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 in the analysis of interviewer effects. The hierarchical regression model offers an elegant way of analyzing the simultaneous 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 rather than experimental control by modeling the interviewer effects conditional on the respondent effects.

Hierarchical Regression Models for Interviewer and Respondent Effects

J. J. HOX
University of Amsterdam

The survey interview is a major source of research data in many social science disciplines. Consequently, there is a large and growing literature on the quality of survey data, as instanced by two special issues of *Sociological Methods & Research* (November 1977 and August 1991; cf. Alwin 1978, 1991), and a series of recent collections on various aspects of survey methodology (Groves et al. 1988; Kasprzik, Duncan, Kalton, and Singh 1989; Biemer, Groves, Lyberg, Mathiowetz, and Sudman 1991).

Discussions of errors in surveys usually distinguish between various error sources. Groves (1989) differentiates two major sources of error: error of nonobservation and observation error. The errors of nonobservation arise because surveys generally fail to cover the complete population; in this category Groves places coverage error, nonresponse error, and sampling error. Observational errors are those errors that would arise even if the survey produced a complete enumeration of the population (Bailar 1987; Groves 1989). Groves (1989,

AUTHOR'S NOTE: *I would like to thank Bob Groves, Frans Kleintjes, 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.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 22, No. 3, February 1994 300-318

pp. 11-12) places observational errors in four categories: interviewer effects, respondent effects, instrument effects, and mode effects (effects of the specific mode of data collection used). This article focuses on the use of hierarchical linear regression models for research on interviewer and respondent effects; extensions that would include instrument and mode effects are mentioned only briefly.

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. 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 their answers to survey questions. The effects of interviewer characteristics are also generally reported as small (Bradburn 1983; Groves 1989). However, even small interviewer effects might have an important impact on the quality of survey data, especially when each interviewer deals with 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 of 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 requirement that interviewer and respondent characteristics are mutually 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, this can be accomplished by sampling both respondents and interviewers at random from a given population and assigning respondents at random to different interviewers. This method is designated as the method of *interpenetrating samples*. In this design, straightforward analysis of variance can be used to estimate the respondent and interviewer effects. More complex designs use reinterviewing by either the same or different interviewers, with reinterviews assigned at random, which leads to more complicated ANOVA models (O'Muircheartaigh 1977; Biemer and Stokes 1985; Groves 1989).

The adequacy of these designs depends to a great extend on the way the respondents are assigned to the interviewers. Much research on respondent and interviewer effects is based on a secondary analysis of survey data collected for a nonmethodological purpose. Because of the cost of an interpenetrating design, the respondents are generally not randomly assigned to the interviewers. For example, respondents might be randomly assigned to interviewers within specific geographic regions to avoid excessive traveling times. Thus respondents in an urban area will mostly be interviewed by urban interviewers, and respondents in a rural area mostly by rural interviewers. Respondents who live close to a university are more likely to be interviewed by students. The result of such nonrandom assignment is that respondent and interviewer characteristics become confounded; in other words, respondent and interviewer variables are correlated. As in other cases where it is not possible to use experimental control by random assignment, statistical control must be used instead to control for confounding variables.

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) or include the interviewers as a random factor (Dijkstra 1983).

For the analysis of interviewer effects, two approaches have become popular. The first is to consider the interviewer effect as a random effect that increases the variance of sample means (and other sample statistics). A random effect ANOVA model is used to estimate the interviewer variance component, and the intraclass correlation is used to estimate the interviewer effect (Hanson and Marks 1953; Kish 1965; Groves 1989).

The intraclass correlation ρ_i is defined as the population proportion of between group variance in a random effects ANOVA model (cf. Hays 1973, p. 535). Thus

$$\rho_I = \sigma_b^2/(\sigma_b^2 + \sigma_w^2), \tag{1}$$

where σ_b^2 is the between-interviewers variance and σ_w^2 the within-interviewers variance. Various estimators are available for ρ_I, depending on different assumptions and on the specific design used. For the simple one-way design that results if respondents are assigned at random to interviewers, the intraclass correlation is estimated as:

$$\rho_I = (MS_B - MS_W)/(MS_B - (n - 1)\, MS_W), \tag{2}$$

where MS_B is the between-interviewer mean square, MS_W is the within-interviewer mean square, and n is the common group size (number of respondents for each interviewer). Equation (2) is equivalent to the formula given by Kish for the intra-interviewer correlation ρ_{int}. It applies to a population with infinitely large groups. For finite groups of unequal size, appropriate formulas are given in the ANOVA literature (e.g., Searle 1971; see Groves 1989, pp. 360-73, for a discussion of the designs and estimates for ρ_I commonly used in interviewer survey research). Shrout and Fleiss (1979) show that most estimators for ρ_I are slightly biased, with the amount of bias varying with the sample size and the variation in group sizes. As we will see later, multilevel modeling offers another way to estimate the intraclass correlation.

In the approach outlined above, the effect of explanatory interviewer variables is investigated by splitting the interviewer sample, (e.g., in male and female interviewers) and estimating the intraclass correlation separately for each subsample. If the intraclass correlation disappears in the subsamples, the explanatory variable used to split the sample of interviewers is assumed to explain the interviewer effect. The general reasoning behind this approach is analogous to the classical elaboration method for contingency tables.

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 level of the dependent variable, which is the respondent level, and both interviewer and respondent variables are combined in one ANOVA or regression model. This violates several assumptions of the ordinary linear model, the most critical assumptions being that the error terms are uncorrelated and that the units of analysis are independently sampled.

Because each interviewer questions several respondents, unmeasured interviewer variation will result in correlated error terms between respondents. The assumption of independent sampling is violated because respondents interviewed by the same interviewer will have values for interviewer variables that are necessarily exactly equal. These violations affect both point estimates of regression parameters and their standard errors. The parameter estimates are generally unbiased but very inefficient, whereas the standard errors are underestimated, which results in a type I error rate that is much higher than the nominal alpha level (de Leeuw and Kreft 1986; Bryk and Raudenbush, 1992; see Hox, Kreft, and Hermkens 1991 for an empirical example). Sometimes even the signs of the regression coefficients might be misleading (Kreft and de Leeuw 1988).

Because the respondents are hierarchically nested within the interviewers, a hierarchical model must be used for the analysis of respondent and interviewer effects. Specialized hierarchical models have been proposed to analyze interviewer effects (Pannekoek 1991; Hill 1991). However, the well-known hierarchical linear regression model is a very useful general model that allows estimation of both the interviewer variance and the effects of explanatory variables measured at the interviewer and the respondent level. This model, also known as the random component model or the random coefficient model, has been described in several review articles (e.g., Mason, Wong, and Entwisle 1984; Raudenbush 1988) and books (Goldstein 1987; Bryk and Raudenbush 1992). It is used extensively in educational research, where pupils, classes, and schools present a natural hierarchical system; applications in the field of interviewer research are still rare (examples are Wiggins, Longford, and O'Muircheartaigh 1990; Hox, de Leeuw, and Kreft 1991).

THE HIERARCHICAL REGRESSION MODEL FOR INTERVIEWER EFFECTS

Suppose we wish to investigate the effect of certain interviewer characteristics on the quality of the data they collect. We also want to know if the interviewer effect is the same for different respondents. Let us assume that we select J interviewers from a large interviewer pool and that each interviewer interviews n_j respondents. The dependent variable Y is a measure that indicates some aspect of the data quality of the responses of a specific respondent, for instance item nonresponse or amount of social desirability. Let $\underline{Y}_{ij}$ be the score of respondent i ($i = 1, \ldots, n_j$), assigned to interviewer j ($j = 1, \ldots, J$). If we have no other information about the interviewers or the respondents, we can apply the following linear model (random variables are underscored):

$$\underline{Y}_{ij} = \underline{\beta}_{0j} + \underline{\varepsilon}_{ij}. \tag{3}$$

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

$$\underline{\beta}_{0j} = \gamma_{00} + \underline{\delta}_{0j}. \tag{4}$$

Substitution of (4) in (3) gives

$$\underline{Y}_{ij} = \gamma_{00} + \underline{\delta}_{0j} + \underline{\varepsilon}_{ij}, \tag{5}$$

where γ_{00} is the overall intercept, $\underline{\delta}_{0j}$ is an interviewer-level residual, which varies randomly between interviewers, and $\underline{\varepsilon}_{ij}$ is the respondent-level residual. It is usually assumed that the $\underline{\varepsilon}_{ij}$ are distributed with expectation zero and have a variance σ_j^2 that is equal for all interviewers, that is: all σ_j^2 are equal to σ_ε^2. The $\underline{\delta}_{0j}$ are assumed to be independent from the $\underline{\varepsilon}_{ij}$, and have a distribution with expectation zero and variance ω_{00}. In this model, the intraclass correlation for the interviewer effect is defined as

$$\rho_I = \omega_{00}/(\omega_{00} + \sigma_\varepsilon^2). \tag{6}$$

If interviewer effects exist, ω_{00} is greater than zero, which introduces a correlation between measurements collected by the same interviewer. The available software for multilevel analysis uses an empirical Bayes maximum likelihood procedure to produce parameter estimates. The estimation procedure iteratively weighs the parameter estimates, with weights depending on the group size and the intraclass correlation. To estimate the intraclass correlation in a multilevel analysis, we replace ω_{00} and σ_ε^2 in equation (6) by their estimates; the effects of finite and varying group sizes are incorporated in the estimation procedure.

Equations (3) to (5) contain no explanatory variables. Suppose 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 Y_{ij} is described by the following linear model:

$$\underline{Y}_{ij} = \underline{\beta}_{0j} + \underline{\beta}_{pj}X_{pij} + \underline{\varepsilon}_{ij}. \tag{7}$$

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

$$\underline{\beta}_{0j} = \gamma_{00} + \gamma_{0q}Z_{qj} + \underline{\delta}_{0j}, \tag{8}$$

$$\underline{\beta}_{pj} = \gamma_{p0} + \gamma_{pq}Z_{qj} + \underline{\delta}_{qj}. \tag{9}$$

Substituting (8) and (9) into (7) gives

$$\underline{Y}_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + \gamma_{pq}Z_{qj}X_{pij} + [\underline{\delta}_{pj}X_{pij} + \underline{\delta}_{0j} + \underline{\varepsilon}_{ij}]. \tag{10}$$

The hierarchical linear model is often presented as a hierarchical system of regressions such as in equations (8) and (9). When more than two levels are involved, it becomes increasingly difficult to visualize the succession of regressions on regressions. Most analysts would prefer to interpret the single equation version of the model (equation 10) directly as a complex regression equation involving multiple explanatory variables and interactions. In equation (10), the part $\gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + \gamma_{pq}Z_{qj}X_{pij}$ is the fixed part, that contains all fixed coefficients. The gammas can be interpreted as raw regression coefficients in a multiple regression. The product terms $Z_{qi}X_{pij}$ in equation (10) arise as a consequence of substituting (9) into (7); they are interaction terms that specify cross-level interactions between interviewer (Z_{qj}) and respondent (X_{pij}) characteristics. The part $\underline{\delta}_{pj}X_{pij}$ + $\underline{\delta}_{0j} + \underline{\varepsilon}_{ij}$, which is written in square brackets in equation (10), contains the random error structure; it is called the random part. The residuals

$\underline{\delta}_{0j}$ and $\underline{\delta}_{pj}X_{pij}$ in (10) are assumed to be independent from the $\underline{\varepsilon}_{ij}$ and to have a joint multivariate distribution with covariance matrix Ω.[1] It should be clear from equation (10) that although the fixed part looks much like an ordinary regression equation, the random part is more complicated, with a random term $\underline{\delta}_{0j}$ in addition to the usual $\underline{\varepsilon}_{ij}$, and for each lower level regression slope a distinct random error term $\underline{\delta}_{pj}$, which also involves the corresponding explanatory variable X_{pij}.

The estimation procedures and programs currently available all produce asymptotic standard errors for the gammas 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 in a manner analogous to the chi-square test for the difference between two nested LISREL models (Bryk and Raudenbush 1992).

Both the regression coefficients and the variance components are conditional on the explanatory variables in the model. This property is very useful if there is no complete orthogonalization of interviewer and respondent variables and statistical control of confounding variables is necessary. The analysis approach follows the general strategy of constructing a model starting at the lowest (respondent) level, and inspecting at each step the size and significance of the regression coefficients and variance components to decide which parameters 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 model fits significantly better than the smaller model. The example in the next section follows this approach.

MODEL SELECTION AND ANALYSIS: AN EXAMPLE

The example data stem from a controlled field experiment on mode effects (de Leeuw 1992). In this 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 pencil-and-paper 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. Due to 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. Because time measures generally have a skewed distribution, an inverse transformation is used, which transforms the variable, time, into the variable, speed. Thus the dependent variable Y_{ij} is the speed of the interview measured in number of questions completed per minute. The research problem is whether interviewers differ in the speed with which they complete an interview. In addition, we wish to analyze which interviewer and/or respondent characteristics influence the interviewing speed. The explanatory variables X_{pij} at the respondent level include two dummy variables indicating the three data collection methods: one contrast variable, tel (coded +1, −1), that compares the two telephone conditions to the face-to-face condition, and one contrast variable, cati, that compares the CATI-condition with the pencil-and-paper telephone condition (cati). The other respondent variables are respondent age (r-age) and loneliness (lonely), as measured by a multi-item scale. The explanatory variables Z_{qj} at the interviewer level are amount of previous 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.).

Because 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. The correlations between respondents and interviewers are generally low, indicating that the partial orthogonalization was successful. But, because the respondent and interviewer effects to be investigated are generally also small, it is safer to take these correlations into account in the analysis by modeling the interviewer effects conditional on the respondent variables.

TABLE 1: Correlations Between Respondent and Interviewer Variables

	Respondent Variable			
Interviewer Variable	*Tel*	*CATI*	*r-age*	*lonely*
Training	–.05	–.10	–.04	–.02
Experience	–.09	–.15	–.12	.05
I-age	–.10	–.09	–.04	.00
Pref.tel.	.10	.03	.08	–.03
Extro	.01	–.14	.03	.05
Friendly	.01	–.01	.00	–.01
Cons.	.03	.08	.02	–.02
Soc.ass.	–.01	–.07	.02	–.02
Term.	.06	–.15	.06	.00

NOTE: Tel = comparison of telephone to face-to-face condition; CATI = computer assisted telephone interviewing; r-age = respondent's age; lonely = loneliness; i-age = interviewer's age; pref.tel. = preference for telephone interviewing; extro = extroversion; friendly = friendly disposition; cons. = conscientiousness; soc.ass. = social assurance; term. = ability to terminate awkward situations.

The starting point for the model construction is the intercept-only model, which is a model with no explanatory variables. By removing all terms involving explanatory variables, the general equation (10) simplifies to equation (5), which is repeated here:

$$\underline{Y}_{ij} = \gamma_{00} + \underline{\delta}_{0j} + \underline{\varepsilon}_{ij}. \tag{5}$$

This model contains the fixed regression coefficient γ_{00} for the grand mean of $\underline{Y}_{ij}$ and the two variance estimates, σ_ε^2 for the residual variance at the respondent level and ω_{00} for the residual variance at the interviewer level.[2] In our example, γ_{00} is estimated as 3.19, indicating an overall interviewing speed of slightly more than three questions per minute. The respondent level variance σ_ε^2 is estimated as 0.68 and the interviewer level variance ω_{00} as 0.11; applying equation (6) produces an estimate for the intra-interviewer correlation ρ_I of 0.14. The parameter estimates for the model corresponding to equation (5) are summarized in Table 2, with the results from other analyses.

The next analysis step analyzes explanatory variables at the lowest (respondent) level as fixed variables; that is, without the corresponding variance components for the regression slopes. Equation (10) now simplifies to

$$\underline{Y}_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + [\underline{\delta}_{0j} + \underline{\varepsilon}_{ij}]. \tag{11}$$

TABLE 2: Parameter Estimates of Selected Models[a]

	Model Equation			
Variable	*(5)*	*(12)*	*(13)*	*(10)*
Fixed part				
Respondent level				
Intercept	3.19	3.73	1.77	1.43
Telephone		.30 (.04)	.30 (.04)	.30 (.04)
r-age		–.01 (.002)	–.01 (.002)	–.01 (.002)
r-lonely		–.04 (.01)	–.04 (.01)	–.04 (.01)
Interviewer level				
Interviewer training			.20 (.10)	.25 (.10)
Interviewer prefers telephone			.25 (.08)	.27 (.08)
Interviewer extrovert			.02 (.006)	.02 (.007)
Interviewer social assurance				–.01 (.007)
Interaction telephone social assurance				.01 (.005)
Random part				
σ_ε^2	.68[b]	.53	.52	.52
ω_{00} (intercept)	.11[b]	.08	.04	.03
ω_{11} (telephone slope)		.02[b]	.02	.01
Proportional decrease of variance				
Of σ_ε^2		23%	23%	23%
Of ω_{00} (intercept)		22%	65%	68%
Of ω_{11} (telephone slope)			0%	22%
Deviance	1294	1173	1161	1155

a. Standard errors between parentheses.
b. This variance estimate is the basis for calculating the proportional decrease of variance in subsequent models.

In the subsequent model, 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 equation (12):

$$\underline{Y}_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + [\underline{\delta}_{pj}X_{pij} + \underline{\delta}_{0j} + \underline{\varepsilon}_{ij}]. \tag{12}$$

In equation (12), each regression slope, γ_{p0}, has a corresponding random error term $\underline{\delta}_{pj}X_{pij}$. 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 might be simplified by dropping the CATI contrast altogether and assuming a random slope only for the telephone contrast.[4] In this model, the respondent level variance σ_ε^2 is 0.53, and the interviewer level intercept variance ω_{00} is 0.08. The parameter estimates of the model derived from equation (12) are summarized in Table 2.

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

$$\underline{Y}_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + [\underline{\delta}_{pj}X_{pij} + \underline{\delta}_{0j} + \underline{\varepsilon}_{ij}]. \quad (13)$$

In our example, only three of the nine interviewer variables are significant: interviewer training, preference for telephone, and extroversion. The parameter estimates for the model based on equation (13), omitting the nonsignificant effects, are again summarized in Table 2.

The between-interviewers variation of the regression slopes for the telephone condition can be modeled by including interactions between the telephone condition variable and explanatory variables at the interviewer level (cf. equations [7], [8], and [9]). This gives the model already formulated in equation (10), which is repeated here:

$$\underline{Y}_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + \gamma_{pq}Z_{qj}X_{pij} + [\underline{\delta}_{pj}X_{pij} + \underline{\delta}_{0j} + \underline{\varepsilon}_{ij}]. \quad (10)$$

The only significant interaction effect is the interaction of the telephone contrast with the interviewer variable, social assurance. Because the interpretation of interaction effects requires that the corresponding simple effects are also included in the model (Jaccard, Turrisi, and Wan 1990), the (nonsignificant) interviewer variable, social assurance, is again added to the model. To aid interpretation, social assurance is centered around its overall mean of 61.8, and the interaction term is computed using the centered variable. The parameter estimates for this final model are summarized in the last column of Table 2. In model (5), the residual variance at the respondent level, σ_ε^2, is 0.68, and the residual variance at the interviewer level, ω_{00}, is 0.11. In model (12), σ_ε^2 is estimated as 0.53; we can express this by saying that the respondent variables reduce the residual variance at the respondent level by 23%. Similarly, the interviewer variables can be said to reduce the residual variance at the interviewer level by 22%. The explanatory interviewer variables added in model (13) reduce the intercept variance by a further 43%. Adding the (nonsignificant) interviewer variable, social assurance, and its interaction with the telephone contrast (model (10) in the last column of Table 2) reduces the intercept variance by 3% and the variance of the regression slope for the telephone contrast by 22%. If we interpret these variance reductions as explained variance, a comparison of the explained variance across 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 (10) does not appear to explain much variance but, in fact, does explain a considerable proportion of the slope variance that appears in the previous model (13).[5]

Using the models' deviances for a chi-square test shows that, in all comparisons of consecutive models, the more complicated models have a significantly better fit. Most of the regression 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. 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 questions per minute; at an overall average of 3.19 questions per minute, this means that telephone interviews are 19% faster. However, because the variable, telephone condition, is involved in an interaction, we cannot interpret the interaction effect and the corresponding simple effects in isolation. 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. Because 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 average 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 et al. 1990). Because the

TABLE 3: Comparisons of Fit Between Successive Models

Model Equation	*Deviance*	*Difference With Previous Model*	df	p
(5)	1294.1	—	—	—
(12)	1172.6	121.5	5	.00
(13)	1161.4	11.2	3	.01
(10)	1155.1	6.3	2	.04

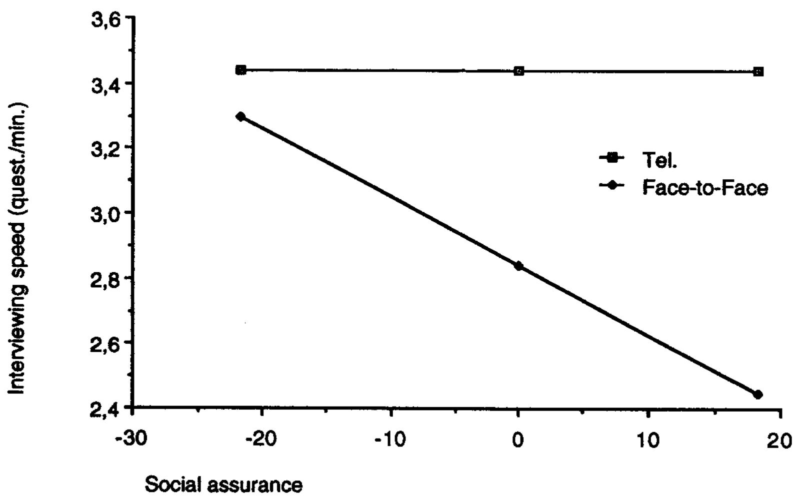

Figure 1: Social Assurance Slopes in Telephone and Face-to-Face Conditions

telephone contrast has only two values, Figure 1 plots the regression slope of the interviewer variable, social assurance, for both values of the telephone contrast.

Figure 1 shows that over the observed range of values for social assurance, telephone interviewing is faster than face-to-face interviewing. In the telephone interview, there is no relationship between social assurance and interviewing time, but in the face-to-face interview, interviewers with a higher social assurance tend to use more 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, whereas 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 assumption does not take place.

DISCUSSION

Extending the hierarchical regression models discussed above to include instrument effects is straightforward. Instrument effects are observation errors that are the effects of differences in the questionnaire used, such as the specific question wording or flow. 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 subsample. Here, question type is coded as an explanatory variable at the respondent level, to be analyzed just as the explanatory variable, interview mode, in the example given above. A different strategy would present all respondents with all different types of questions. Then, 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 Bryk and Raudenbush 1987; Goldstein 1987).

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 complicated to organize. This makes it difficult to use such studies for methodological research because, as a result, interviewer and respondent characteristics might be confounded. Multilevel analysis, as outlined above, offers some remedies for this situation. If the relevant respondent variables are known, they can be put in the regression model to equalize interviewers by statistical means. If, after controlling respondent variables, interviewer variables explain significant variance, we might conclude that this reflects real interviewer effects. Conversely, if we are primarily interested in respondent effects, we can control for interviewer differences and investigate whether 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 that it relies on statistical control instead of experimental control. It 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.

An ordinary least squares analysis of multilevel data produces unbiased but inefficient estimates and downwardly biased standard errors. Consequently, an OLS analysis could still be used to explore the data before a full hierarchical model is applied (the multilevel programs HLM and ML3 even offer this as an option). Furthermore, the criticism formulated in this article does not mean that the results of earlier analyses of interviewer and respondent data are totally invalid. The regression coefficients (or ANOVA effects) reported in the literature are still unbiased indicators of the population effects. However, the associated *p* values should be regarded with extreme caution; they are likely to be much too low.

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 intraclass correlation for the interviewer effect ρ_I is part of the equation that determines the appropriate standard error for many statistical tests (cf. Kish 1987). Even small values for this intraclass correlation might result in a large bias in the standard errors produced by ordinary statistical tests, 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 intrainterviewer correlation and a large interviewer load can still be a large bias of the standard error (for examples see Groves 1989), producing spuriously significant statistical test results. The effect of the intrainterviewer 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). Calculating this design effect can be complicated. The hierarchical linear regression model is an alternative way to accommodate the design effect in such designs, because the clustering is always a part of the model (Goldstein and Silver 1989; for an application, see Hox et al. 1991). For the more complex covariance structure models, the approach outlined by Muthén in this issue offers a simple procedure to correct for the bias introduced by clustering.

NOTES

1. If there are more than two levels, the covariance matrices Ω should be given a superscript indicating to which level they belong. As long as there is no risk of confusion, the simpler notation is used.

2. All calculations were done with Longford's program VARCL (Longford 1990).

3. All *p* values are derived from a two-sided normal approximation of Z = (coefficient/SE). Because the SEs are asymptotic and the normal distribution is 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.

5. Interpretation of the variance components in a model with random slopes is difficult because they are not generally invariant under admissible linear transformations of the explanatory variables. Bryk and Raudenbush (1992) present a more sophisticated way to calculate explained variance. The article by Snijders and Bosker (1994 [this issue]) offers a detailed criticism of reasoning-by-analogy to the ordinary multiple regression model. In my view, calculating an analogue of a multiple R^2 is, nevertheless, useful because it provides some indication of how well the model predicts various effects.

REFERENCES

Alwin, D. F. 1978. "Making Errors in Surveys: An Overview." Pp. 7-26 in *Survey Design and Analysis. Current Issues*, edited by D. F. Alwin. Beverly Hills, CA: Sage.

———. 1991. "Research on Survey Quality." *Sociological Methods & Research* 20:3-29.

Alwin, D. A. and J. A. Krosnick. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods & Research* 20:139-80.

Bailar, B. A. 1987. "Nonsampling Errors." *Journal of Official Statistics* 3:323-25.

Bailar, B. A., L. Bailey, and L. Stevens. 1977. "Measures of Interviewer Bias and Variance." *Journal of Marketing Research* 14:337-43.

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

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

Biemer, P. P. and S. L. Stokes. 1985. "Optimal Design of Interviewer Variance Experiments in Complex Surveys." *Journal of the American Statistical Association* 80:158-66.

Bradburn, N. 1983. "Response Effects." Pp. 289-328 in *Handbook of Survey Research*, edited by P. H. Rossi, J. D. Wright, and A. B. Anderson. San Diego: Academic Press.

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

———. 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 Leeuw, E. D. 1992. *Data Quality in Mail, Telephone, and Face to Face Surveys*. Amsterdam: TT-Publikaties.

de Leeuw, J. and Ita G. G. Kreft. 1986. "Random Coefficient Models for Multilevel Analysis." *Journal of Educational Statistics* 11:57-86.

Dijkstra, W. 1983. "How Interviewer Variance Can Bias the Results of Research on Interviewer Effects." *Quality and Quantity* 17:179-87.

Freeman, J. and E. Butler. 1976. "Some Sources of Interviewer Variance in Surveys." *Public Opinion Quarterly* 40:79-91.

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

Goldstein, H. and R. Silver. 1989. "Multilevel and Multivariate Models in Survey Analysis." Pp. 221-35 in *Analysis of Complex Surveys*, edited by C. J. Skinner, D. Holt, and T.M.F. Smith. New York: Wiley.

Groves, R. M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.

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

Hagenaars, J. A. and T. G. Heinen. 1982. "Role-Independent Interviewer Characteristics." Pp. 91-130 in *Response Behaviour in the Survey Interview*, edited by W. Dijkstra and J. van der Zouwen. London: Academic Press.

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

Hays, W. L. 1973. *Statistics for the Social Sciences*. New York: Holt, Rinehart & Winston.

Hill, D. H. 1991. "Interviewer, Respondent, and Regional Office Effects on Response Variance: A Statistical Decomposition." Pp. 463-83 in *Measurement Errors in Surveys*, edited by P. P. Biemer, R. A. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman. New York: Wiley.

Hox, J. J., E. D. de Leeuw, and Ita G. G. Kreft. 1991. "The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model." Pp. 439-61 in *Measurement Errors in Surveys*, edited by P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman. New York: Wiley.

Hox, J. J., Ita G. G. Kreft, and P.L.J. Hermkens. 1991. "The Analysis of Factorial Surveys." *Sociological Methods & Research* 19:493-510.

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

Kasprzyk, D., G. J. Duncan, G. Kalton, and M. P. Singh, eds. 1989. *Panel Surveys*. New York: Wiley.

Kish, L. 1965. "Studies of Interviewer Variance for Attitudinal Variables." *Journal of the American Statistical Association* 57:92-115.

———. 1987. *Statistical Design for Research.* New York: Wiley.

Kreft, Ita G. G. and E. D. de Leeuw. 1988. "The Seesaw Effect, a Multilevel Problem." *Quality and Quantity* 22:127-37.

Longford, N. T. 1990. *VARCL. Software for Variance Component Analysis of Data with Nested Random Effects (Maximum Likelihood).* Princeton, NJ: Educational Testing Service.

Mason, W. M., G. Y. Wong, and B. Entwisle. 1984. "Contextual Analysis Through the Multilevel Linear Model." Pp. 72-103 in *Sociological Methodology 1983-84,* edited by S. Leinhardt. San Francisco: Jossey-Bass.

O'Muircheartaigh, C. A. 1977. "Response Errors." Pp. 193-239 in *The Analysis of Survey Data. Vol. II,* edited by C. A. O'Muircheartaigh and C. Payne. London: Wiley.

Pannekoek, J. 1991. "A Mixed Model for Analyzing Measurement Errors for Dichotomous Variables." Pp. 517-31 in *Measurement Errors in Surveys,* edited by P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman. New York: Wiley.

Raudenbush, S. W. 1988. "Educational Applications of Hierarchical Linear Models: A Review." *Journal of Educational Statistic* 13:769-76.

Searle, S. R. 1971. *Linear Models.* New York: Wiley.

Shrout, P. E. and Fleiss, J. L. 1979. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin* 86:420-28.

Stokes, L. and M.-Y. Yeh. 1987. "Searching for Causes of Interviewer Effects in Telephone Surveys." Pp. 357-73 in *Telephone Survey Methodology,* edited by R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, and J. Waksberg. New York: Wiley.

Sudman, S. and N. Bradburn. 1974. *Response Effects in Surveys.* Chicago: Aldine.

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

Joop J. Hox is an associate professor of methods and statistics at the Department of Education, University of Amsterdam, the Netherlands. He is a staff member of the Dutch Interuniversity Graduate School of Psychometrics and Sociometrics and teaches courses on multivariate methods and modeling at the Graduate School of Education of the University of Amsterdam. His main research interests are multilevel modeling, survey research methodology, and problems of operationalization and measurement. He has edited several books and published in both Dutch and international journals.