

How are Variables at Different Levels Related?

Sequences of individual/group processes and their outcomes

Ming Ming Chiu
Purdue University

Dan Suthers asked me this question at an Alpine Workshop in 2009.

I appreciate the research assistance of
Choi Yik Ting



Research Questions

- Among groups of students working on an algebra problem, do groups with more sequences of *correct evaluations* → *correct, new idea (micro-creativity)* have better group solutions?
- How are sequences of talk / online messages related to group outcomes?
- More generally, how are variables at different levels related to one another?

Data at Different Levels

- Country
- Province
- School
- Classroom
- Group
- Individual
- Time periods
- Turns of talk, messages, codes, etc.

Two Approaches

Choose unit of analysis

- **Higher-level** variables
 - Group level
 - e.g., group solution score
- **Lower-level** variables
 - Turn of talk level
 - e.g., micro-creativity

Advantages of Higher-level Analyses

- Simple to implement via regressions
 - OLS, Logit/Probit, ordered Logit/Probit
 - For continuous, binary, ordered outcomes
- Enter lower-level variables as percentages
 - % of $A \rightarrow B$ sequences
 - $\frac{\# \text{ (Correct evaluations} \rightarrow \text{Micro-creativity)}}{\text{Total sequences (2 turns long)}}$
- Control for other variables (e.g., math grade)

Videotape Group Problem Solving

- 80 average ability, 9th grade students in US city
 - Worked in **20** groups of 4
 - **3,234** turns of talk
- Introduce 2 variable algebraic equations
 - 1st day of group work
 - No group work preparation
 - Work on problem for 30 minutes
- Videotape & Transcripts
 - Two RAs coded each turn of talk
 - High inter-rater reliability: Krippendorff's ✓



Algebra Problem: Texting Plans

Under the *Universal Texting* plan, each text message costs \$.10. *Budget Texting* costs \$.01 per text message, but charges a monthly fee, \$18.

- 1) How many text messages do you send each month?
- 2) Which company costs less for you?
- 3) How many texts should you send for the *Universal* plan and the *Budget* plan to cost the same?



Higher-level Analysis

Outcome

- Solution Score (0-3)

Explanatory variables

- Group's average mathematics grade
- % of Correct evaluation → Micro-creativity sequences

Higher-level Analysis

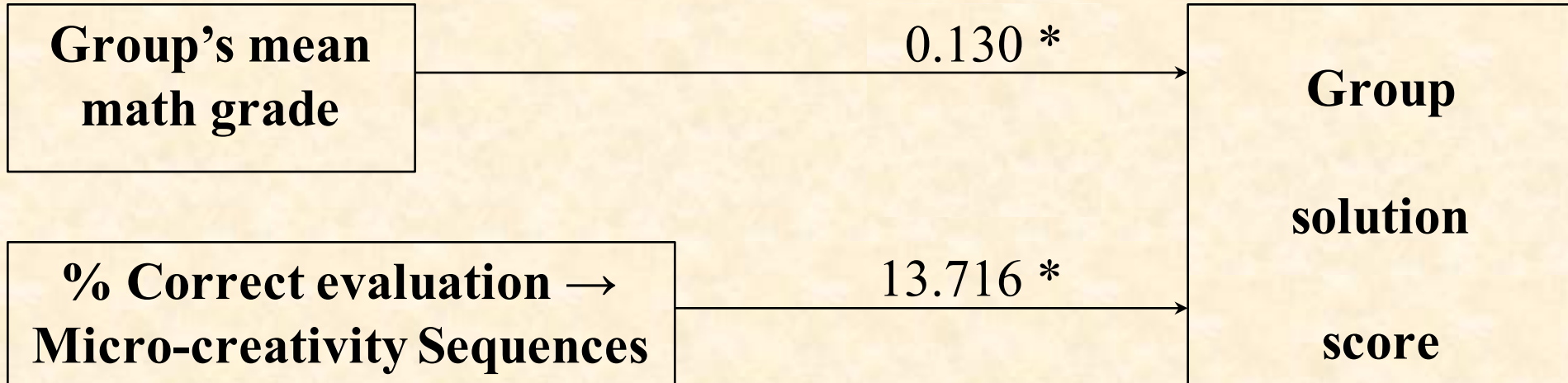
**Group's mean
math grade**

0.130 *

**Group
solution
score**

**% Correct evaluation →
Micro-creativity Sequences**

13.716 *



Disadvantages of Higher-level Analyses

- Coarser analysis
- Hard to model lower-level variables
 - especially their interactions
- Fewer observations → Less precise
- Requires large data set
 - Medium effect (0.3) requires
 - 84 groups for 80% statistical power
 - 112 groups for 90% statistical power

Higher-level Analysis

Outcome

- Solution Score (0-3)

Explanatory variables

- Group's average mathematics grade
- % of Correct evaluation → Micro-creativity sequences
- **% of Correct evaluations**
- **% Micro-creativity**

Higher-level Analysis

**Group's mean
math grade**

0.074

**% Correct evaluation →
Micro-creativity sequences**

7.308

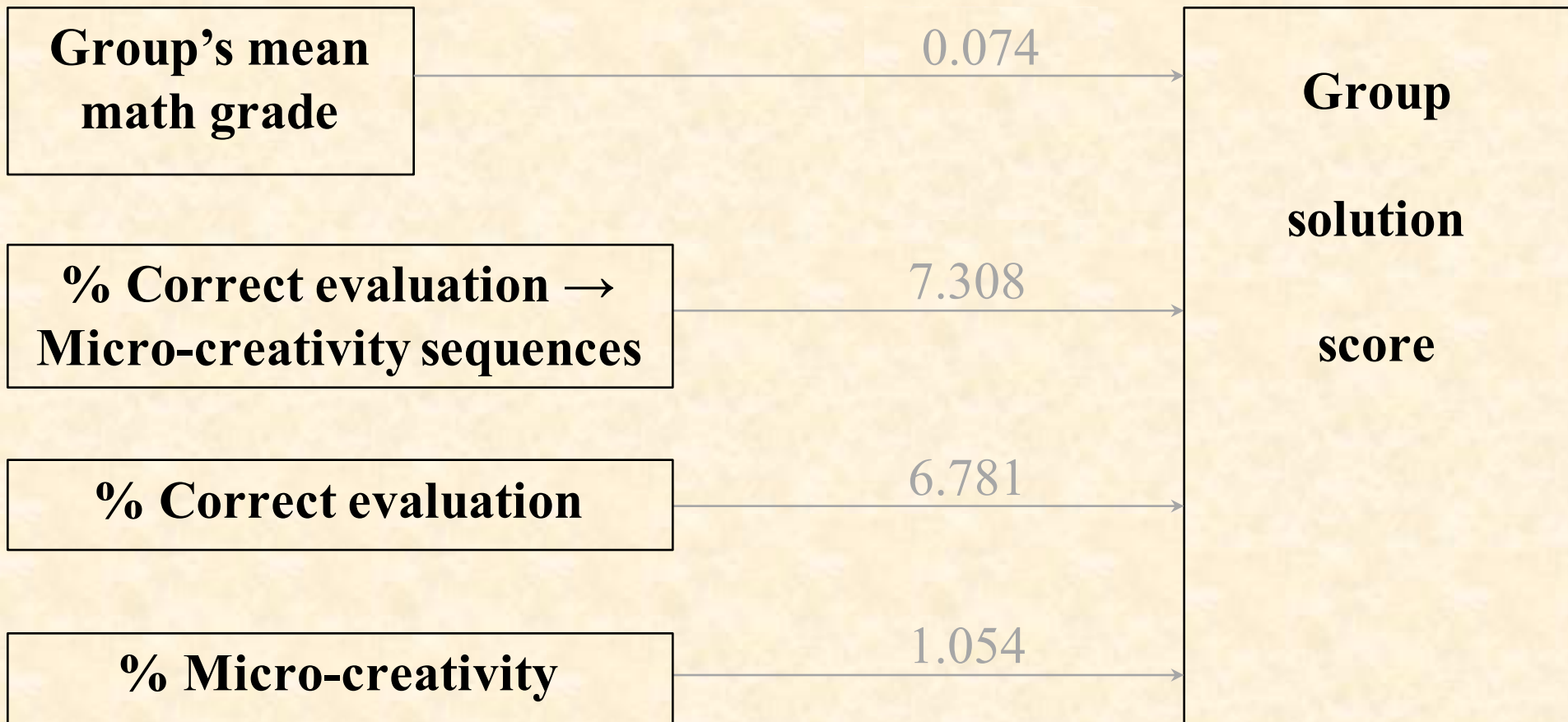
% Correct evaluation

6.781

% Micro-creativity

1.054

**Group
solution
score**



Advantages of Lower-level Analyses

- Finer-grained analyses
- Directly model lower-level variables
- More observations → More precise
- 3,234 turns of talk yield $> 99\%$ statistical power for even a small effect (0.1)
- Can test complex explanatory models

Disadvantages of Lower-level Analyses

- Harder to implement
 - Multilevel analysis
 - To also model time,
 - Time-series analysis
 - Statistical Discourse Analysis (SDA)

Statistical Discourse Analysis

Analytical Difficulty

🎬 Differences across topics

🎬 Time periods differ ($T_2 \rightarrow T_4$)


🎬 Serial correlation ($t_8 \rightarrow t_9$)

🎬 Parallel talk ($\rightarrow \rightarrow \Rightarrow \Rightarrow$)

🎬 Discrete outcomes (Yes / No)

🎬 Multiple outcomes (Y_1, Y_2)

🎬 Infrequent outcomes (00010)

🎬 People & Groups differ   

🎬 Mediation effects ($X \rightarrow M \rightarrow Y$)

🎬 False positives ($++ \oplus ++$)

🎬 Missing data (101?001?10)

🎬 Robustness

Strategy

🎬 Multilevel analysis

🎬 Breakpoint analysis & Multilevel analysis

🎬 χ^2 index of Q-statistics; Model with lag variables

🎬 Store path: ID prior turn, Vector Auto-Regression

🎬 Logit / Probit

🎬 Multivariate outcome models

🎬 Logit bias estimator

🎬 Multilevel analysis

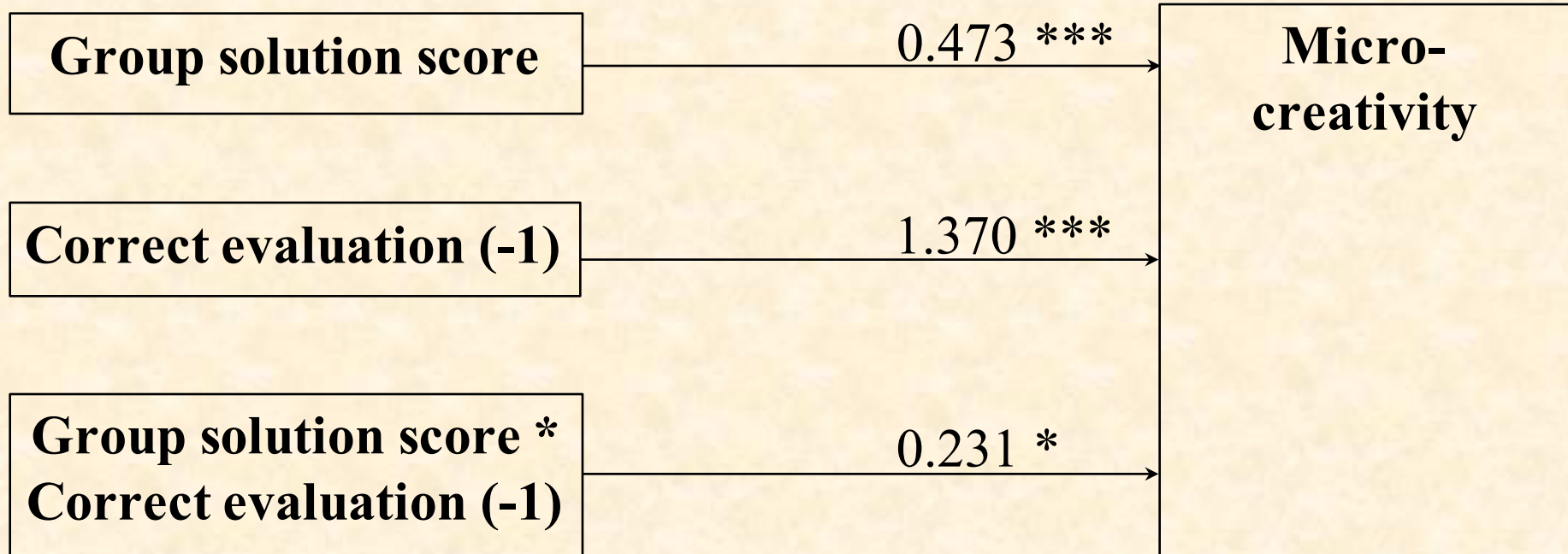
🎬 Multilevel mediation tests

🎬 2-stage linear step-up procedure

🎬 Markov Chain Monte Carlo multiple imputation

🎬 Separate outcome models;
Data subsets & unimputed data

Lower-level Analysis



Disadvantages of Lower-level Analyses

- Hard to model control variables for the higher-level outcome *solution*
- To do so, we identify the portion of the outcome *solution* that is explained by control variables, and we remove it

Remove Control Variables' Effects on *Solution*

- Higher-level regression on the outcome *solution* with all control variables
- Store unexplained part of solution score (*solution residual*)

Unexplained Residual after Controlling for Other Variables

**Group's Mean
Math Grade**

0.253 **

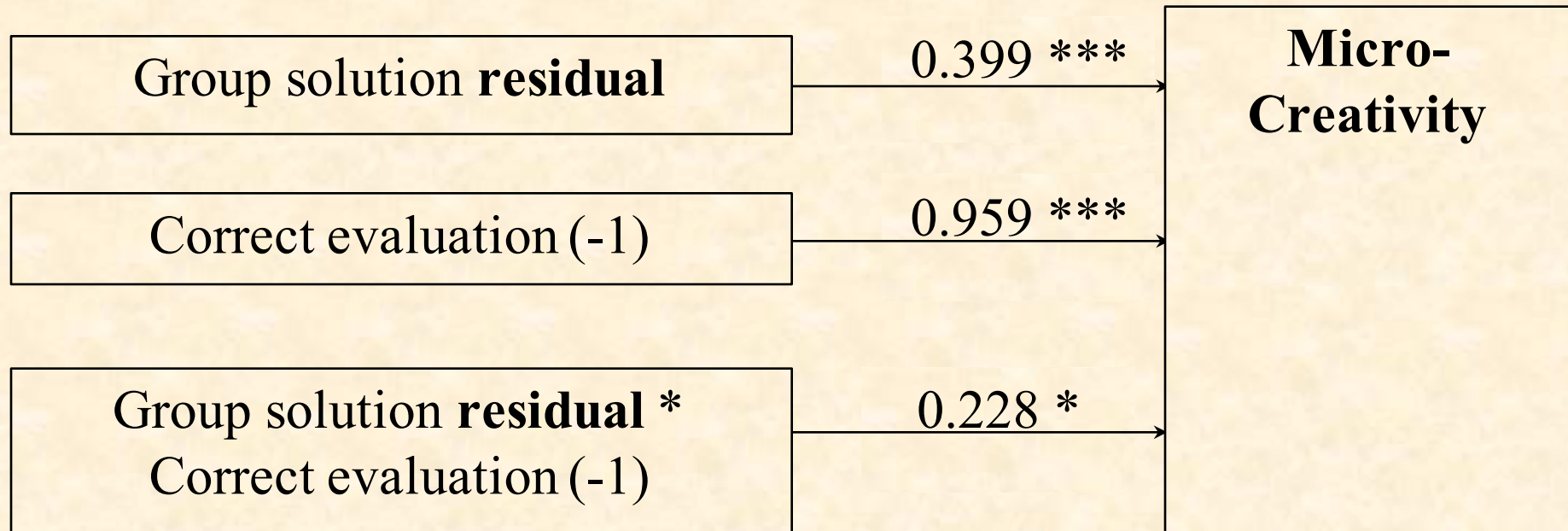
**Group
Solution
Score**

Store unexplained component in
Solution residual

Remove Control Variables' Effects on *Solution*

- Higher-level regression on the outcome *solution* with all control variables
- Store unexplained part of solution score (*solution_residual*)
- **Lower-level regression on micro-creativity**
 - **Replace all instances of *solution* with *solution_residual***

Lower-level Analysis, adjusted



Low-level Analysis Limitation

- Multi-stage regressions often ignore uncertainty / error in earlier stages
 - Compute residuals without estimating uncertainty / error

Test Links among Variables across Levels

- Run simpler, less precise, higher-level regressions
- If we need more precision or a more complex model, run lower-level regressions
- If controls for higher-level outcome(s) are needed, run multi-stage regressions
- Ideally, results are consistent
 - Directions of the effects do not flip



Thank you!