# An Introduction to Hierarchical Linear Modeling for Marketing Researchers

*Barbara A. Wech and Anita L. Heck*

Organizations are hierarchical in nature. Specifically, individuals in the workplace are entrenched in work groups, which are entrenched in departments, which are entrenched in organizations, which are entrenched in the larger environment. Hence, hierarchical linear modeling (HLM) is a statistical technique available to researchers that is ideally suited for the study of cross-level issues. The purpose of this article is to provide marketing researchers with an overview and detailed description of HLM, as well as a practical illustration of HLM. Researchers in marketing, particularly those interested in the study of teams and cross-level research questions, should find HLM notably advantageous in their research due to the ability of HLM to simultaneously investigate relationships within a particular hierarchical level, as well as relationships between or across hierarchical levels.

Keywords: Hierarchical Linear Modeling, HLM, Marketing research, Groups, Teams

## Introduction

Organizations are a multi-level, hierarchical phenomena. That is, individual employees are typically embedded in teams, such as sales teams. Teams, in turn, are nested in departments. Departments are housed in facilities. Facilities are components of organizations that are entrenched in industries. Finally, industries are embedded in larger environments. If an organization is comprised of teams or groups, such groups are not only nested within departments, but may also include employees from different departments, such as in cross-functional product development teams. In large organizations, departments or facilities may be embedded in divisions that are, in turn, nested in organizations. As marketing research matures, it is clear that researchers cannot ignore the complex, cross-level nature of the organizations they study (Bryk & Raudenbush 1989).

The purpose of this article is to present an overview, detailed description, and simulated example of HLM, a statistical technique capable of analyzing hierarchical, cross-level data. Marketing researchers are actively examining cross-level situations (Arora & Allenby 1999; Sethi 2000; Sethi, Smith & Park 2001). HLM can aid researchers in more appropriately analyzing cross-level data and in opening an avenue of new research questions dealing with the hierarchical nature of organizations.

## Analyzing Cross-Level Data

There are essentially three statistical techniques that can be used to analyze cross-level data. The first technique is disaggregation. For example, a researcher interested in assessing the effect of the cohesiveness of new product development teams on customer satisfaction levels can disaggregate the data by assigning each lower level unit a score representing the higher level unit within which it is nested (Bryk & Raudenbush 1992). That is, all individuals would receive a score representing their team's level of cohesiveness. The analysis of the relationship between team cohesiveness and customer satisfaction would then be conducted at the individual level. However, there are limitations with this approach. First, it would be difficult to satisfy the independence of observations assumption underlying traditional

statistical methods. Individuals within a team would likely be similar due to their experiencing similar stimuli. Thus, by ignoring this dependence, standard errors may be underestimated leading to the probability that the researchers would be more likely to find significant results.

A second technique involves the aggregation of lower level units so that relationships at the aggregate level can be examined. For example, researchers might aggregate the individual level data (e.g., customer satisfaction) to the group level. This approach, however, ignores potentially meaningful variance at the lower, individual level (Hofmann 1997).

The third approach to analyzing cross-level data is hierarchical linear modeling or HLM. HLM overcomes the limitations found with the first two approaches. It simultaneously assesses relationships within a certain level and between or across hierarchical levels. HLM recognizes that individuals may not supply independent observations because employees within a particular group are likely to be more similar to each other than to individuals in other groups. Unlike the above two methods in which individual and group level residuals are not separately evaluated, HLM models both individual level and group level responses, acknowledging the partial interdependence of individuals within a team (Hofmann 1997). In addition, the technique assesses both lower and higher level variance in the dependent variable, while simultaneously preserving the correct level of analysis for the independent variable. This allows researchers to model individual and group level variance in the dependent variable while using individual independent variables at the individual level and group independent variables at the group level. HLM 5 (Raudenbush, Bryk, Cheong & Congdon 2001) is a user-friendly software package designed to test hierarchical linear models.

## HLM Analyses

To better facilitate a basic understanding of the HLM technique, a simulated example along with a detailed explanation, is presented. The simulated data was obtained from a banking sub-sample of a multiple-industry data set. Suppose researchers wish to study cross-functional product development teams. Specifically, they wish to assess the following hypotheses:

Hypothesis 1: Individual effort is positively related to customer satisfaction.

Hypothesis 2: Team cohesiveness of product development teams is positively related to customer satisfaction above and beyond individual effort.

Hypothesis 3: Team cohesiveness of product development teams moderates the effort-customer satisfaction relationship (i.e., it is suggested that the relationship between effort and customer satisfaction is stronger in situations where members of product development teams are more cohesive).

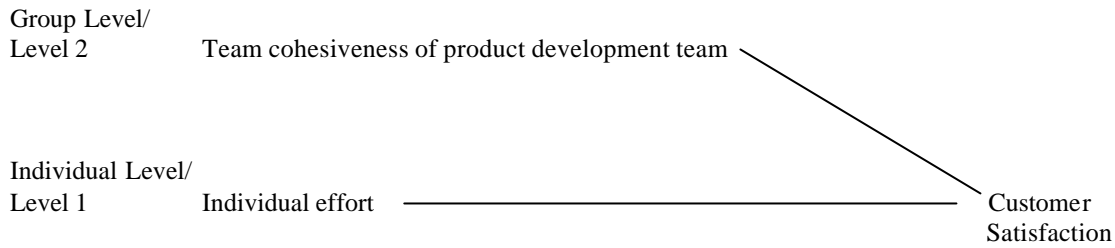## Figure 1. Pictorial representation of Hypotheses 1 and 2

Group Level/
Level 2        Team cohesiveness of product development team

Individual Level/
Level 1        Individual effort ——————————————— Customer
                                                      Satisfaction

## Figure 2. Pictorial representation of Hypothesis 3

Group Level/
Level 2                        Team cohesiveness of product development team

Individual Level/
Level 1        Individual effort ————————————— Customer Satisfaction
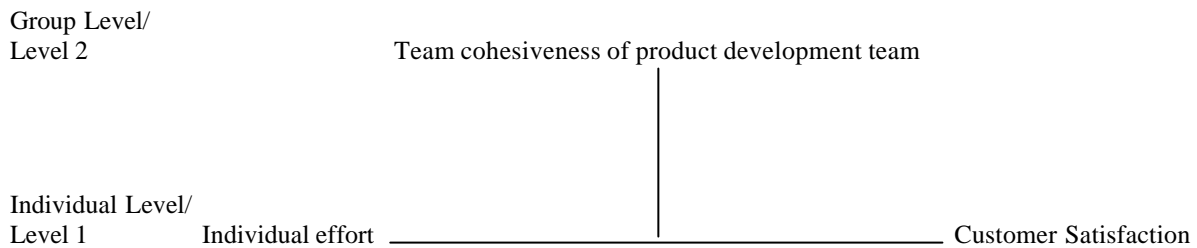
Next, a descriptive summary of the cross-level analysis or HLM process is offered. In general, HLM simultaneously assesses relationships both within and across (or between) levels. To model both within level and between level relationships, two models must be simultaneously estimated. HLM achieves this process by performing regressions of regressions (Hofmann, 1996). Conceptually a two-step procedure is used, wherein two models are estimated (Arnold, 1992). The first step involves estimating within-unit models. Separate regression models are estimated for each group to examine relationships among variables within each of the lower level units (i.e., groups) separately, producing intercept and slope parameters linking the individual level independent variable (e.g., individual effort) to the individual level outcome variable (e.g., customer satisfaction) for each group. The second step involves estimating relationships between levels, using the randomly varying intercepts and slope parameters from the within-unit models as outcome variables (i.e., dependent variables) and regressing them on a Level-2 or group level predictor (i.e., team cohesiveness). These regression equations represent the between-unit models (Arnold, 1992).

To assess the three hypotheses, a sequence of models is required: the null (one-way ANOVA), random-coefficient regression, intercepts-as-outcomes, and slopes-as-outcomes models.

## Null Model

Certain prerequisites must be satisfied to conduct cross-level analyses. First, there must be systematic within- and between-group variance in the dependent variable. This condition is necessary because the dependent variable (customer satisfaction) is hypothesized to be significantly related to both an individual level variable (individual effort) and group level variable (team cohesiveness). This is assessed in HLM using a one-way analysis of variance. Unless there is significant between-group variance in the dependent variable, team

cohesiveness would not have an opportunity to explain significant amounts of such variance. A null model with no independent variables at Level-1 or Level-2 estimates the following equations:

Level-1: $DV_{ij} = \beta_{0j} + r_{ij}$
Level-2: $\beta_{0j} = \gamma_{00} + U_{0j}$

where

$DV$ = customer satisfaction
$\beta_{0j}$ = mean customer satisfaction for group j
$\gamma_{00}$ = grand mean customer satisfaction
$r_{ij} = \sigma^2$ = within-group variance in customer satisfaction
$U_{0j} = \tau_{00}$ = between-group variance in customer satisfaction

The Level-1 equation does not include an independent variable, therefore the regression equation includes only an intercept estimate. The Level-2 model regresses each group's mean dependent variable onto a constant. In other words, $\beta_{0j}$ is regressed onto a unit vector which results in a $\gamma_{00}$ parameter equal to the grand mean[1] of the dependent variable (i.e., the mean of group means, $\beta_{0j}$).

The one-way ANOVA provides information regarding the amount of variance in the dependent variable that is within and between groups. Whereas there is no significance test for within group variance, the HLM program produces a chi-square statistic to test the significance of the between-group variance. A significant chi-square for the dependent variable shows that between-group variance is significantly different from zero, indicating that the intercept term varies across groups. A significant chi-square for customer satisfaction was obtained ($?^2 = 144.01$, $\underline{p} < .001$). In addition, using information estimated in the null model, an intraclass correlation coefficient (ICC) can be computed that represents the percent of the total variance in the dependent variable that is between groups (cf. Bryk & Raudenbush, 1992). The ICC indicates the amount of variance that could potentially be explained by the Level-2 predictor, team cohesiveness. The following equation was used: $ICC = t_{00} / (t_{00} + s^2)$ resulting in an ICC = .30.

---

[1] When using HLM, "centering" variables in various ways affects results (Hofmann & Gavin 1998; Kreft; de Leeuw & Aiken 1995). Because the intercept term ($\beta_{0j}$) is critical in the Level-2 analyses, it must be interpreted in a meaningful way. However, the intercept is difficult to interpret when the Level-1 independent variable (individual effort) is not in ratio scale. The intercept is the expected value of the dependent variable when individual effort is zero. Zero cannot, however, be a score that represents individual effort as the items are rated from 1 to 5. For this reason, grand mean centering is one method that can be employed to make intercepts more interpretable. Using this method the grand mean is subtracted from each individual's score on the predictor (individual effort) (Hofmann & Gavin 1998). Thus, the intercept $\beta_{0j}$ is rendered to equal the expected value of the dependent variable when individual effort is at the sample mean. For a detailed explanation of centering options, see Hofmann and Gavin (1998).

**Random-Coefficent Regression Model**

Next, researchers can assess whether there is significant between-group variance in the intercepts and slopes using a random-coefficient regression model. To find support for Hypothesis 2, there must be significant variance in intercepts across groups, and for Hypothesis 3 to be supported, there must be significant variance in slopes across groups. This model tests the significance of Hypothesis 1. The random-coefficient regression model estimates the following equations:

Level-1: $DV_{ij} = \beta_{0j} + \beta_{1j}(IV) + r_{ij}$
Level-2: $\beta_{0j} = \gamma_{00} + U_{0j}$
$\qquad \beta_{1j} = \gamma_{10} + U_{1j}$

where

DV = customer satisfaction
IV = individual effort
$\beta_{0j}$ = mean customer satisfaction for group j
$\beta_{1j}$ = grand mean individual effort for group j
$\gamma_{00}$ = mean of the intercepts across groups
$\gamma_{10}$ = mean of the slopes across groups (Hypothesis 1)
$r_{ij} = \sigma^2$ = Level-1 residual variance
$U_{0j} = \tau_{00}$ = variance in the intercepts
$U_{1j} = \tau_{11}$ = variance in the slopes

The Level-2 regression equation is equal to an intercept term and a residual since there are no Level-2 predictors of $\beta_{0j}$ or $\beta_{1j}$. The $\gamma_{00}$ and $\gamma_{10}$ parameters denote the Level-1 coefficients averaged across groups (i.e., they are the pooled $\beta_{0j}$ and $\beta_{1j}$ parameters). Since $\beta_{0j}$ and $\beta_{1j}$ are regressed onto constants, the variance of the Level-2 residual terms (i.e., $U_{0j}$ and $U_{1j}$) represents the between-group variance in the Level-1 parameters.

A t-test is used to test the significance of $\gamma_{10}$. This provides evidence of whether the pooled Level-1 slopes between the independent variable and the dependent variable differ from zero. Thus, this test assesses whether, on average, the relationship between the independent variable (individual effort) and the dependent variable (customer satisfaction) is significant or whether Hypothesis 1 is supported. Hypothesis 1 was supported by the simulated data ($\underline{t} = 3.687$, $\underline{p}$ , .001).

In order to test the cross level hypotheses, the HLM procedure states that there must be significant variance across groups in the Level-1 intercepts ($\beta_{0j}$). The intercept terms represent the between-group variance in the dependent variable after controlling for the independent variable. Chi-square tests for the estimates of the intercept ($\tau_{00}$) and slopes ($\tau_{11}$) are performed to confirm that the variance in the intercepts and slopes for the dependent variable across groups is significant. If there is not significant between group variance, then a group effect would not exist. The simulated data indicated that there was significant between group variance ($?^2 = 75.34$, $\underline{p} < .001$).

With information provided from the null and random-coefficients regression models, researchers can calculate an $R^2$ for the relationship between individual effort and customer

satisfaction. This $R^2$ is the percentage of the individual variance in customer satisfaction that is explained by individual effort. $R^2$ is calculated using the following equation: $s^2_{null}$ - $s^2_{random\ regression}$ / $s^2_{null}$ = .284-.278 / .284 = .02.

## Intercepts-as-Outcomes Model

After establishing that there is significant variance across groups in the Level-1 intercepts, then the cross level hypothesis (Hypothesis 2) can be directly tested. It is tested using the following equations:

Level-1: $DV_{ij} = \beta_{0j} + \beta_{1j}$ (individual effort) + $r_{ij}$
Level-2: $\beta_{0j} = \gamma_{00} + \gamma_{01}$ (team cohesiveness) + $U_{0j}$
$\qquad \beta_{1j} = \gamma_{10} + U_{1j}$

where

$\quad$ DV = customer satisfaction
$\quad$ IV = individual effort
$\quad \gamma_{00}$ = Level-2 intercept
$\quad \gamma_{01}$ = Level-2 slope (Hypothesis 2)
$\quad \gamma_{10}$ = mean (pooled) slopes
$\quad r_{ij}$ = Level-1 residual variance
$\quad U_{0j}$ = Residual intercept variance
$\quad U_{1j}$ = Variance in slopes

A $\underline{t}$-test is performed to test the significance of $\gamma_{01}$. Results show whether the group level variable (team cohesiveness) has a significant effect on the dependent variable (customer satisfaction). The results from the simulated data support Hypothesis 2 ($\underline{t}$ = 2.423, $\underline{p}$ < .02).

Using information from the HLM intercepts-as-outcomes analyses, an overall $\underline{R}^2$ for the respective Level-2 equations can be computed. Given the $\underline{R}^2$, one can determine how much of the independent variables' variance is between groups, and subsequently how much of the total variance, can be attributed to team cohesiveness. The $\underline{R}^2$ equation is: $\tau_{00\ random\ regression}$ - $\tau_{00\ intercepts\ as\ outcomes}$ / $\tau_{00\ random\ regression}$ = .099 - .088 / .099 = .11.

The intercepts-as-outcomes model also produces a chi-square test that indicates whether after including team cohesiveness, there still remains significant variance in the intercept term across groups that could be explained by additional group level variables. A significant condition must exist in order to test for a moderator. The simulated data indicated that a moderator could be tested ($?^2$ = 70.68, $\underline{p}$ < .001).

## Slopes-as-Outcomes Model

Finally, after establishing that significant group variance in the slopes was present in the random coefficient regression model, the researcher can then examine whether the variance in the slope across groups is significantly related to the group level independent variable (team cohesiveness). This is a direct test for the cross-level moderator (Hypothesis 3). The slopes-as-outcomes model is employed for this step as follows:

Level-1: $DV_{ij} = \beta_{0j} + \beta_{1j} (IV) + r_{ij}$
Level-2: $\beta_{0j} = \gamma_{00} + \gamma_{01}$ (team cohesiveness) $+ U_{0j}$
          $\beta_{1j} = \gamma_{10} + \gamma_{11}$ (team cohesiveness) $+ U_{1j}$

where

DV = customer satisfaction
IV = individual effort
$\gamma_{00}$ = Level-2 intercept
$\gamma_{01}$ = Level-2 slope (Hypothesis 2)
$\gamma_{10}$ = Level-2 intercept
$\gamma_{11}$ = Level-2 slope (Hypothesis 3)
$r_{ij}$ = Level-1 residual variance
$U_{0j}$ = Residual intercept variance
$U_{1j}$ = Residual slope variance

The t-test associated with $\gamma_{11}$ provides the test of Hypothesis 3. Hypothesis 3 was supported by the simulated data ($t = 2.118$, $p < .05$). Information provided in the HLM output for the intercepts-as-outcomes and slopes-as-outcomes models can be used to calculate an $R^2$ for the moderator, team cohesiveness. The $R^2$ indicates the percentage of variance in the relationship between individual effort and customer satisfaction that is accounted for by team cohesiveness. The $R^2$ equation is: $\tau_{11 \text{ intercepts as outcomes}} - \tau_{00 \text{ slopes as outcomes}} / \tau_{00 \text{ intercepts as outcomes}} = .004 - .003 / .004 = .25$.

## Conclusion

HLM is an invaluable technique available for assessing multi-level relationships in marketing research. It is our hope that this overview of the technique will encourage researchers to use this statistical technique as it offers many advantages in interpreting data within and across groups than other typically used techniques, such as ordinary least squares. See Bryk and Raudenbush (1992) for a detailed explanation for more complex hierarchical models that can be analyzed by the HLM software.

## References

Arora N & Allenby GM (1999). Measuring the influence of individual preference structures in group decision making. *Journal of Marketing Research, 36*, 476-487.

Bryk AS & Raudenbush SW (1989). Methodology for cross-level organizational research. *Psychological Bulletin, 101,* 147-158.

Bryk AS & Raudenbush SW (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Hofmann DA (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23,* 723-744.

Hofmann DA & Gavin MB (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 23*, 723-744.

Kreft I. GG ; de Leeuw J & Aiken LS (1995). The effect of different forms of centering in Hierarchical Linear Models. *Multivariate Behavioral Research, 30*, 1-21.

Raudenbush SW; Bryk AS; Cheong YF & Congdon RJ (2001). *HLM 5: Hierarchical linear and nonlinear modeling* (2nd edition). Chicago: Scientific Software.

Sethi R (2000). New product quality and product development teams. *Journal of Marketing, 64*, 1-14.

Sethi R ; Smith DC & Park CW (2001). Cross-functional product development teams, creativity, and the innovativeness of new consumer products. *Journal of Marketing Research, 38*, 73-85.

**Barbara A. Wech is an Assistant Progfessor in the Department of Management, Marketing, and Industrial Distribution at the University of Alabama-Birmingham, and Anita L. Heck is an Assistant Professor in the Department of Business, Troy State University.**