KM 41 (1992)
pg.59-74

Modeling Interviewer Effects with Multilevel Models

J.J. Hox[1]
University of Amsterdam

Summary

It is generally recognized that interviewers may have an important effect on the quality of the data collected in survey research. This article presents an application of the hierarchical regression model (multilevel regression model) in the analysis of interviewer effects. The hierarchical regression model offers an elegant way to analyze the effects of specific interviewer and respondent characteristics. It is especially attractive if the research design does not provide for a random assignment of respondents to interviewers, because it allows the researcher to use statistical instead of experimental control by modeling the interviewer effects conditional on the respondent effects.

Faculty of Educational Sciences, University of Amsterdam
IJsbaanpad 9, 1076 CV Amsterdam
(020) 6644331 E-mail: A716HOX @ HASARA11

[1] I thank Bob Groves, Ita Kreft, Edith de Leeuw, and an anonymous reviewer for their comments on earlier versions, and Edith de Leeuw for providing me with the data set.

The survey interview is a major source of research data in many social science disciplines (cf. Brown and Gilmartin, 1969; Presser, 1984). As a result, there is a large and still growing literature about the quality of survey data (for example, Alwin, 1978, 1991; De Jong-Gierveld & Van der Zouwen, 1987; Groves et al., 1988; Van der Zouwen & Dijkstra, 1989; Biemer et al., 1991). Groves (1989) differentiates between two major sources of error: error of nonobservation and observation error. The errors of nonobservation arise because surveys as a rule do not register the complete population; in this category Groves puts coverage error, nonresponse error, and sampling error. Observational errors are those errors that would arise even if the survey produces a complete enumeration of the population (Bailar, 1987; Groves, 1989; O'Muircheartaigh, 1977). Groves (1989) categorizes observational errors according to their source into four categories: *interviewer* effects, *respondent* effects, *instrument* effects, and *mode* effects (effects of the specific mode of data collection used.) This article focuses specifically on the use of hierarchical linear regression models for research on interviewer and respondent effects.

Both *respondents* and *interviewers* have long been recognized as a potential source of error in survey interview data. Various respondent characteristics such as age and education have been thought to affect data quality (Sudman and Bradburn, 1974; Groves, 1989). In general, the literature is somewhat equivocal; respondent effects are generally reported to be small (Groves, 1989), although Alwin and Krosnick (1991) report fairly large effects of respondents' education on the reliability of survey questions. The effects of interviewer characteristics are also generally reported as small (Sudman and Bradburn, 1974; Bradburn, 1983; Groves, 1989). Still, even small interviewer effects may have an important effect on the quality of survey data, especially when each interviewer interviews a large number of respondents (Kish, 1965). Finally, there is some evidence for interaction effects between respondent and interviewer characteristics (Freeman and Butler, 1976), especially with respect to race and gender (Collins, 1980; Stokes and Yeh, 1987).

Studies on respondent and interviewer effects generally combine both respondent and interviewer variables. The design of such studies reflects several specific methodological problems; for a concise review see Hagenaars and Heinen (1982). For our purpose, two factors are important: the necessity to use a design in which the interviewer and respondent characteristics are experimentally independent, and the hierarchical structure of the data.

To investigate the independent (additive and interaction) effects of respondent and interviewer characteristics, a design must be used that leads to low or preferably zero correlations between interviewer and respondent characteristics. In its simplest form, both respondents and interviewers are sampled at random from some population, and respondents are assigned at random to different interviewers. In this method, described by Mahalanobis (1946) as the method of 'interpenetrating samples,' straightforward analysis of variance can be used to estimate the effect of various independent variables. More complicated designs use reinterviewing by either the same or different interviewers, with reinterviews assigned at random, which leads to more complicated ANOVA models (cf. Hanson, Hurwitz, and Bershad, 1961; O'Muircheartaigh, 1977; Biemer and Stokes, 1985; Groves, 1989).

The adequacy of the research design depends critically upon the way respondents are assigned to interviewers. In fact, much research on respondent and interviewer effects is based on a secondary analysis of survey data collected for a non-methodological purpose. Because of the expense, respondents are generally not randomly assigned to the interviewers. In such cases, respondents may be randomly assigned to interviewers within specific geographic regions, which avoids excessive traveling times, and at least partially controls the confounding of respondent and interviewer variables (see Bailar, 1983, for some common fractional interpenetration designs). However, if the assignment of respondents to interviewers is not completely random, but decided by convenience, interviewer characteristics are to an unknown degree correlated with respondent characteristics, and the estimates of interviewer effects are confounded by respondent effects, and vice versa. As a result, statistical control must be used by conditioning the analysis of interviewer effects on confounding respondent effects.

The analysis of respondent effects is simple, although in the presence of a significant interviewer effect the analysis should include this as a design effect (Kish, 1987), for instance by including the interviewers as a random factor (Dijkstra, 1983). For the analysis of interviewer effects two approaches have been popular. The first is to consider the interviewer effect as a random effect, which increases the variance of sample means (and other sample statistics). A random effect ANOVA model can then be used to estimate the interviewer variance component, and the intra class correlation can be used to estimate the population interviewer effect (cf. Hanson and Marks, 1953; Kish, 1965, 1987; Groves, 1989). In this approach, the effect of explanatory interviewer variables is investigated by splitting the interviewer sample, e.g., in male and female interviewers, and estimating the intra class correlation separately for each subsample. If the intra class correlation vanishes in the subsamples, the explanatory variable used to split the sample of interviewers is assumed to explain the interviewer effect.

The second approach to the analysis of interviewer effects is to assess the effect of explanatory variables measured at the interviewer level, such as the interviewer's sex, age, or experience (Sudman and Bradburn, 1974; Bailar, Bailey, and Stevens, 1977; Berk and Bernstein, 1980). Typically, interviewer variables are disaggregated to the respondent level, and both interviewer and respondent variables are combined in one ANOVA or regression model.

Such a single level analysis combines interviewer and respondent data in one regression equation. However, this can be shown to violate a number of important assumptions of ordinary multiple regression analysis. Here, the most critical assumptions are that the error terms are uncorrelated, and that the units of analysis are independently sampled. Since a number of respondents is interviewed by each single interviewer, unmeasured interviewer variation will, to an unknown degree, cause correlated error terms within respondents. The assumption of independent sampling is violated because respondents interviewed by the same interviewer will have values for interviewer variables which are necessarily exactly equal. These violations affect both point estimates of regression parameters and their standard errors. The standard errors are underestimated, particularly for the interviewer variables (cf. Kreft, 1987). Furthermore, the point estimates, while generally unbiased (Tate & Wongbundhit, 1983; De Leeuw & Kreft, 1986), are inefficient (see Hox, Kreft & Hermkens, 1991, for an empirical example.) In some cases, even the signs of the regression coefficients may be misleading (cf. Kreft & De Leeuw, 1988).

Because the respondents are hierarchically nested within the interviewers, a hierarchical analysis model must be used for the analysis of respondent and interviewer effects. Specialized hierarchical models have been proposed to analyze interviewer effects (for instance Pannekoek, 1988, 1991; Hill, 1991). However, the well known hierarchical linear regression model is a very useful general model that permits the estimation of both the interviewer variance and of the effects of explanatory variables measured at the interviewer and the respondent level. The hierarchical linear regression model, also known as the random component model (Longford, 1986), or the random coefficient model (De Leeuw & Kreft, 1986), has been described in a number of review articles (for example, Mason, Wong, & Entwisle, 1984; Raudenbush & Bryk, 1986, 1988; Raudenbush, 1988) and books (Goldstein, 1987; Bock, 1989; Bryk & Raudenbush, 1992). For the statistical and computational details of these *multilevel* models I refer to this literature. The multilevel regression model has been used extensively in educational research, where pupils, classes, and schools present a 'natural' hierarchical system (cf. Bock, 1989); applications in the field of interviewer research are still rare (examples are Wiggins, Longford, and O'Muircheartaigh, 1990; Hox, De Leeuw, and Kreft, 1991; Van den Eeden, 1991).

The Hierarchical Regression Model for Interviewer Effects

Suppose we are interested in the effect of certain interviewer characteristics on the quality of the data obtained when they interview specific respondents. Let us assume that we select J interviewers at random from a large interviewer pool, and that each interviewer interviews n_j randomly selected respondents. The dependent variable $\mathbf{Y}_{ij}$ is a measure that indicates some aspect of the data quality of the responses of a specific respondent, for instance item nonresponse or social desirability. Thus, $\mathbf{Y}_{ij}$ is the score of respondent i $(i = 1,..,n_j)$, assigned to interviewer j $(j = 1,..,J)$. If we have no other information about the interviewers or respondents, we can apply the following linear model (stochastic variables are written in bold):

$$\mathbf{Y}_{ij} = \boldsymbol{\beta}_{0j} + \boldsymbol{\varepsilon}_{ij} \qquad (1)$$

$\boldsymbol{\beta}_{0j}$ is the intercept (the expected value of $\mathbf{Y}$) for interviewer j, and $\boldsymbol{\varepsilon}_{ij}$ is the residual for respondent i for interviewer j, which varies between respondents. The intercept $\boldsymbol{\beta}_{0j}$ is treated as a stochastic variable at the interviewer level, which can in turn be written as:

$$\boldsymbol{\beta}_{0j} = \gamma_{00} + \boldsymbol{\delta}_{0j} \qquad (2).$$

Substitution of (2) in (1) gives

$$\mathbf{Y}_{ij} = \gamma_{00} + \boldsymbol{\delta}_{0j} + \boldsymbol{\varepsilon}_{ij} \qquad (3)$$

where γ_{00} is the overall intercept, $\boldsymbol{\delta}_{0j}$ is an interviewer level residual which varies between interviewers, and $\boldsymbol{\varepsilon}_{ij}$ is the respondent level residual. It is assumed that the $\boldsymbol{\varepsilon}_{ij}$ are distributed within each interviewer with expectation zero and variance σ_j^2; in most applications it is assumed that the respondent level residual is equal for all interviewers, that is: all σ_j^2 are equal to σ_ε^2. The interviewer level residuals $\boldsymbol{\delta}_{0j}$ are assumed to be independent from the $\boldsymbol{\varepsilon}_{ij}$, having a distribution with expectation zero and variance σ_0^2. The intra class correlation ρ for the interviewer effect can now be calculated as:

$$\rho = \sigma_0^2 / (\sigma_0^2 + \sigma_\varepsilon^2). \qquad (4)$$

If there is no variation at the interviewer level, σ_0^2 equals zero, which shows that the assumption of nonzero interviewer effects introduces a correlation between measurements collected by the same interviewer, but not between measurements collected by different interviewers.

In the complete hierarchical regression model, we have *P* explanatory variables X_{pij} (p=1..P) at the respondent level (e.g., respondents' age or education) and *Q* explanatory variables Z_{qj} (q=1..Q) at the interviewer level (e.g., interviewers' age or experience). The effect of the respondent variable X_{pij} on the dependent variable $\mathbf{Y}_{ij}$ can be described by the following linear model:

$$\mathbf{Y}_{ij} = \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}_{pj} X_{pij} + \boldsymbol{\varepsilon}_{ij} \qquad (5)$$

The intercept $\boldsymbol{\beta}_{0j}$ and the slopes $\boldsymbol{\beta}_{pj}$ are treated as random variables at the interviewer level that can be modeled by the interviewer variables Z_{qj}:

$$\boldsymbol{\beta}_{0j} = \gamma_{00} + \gamma_{0q} Z_{qj} + \boldsymbol{\delta}_{0j} \qquad (6)$$

$$\boldsymbol{\beta}_{pj} = \gamma_{p0} + \gamma_{pq} Z_{qj} + \boldsymbol{\delta}_{qj} \qquad (7)$$

Substituting (6) and (7) into (5) gives

$$\mathbf{Y}_{ij} = \gamma_{00} + \gamma_{p0} X_{pij} + \gamma_{0q} Z_{qj} + \gamma_{pq} Z_{qj} X_{pij} + [\, \boldsymbol{\delta}_{pj} X_{pij} + \boldsymbol{\delta}_{0j} + \boldsymbol{\varepsilon}_{ij} \,]. \qquad (8)$$

In equation (8) the part $\gamma_{00} + \gamma_{p0} X_{pij} + \gamma_{0q} Z_{qj} + \gamma_{pq} Z_{qj} X_{pij}$ contains only fixed coefficients; it is called the *fixed part*. The gamma's in this part can be interpreted as raw regression coefficients in a multiple regression. The product $Z_{qi} X_{pij}$ that arises as a consequence of substituting (7) into (5) is an interaction term, which specifies specific cross level interactions between interviewer and respondent variables. The part $\boldsymbol{\delta}_{pj}X_{pij} + \boldsymbol{\delta}_{0j} + \boldsymbol{\varepsilon}_{ij}$ that is written in square brackets in equation (8) contains the random error structure; it is called the *random part*. The residuals $\boldsymbol{\delta}_j$ are assumed to be independent from the $\boldsymbol{\varepsilon}_{ij}$, and to have a joint multivariate distribution with covariance matrix Ω. It should be clear from equation (8) that even while the fixed part may look much like an ordinary regression equation, the random part is more complicated, with random terms $\boldsymbol{\delta}_{0j}$ in addition to the usual $\boldsymbol{\varepsilon}_{ij}$, and each first level regression slope having a distinct random error term $\boldsymbol{\delta}_{pj}$, which also involves the corresponding first level explanatory variable X_{pij}. The estimation procedures and programs currently available all produce asymptotic standard errors for the gamma's and the variance components. They also produce an overall measure of the fit of a specific model, the *deviance*. The difference between the deviances of two nested models is distributed as a chi-square variate, with degrees of freedom equal to the difference between the number of parameters estimated by both models. Thus, the deviance can be used to compare the fit of different submodels (cf. Kreft, De Leeuw and Kim, 1990) in a manner analogous to the chi-square test for the difference between two nested Lisrel models.

It should be noted that both the regression coefficients gamma and the variance components sigma are conditional upon the explanatory variables in the model. This property of the random coefficient model is very useful if there is no complete orthogonalization of interviewer and respondent variables, and statistical control of confounding variables is necessary. For example, it may be useful to compare both regression coefficients and variance components of a model with only interviewer variables to the corresponding estimates obtained in a model that also incorporates respondent variables. The differences between these estimates would indicate how much of the alleged interviewer effects could actually be explained by systematic differences between the respondents interviewed by the different interviewers. This approach follows the general strategy of constructing a model starting at the lowest level, and inspecting at each level the size and significance of the regression coefficients and variance components to decide which variables must remain in the model. In addition to the standard errors of the parameters in the model, the deviances of two nested models can be used to decide if the larger models fits significantly better than the smaller model. The example in the next paragraph follows this approach in the specific case of an analysis of interviewer effects; Raudenbush and Bryk (1986) present an example of a similar analysis strategy in educational research.

Model Selection and Analysis; an Example

The example data concern a controlled field experiment on mode effects (De Leeuw, 1992), in which interviewer and respondent effects were also investigated (see Hox, De Leeuw, and Kreft, 1991, for detailed results). In the present example, data are analyzed from 515 respondents, who were questioned by 20 interviewers. Three data collection methods are compared: 221 of the interviews were conducted face to face, 219 by telephone using a paper and pencil questionnaire, and 75 by telephone using Computer Assisted Telephone Interviewing (CATI), all three using the same interviewers. The respondents were randomly assigned to the different data collection methods; in both telephone conditions they were also randomly assigned to interviewers. Because of financial constraints, in the face to face condition random assignment of respondents to interviewers was used within four broad geographical regions.

The dependent variable in the analysis is the total time needed for an interview. Since time measures generally have a skewed distribution, an inverse transformation is used ($f(x)=1/x$; see Kirk, 1968), which transforms the dependent variable *time* into the variable *speed*. Thus, the dependent variable $\mathbf{Y}_{ij}$ is the speed or the *pace* of the interview, measured in number of questions completed per minute. The explanatory variables X_{pij} at the respondent level include two dummy variables that indicate the three data collection methods used: one contrast variable (coded +1, -1; Cohen, 1983) contrasting the telephone condition with the face to face condition *(tel)* and one contrast variable contrasting the CATI with the paper and pencil telephone condition *(cati)*. The other respondent variables are respondent age *(r-age)*, and loneliness *(lonely)*, as measured by the De Jong-Gierveld loneliness scale (De Jong-Gierveld, 1987). The explanatory variables Z_{qj} at the interviewer level are: amount of earlier interviewer *training*, amount of interviewing *experience*, interviewer age *(i-age)*, interviewer preference for telephone interviewing *(pref.tel.)*, and the interviewer's score on five personality scales: extroversion *(extro)*, friendly disposition *(friendly)*, conscientiousness *(cons.)*, social assurance *(soc.ass.)*, and ability to terminate awkward situations *(term.)*.

Since the design is not completely orthogonal, the first step in the analysis is to inspect the correlations between respondent and interviewer explanatory variables. These are given in Table 1:

Table 1. Correlations between Respondent and Interviewer Variables.

Int. Var. /// Resp. Var.	Tel.	CATI	r-age	lonely
training	-.05	-.10	-.04	-.02
experience	-.09	-.15	-.12	.05
i-age	-.10	-.09	-.04	.00
preftel.	.10	.03	.08	-.03
extro	.01	-.14	.03	.05
friendly	.01	-.01	.00	-.01
consc.	.03	.08	.02	-.02
soc. ass.	-.01	-.07	.02	-.02
term.	.06	-.15	.06	.00

The correlations between respondents and interviewers are low, indicating that the partial orthogonalization was successful. Yet, since the respondent and interviewer effects to be investigated are generally also small, it is safer to take these correlations into account in the analysis. Modeling the interviewer effects conditional upon the respondetn variables will accomplish this.

The starting point for the model construction is the model with no explanatory variables, also known as the *intercept-only* model. This is given by equation (3), which is repeated here:

$$Y_{ij} = \gamma_{00} + \delta_{0j} + \varepsilon_{ij}. \tag{3}$$

This model provides us with an estimate of the global intercept and the two variance estimates σ_ε^2 for the residual variance at the respondent level and σ_0^2 for the intercept variance at the interviewer level.[2] In our example the intercept is 3.2, indicating an overall interviewing speed of slightly more than three questions per minute. The total variation is decomposed into a respondent level variance σ_ε^2, which equals 0.68, and an interviewer level variance σ_0^2, which equals 0.11; this estimates the intra interviewer correlation ρ as 0.14. The results for the model corresponding to equation (3) are summarized in Table 2 below.

The next analysis step examines the explanatory variables at the lowest (respondent) level. First, they are added as fixed variables, or, in other words: without the corresponding variance components for the regression slopes. This model is given by:

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + [\, \delta_{0j} + \varepsilon_{ij} \,]. \tag{9}$$

In the next analysis step, the regression coefficients of the respondent variables are assumed to be random, that is: they are assumed to vary between interviewers. This is described by the following equation:

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + [\, \delta_{pj}X_{pij} + \delta_{0j} + \varepsilon_{ij} \,]. \tag{10}$$

In equation (10), each random regression slope γ_{p0} has a corresponding error term $\delta_{pj}X_{pij}$. In our example data, the effect of the CATI contrast turns out to be not significant (p=.25).[3] The other explanatory respondent variables are all significant (largest p<.00). Only the regression slope for the telephone contrast has a significant variance component· The conclusion is that the model for the respondent effects may be simplified by dropping the CATI contrast altogether, and assuming a random slope only for the telephone contrast[4] The results of the simplified model derived from equation (10) are summarized in Table 2 below.

---

[2] All calculations were done with Longford's program VARCL, which produces Full Maximum Likelihood (FML) estimates (cf. Kreft et al., 1990).

[3] All p-values are derived from a two sided normal approximation of Z=(coefficient/SE). Since the SE's are asymptotic, and the normal distribution not always proper (for instance when variances are tested), the resulting p-values are approximate.

[4] The random slope effects were examined by allowing all four slopes to be random and examining their standard errors. There is a penalty to this strategy, because including many random effects implies statistical models which are *much* more complex. This may lead to unstable or improper estimates, and slows down the computations, especially if some effects are near zero. If there are many respondent variables, the preferred strategy would be to make the respondent variables random one by one, and finally estimate a model including random effects for only those slopes which had a significant variance. Another strategy would be to inspect simple regression coefficients, an option offered by the HLM program.

According to this model, interviews go faster in the telephone condition (γ=0.30), whether they were by paper and pencil or by the CATI method, and older respondents are a bit slower (γ=-.01). Also, lonely respondents take longer to interview (γ=-.04). In this model, the respondent level variance σ_{ε}^2 is 0.53, and the interviewer level intercept variance σ_0^2 is 0.08. The total variance in the intercept-only model analyzed earlier is 0.79; including respondent level variables decreases this total variance to 0.61. Both the respondent level variance and the interviewer level variance are decreased by about 22 percent. Dividing the interviewer level variance by the total variance still gives 0.14, but this figure is no longer the intra interviewer correlation. (When the model includes random slope effects, this figure no longer has a simple interpretation, and will not be calculated.)

The next analysis step adds explanatory variables at the interviewer level, giving:

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q} Z_{qj} + [\, \delta_{pj}X_{pij} + \delta_{0j} + \varepsilon_{ij} \,]. \qquad (11)$$

In our example, only three of the nine interviewer variables turn out to be significant: interviewer training, preference for telephone, and extroversion. The results for the final model based on equation (11), dropping the nonsignificant effects, are again summarized in Table 2 below.

The significant between-interviewers variation of the regression slopes for the telephone condition can be modeled by including interactions between the respondent level variable 'telephone condition' and explanatory variables at the interviewer level (cf. equations 5, 6 and 7). This gives the full model already formulated in equation (8), which is repeated here:

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q} Z_{qj} + \gamma_{pq} Z_{qj}X_{pij} + [\, \delta_{pj}X_{pij} + \delta_{0j} + \varepsilon_{ij} \,]. \qquad (8)$$

In our case, there is only one significant cross-level interaction, which is the interaction of the telephone contrast with the interviewer variable social assurance. Since interaction effects should only be examined if the corresponding main effects are also included in the model (Jaccard, Turrisi, and Wan, 1990), the (nonsignificant) interviewer variable social assurance is added to the model with its interaction with the telephone contrast. To ease interpretation, 'social assurance' is centered around its overall mean of 61.8, and the interaction variable is computed using the centered variable (cf. Aiken & West, 1991). The various coefficients of this final model are summarized in the last column of Table 2:

Table 2. Results of selected models[a]

Variables in // Model eq.:	(3)	(10)	(11)	(8)
Fixed part (gamma's):				
intercept	3.19	3.73	1.77	1.43
tel.		.30 (.04)	.30 (.04)	.30 (.04)
resp-age		-.01 (.002)	-.01 (.002)	-.01 (.002)
resp-lonely		-.04 (.01)	-.04 (.01)	-.04 (.01)
int. training			.20 (.10)	.25 (.10)
int. pref. tel.			.25 (.08)	.27 (.08)
int. extro.			.02 (.006)	.02 (.007)
int. soc. ass. (n.s.)				-.01 (.007)
interaction tel.* soc.ass.				.01 (.005)
Random part (sigma's):				
σ_ε^2	.68[b]	.53	.52	.52
σ_0^2 (interc.)	.11[b]	.08	.04	.03
σ_1^2 (tel. slope)		.02[b]	.02	.01
Explained variance:				
of σ_ε^2		23%	23%	23%
of σ_0^2 (interc.)		22%	65%	68%
of σ_1^2 (tel. slope)			0%	22%
Deviance	1294	1173	1161	1155

[a] Standard errors between parentheses.

[b] This variance estimate is the basis for the explained variance in subsequent models.

In model (10), the respondent variables explain 23% of the residual variance at the respondent level, and 22% of the intercept variance at the interviewer level. The explanatory interviewer variables added in model (11) explain a further 43% of the intercept variance at the interviewer level. Adding the (nonsignificant) interviewer variable social assurance and its interaction with the telephone contrast (model (8) in the last column of Table 2) explains a further 3% of the intercept variance, and 22% of the variance of the regression slope for the telephone contrast. The decomposition of the explained variance between the different models in Table 2 shows that both the respondent and the interviewer variables explain a significant portion of the initial variance in interview speed. The interaction that is added in model (8) does not appear to explain much variance, but in fact it does explain a considerable proportion of the slope variance that appears in the previous model (11).

Using the model's deviances for a chi-square test shows that in all comparisons of consecutive models, the more complicated models have a significantly better fit:

Table 3. Comparisons of fit between succesive models.

Model eq.:	Deviance	Dif. w. prev. model	df	p
(3)	1294.1	--	--	--
(10)	1172.6	121.5	5	.00
(11)	1161.4	11.2	3	.01
(8)	1155.1	6.3	2	.04

Most of the gamma coefficients in Table 2 are stable between different models. Although interviewer and respondent variables are correlated, adding the interviewer variables to the model does not appreciably change the regression slopes for the respondent variables. Only the intercept changes. Since the intercept reflects the expected value of the dependent variable if all explanatory variables equal zero, this is merely the consequence of adding explanatory (interviewer) variables to the model that are not all centered around their overall means. The interpretation of the regression slopes is straightforward. Interviews take longer with older and lonely respondents, previously trained and extrovert interviewers are faster, and interviewers that have expressed a preference for using the telephone are also faster. The regression contrast for the telephone condition is coded -1 for the face to face condition, and +1 for both telephone conditions. Its slope coefficient of 0.30 means that the telephone condition is faster by (2×0.30=) 0.6 question per minute; at an overall average of 3.18 questions per minute this means that telephone interviews are 19% faster. However, since the variable 'telephone condition' is involved in an interaction, we cannot interpret the interaction effect and the corresponding simple (main) effects in isolation cf. Kerlinger, 1986). When an interaction between two explanatory variables is involved, the simple regression coefficients for either of these variables reflect a conditional relationship, which is the relationship that holds when the other explanatory variable has the value zero. Since social assurance is centered around its overall mean of 61.8, the regression slope for the telephone contrast reflects the effect of this explanatory variable for interviewers with an avarage social assurance. To interpret the interaction, it is convenient to plot the regression slope of one explanatory variable at various values of the other (Jaccard, Turrisi, and Wan, 1990). Since the telephone contrast has only two values, the best graphical representation here is to plot the slope of the interviewer variable social assurance for both values of the telephone contrast, which is displayed in Figure 1 below:

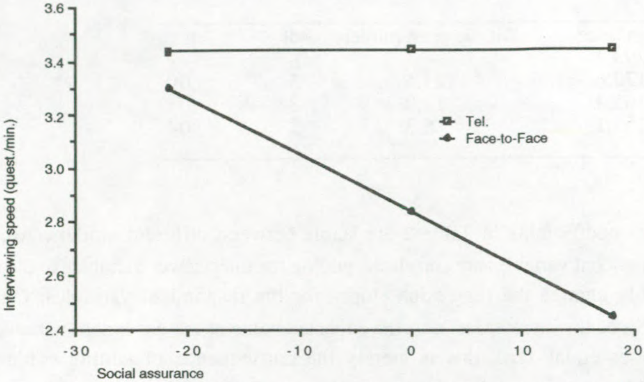

Fig.1. Social assurance slopes in telephone and face-to-face conditions.

Figure 1 shows that over the observed range of values for social assurance, telephone interviewing is faster than face-to-face interviewing. Interviewers with a higher social assurance tend to use more time for the face to face interview, while for the telephone interview there is no relationship between social assurance and interviewing time. For an explanation, it could be hypothesized that the more personal situation in the face to face interview leads the less socially assured interviewers to adopt a task oriented role, while the more socially assured interviewers adopt a social role, which uses up more time. In the more businesslike situation of the telephone interview, this differential role taking does not take place.

Discussion

The extension of the hierarchical regression models discussed above to include instrument effects is straigthforward. *Instrument effects* are observation errors that are the effects of differences in the questionnaire used, such as the specific question wording or flow (Groves, 1989, p. 12). Two research designs are commonly used to investigate instrument effects. One strategy is to use a split sample (split ballot) design, which divides the respondent sample at random into subsamples, and presents different variations of the questionnaire to each of the subsamples. The type of question presented can then be coded as an explanatory variable at the respondent level, which can be analyzed just as the explanatory variable 'interview mode' in the example reported above. A different strategy would present all respondents with all different types of questions. In this case, the questions can be viewed as repeated measures nested within respondents, and a three level analysis can be used to analyze the effects of explanatory variables at the interviewer, respondent, and question level (for a discussion of multilevel analysis as a tool for analyzing repeated measurements see Goldstein, 1986, 1989; Bryk and Raudenbush, 1987; DiPrete and Grusky, 1990). The question whether a given variable should be included as an explanatory variable, or as a grouping variable defining a number of groups at a separate level, can be quite subtle. For instance, take the example analyzed in the previous section. Should the variable 'mode' be taken as an explanatory variable at the respondent level, or as a grouping variable defining three groups of respondents at a 'mode' level, which is between the interviewer and the respondent level? The answer is that applying random coefficient models implies the notion of a hierarchically structured population, and of taking a sample from the observation units at each of the appropriate levels. Thus, including 'mode' as a grouping variable defining groups at a separate 'mode' level, implies the conception of a large population of possible data collection modes, of which the three modes included here are just a sample. Including 'mode' at the respondent level as an explanatory variable with three categories (in this case recoded as contrast variables) implies that the three modes are *fixed;* they constitute all modes of interest in the study. Since it is difficult to conceptualize the three modes as a random sample from a large population of possible modes, 'mode' is included as an explanatory variable in the analyses above. The situation is, however, not always clear. If a large number of different question types is presented to the respondents, it may be appropriate to conceptualize this as sample from the population of all possible question types.

Ideally, in interviewer effect studies, respondents should be assigned to interviewers at random. In large scale studies this is seldom done, because it is expensive and difficult to organize. This makes it difficult to use such studies for methodological research, because interviewer and respondent characteristics are confounded. Multilevel analysis as outlined above offers some remedies to this situation. If the relevant respondent variables are known, they can be put in the hierarchical regression model to equalize interviewers by statistical means. If, after controlling respondent variables, interviewer variables explain significant variance, we may conclude that this reflects real interviewer effects. Conversely, if we are primarily interested in respondent effects, we can control for interviewer differences and investigate if adding respondent variables to a regression equation containing the interviewer variables explains additional variance. The procedure is similar to analysis of covariance, with one set of variables as the explanatory variables of interest, and the other set as the covariates to be adjusted for, but the assumptions of the multilevel model are much more realistic than those of analysis of covariance. The limitation of this approach is of course that it relies on statistical control instead of experimental control. The adequacy of theb statistical control depends on the assumption that all relevant covariates have been included, and have been correctly modeled. Without randomization, it is impossible to conclude that the influence of all confounding variables has been eliminated.

Finally, even researchers who are not interested in interviewer effects may find it useful to use hierarchical analysis models to include interviewer effects in the analysis, to *control* for potential interviewer effects. If there are non-zero interviewer effects, the value of the intra interviewer correlation r enters the equation that determines the appropriate standard error for many statistical tests (cf. Skinner, Holt and Smith, 1989). Even small values for the intra interviewer correlation may result in a large bias in the standard errors, because the *interviewer load* is also a factor. If the interviewer load is high, meaning that a small number of interviewers conducts a large number of interviews (not unusual in large scale telephone surveys), the combined result of a small intra interviewer correlation and a large interviewer load can be a serious statistical bias of the standard error (for examples see Groves, 1989). Statistical procedures that do not take this bias into account may result in spuriously significant statistical tests. The effect of the intra interviewer correlation is comparable to the bias that results from cluster sampling; survey statisticians generally model this by including it as a 'design effect' in the statistical model (Kish, 1987; Lee, Forthofer and Lorimor, 1989). Few substantive studies will actually incorporate random assignment of respondents to the interviewers, because interviewer effects have generally been shown to be low. The hierarchical linear regression model is an effective way to accommodate the design effect in such designs (Goldstein and Silver, 1989; Hox, De Leeuw and Kreft, 1991), while for the more complex covariance structure models the approach outlined by Muthén (Muthén, 1989) appears useful.

ontvangen 4 - 2 -1992
geaccepteerd 8 - 7 -1992

References

Aiken, L.S. & West, S.G. (1991). *Multiple Regression: Testing and Interpreting Interactions.* Newbury Park, CA: Sage.

Alwin, D.F. (1978). Making Errors in Surveys: An Overview. In D.F. Alwin (ed.), *Survey design and Analysis. Current Issues*. Beverly Hills: Sage.

Alwin, D.F. (1991). Research on Survey Quality. *Sociological Methods and Research,* 20, 3-29.

Alwin, D.A. & Krosnick, J.A. (1991). The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes."*Sociological Methods and Research,* 20, 139-180.

Bailar, B.A. (1987). Nonsampling Errors. *Journal of Official Statistics,* 3, 323-325.

Bailar, B., Bailey, L. & Stevens, L. (1977). Measures of Interviewer Bias and Variance. *Journal of Marketing Research,* 14, 337-43.

Berk, M.L., & Bernstein, A.M. (1980). Interviewer Characteristics and Performance on a Complex Health Survey. *Social Science Research,* 17, 239-51.

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., & Sudman, S. (eds.) (1991). *Measurement Errors in Surveys.* New York: Wiley.

Bock, R.D. (ed). (1989). *Multilevel analysis of educational data.* San Diego: Academic Press.

Bradburn, N. (1983). Response Effects. In: P.H. Rossi, J.D. Wright, & A.B. Anderson (eds). *Handbook of Survey Research.* San Diego: Academic Press.

Brown, J.J., & Gilmartin, B.G. (1969). Sociology Today: Lacunae, Emphases, and Surfeits. *American Sociologist* 4, 283-291.

Bryk, A.S. & Raudenbush, S.W.. (1987). Applying the Hierarchical Linear Model to Measurement of Change Problems. *Psychological Bulletin,* 101, 147-58.

Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical Linear Models.* Newbury Park, CA: Sage.

Collins, M. (1980). Interviewer Variability: A review of the Problem. *Journal of the Market Research Society,* 22, 75-95.

De Jong-Gierveld, J, & Van der Zouwen, J. (1987). *De vragenlijst in het sociaal onderzoek.* Deventer: Van Loghum Slaterus.

De Jong-Gierveld, J. (1987). Developing and testing a model of loneliness. *Journal of Personality and Social Psychology,* 53, 119-128.

De Leeuw, E.D. (1992). *Data Quality in Mail, Telephone, and Face to Face Surveys.* Amsterdam, Vrije Universiteit: Unpublished Doctoral Dissertation.

De Leeuw, J., & Kreft, Ita G.G. (1986). Random Coefficient Models for Multilevel Analysis. *Journal of Educational Statistics,* 11: 57-86.

DiPrete, T.A. & Grusky, D.B. (1990). The Multilevel Analysis of Trends with Repeated Cross-sectional Data. In: *Sociological Methodology 1990,* edited by C.C. Clogg. London: Blackwell.

Freeman, J. & Butler, E. (1976). Some Sources of Interviewer Variance in Surveys. *Public Opinion Quarterly,* 40, 79-91.

Goldstein, H. (1986). Efficient Statistical Modelling of Longitudinal Data. *Annals of Human Biology,* 13, 129-141.

Goldstein, H. (1987). *Multilevel Models in Educational and Social Research.* London: Griffin/New York: Oxford University Press.

Goldstein, H. (1989). Models for Multilevel Response Variables with an Application to Growth Curves. In R.D. Bock (ed). *Multilevel Analysis of Educational Data*. San Diego: Academic Press.

Goldstein, H. & R. Silver. (1989). Multilevel and Multivariate Models in Survey Analysis. In C.J. Skinner, D. Holt, & T.M.F. Smith (eds.) *Analysis of Complex Surveys*. New York: Wiley.

Groves, R.M., P.P Biemer, L.E. Lyberg, J.T. Massay, W.L. Nicholls II, & J. Waksberg (eds). (1988). *Telephone Survey Methodology.* New York: Wiley.

Hagenaars, J.A. & Heinen, T.G. (1982). Role-independent Interviewer Characteristics. In W. Dijkstra & J. van der Zouwen (eds.) *Response Behaviour in the Survey Interview*. London: Academic Press.

Hanson, R.H., & Marks, E.S. (1953). Influence of the Interviewer on the Accuracy of Survey Results. *Journal of the American Statistical Association,* 53, 635-55.

Hanson, M., Hurwitz, H.W. & Bershad, M. (1961). Measurement Errors in Censuses and Surveys. *Bulletin of the International Statistical Institute,* 38, 359-74.

Hill, D.H. (1991) Interviewer, Respondent, and Regional Office Effects on Response Variance: A Statistical Decomposition. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, & S. Sudman (eds.) (1991). *Measurement Errors in Surveys.* New York: Wiley.

Hox, J.J., Kreft, Ita G.G. & Hermkens, P.L.J. (1991). The analysis of factorial surveys. *Sociological Methods and Research,* 19, 493-510.

Hox, J.J., E.D. De Leeuw, & Ita G.G. Kreft. (1991). The Effect of Interviewer and respondent Characteristics on the Quality of Survey Data: A Multilevel Model. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, & S. Sudman (eds.) (1991). *Measurement Errors in Surveys.* New York: Wiley.

Jaccard, J., Turrisi, R., & Wan, C.K. (1990). *Interaction Effects in Multiple Regression.* Newbury Park, CA: Sage.

Kerlinger, F.N. (1986). *Foundations of Behavioral Research.* London: Holt, Rinehart & Winston.

Kirk, R.E. (1968). *Experimental Design. Procedures for the Behavioral Sciences.* Belmont, CA: Wadsworth.

Kish, L. (1965). Studies of Interviewer variance for Attitudinal Variables. *Journal of the American Statistical Association,* 57, 92-115.

Kish, L. (1987). *Statistical Design for Research.* New York: Wiley.

Kreft, Ita G.G. (1987). *Models and Methods for the Measurement of Schooleffects.* Amsterdam: Ph.D. Dissertation.

Kreft, Ita, G.G. & De Leeuw, E.D. (1988). The seesaw effect, a multilevel problem. *Quality and Quantity,* 22, 127-137.

Kreft, Ita G.G., De Leeuw, J, & Kim, K.-S. (1990). *Comparing Four Different Statistical Packages for Hierarchical Linear Regression. Genmod, Hlm, ML2, and Varcl.* Los Angeles: UCLA CSE Technical Report.

Lee, E.S., Forthofer, R.N. & Lorimor, R.J. (1989). *Analyzing Complex Survey Dara.* Newbury Park, CA: Sage.

Longford, N.T. (1986). *Statistical Modelling of Data from Hierarchical Structures Using Variance Component Analysis.* Lancaster: Center for Applied Statistics, University of Lancaster.

Mahalanobis, P.C. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society,* 109, 325-78.

Mason, W.M., Wong, G.Y., & Entwisle, B. (1984). Contextual Analysis through the Multilevel Linear Model. In S. Leinhardt (ed.) *Sociological Methodology 1983-84*. San Francisco: Jossey- Bass.

Muthén, B.O. (1989). Latent Variable Modeling in Heterogeneous Populations. *Psychometrika,* 54, 557-85.

O'Muircheartaigh, C.A. (1977). Response Errors. In C.A. O'Muircheartaigh & C. Payne (eds.) *The Analysis of Survey Data. Vol. II.* London: Wiley.

Pannekoek, J. (1988). Interviewer Variance in a Telephone Survey. *Journal of Official Statistics,* 4, 375-84.

Pannekoek, J. (1991). A Mixed Model for Analyzing Measurement Errors for Dichotomous Variables. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, & S. Sudman (eds.) (1991). *Measurement Errors in Surveys.* New York: Wiley.

Presser, S. (1984). The Use of Survey Data in Basic Research in the Social Sciences. In C.F. Turner & E. Martin (eds.) *Surveying Subjective Phenomena, Vol II.* New York: Russel Sage Foundation.

Raudenbush, S.W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics,* 13, 769-776.

Raudenbush, S.W., & Bryk, A.S. (1986). A Hierarchical Model for Studying School Effects. Sociology of Education, 59, 1-17.

Raudenbush, S.W. & Bryk, A.S. (1988). Methodological advances in studying effects of schools and classrooms on student learning. *Review of Research on Education.*

Stokes, L., & Yeh, M.-Y. (1987). Searching for Causes of Interviewer Effects in Telephone Surveys. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (eds.) *Telephone Survey methodology*. New York: Wiley.

Sudman, S. & Bradburn, N. (1974). *Response Effects in Surveys.* Chicago: Aldine.

Tate, R.L. and Wongbundit, Y. (1983). Random versus nonrandom coefficient models for multilevel analysis. *Journal of Educational Statistics,* 8, 103-120.

Van den Eeden, P. (1991). Interviewer effects and multilevel analysis: The responding environment theory reconsidered. Colloquiumpaper, Vrije Universiteit.

Van der Zouwen, J., & Dijkstra, W. (1989). *Sociaal-wetenschappelijk onderzoek met vragenlijsten.* Amsterdam: VU Uitgeverij.

Wiggins, R.D., Longford, N.T., & C.A. O''Muircheartaigh. (1990). A Variance Components Approach to Interviewer Effects. Manuscript submitted for publication