18. *Writing about Hierarchical Linear Models*

SOLUTIONS

1. a. i. $BMI_{ia} = \alpha_{0a} + \alpha_{1a}AGE + \alpha_{2a}MALE + \alpha_{3a}INC\ HIGH\text{-}SCHOOL + \alpha_{4a}MARRIED + \alpha_{5a}SMOKER + \alpha_{6a}SEDENTARY + \varepsilon_{ia}$, where the subscript *i* is used to index individuals and *a* to index neighborhoods.
   ii. $\alpha_{0a} = \eta_{00} + \eta_{01}LOW\ DWELLING\ VALUE + \eta_{02}MID\ DWELLING\ VALUE + \gamma_{0a}$ The title of table 18A indicates that a random intercept model was estimated. Only the intercept in the level-1 equation is permitted to vary randomly as a function of area-level characteristics. The slopes are treated as fixed effects.
   b. Even after accounting for a variety of individual-level characteristics (e.g. age, sex, education level), there is still statistically significant random variation in the BMI between individuals. There remains random variation across areas/neighborhoods in BMIs after accounting for the average dwelling value of areas.
   c. The intraclass correlation can be calculated level-2 variance / (level-1 variance + level-2 variance). Substituting values from table 8A, we obtain = 0.90 / (19.13 + 0.90) = 4.45%. Approximately 4.5% of the total variation in BMI across individuals can be explained by differences in the neighborhoods in which they live.
   d. i. Regular smokers were estimated to have a BMI that is, on average, 0.82 kg/m^2 lower than a nonsmoker, adjusting for other individual and neighborhood characteristics ($p < 0.05$).
   ii. Average BMI was positively associated with extent of neighborhood disadvantage. Individuals living in the mid-dwelling-value areas had 1.28 kg/m^2 higher BMI, and those in the low-dwelling-value areas 1.93 kg/m^2 higher, when each was compared with those living in the high-dwelling-value areas (both $p < 0.05$).

3. a. Level-1 = student; Level-2 = teacher; Level-3 = school
   b. "Hierarchical data structures are present in education settings where students are nested within a teacher and teachers are nested within a school. The nesting form of the data structure generates a hierarchical linear model (HLM). In other words, models at different levels can be built based on a specific number of lower level units nested within upper level units, eventually forming a HLM design. . . . Thus, in such situations, students' gain scores

in mathematics from one year to the next can be predicted based on characteristics not only of the student, but also of the teacher (e.g., teacher qualifications and experience) and of the school (e.g., poverty)." (Adapted from Subedi et al. [2011], p. 4.)

c. The positive effect on mathematics gains scores from having a teacher with content certification in mathematics may be enhanced among students of lower student socioeconomic status.

5. a. When other student, teacher, and school-level characteristics were taken into account, each one-point increase in baseline (pretest) math scores was associated with approximately a quarter of a point increase in math gain scores between rounds ($p < 0.01$).

b. Each additional year of teaching experience was associated with a four-tenths of a point increase in students' math gain scores ($p < 0.01$).

c. Student SES and teacher mathematics content certification interacted in their effect on students' math gain scores: Low SES students (as identified by participation in the free lunch program) had lower mean math gains than high SES students. Moreover, although having a math content certified teacher was associated with higher math gain scores regardless of student SES level, the effect was amplified for low SES students. Having a math certified teacher was associated with a mean math gain score of 2.88 points among low SES students, versus a mean increase of 1.97 points among higher SES students ($p < 0.01$).

7. The unconditional models estimate the amount of total variation in students' mathematics gains scores that is found between teachers (level-2) and between schools (level-3). Statistically significant level-2 and level-3 variance components in the unconditional models would indicate that it is important to consider teacher-level and school-level factors, and that an HLM is appropriate. The conditional models, which add student, teacher, and school characteristics, are estimated in order to assess the degree to which the random variation across levels can be accounted for by the included factors.

9. In the study by Pan et al. (2005) the level-1 unit of analysis was a time point at which child's vocabulary was measured, and level-2 unit of analysis was the mother/child dyad. In the longitudinal study, time points were nested within children.