociological methodology 1981.* San Francisco, Jossey-Bass, 240-280.

Dayton, C.M., & Macreay, G.B. (1980). A scaling model with response errors and intrinsically unscalable respondents. *Psychometrika*, 45, 343-356.

De Leeuw, E.D. & Hox, J.J. (1994). Are inconsistent respondents consistently inconsistent? A study of several nonparametric person fit indices. In: J.J. Hox & W. Jansen (Eds). *Measurement problems in the social sciences.* Amsterdam: SISWO.

Goodman, L.A., (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.

Goodman, L.A., (1975). A new model for scaling response patterns: an application of the quasi-independence concept. *Journal of the American Statistical Association*, 70, 755-768.

Hagenaars, J.A., & Luijkx, R., (1987). *LCAG; latent clas models and other log- linear models with latent variables. User's manual.* Working Paper Series #17, Tilburg University, Department of Sociology.

Harnish, D.L., & Linn, R.L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement*, 18, 133-146.

Langeheine. R. (1988). New developments in latent clas theory. In: R. Langeheine & J. Rost. *Latent trait and latent class models.* New York, Plenum Press.

Lazarsfeld, P.F., (1950). The interpretation and computation of some latent structures. In: S.A. Stouffer, L. Guttman, E. Suchman, P.F. Lazarsfeld, & J. Clausen (Eds). *Measurement and prediction.* Princeton, Princeton, University Press, 413-472.

Lazarsfeld, P.F., & Henry, N.W., (1968). *Latent structure analysis.* Boston, Houghton Mifflin.

Levine, M.V., & Rubin, D.B. (1979). Measuring the appropriateness of multiple-choice test score. *Journal of Educational Statistics*, 4, 269-290.

McCutcheon, A.L., (1987). *Latent class analysis.* Newbury Park, Sage.

Meijer, R.R. (1994) *Nonparametric person fit analysis.* Ph.D.-thesis, Vrije Universiteit, Amsterdam.

Meijer, R.R., & De Leeuw, E.D. (1993). Person fit in survey research: The detection of respondents with unexpected response patterns. In J. H.L. Oud & R.A.W. van Blokland-Vogelesang (Eds.), *Advances in longitudinal and multivariate analysis in the behavioral sciences.* Nijmegen: ITS.

Molenaar, I.W., & Hoytink, H. (1990). The many null distributions of person fit indices. *Psychometrika*, 55, 75-106.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.

Sijtsma, K. (1988). *Contributions to Mokken's nonparametric item response theory.* Amsterdam, NL: Free University Press.

Tarnai, C., & Rost, J. (1990). *Identifying aberrant response patterns in the Rasch model.* Munster: Sozialwissenschaftliche Forschungsdokumentationen.

Van der Flier, H., (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-cultural Psychology*, 13, 267-298.

Van der Pol, F., Langeheine, R., & De Jong, W. (1991). *Panmark user manual; panel analysis using Markov chains.* Version 2.2. Voorburg, CBS, Department of Statitical Methods.

Van Tilburg, T.G., & De Leeuw, E.D. (1991). Stability of scale quality under various data collection procedures: A mode comparison on the „De Jong-Gierveld loneliness scale“. *International Journal of Public Opinion Research*, 3, 69-85.

Van Tilburg, T.G. (1988). *Verkregen en gewenste ondersteuning in het licht van eenzaamheidservaringen* [Obtained and desired social support in association with loneliness]. Amsterdam, Vrije Universiteit, Ph.D. Thesis.