Center for Educational Research

Max-Planck-Institut für Bildungsforschung
Max Planck Institute for Human Development

Methodological and Statistical Aspects of Assessing Educational Environments

Oliver Lüdtke

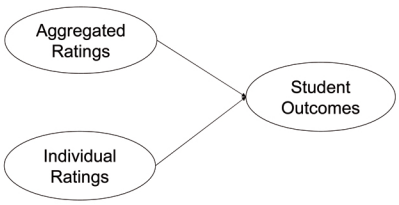
Theoretical Background

A key assumption of most research in the educational context is that cognitive, motivational, emotional, and behavioral student outcomes are substantially shaped by features of the learning environment.
In educational research, characteristics of the learning environment are generally assessed by asking students to evaluate features of their lessons. The student ratings produced by this simple and efficient research strategy can be analyzed from two different perspectives. At the ***individual level*** they represent the individual student's perception of the learning environment. Scores aggregated to the ***classroom level*** reflect perceptions of the shared learning environment, corrected for individual idiosyncrasies

Research Questions

1. What is primary unit of analysis: Student or class?
2. Assessing the psychometric properties of aggregated ratings
3. Centering of individual ratings in multilevel analyses
4. How to take into account the unreliability of group averaged scores?

Multilevel Analysis



1. Level of Analysis

Which level of analysis should be used when students' ratings of their learning environment are analyzed: the individual level, the class level or both?

In educational studies students are usually nested within environments (e.g., students within teachers).

Nested Design

	Teacher 1	Teacher 2	Teacher 3
Student 1	x		
Student 2	x		
Student 3		x	
Student 4		x	
Student 5			x
Student 6			x

Crossed Design

	Teacher 1	Teacher 2	Teacher 3
Student 1	x	x	x
Student 2	x	x	x
Student 3	x	x	x
Student 4	x	x	x
Student 5	x	x	x
Student 6	x	x	x

In a fully crossed design the variance in student ratings can be decomposed into 4 components:

$$\sigma^2(X_{rt}) = \sigma^2(t) + \sigma^2(r) + \sigma^2(d) + \sigma^2(\varepsilon)$$

- $\sigma^2(t)$: Target variance: (e.g., differences among teachers)
- $\sigma^2(r)$: Rater variance: (e.g., students' rater tendency)
- $\sigma^2(d)$: Dyad variance: (e.g., specific interaction student teacher)
- $\sigma^2(\varepsilon)$: Measurement error variance

In a nested design dyad variance and rater variance are completely confounded. Hence, individual ratings reflect the specific interaction of the student and the environment as well as his/her overall tendency to evaluate environments as positive or negative. In contrast, students averaged ratings represent differences between environments (target variance), because the other variance components are averaged out by the aggregation process.
→ class/school level should be primary unit of analysis!

2. Psychometric Quality of Student Ratings

Two aspects need to be distinguished when psychometric properties of aggregated student ratings are assessed:

Reliability

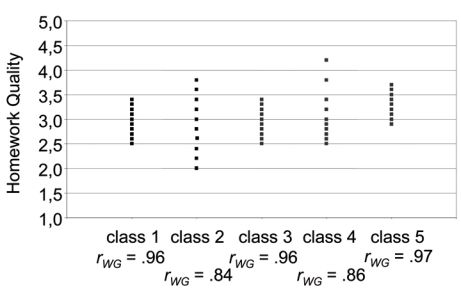
Do the student ratings differentiate among classes?
Proportion of the total that is located between classes (ICC_1 and ICC_2; Bliese, 2000).

Within-group agreement

Do students within a class show similar ratings?

Variation of student ratings within a single class (r_{WG}; James et al., 1984).

Example:
Students rating the quality of their homework assignments



ICC_1 = .08

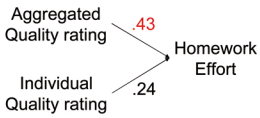
ICC_2 = .47

3. Centering Student Ratings in Multilevel Modeling

In multilevel analyses, student ratings are usually included at the individual and class level. Basically, two different options are discussed for centering student ratings at the individual level.

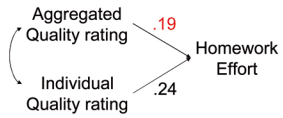
Centering within cluster (CWC): $X_{ij} - \overline{X}_j$

Differences across classes in individual ratings will be removed.
Not controlling for interindividual differences in individual ratings.



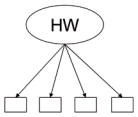
Centering at the grand-mean (CGM): $X_{ij} - \overline{X}$

Differences across classes in individual classes are not removed.
Controlling for interindividual differences in individual ratings.

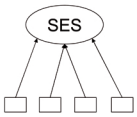


There is no consensus among researchers whether CWC (e.g., Trautwein et al., 2006; Urdan, 2004; Wendorf & Alexander, 2004) or CGM (e.g., Church et al., 2001; Kaplan et al., 2002; Karabenick, 2004) should be used.
To address this question, it is helpful to distinguish two different processes of aggregating individual scores:

Reflective aggregation



Formative aggregation



In research on learning environments, student ratings are used to assess a generic group-level construct by utilizing a reflective aggregation process. In that case CWC is strongly recommend. The rationale of the classical contextual analysis model (i.e., to adjust for interindividual differences based on CGM) does not seem appropriate for student ratings of the learning environment. Controlling for differences in individual ratings would partial out a central component of the between-group relationships.

4. Controlling for Unreliable Group Measure

Aggregated student rating can show low reliability, depending on the number of individuals in each of the groups and the intraclass correlation. Hence, multilevel analysis based on aggregated scores can result in substantially biased estimates of contextual effects.
A new ***multilevel latent covariate (MLC) approach*** is introduced that corrects for unreliability at the group level (implemented in M*plus*).

Comparing the MLC Approach with the Traditional Manifest Approach: Effects of Homework Quality on Homework Effort

HW Effort student	Latent Appoach		Manifest Aproach	
	b	S.E.	*b*	S.E.
Fixed Effect				
$\hat{\gamma}_{00}$ Intercept	0.00	0.03	0.04	0.03
$\hat{\gamma}_{01}$ HW Quality (class average)	0.71	0.07	0.43	0.05
$\hat{\gamma}_{10}$ HW Quality (student)	0.24	0.02	0.24	0.02
	Var Comp		Var Comp	
Random Effect				
$Var(u_{0j})$	0.02		0.03	
$Var(r_{ij})$	0.60		0.60	

Note. *N*(Level1) = 1501. *N*(Level 2) = 93. Average Cluster Size = 16.14.

Application of the manifest approach can result in underestimating the effect of learning environments.

Conclusion

First, researchers interested in identifying effects of learning environments should not use individual student ratings, but focus on ratings aggregated at the relevant level (e.g., class or school level).
Second, it is imperative to assess the reliability and within-group agreement of aggregated student ratings before relating these perceptions to outcome variables.
Third, when individual and aggregated perceptions are simultaneously included as predictor variables, group-mean centering is typically the appropriate option.
Fourth, the new multi-level latent covariate approach allows researchers to control for the unreliability of group-levels variables that are based on aggregated students ratings.

Reference

Lüdtke, O. (2007). *Methodological and statistical aspects of assessing educational contexts: Background and applications* [Methodische und statistische Aspekte der Erfassung von schulischen Kontextmerkmalen: Grundlagen und Anwendungen]. Unveröffentlichte Habilitationsschrift, FU Berlin.