



The Future of Rural School Research and Development: Methodological Innovations for the Next Generation

James A. Bovaird

University of Nebraska-Lincoln

Special Thanks

- Members of the CYFS Statistics & Research Methodology Unit:
 - Dr. Greg Welch, Dr. Leslie Shaw, Natalie Koziol, Frances Chumney, Chaorong Wu, Carina McCormick, Ann Arthur, Elizabeth Timberlake
 - Dr. Kevin Kupzk, University of Nebraska Medical Center
 - Dr. Michael Toland, University of Kentucky
- Colleagues in the National Center for Research on Rural Education

Central Tenets

- Rural education research has and will continue to overcome perceived limitations to true experimentation, yet still approximate the level of knowledge available through random assignment and explicit environmental control.
- Educational policy must be utilitarian, consequently, research impacting educational policy in rural settings must focus on systems-level applications.
- Viable solutions exist in other disciplines and can be translated to rural education research.

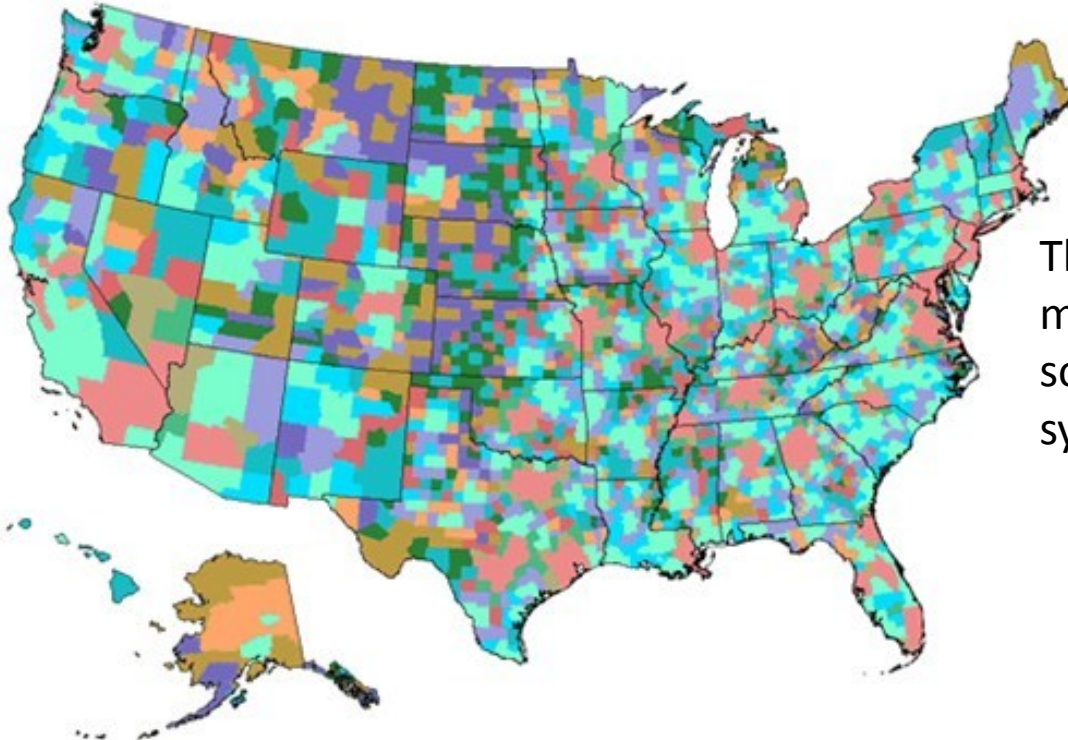
Topics to be Discussed Today

- Operational definition of "rural" and its impact on inferences
- Preserving and featuring the uniqueness of rural settings in systems level investigations through advanced statistical modeling
- Quasi-experimentation as an alternative to traditional random assignment
- Efficiency of measurement paradigms to reduce the amount of data necessary for valid inferences
- Innovations in small sample inferential testing

Measuring “Rural”

Some Classifications Schemes Can Get Ugly

2003 Urban influence codes

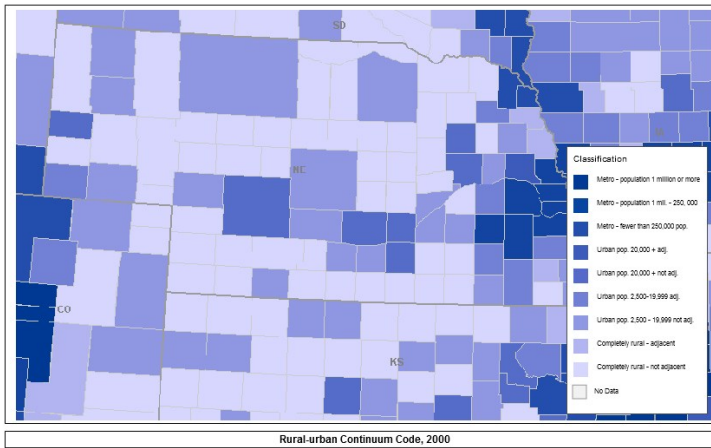
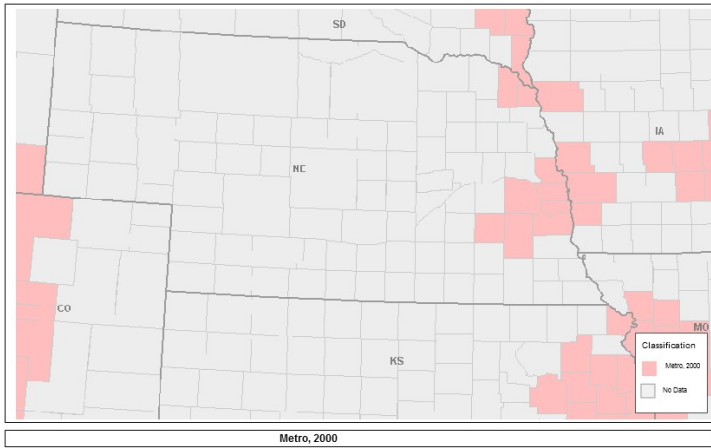


- | | | |
|---|---|---|
|  Large metro |  Micro adj large |  Noncore adj large |
|  Small metro |  Micro adj small |  Noncore adj small w/ town |
| |  Micro nonadj |  Noncore adj small no town |
| | |  Noncore adj micro w/ town |
| | |  Noncore adj micro no town |
| | |  Noncore nonadj w/ town |
| | |  Noncore nonadj no town |

Source: USDA, Economic Research Service.

This part of the presentation is meant to help make sense of some of the most common systems & their implications.

Establishing Organization (Researcher)	Classification	Unit (Level) of Classification	Description
Office of Management & Budget (OMB)	Metropolitan and Micropolitan Statistical Areas	County	<p><u>METRO</u>: Areas are based on the presence of an urbanized area with a population of at least 50,000.</p> <p><u>MICRO</u>: Areas are defined as an urban cluster with a population of at least 10,000 but no more than 50,000. Counties that do not fit into either of these definitions are classified as "Outside Core Based Statistical Areas."</p>
United States Census Bureau (Census)	Urban (Urbanized Areas or Urban Clusters) and Rural	Census tract and/or block (county based)	<p><u>URBAN</u>: Urbanized Areas are defined as having 50,000 or more people and Urban Clusters at least 2,500 and less than 50,000 people.</p> <p><u>RURAL</u>: Rural areas consist of all territory located outside of urbanized areas and urban clusters, thus populations of fewer than 2,500 residents.</p>
National Center for Education Statistics (NCES)	Urban-Centric Locale Codes	School and school- district	Urban-centric locale codes are based on a combination of proximity to an urbanized area and population size. School districts are classified based on the locale code at which the majority of the district's students are enrolled.
Economic Research Service, United States Department of Agriculture (USDA)	Urban Influence Codes	County	The OMB's metropolitan and micropolitan statistical areas are divided into smaller groups based on population, adjacency to metro areas, and commuting patterns.
Economic Research Service, United States Department of Agriculture (USDA)	Rural-Urban Commuting Area (RUCA)	Census tract and ZIP code	RUCA codes combine measures of population density, urbanization, and commuting patterns at the census tract and ZIP code levels. Coding scheme enables researchers to create both primary and secondary codes for areas.



Nebraska

Three rural definitions based on Census Urban Areas

Rural locations are those outside Census Urban Areas with a population...

...greater than or equal to 2,500

Outside Census Urban Areas $\geq 2,500$

...greater than or equal to 10,000

Outside Census Urban Areas $\geq 2,500$

Census Urban Areas: 2,500 - 9,999

...greater than or equal to 50,000

Outside Census Urban Areas $\geq 2,500$

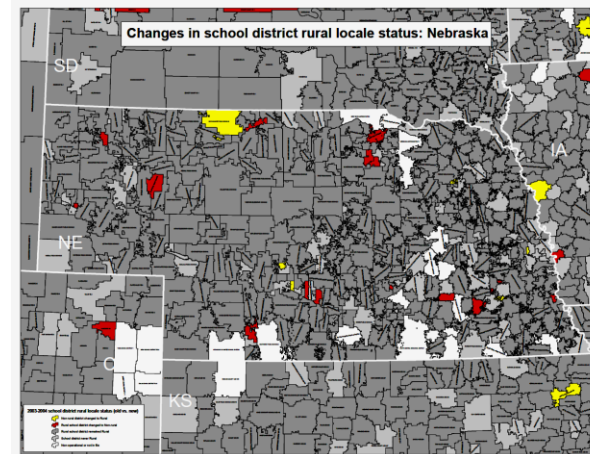
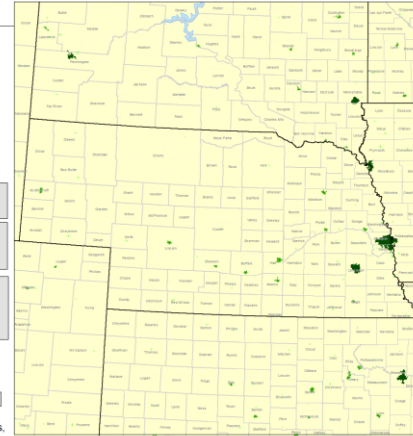
Census Urban Areas: 2,500 - 9,999

Census Urban Areas: 10,000 - 49,999

Urban locations under all three definitions:

Census Urban Areas: $\geq 50,000$

For more information on definitions, see documentation



Current Work

- Summary and synthesis of current coding strategies
- Empirical evaluation of the impact of different coding strategies
- Empirical evaluation of continuous variable and latent variable approaches to modeling rurality

Modeling the Rural Context

Kansas Kindergarten Readiness Project: Student Readiness for School

- Cross-sectional contextual (county) model of (student) school readiness
 - Bovaird (2005); Bovaird, Martinez, & Stuber (2006)
 - Multilevel model is appropriate
 - *students nested within county*
- Goal:
 - To describe the relationship between county-level contextual characteristics and kindergarten preparedness, controlling for student-level characteristics.

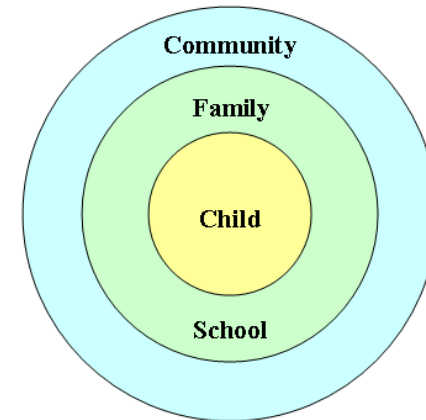
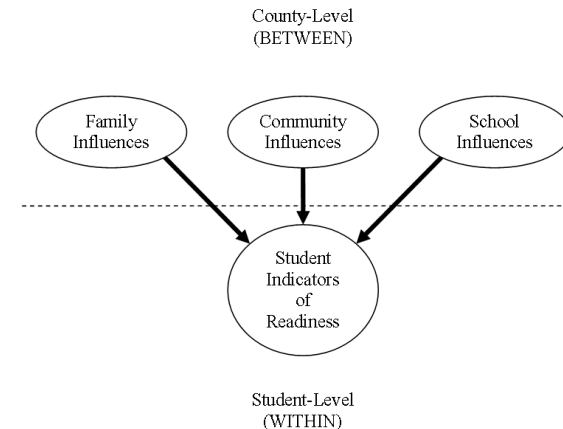


Figure 1. Conceptual drawing of the model of student readiness.



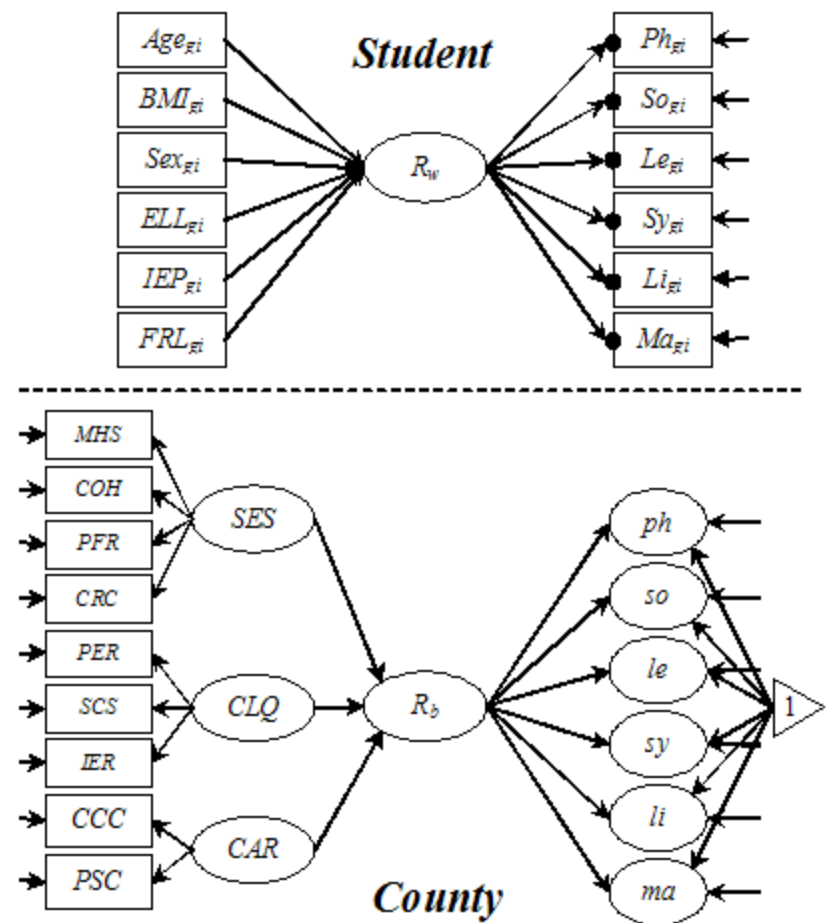
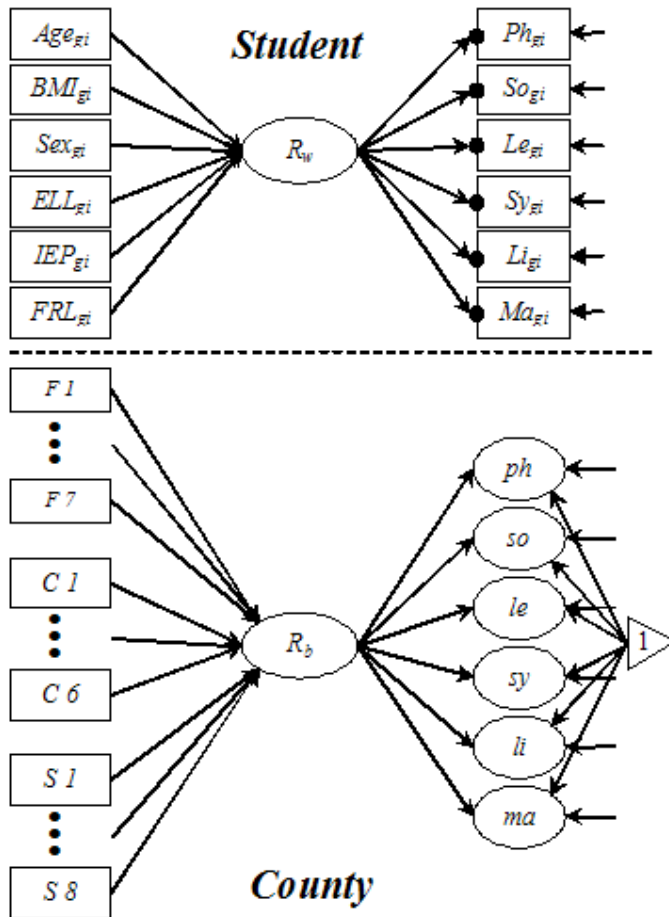
Kansas Kindergarten Readiness Project: The Measures

- *Kansas School Entry Assessment*
 - Teacher-completed measure of kindergarten preparedness
 - 41 items in 6 areas of school readiness: symbolic development (Sy), literacy development (Li), mathematical knowledge (Ma), social skills development (So), learning to learn (Le), physical development (Ph)
- Student-level predictors:
 - age (Age), body-mass index (BMI), gender (Sex), language status (ELL), eligibility for free or reduced lunch (FRL), IEP status (IEP)
- 21 county-level contextual variables supplied by state agencies grouped into three goal areas:
 - Family Goal - children live in safe and stable families that support learning
 - Community Goal - children live in safe and stable communities that support learning, health, and family services
 - School Goal - children attend schools that support learning

Kansas Kindergarten Readiness Project: The Data

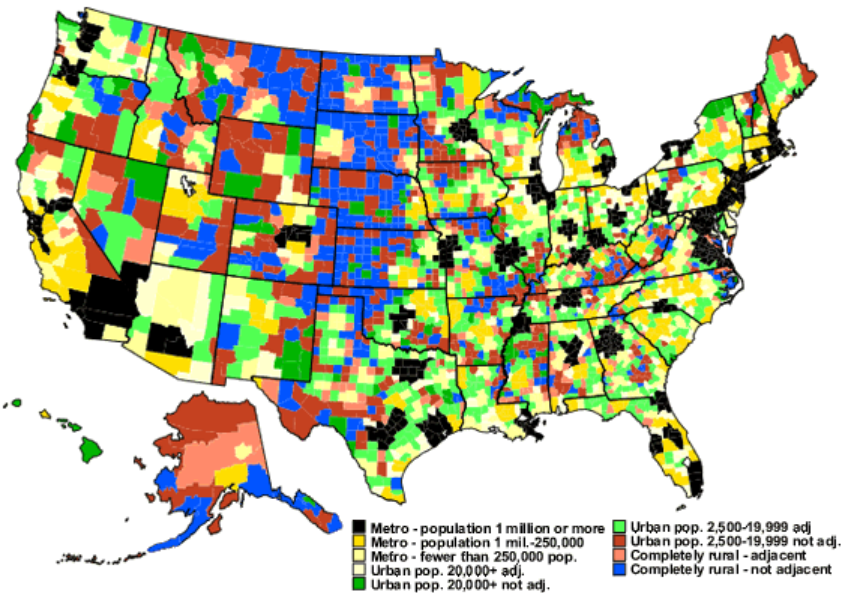
- **N = 1,997 kindergartners**
- 1-2 kids per teacher (teacher IDs not tracked)
- 233 schools (1-22 kids/school, avg = 6 kids/school)
- 154 districts (1-35 kids/district, avg = 12 kids/district)
- **J = 95 counties** (1-46 kids/county, avg = 21 kids/county)
 - Out of a possible 105 counties in Kansas

Multilevel MIMIC Model

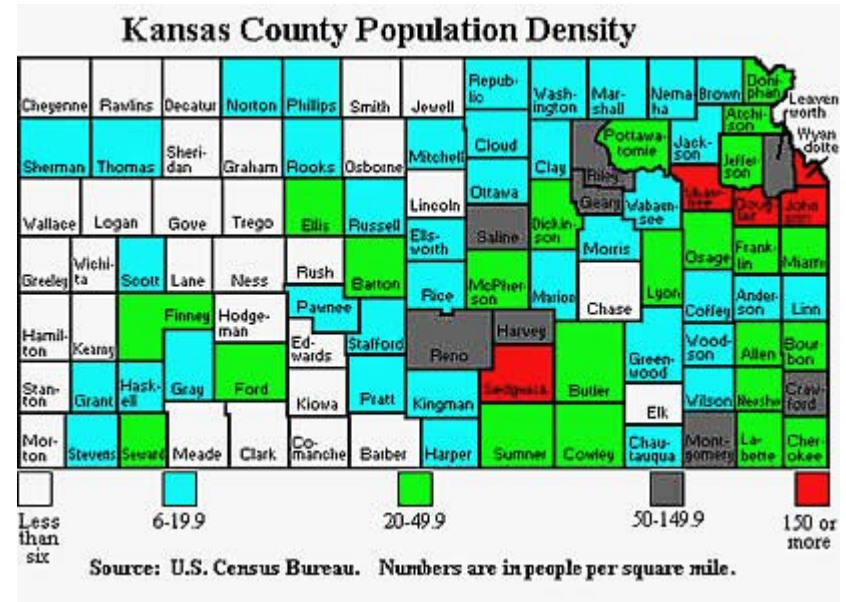


U.S. Population Density by County

Rural-urban continuum codes, 2003



Kansas Population Density by County



How Does the Definition of Rural and its Measurement Impact Inference?

- Measurement
 - Continuous variable
 - Population
 - **Population per square mile**
 - Categorical variable
 - Johnson Codes
 - Beale RUC Codes
 - **Other ad hoc categorization (median split)**
 - Other
- Statistical modeling
 - Interaction (continuous or categorical variables)
 - **Multiple groups SEM**
 - Spatial nesting

	Average	Average		
	# kindergarten (2003)	# sampled	Count	% Sampled
Population per Square Mile				
1	32.5	13.5	31	41.5
2	86.5	19.5	38	22.5
3	296.3	23.5	23	7.9
4	552.5	18.1	8	3.3
5	3182.6	30.4	5	1.0
Beale/RUC Codes				
1	1520.5	22.3	6	1.5
2	1766.5	29.0	4	1.6
3	479.3	29.9	7	6.2
4	532.7	28.3	3	5.3
5	493.8	18.3	8	3.7
6	155.5	16.8	11	10.8
7	124.7	19.5	23	15.6
8	60.0	17.5	4	29.2
9	39.3	15.5	39	39.4
Metro/Town/Rural				
1	1149.6	27.0	17	2.3
2	225.0	19.2	45	8.5
3	41.3	15.7	43	38.0

***Quasi-Experimental Design
Alternatives***

Sequentially Designed Experiments: Fixed vs. Sequential Designs

- **Fixed** experimental design:
 - Typical design in education and the social and behavioral sciences
 - Sample size and composition (e.g., experimental group allocation) determined prior to conducting the experiment
- **Sequential** experimental design:
 - **Sample size treated as a random variable**
 - Allows sequential interim analyses and decision-making
 - Based on cumulative data and previous design decisions
 - While maintaining appropriate Type I (α) & Type II (β) error rates

Sequentially Designed Experiments: Primary Benefits & Limitations

- Benefits:
 - Early termination
 - Unnecessary exposure
 - Prevent unnecessarily withholding administration
 - Financial savings
- Limitations:
 - Increased design complexity
 - Increased computational burdens
 - Threat to validity due to ability for early termination
 - Early termination for efficacy, futility, or participant safety
 - Early termination decision is more complex than just a statistical criterion
 - Consistency across both primary and secondary outcomes, risk groups, etc.

Sequentially Designed Experiments: Characteristics

- ***At least 1 interim analysis*** at a pre-specified interim stage prior to formal completion of the experiment
- Statistical details are determined a priori (***there's a protocol***)
 - # interim stages, n at each stage, desired nominal α and β levels
 - Critical values (*boundary values*) are computed for each interim stage
 - All available data is analyzed (data from that stage + all previous stages)
 - The appropriate test statistic and the Fisher information level (the inverse of the squared standard error) are computed.
 - The test statistic is then compared with critical boundary values determined *a priori* to maintain appropriate nominal experiment-wise Type I and Type II error rates given the occurrence of multiple statistical tests at interim stages.
 - If the test statistic falls within a decision region, the experiment stops.
 - Otherwise, the experiment continues to the next stage or until the maximum sample size is reached.

Sequentially Designed Experiments: Types

- 3 General Types:
 - **Fully** sequential designs
 - Continuous monitoring - updated after every observation or after every participant completes the study
 - **Group** sequential designs
 - Considered analogous to fully sequential designs EXCEPT that boundary values are computed for a predetermined number of equally spaced **stages** rather than after each participant
 - **Flexible** sequential designs
 - Can be viewed as a compromise between fully sequential and group sequential designs
- Differ based on sample recruitment and decision-making criteria.

Sequentially Designed Experiments:

Empirical Example

- CBC in the Early Grades (Sheridan *et al*, 2011)
- 4-cohort fixed-design cluster randomized trial to evaluate the effectiveness of a school-based consultation (CBC) approach for students with challenging classroom behaviors
 - 22 schools, 90 classrooms/teachers, 270 K-3rd grade students & parents
 - Randomly assigned as small (2-3) parent-teacher groups to:
 - business-as-usual control condition
 - experimental CBC condition.
- Study designed to detect a medium standardized effect ($ES = .38$).

Sequentially Designed Experiments: Methodological Study

- Bovaird *et al*, 2009; Bovaird, 2010
- Procedures
 - Implemented a post hoc application of a sequential design and analysis strategy
 - Cohort (4) = “Group”
 - Assuming eventual “known” fixed design conclusions as true...
 - *What is the degree to which sample size savings may have been realized if we had implemented a group sequential design rather than a fixed design?*

Sequentially Designed Experiments: Sequential vs. Fixed Design Results

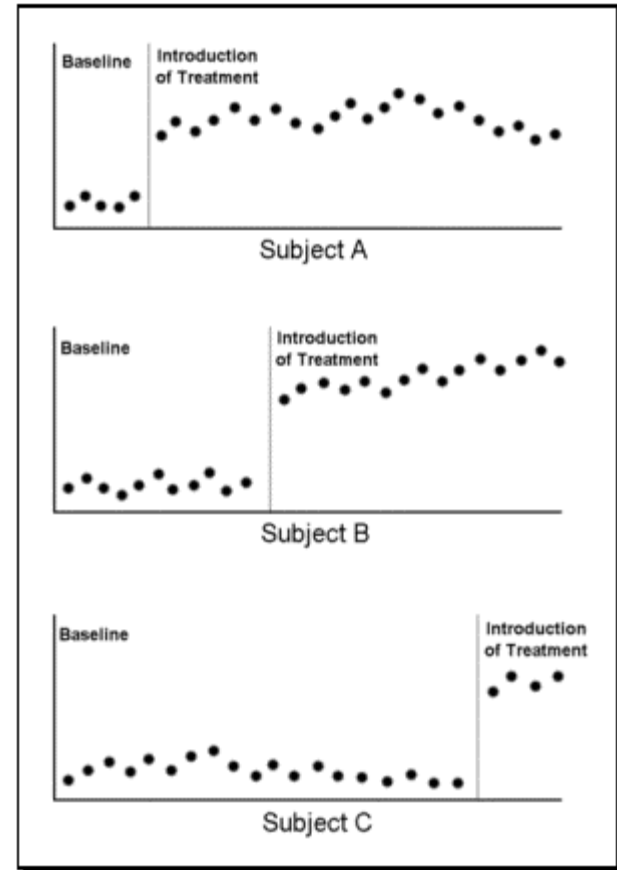
Table 1. Parameter Estimates (*Est*), Standard Errors (*SE*), and Hypothesis Test (*t*) Decisions for Fixed (*p*) and Sequential (*Dec*) Analyses

		Stage						Stage			
		1	2	3	4			1	2	3	4
		($N_{teach}=25$)	($N_{teach}=54$)	($N_{teach}=80$)	($N_{teach}=90$)			($N_{teach}=25$)	($N_{teach}=54$)	($N_{teach}=80$)	($N_{teach}=90$)
BASC – Adaptive Skills (Parent)	<i>Est.</i>	0.34	-0.03	-0.41	-0.36	<i>Est.</i>	3.63	2.59	2.39	2.20	
	<i>SE</i>	1.80	1.15	0.90	0.88	<i>SE</i>	1.41	1.12	0.91	0.89	
	<i>t</i>	0.19	-0.02	-0.45	-0.41	<i>t</i>	2.57	2.32	2.63	2.46	
	<i>p</i>	0.43	0.49	0.33	0.34	<i>p</i>	<u>0.01</u>	<u>0.01</u>	<u>0.00</u>	<u>0.01</u>	
	<i>Dec.</i>	Continue	Accept H_0			<i>Dec.</i>	Reject H_0				
BASC – Externalizing Behavior (Parent)	<i>Est.</i>	-3.11	-0.55	0.58	0.39	<i>Est.</i>	-0.82	-1.02	-2.23	-2.43	
	<i>SE</i>	2.31	1.43	1.27	1.26	<i>SE</i>	2.11	1.59	1.29	1.27	
	<i>t</i>	-1.35	-0.39	0.46	0.31	<i>t</i>	-0.39	-0.64	-1.73	-1.91	
	<i>p</i>	0.09	0.35	0.32	0.38	<i>p</i>	0.35	0.26	<u>0.04</u>	<u>0.03</u>	
	<i>Dec.</i>	Continue	Accept H_0			<i>Dec.</i>	Continue	Continue	Continue	Reject H_0	
SSRS – Social Skills Score (Parent)	<i>Est.</i>	-1.05	0.33	1.73	0.94	<i>Est.</i>	3.13	3.16	3.90	3.72	
	<i>SE</i>	3.22	2.05	1.91	1.89	<i>SE</i>	2.55	1.98	1.51	1.49	
	<i>t</i>	-0.33	0.16	0.90	0.50	<i>t</i>	1.22	1.60	2.58	2.50	
	<i>p</i>	0.37	0.44	0.18	0.31	<i>p</i>	0.11	0.06	<u>0.01</u>	<u>0.01</u>	
	<i>Dec.</i>	Continue	Accept H_0			<i>Dec.</i>	Continue	Continue	Reject H_0		
Parent- Teacher Relationship (Parent)	<i>Est.</i>	0.03	0.05	0.05	0.03	<i>Est.</i>	0.13	0.23	0.20	0.19	
	<i>SE</i>	0.24	0.14	0.11	0.11	<i>SE</i>	0.15	0.09	0.08	0.07	
	<i>t</i>	0.13	0.34	0.45	0.26	<i>t</i>	0.82	2.69	2.67	2.57	
	<i>p</i>	0.45	0.37	0.33	0.40	<i>p</i>	0.21	<u>0.00</u>	<u>0.00</u>	<u>0.01</u>	
	<i>Dec.</i>	Accept H_0				<i>Dec.</i>	Continue	Reject H_0			

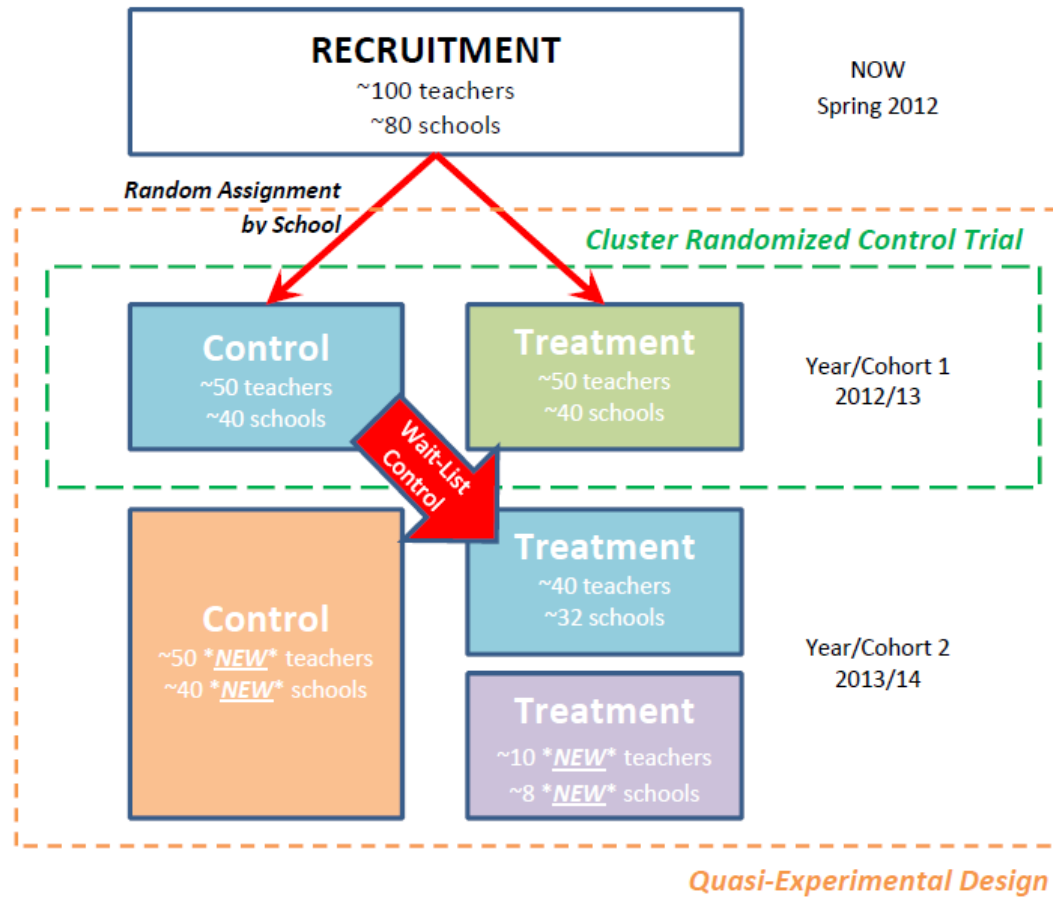
Stepped Wedge Designs:

Wait-List Control Designs & Multiple Baseline Designs

<u>Parallel</u>		<u>Crossover</u>		<u>Stepped Wedge</u>					
Time		Time		Time					
1		1 2		1 2 3 4 5					
Cluster	1	1	1 0	1	0	1	1	1	1
Cluster	2	1	1 0	2	0	0	1	1	1
Cluster	3	0	0 1	3	0	0	0	1	1
Cluster	4	0	0 1	4	0	0	0	0	1



CSI: Coaching Science Inquiry in Rural Schools



CSI:

Data Analysis

- The sample size is considered fixed and is determined by the maximum capacity of participating teachers in the treatment condition during a given year. This capacity is fixed at 50 treatment teachers per cohort.
 - The number of schools is the important sample size in terms of statistical power.
- Traditional evaluation of the **Cluster RCT** via MLM/HLM
- Re-evaluation of the full 2-year data set as a quasi-experimental **Stepped-Wedge** design
- Re-evaluation of the full 2-year dataset as a quasi-experimental design with **propensity score matching**.

More-Efficient Measurement

Planned Missing Data Designs (PMDDs)

- “Efficiency-of-measurement design” (Graham, Taylor, Olchowski, & Cumsille, 2006)
 - Random sampling
 - Optimal Designs
 - See Allison, Allison, Faith, Paultre, & Pi-Sunyer (1997)
 - Balance cost (\$) with statistical power
 - Fractional Factorial Designs
 - See Box, Hunter, & Hunter (2005)
 - Carefully chosen subset of cells from a factorial design focus “information” on most important conditions while minimizing resources
 - Not so different from adaptive testing...
 - Measurement Models

Measurement PMDDs

- Simple matrix sampling (Shoemaker, 1973)
 - Useful for means, but not correlations
- Fractional block design (McArdle, 1994)
- Allows means + SOME correlations
 - Requires multiple-group SEM for analysis
- Balanced incomplete blocks (spiral) designs (Johnson, 1992)
 - Means & correlations available
 - Same number of Ss respond to each item

Table 1
Example of Simple Multiple Matrix Design

Form	Blocks of items						
	A	B	C	D	E	F	G
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0
5	0	0	0	0	1	0	0
6	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1

Note. 1 = questions asked; 0 = questions not asked. Letters A–G refer to different sets of items.

From Graham et al. (2006)

	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆	G ₇	G ₈
1	1	○	○	○	○	1	1	1
2	2	2	○	○	○	○	2	2
3	3	3	3	○	○	○	○	2
4	4	4	4	4	○	○	○	○
○	○	5	5	5	5	○	○	○
○	○	○	6	6	6	6	○	○
○	○	○	○	7	7	7	7	○
○	○	○	○	○	8	8	8	8

From McArdle (1994)

Measurement PMDDs (cont.)

- 3-form design (Graham, Hofer, & Piccinin, 1994; others)
 - Items split into 4 sets (X, A, B, C)
 - All Ss get X + 2 additional sets (XAB, XAC, XBC)
 - More hypotheses testable [$k(k-1)/2$ two-variable effects w/in each set + $2k$ two-variable effects across two sets)
 - Don't forget multiplicity!
- Split questionnaire survey design (SQSD; Raghunathan & Grizzle, 1995)

Table 2
The 3-Form Design, With X Set

Form	Item set			
	X	A	B	C
1	1	1	1	0
2	1	1	0	1
3	1	0	1	1

Note. 1 = questions asked; 0 = questions not asked.

From Graham et al. (2006)

Table 3
Ten-Form, Six-Set Variation of the Split Questionnaire Survey Design, With X Set

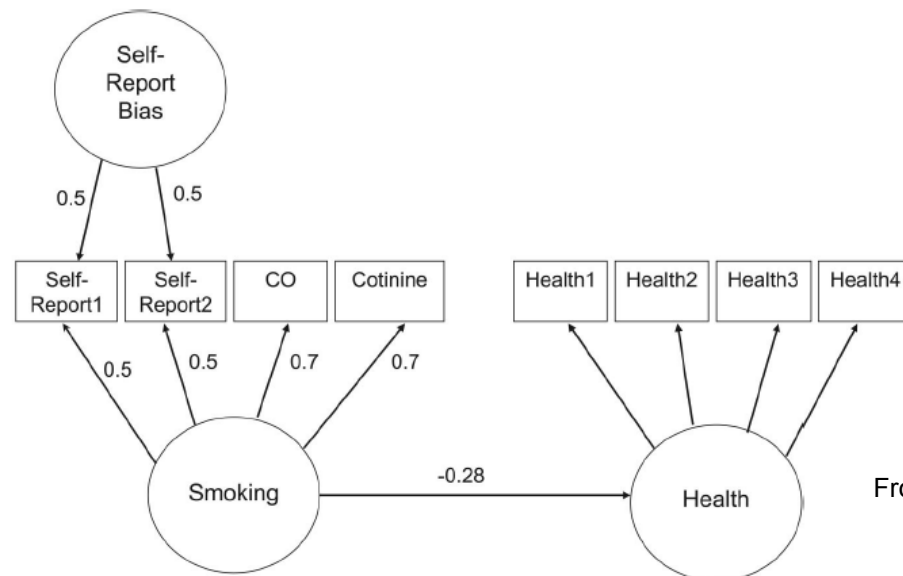
Form	Item set					
	X	A	B	C	D	E
1	1	1	1	0	0	0
2	1	1	0	1	0	0
3	1	1	0	0	1	0
4	1	1	0	0	0	1
5	1	0	1	1	0	0
6	1	0	1	0	1	0
7	1	0	1	0	0	1
8	1	0	0	1	1	0
9	1	0	0	1	0	1
10	1	0	0	0	1	1

Note. 1 = questions asked; 0 = questions not asked.

From Graham et al. (2006)

Measurement PMDDs (cont.)

- 2-method measurement
 - Many cases w/ cheap, relatively noisy (lower reliability) measure
 - i.e. self-report
 - May require a response bias correction model
 - Few cases w/ both cheap and expensive, more reliable measure
 - i.e. biological markers



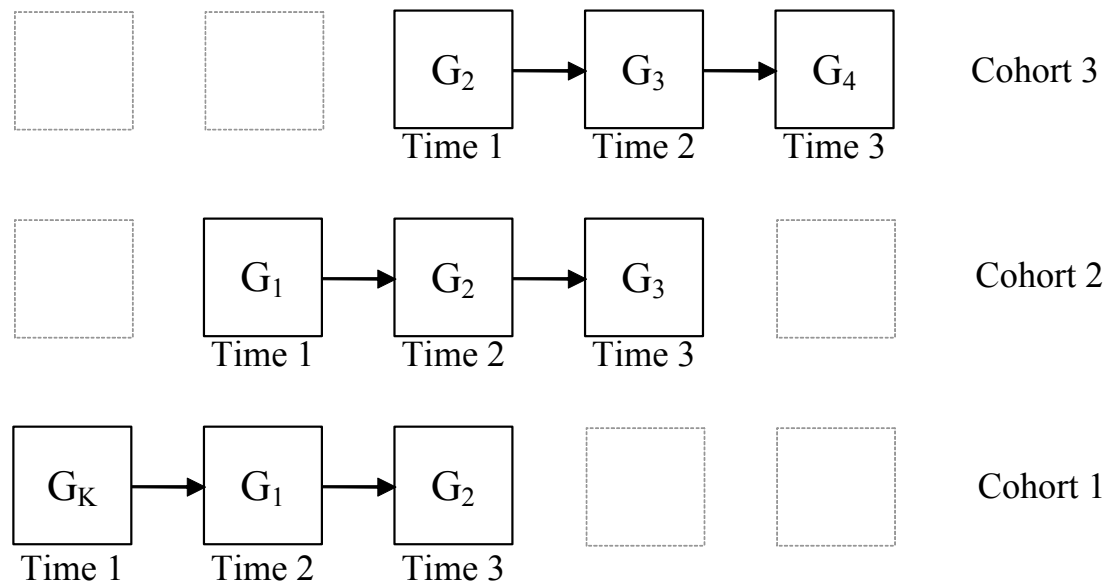
From Graham et al. (2006)

Accelerated Longitudinal Designs

- Convergence design
 - Bell (1953)
- Cross-sequential design
 - Schaie (1965)
- Cohort-sequential design
 - Nesselroade & Baltes (1979)
- Accelerated longitudinal design
 - Tonry, Ohlin, & Farrington (1991)

What Does Accelerated Mean?

- Overlapping 'Cohorts'
 - A cohort is a group of participants that begin a study at a common age or grade in school
- Tracked for a limited number of measurement occasions
- Groups are linked at their overlapping time points to approximate the true longitudinal curve/trajectory



Accelerated Longitudinal Design

- Advantages
 - Allows for assessment of intra-individual change
 - Takes less time than a purely longitudinal design
 - Subject attrition and cumulative testing effects are not as prevalent
- Possible applications
 - Any longitudinal research setting
 - Developmental research
 - Educational or Classroom studies
 - Gerontology or aging research

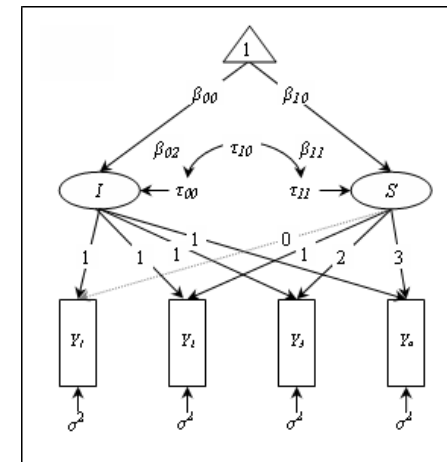
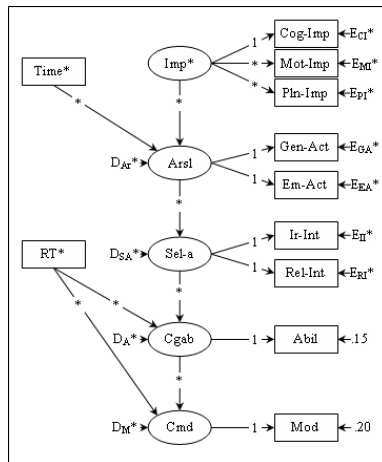
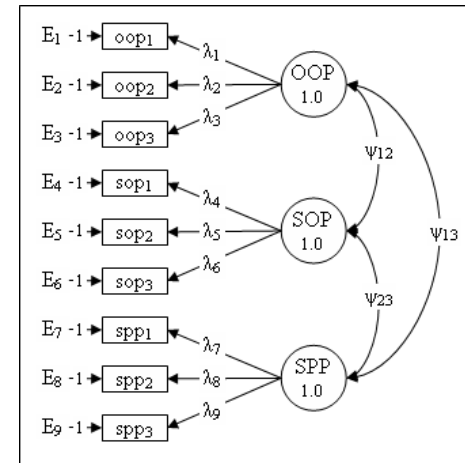
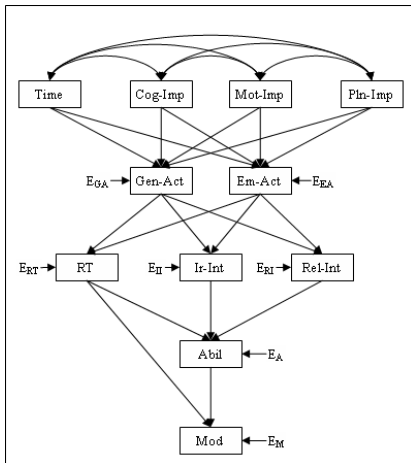
Small Samples but Big Models

Structural Equation Modeling

- A collection of techniques
 - Allow relationships between:
 - 1 or more IVs (continuous or discrete)
 - One or more DVs (continuous or discrete)
 - IVs & DVs can be either measured variables or latent *constructs*
- **Construct:** term often used to refer to some attribute we want to measure
 - In social sciences, most constructs are latent (i.e., unobservable) traits
 - How do you impart clear meaning to scores that measure an unobservable trait?
 - Measurement instruments cannot exactly represent the attributes we seek to measure

Structural Equation Modeling

4 Common Types of SEM Models



Structural Equation Modeling

Research Questions

- Model Quality
 - Does the model fit the data? Does the model produce an estimated population covariance matrix that is consistent with the sample (observed) covariance matrix?
 - Which theory (model) produces an estimated population covariance matrix that is most consistent with the sample covariance matrix?
- Model Parameters
 - How much variance in the DV(s) is accounted for by the IV(s)?
 - What is the value of the path coefficient? Is it significantly different from zero? Which paths are more/less important?
 - Does an IV directly affect a specific DV or does the IV affect the DV through an intermediary, or mediating, variable?
- Special Models
 - Do two or more groups differ in the covariance matrices, regression coefficients, or means?
 - Does a variable change over time? What is the shape of the change? Do individuals vary in their initial level or rate of change?
 - How reliable are each of the measured variables?

Alternatives to SEM:

Factor Analysis

- Exploratory
 - Principal components analysis (PCA)
 - Parceling
 - Obtain factor scores for higher order constructs
- Confirmatory
 - Model fit
 - Larger # items = higher power w/ smaller N's
 - Parameter significance
 - Larger # items = lower power w/ smaller N's
 - Parameter precision
 - Smaller SE's are easier to obtain by increasing the sample size rather than reducing variability

Alternatives to SEM: Factor Scores

- Add (or average) variables together that load highly on a factor
 - variables with large SDs contribute more heavily to the solution
 - standardize first
 - in many cases, this will be adequate
 - Recommended in small samples
- Regression approach
 - Capitalizes on chance relationships, so factor-scores are biased
 - Often correlations among scores for factors even if supposed to be orthogonal
 - Best overall non-Bayesian method
- Bartlett method
 - Factor scores only correlate with their own factors & scores are unbiased
 - Factor scores may still be correlated with each other even when “orthogonal”
- Anderson-Rubin approach
 - Uncorrelated factor scores even if factors are correlated
 - Best if goal is an orthogonal score
- Empirical Bayes Estimation
 - Implemented in SEM & MLM programs
 - Ideal approach, but can suffer from shrinkage in small samples

Alternatives to SEM:

Others

- Unweighted (ordinary) Least squares
- Reduce to path analysis
 - Create summary score for each construct
 - PCA vs. FA vs. avg/sum of z-scores
 - Analyze as ML with a reduced model
 - Analyze with ULS/OLS as a multi-step regression
- Canonical correlation
- ML w/ Re-sampling procedures
 - Use bootstrapping to obtain empirical standard errors

Re-sampling Methods

- Use the obtained data to generate a simulated population distribution
 - Resample hundreds/thousands of times
 - Defines an empirical sampling distribution
 - Distributional assumptions are now irrelevant
 - Validity
 - internal – random sampling from the population is not necessary
 - external – random sampling is essential
- Uses:
 - Traditional hypothesis testing
 - Defining confidence intervals
 - Computing estimate stability
 - Measure estimate bias
- Types:
 - Randomization tests, Jackknife, Bootstrapping

Partial Least Squares

PLS; Wold (1966, 1973, 1982)

- Uses fixed point (FP) algorithm for parameter estimation
 - Model parameters divided into subsets
 - Each set is partially estimated w/ OLS with other subset fixed
 - Switch & cycled through until convergence
- Avoids improper solutions by replacing factors w/ linear composites of observed variables like in PCA
- Does not rely on distributional assumptions
- Does not solve global optimization problem
 - No single criterion consistently minimized/maximized to determine model parameter estimates (i.e. fit function in SEM)
 - Difficult to evaluate PLS procedure
 - No mechanism to evaluate the overall goodness of fit

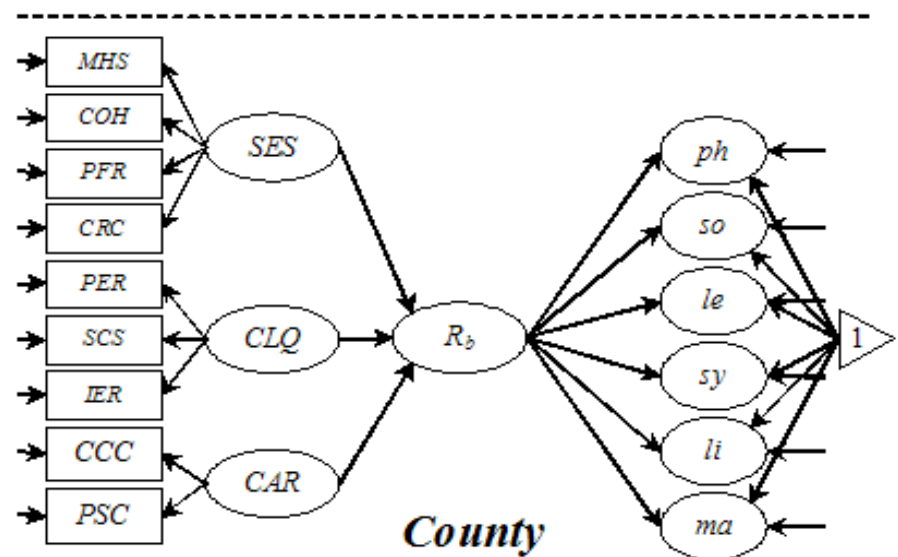
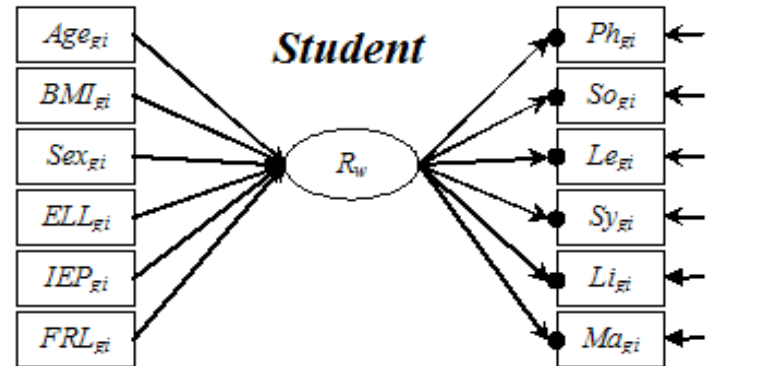
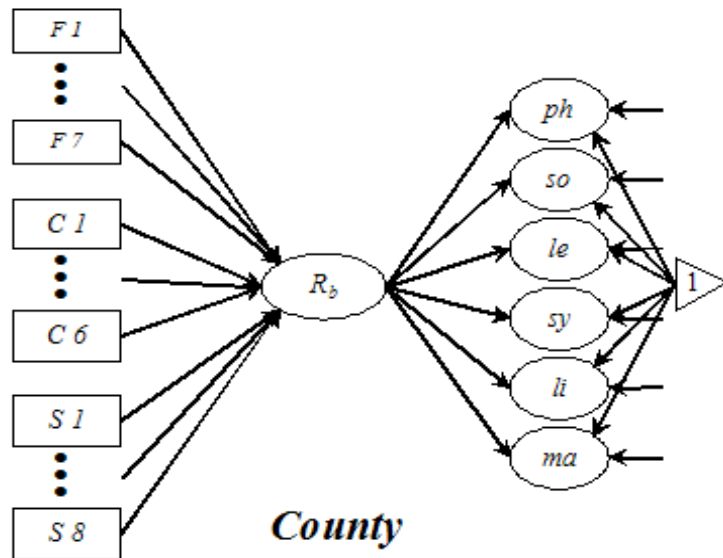
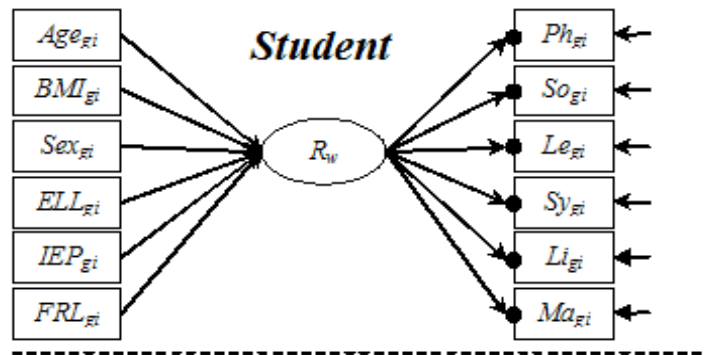
Generalized Structured Component Analysis *GSCA; Hwang & Takane (2004)*

- Avoids principal limitation of PLS – lack of global optimization procedure
- Retains advantages of PLS – less restricted distributional assumptions, no improper solutions
- Also versatile – higher order constructs, multiple groups, etc.
- Substitutes components for factors (like PLS)
- Offers a global least squares optimization criterion (not like PLS)
- Consistently minimized to obtain estimates of model parameters
- Monte Carlo study
 - Even with $N = 10$, mean congruence coefficient between parameters & estimates is greater than .90 (.908, min = .432), thus acceptable (Mulaik, 1972)
 - With $N = 50$, minimum congruence coefficient is close to .90 (.848, mean = .981)

Bayesian Methods

- Bayesian estimate of unknown parameter θ = mean of the posterior distribution $p(\theta | Y)$
- Parameter vector $\underline{\theta}$ considered random
- (at least) 2 advantages
 - informative priors can lead to more accurate estimates of parameters of interest
 - previous research can inform current research
 - sampling-based Bayesian methods (i.e. Markov Chain Monte Carlo; MCMC) do not rely on asymptotic theory
 - can be useful for smaller samples
- Monte Carlo study in Lee & Song (2004)
 - Considered sample sizes based on $n = da$ where $d = 2, 3, 4, 5$ and $a = \#$ parameters
 - Summary:
 - For data that are normally distributed, Bayesian approach can be used with small sample sizes while ML cannot – even when $n = 2a$ to $3a$

Finite Population Corrections: Motivating Example



Finite Population Corrections: Some Limiting Conditions?

- Problem(s)
 - Small N - but almost all of the available data
 - Small effects – indirect/proxy effects
- Sampling in MLM
 - possible to obtain proportionally large samples or near-census sampling at the macro-levels
 - especially when sampling from finite geographical locations
- Relevance:
 - educational testing
 - cross-cultural research
 - behavioral ecosystems modeling
- Potential solution??
 - finite population correction (fpc)

Finite Population Correction (fpc)

- Definition
 - Reduces sampling error by decreasing the variance related to the sampling method (sampling without replacement)
 - Adjustment factor varies with the sample size, and is directly related to the proportion of the population sampled
- Usefulness
 - When finite population corrections are omitted, the standard errors are overestimated
 - Standard formulas assume sample taken from a population so large that it may as well be infinite
 - The fpc factors may be used to develop confidence estimates or in sample size estimation

Finite Population Correction (fpc)

- Guidelines for applying fpc
 - May be applied to either the variance or the standard error
 - Formula for variance: $(N-n) / (N-1)$
 - Formula for standard error: $\sqrt{((N-n) / (N-1))}$
 - Proportion of population that may be sampled without application of fpc depends on the research question and the size of effects expected
 - When less than 5% of population has been sampled, fpc factor is negligible
 - Proportion of population for which fpc should be applied is not completely agreed upon – generally 5% - 10%

Finite Population Corrections: Primary Model Results

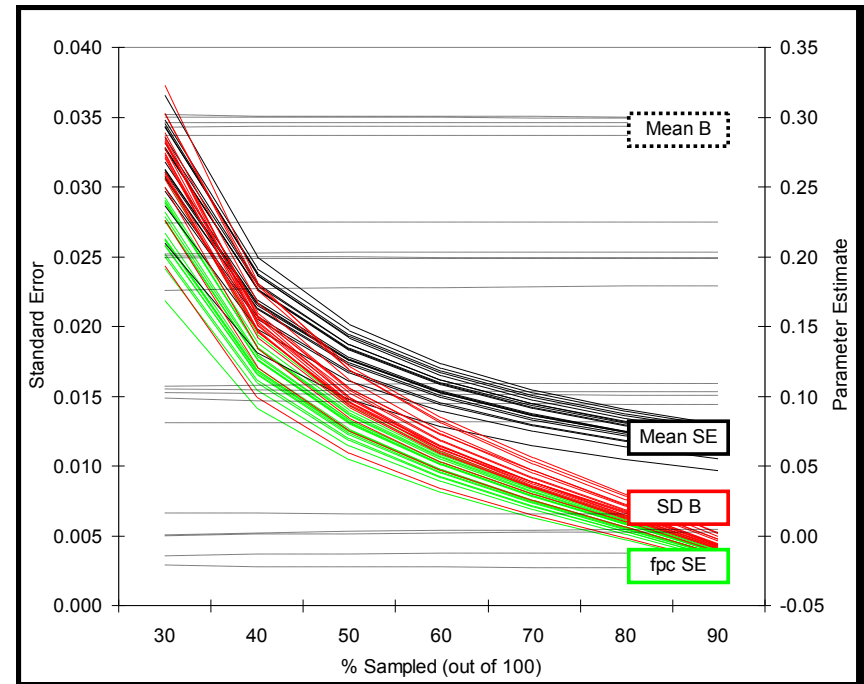
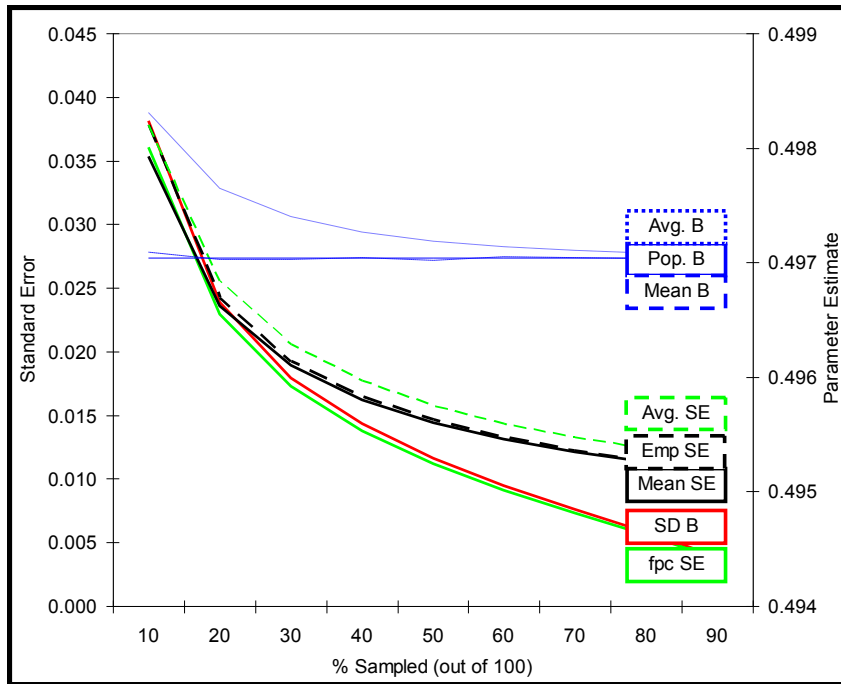
Predictor	Estimate	Standard Errors		Confidence Intervals				t Ratio (Est./SE)		p values	
		Estimated	Corrected	Original	Corrected	Estimated	Corrected	Estimated	Corrected		
F1	0.0001	0.0014	0.0004	-0.0025	0.0028	-0.0007	0.0010	0.11	0.35	p > .05	p > .05
F2	-0.0018	0.0014	0.0004	-0.0046	0.0009	-0.0027	-0.0010	-1.30	-4.20	p > .05	p < .05
F3	0.0027	0.0017	0.0005	-0.0007	0.0061	0.0016	0.0037	1.54	4.98	p > .05	p < .05
F4	-0.0348	0.0207	0.0064	-0.0753	0.0057	-0.0473	-0.0222	-1.68	-5.43	p > .05	p < .05
F5	0.0015	0.0038	0.0012	-0.0059	0.0089	-0.0008	0.0038	0.41	1.31	p > .05	p > .05
F6	-0.0008	0.0016	0.0005	-0.0039	0.0023	-0.0018	0.0002	-0.50	-1.60	p > .05	p > .05
F7	0.0014	0.0017	0.0005	-0.0020	0.0047	0.0003	0.0024	0.79	2.55	p > .05	p < .05
C1	0.0002	0.0004	0.0001	-0.0006	0.0011	0.0000	0.0005	0.56	1.82	p > .05	p > .05
C2	-0.0096	0.0083	0.0026	-0.0259	0.0068	-0.0146	-0.0045	-1.15	-3.70	p > .05	p < .05
C3	-0.0001	0.0006	0.0002	-0.0011	0.0010	-0.0004	0.0003	-0.09	-0.30	p > .05	p > .05
C4	0.0006	0.0022	0.0007	-0.0036	0.0049	-0.0007	0.0019	0.29	0.94	p > .05	p > .05
C5	-0.0017	0.0055	0.0017	-0.0124	0.0091	-0.0050	0.0017	-0.31	-0.99	p > .05	p > .05
C6	0.0031	0.0008	0.0003	0.0015	0.0048	0.0026	0.0037	3.70	11.95	p < .05	p < .05
S1	-0.0145	0.0056	0.0018	-0.0256	-0.0034	-0.0179	-0.0111	-2.56	-8.27	p < .05	p < .05
S2	0.0003	0.0003	0.0001	-0.0003	0.0009	0.0001	0.0005	1.08	3.48	p > .05	p < .05
S3	-0.0008	0.0006	0.0002	-0.0019	0.0003	-0.0012	-0.0005	-1.46	-4.72	p > .05	p < .05
S4	0.0014	0.0049	0.0015	-0.0081	0.0110	-0.0015	0.0044	0.29	0.94	p > .05	p > .05
S5	0.0011	0.0049	0.0015	-0.0085	0.0108	-0.0019	0.0041	0.23	0.74	p > .05	p > .05
S6	-0.0036	0.0049	0.0015	-0.0132	0.0059	-0.0066	-0.0007	-0.75	-2.41	p > .05	p < .05
S7	0.0005	0.0006	0.0002	-0.0007	0.0016	0.0001	0.0008	0.80	2.59	p > .05	p < .05
S8	-0.0527	0.0111	0.0035	-0.0745	-0.0309	-0.0595	-0.0460	-4.74	-15.27	p < .05	p < .05

Finite Population Corrections: Sub Model Results

Predictor	Estimate	Standard Errors		Confidence Intervals		t Ratio (Est./SE)		p values			
		Estimated	Corrected	Original	Corrected	Estimated	Corrected	Estimated	Corrected		
SES	0.0004	0.0022	0.0007	-0.0033	0.0040	-0.0008	0.0015	0.16	0.50	p > .05	p > .05
CLQ	0.0067	0.0126	0.0039	-0.0140	0.0273	0.0003	0.0131	0.53	1.71	p > .05	p < .05
CAR	-0.0023	0.0056	0.0017	-0.0115	0.0070	-0.0051	0.0006	-0.41	-1.31	p > .05	p > .05

Using a 1-tailed hypothesis test.

Finite Population Corrections: Simple Regression vs. Multiple Regression



Finite Population Corrections: Relevance

- How realistic, or meaningful are finite samples?
 - Very realistic for upper hierarchical levels in education
 - School districts (NCLB), counties (state ed. depts.), states (NAEP/NCLB), countries (PISA)
 - Moderately realistic for cross-cultural studies when assessing culture/country-level variables
 - Potentially realistic for small clinical, under-represented, or geographically isolatable populations



THANK YOU!

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant # R305C090022 to the University of Nebraska-Lincoln. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.