

NEBRASKA ACADEMY FOR METHODOLOGY, ANALYTICS & PSYCHOMETRICS

Methodology Applications Series 2018-2019





Adaptation in the Social, Behavioral, and Education Sciences: Implications for Measurement, Intervention, & Evaluation

James A. Bovaird, PhD

Founding Director, Nebraska Academy for Methodology, Analytics & Psychometrics Associate Professor of Educational Psychology Program Director, Quantitative, Qualitative & Psychometric Methods Program Courtesy Associate Professor of Survey Research & Methodology



Introduction

- Adaptive methodologies have a long, but often unfamiliar, history.
 - Trace back to the early 1900's
- Three (3) broad categories:
 - Adaptive testing
 - Computerized adaptive testing (CAT) adapts the assessment tool to refine *measurement*.
 - Adaptive interventions
 - The sequential, multiple assignment, randomized trial (SMART) adapts the *intervention* itself to refine the treatment.
 - Adaptive designs
 - Sequentially-designed experiments adapt the *evaluation* context of an existing intervention to refine the resources necessary to make a valid inference.



History

- 1905
 - Start of adaptive individualized intelligence testing. (Alfred *Binet*)
- 1929
 - Development of a double sampling inspection procedure for the purpose of industrial quality control. (Harold F. *Dodge* and Harry G. *Romig*)
- 1938
 - Census of Bengalese jute area (Prasanta Chandra Mahalanobis)
- 1943
 - Sequential probability ratio test for military armament testing. (Abraham Wald; Statistical Research Group at Columbia University: Milton Friedman, W. Allen Wallis)
 - Launched the complementary field of *sequential analysis*.
 - Statistical hypothesis testing procedures which allow a statistical test to be calculated at any stage of the experiment prior to completion
 - 3-alternative rule for inferential decision-making: FTR H₀, reject H₀, or continue experiment
- 1960
 - Book on sequential medical trials effectively introduced the sequential design of randomized clinical trials (RCT). (Peter *Armitage*)
- 1980's
 - Computerized adaptive testing procedures for educational and psychological testing based on the principles of sequential design of experiments.
- 2000's
 - Introduction & continued development of SMART and other frameworks for developing adaptive interventions.





ADAPTIVE TESTING



Computer Adaptive Testing (CAT)

- A CAT administers items that are most appropriate for a given ability level
- For example, higher-ability examinees will be administered harder items
- Items are essentially weighted according to their difficulty, making test scores comparable
- A CAT can often achieve the precision of a fixedlength test using half as many items
- Made practical through Item Response Theory (IRT)

$$P(X_{is} = 1 | \theta_s, b_i, a_i) = \frac{e^{Da_i(\theta_s - b_i)}}{1 + e^{Da_i(\theta_s - b_i)}}, \quad (2) \quad \text{MAP ACADEMY}$$

IRT: Item Response Function



Figure 16.1 Item response functions for six hypothetical items. A, B, and C are 1PL models; D and E are 2PL models, and F is a 3PL model. The numbers in parentheses correspond with the discrimination, difficulty, and guessing parameter estimates, respectively



IRT: Item Information



Figure 16.2 Item information functions contrasted with their corresponding item response functions for three of the items in Figure 16.1 differing in discrimination and difficulty



IRT: Test Information



Figure 16.3 Item information functions, scale information function, and test standard error for a hypothetical test that includes the six items originally presented in Figure 16.1. Note that item E has the highest discrimination and thus the most information. Even though the average difficulty is 0.0, maximum precision is obtained for examinees that are 0.75 standard errors below "average"



How CAT Works

- To begin, all examinees are administered moderately difficult items
 - Missing an item will result in a lower ability estimate, and the computer will administer an easier item
 - Answering an item correctly will increase one's ability estimate, and the computer will administer a more difficult item
- Using IRT, the computer estimates the respondent's ability level after each item is administered
 - Subsequent items are tailored to the respondent's ability level
- Testing continues until the algorithm identifies the difficulty level at which the respondent will miss about 50% of the items
 - Information is concentrated and maximized at this most-appropriate difficulty level
 - Stopping rules are based on EITHER *logistical convention* (fixed # of items) OR a sufficiently *small standard error*



How CAT Works (cont.)



Image from http://www.nlsinfo.org/nlsy97/nlsdocs/nlsy97/codesup/mapp10.html





ADAPTIVE IINTERVENTIONS



- Utilize individual variables to adapt the intervention and then dynamically utilize individual outcomes to readapt the intervention.
- Type or dosage of the intervention offered to patients is individualized on the basis of patients' characteristics or clinical presentation (*tailoring variables*)
 - Then repeatedly adjusted over time in response to their ongoing performance (like a CAT)
- Multistage process
 - Operationalized via a sequence of *decision rules* that determine when and how the intervention should be modified to maximize long-term primary outcomes
- Recommendations based on characteristics PLUS intermediate outcomes
- Also known as
 - Dynamic treatment regimens (Murphy et al. 2001, Robins 1986)
 - Adaptive treatment strategies (Lavori & Dawson, 2000; Murphy, 2005)
 - Multi-stage treatment strategies (Thall et al. 2002, Thall & Whathen, 2005)
 - Treatment policies (Lunceford et al. 2002, Wahed & Tsiatis, 2004, 2006)



- 4 key elements
- (1) Sequence of critical decisions in a patient's care
 - Which intervention/dosage do they get first?
 - What intervention/dosage do they change to if the initial is unsuccessful?
- (2) Set of possible intervention/dosage options at each critical decision point
 - Different types, model of delivery, combinations , approaches to enhance engagement & adherence
- (3) Set of tailoring variables to pinpoint when the intervention should be altered and ID which option is best for whom
 - Early signs of nonresponse, adherence, side effects, burden
 - Contextual information
- (4) Sequence of decision rules, 1 rule per critical decision
 - Links individual characteristics & ongoing performance with specific intervention options
 - Inputs a tailoring variable and outputs 1 or more intervention options
 - IF (tailoring variable = X); THEN (intervention = Y).



- Statistical analysis
 - Essentially the same as for fixed interventions
 - Sampling, control groups, assignment, multiple cohorts, statistical power, timing/spacing of repeated measurements
 - Research questions are essentially the same
 - Important to maintain random assignment for causal inferences
 - Replicability is closely linked to fidelity of implementation of decision rules



- Why consider an adaptive design?
 - Variable responsiveness to treatment
 - Changing effectiveness
 - Emerging or evolving comorbidities
 - Potential for relapse
 - High cost of intensive interventions + burden or side effects motivate development of interventions that can be scaled when needed
 - Difficulty in maintaining adherence



Response to Intervention (RTI) A Case of Adaptive Design

- Approach to academic intervention to provide early, systematic, and/or appropriate grade- or age-level standards.
- Promotes academic success through niversal screening, early intervention, frequent progress monitoring, & increasingly intensive research-based instruction or interventions for those students who struggle.
- Multilevel approach that is adjusted & modified as needed.
- Special case of Multi-tiered System of Support (MTSS)



Response to Intervention (RTI) A Case of Adaptive Design

- Think of the RTI framework as a pyramid:
 - Tier 1: Research-based core instruction
 - Base or primary level of prevention/intervention
 - Most commonly used teaching strategies & interventions
 - Tier 2: Targeted intervention
 - Middle or secondary level
 - Interventions are more intensive bc the students are considered to be at greater risk
 - Tier 3: Intensive intervention
 - Top or tertiary level
 - Students receive the most intense and consistent interventions



Response to Intervention (RTI) A Case of Adaptive Design

- Screening:
 - Students are tested to determine their baseline & identify weaknesses
 - Cut points or cut scores are used to determine whether additional testing or intervention is needed
- Data-based Decisions:
 - Data is used to determine intensity & duration or any needed intervention
- Monitoring:
 - Assess; keep records; monitor student progress & responsiveness



SMART

A Method to Develop Adaptive Interventions

- <u>Sequential</u> <u>Multiple</u> <u>Assignment</u> <u>Randomized</u> <u>Trial</u>
 - Provides high-quality data that can be used to *construct* adaptive interventions.
- Multiple intervention stages
 - Each stage corresponds to a critical decision in adaptive intervention
 - Each participant moves through multiple stages
 - At each stage, Ss are <u>randomly</u> re-assigned to one of several intervention options
 - Allows valid causal inferences
- Used to **DEVELOP** adaptive interventions rather than **EVALUATE** whether it is better than control
 - Could be used to empirically develop tiered interventions like RTI and MTSS
 - Should be followed by a RCT
- Example
 - Stage 0: multiple intervention options are IDed and ranked in order of increasing intensity and/or scope (say A, B, C, D)
 - Stage 1: Ss randomized to 2 or more possible initial interventions (say B, C)
 - After X period of time, Ss are classified as either responsive or nonresponsive
 - Stage 2: nonresponders re-randomized to 2 or more treatment conditions that are as/equal or more intensive than the current condition
 - Stage 3: repeat stage 2
 - Stage 4: etc.



SMART: An Example

- The next slide illustrates one potential SMART design for evaluating adaptations within a hypothetical home-school consultation intervention that targets parents with low engagement in their child's behavioral plan.
 - This design would be useful for optimizing adaptations related to different courses of action.
- Four adaptive interventions could be compared:
 - Start with *in-person* consultation;
 - continue with <u>in-person</u> consultation if the parent is responding to the intervention
 - <u>intensify in-person</u> consultation if the parent is not responding;
 - Start with <u>in-person</u> consultation;
 - change to *distance-based* consultation if the parent is responding
 - <u>intensify in-person</u> consultation if the parent is not responding;
 - Start with <u>distance-based</u> consultation;
 - continue with *distance-based* consultation if the parent is responding
 - change to <u>in-person</u> consultation if the parent is not responding;
 - Start with <u>distance-based</u> consultation;
 - continue with *distance-based* consultation if the parent is responding
 - change *to intensified in-person* consultation if the parent is not responding.
- Upon identifying the optimal adaptive intervention, the efficacy of the intervention should be tested by comparing it to a control or BAU group within the context of a traditional RCT.

From: Koziol, N.A., Witte, A., & Garbacz, S.A. (2018). *Research design and evaluation*. Unpublished manuscript.



SMART: An Example



From:

Koziol, N.A., Witte, A., & Garbacz, S.A. (2018). *Research design and evaluation*. Unpublished manuscript.

MAP ACADEMY

SMART vs Sequential/Adaptive Designs

- Adaptive experimental design (versus SMART)
 - Multistage design
 - Accumulating data are used to decide how to modify aspects of the study while preserving validity & trial integrity
- STAGE
 - SMART: intervention stage corresponding with a critical care decision and a new intervention level
 - SEQ: experimental stage that involves new Ss
- ADAPTIVE
 - SMART: intervention is adapted based on response to prior intervention
 - SEQ: randomization probabilities are adapted based on other Ss





ADAPTIVE EXPERIMENTAL DESIGNS



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



Sequential Designs

- Also referred to as *adaptive* or *flexible* designs
- Current design decisions are sequentially selected according to previous design points
- Fixed design = sample size and composition determined a priori
- Sequential design = the number of observations/participants is not predetermined
 - Sample size and composition are considered random due to decision dependence on previous observations.
 - A finite upper limit is often set in practice.



Primary Benefits of Sequential Designs

- Allow for early termination of experiments if cumulative evidence suggests a clear effect or lack thereof
- Ethical perspectives:
 - Prevent unnecessary exposure to unsafe experimental conditions in terms of both length of exposure and the number of participants exposed
 - Prevent unnecessarily withholding administration when the experimental condition is clearly beneficial
- Logistical perspectives:
 - Financial savings due to reduced sample sizes
 - Fail to Reject H₀: early termination for lack of effectiveness at a total sample size smaller than would be the case with a fixed design
 - Reject H₀: a similar savings is observed in the total sample size required,
 - Sample size savings typically reported as greater under H_A than under H₀
 - Actual sample savings generally reported to be as large as 10% under $\rm H_0$ & as large as 50% under $\rm H_A$



Sequential Design 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,
 - # 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.



Multiplicity

- Traditional fixed design methods for determining critical or boundary values cannot be applied in sequential designs due to inflation in the Type I error rate.
- Armitage, McPherson, & Rowe (1969) showed that repeated significance tests at a fixed level on accumulating data increase the probability of obtaining a significant result under H₀
 - With α = .05 in a 2-tailed fixed-sample test, the z_{cv} = ±1.96
 - 2-stage sequential design carried to completion, resulting in the same N, nominal $\alpha = .083$
 - If $z_{cv} = \pm 1.96$ is used in a 5-stage group sequential trial with early stopping to reject H₀, then $p(\alpha) = 0.14169$ to reject H₀ at or before the 5th stage
- 3 general boundary method categories:
 - Fixed boundary shape methods
 - Whitehead methods
 - Error spending methods



Boundary Values

- Similar to conventional critical values are set up for each interim stage
- Derived to maintain experiment-wise Type I and Type II error rates at specified nominal levels.
- Up to 4 boundary values are derived *a priori* depending on the number of tails in the statistical hypothesis test.
 - 1-tailed test: up to 2 boundary values (1 FTR H_0 , 1 Reject H_0)
 - 2-tailed test: up to 4 boundary values (2 FTR H_0 , 2 Reject H_0)
- If the test statistic:
 - Exceeds the upper critical value, H₀ is rejected for efficacy
 - Exceeds the lower critical value, H_0 is rejected for a harmful effect
 - Falls within the acceptance region, FTR H₀ and the experiment stops
 - Falls in between the rejection and acceptance regions, the experiment continues to the next interim stage.
- At the last stage, the boundaries of the rejection and acceptance regions are identical, the experiment stops, and the decision is made to either reject or fail to reject H₀.
- If the protocol only allows early stopping to reject H₀ no FTR H₀ boundaries at early stages
- With early stopping only to FTR H₀, there are no interim rejection boundaries.



Boundary Methods

- Fixed boundary shape methods
 - Pocock method
 - Used when equally spaced information levels derive a constant boundary value
 - Nominal alpha level is smaller than the desired α = .05 level
 - Overall design more statistically conservative.
 - A constant boundary value at all 5 stages would be 2.41 rather than 1.96
 - Nominal alpha level at the final stage ($\alpha = .032$) would be smaller than the overall alpha-level of the design ($\alpha = .050$).
 - O'Brien-Fleming method
 - Requires overwhelming evidence to reject the null hypothesis at early stages by implementing initially conservative, but successively decreasing, boundary values.
 - For instance, a 5-stage design would use critical values of 4.56, 3.23, 2.63, 2.28, and 2.04 where the final critical value is then close to the fixed design critical z-value of 1.96.



Boundary Methods (2)

- Whitehead methods
 - Appropriate for designs with discrete monitoring
 - i.e., after groups of participants complete the study
 - Generalized from designs requiring continuous monitoring

 i.e., after each participant completes the study
 - Result in triangular or straight-line boundaries
 - Require less computation that the fixed boundary shape methods
 - Maintain Type I error rates and statistical power that is close to, but does differ slightly from, nominal values.



Boundary Methods (3)

- Error spending methods
 - Requires the most computation when a large number of interim stages are utilized.
 - Relative to Whitehead and fixed boundary shape methods
 - Use an error spending function
 - Function specifies an amount of the pre-specified nominal α and β rate error to be spent at each stage
 - Boundary values derived from this amount.



1-Sided Boundary Plots











2-Sided Boundary Plots











- 3 General Types:
 - Fully sequential designs
 - Group sequential designs
 - Flexible sequential designs
- Differ based on sample recruitment and decision-making criteria.



• Fully sequential designs

- Continuous monitoring updated after every observation or after every participant completes the study
- Critical boundary values change as the experiment progresses
- Require that the outcome of the experiment is knowable in a relatively short period of time relative to the length of time required for recruitment or follow-up.
- Long periods of "no effect" or continuation might be falsely interpreted as equivalence of experimental conditions.
- Originated w/ Wald's *sequential probability ratio*, but not widely applied historically
 - EXCEPT computerized adaptive testing (CAT) & computerized classification testing (CCT)
 - question selection depends upon the response to the previous question
 - Each question is considered an observation
 - The test (i.e., the experiment) concludes when a pre-specified degree of precision or consistency of responses is attained rather than a statistically significant difference between two experimental conditions.



- 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
 - Likewise, fully sequential designs can be viewed as group sequential designs where the size of the participant group is n = 1.
 - More practical than fully sequential designs
 - Evaluation after each participant completes the study may not be feasible
 - Early termination after small n may not be viewed as persuasive evidence
 - 4-5 interim stages or data inspections are typically recommended
 - larger number of interim analyses do not lead to extra benefits



- Flexible sequential designs
 - Allow flexible modification during the experiment through an alpha spending function
 - Allocates the amount of the nominal experiment-wise error to be spent at each interim
 - Maintains the overall nominal Type I error rate at the α = .05 level
 - Alpha spending function is specified during the study design phase
 - Cannot be changed mid-experiment
 - Protects against potential researcher abuse
 - Allow periodic interim evaluation as do other types of sequential designs
 - But flexible designs also may allow the interim evaluations to become more frequent as the decision point becomes closer
 - Can be viewed as a compromise between fully sequential and group sequential designs



Benefits of Sequential Designs

- Ability to make use of existing information at interim stages of the experiment to inform future design decisions, especially in regards to participant recruitment.
- When the researcher has a limited amount of theoretical or empirical knowledge that would otherwise prevent well-informed decision-making re: optimal design of experiments
 - Traditional (fixed) designs are planned to have sufficient power to detect an effect size specified a priori
 - To the extent that this effect size is incorrect in the population, the researcher has inadvertently over- or under-powered the study
 - Early stopping in the event the effect size is larger than previously estimated
 - Probability of early study termination is higher and savings may be obtained in terms of decreased resource allocation, regardless of the general type of sequential design employed.
 - Further data collection in the event the effect size is smaller than expected
 - Fully sequential and group sequential designs may result in a decrease in power due to the additional analyses
 - Can be ameliorated to an extent by employing a flexible sequential design



Benefits of Sequential Designs (2)

- Some may argue that adopting a flexible design from the onset of a study is a reasonable approach when:
 - The magnitude of the true treatment effect is not clearly known
 - If there is a discrepancy between a clinically-meaningful effect and an observable effect
 - If it is not clear that the potential effectiveness warrants using up what are otherwise limited resources



Sequential Designs & RCTs

- Sequential designs are particularly applicable to medical and pharmaceutical trials generally referred to as randomized clinical trials, or RCTs
 - Type of *randomized experiment* in which the effectiveness of a new drug, intervention, or other medical procedure is evaluated by comparing a treatment group with a control group.
 - Conducted according to a plan, or *protocol*, which details the RCT's objectives, data collection procedures, and data analysis framework
 - RCTs typically require a *sequentially recruited participant stream* and lend themselves well to sequential designs
 - Particularly useful due to the need for *trial monitoring* to minimize exposure to unacceptable toxicity or potential harm of new interventions or to minimize continuation after the benefit or risk is clearly apparent
 - Particularly when the clinical outcome is considered irreversible, fixed designs are not acceptable, thus the need for monitored sequential designs.



Sequential Designs & RCTs (2)

- Flexible sequential designs are particularly useful in Phase II clinical trials.
 - Phase I trial is used to determine whether a new drug or treatment is safe for humans and to estimate an initial effect size
 - Phase II trial is conducted on a larger sample of participants to determine how well the drug works while continuing to monitor participant safety
 - Phase I effect size estimates are sometimes found to be too optimistic, so a Phase II sample size based on an inflated effect size may be smaller than required
 - A flexible sequential design would allow for interim sample size augmentation to accommodate the smaller-thanexpected effect size.



Limitations of Sequential Designs

- Increased design complexity
- Increased computational burdens
 - Determining boundary values
 - Controlling the experiment-wise error rate
- Threat to validity due to ability for early termination
 - Early termination for efficacy, futility, or participant safety
 - Smaller sample sizes can lead to a distrust of the findings
 - Some analytic assumption problems due to asymptotic principles (i.e. ML)
 - Early termination decision is more complex than just a statistical criterion
- Consistency across both primary and secondary outcomes, risk groups, etc.



Group Sequential Clinical Trials

- A group sequential trial generally involves 6 distinct steps:
 - 1. Specify statistical details of the design
 - $H_0 \& H_A$, test statistic
 - Type I & Type II error rates
 - stopping criterion
 - # stages
 - relative information level at each stage
 - 2. Compute the boundary values & required N at each stage based on the design specifications
 - 3. At each stage, additional data with the required sample sizes are collected
 - 4. At each stage, the available data (data collected at the current stage as well as previous stages) are analyzed & the test statistic is computed
 - 5. At each stage compare the test statistic with the corresponding boundary values.
 - Either the trial is stopped to reject or fail to reject H₀, or the trial is continued to the next stage
 - If the trial continues to the final stage, H₀ is either rejected or you must fail to reject
 - 6. After the trial stops, compute parameter estimates, confidence limits for the parameter, and a p-value for the hypothesis test.



Substantive Context: CBC in the Early Grades

- Significant behavioral problems of student with disruptive behaviors are often both a result of and contributor to ineffectual parent-child interactions.
- Negative behaviors exhibited by such students are manifest across the multiple settings in which they function (i.e., school, home, community).
- Thus, the responsibility for children's education and behavioral well-being is dispersed across multiple systems (i.e., schools, families, and community support systems).
- However, families of students with social-emotional and behavioral problems are often poorly connected with schools or other service-delivery systems (Dishion & Stromshak, 2006).
- Interventions that support family members as *partners* in (rather than merely recipients of) services improve retention, satisfaction, and levels of active participation in treatment (Hoagwood, 2005).
- The identification of strategies that link families of children with severe behavioral challenges with high quality treatments through a partnership approach (i.e., bringing together major systems influencing children's behavioral functioning) is the focus of **conjoint (family-school) consultation**.



Methods: Participants

- 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
 - 207 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).
 - Fixed sample size of N = 270 children (k = 90 classrooms w/ 3 kids/class) was determined through an *a priori* power analysis using Optimal Design.



Methods: Procedure

- 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?
- All analyses implemented in SAS:
 - PROC SEQDESIGN design the boundary values
 - Essentially an *a priori* power analysis
 - PROC MIXED analytic model
 - Can be whatever program or procedure is appropriate given the study design and the hypotheses being tested
 - PROC SEQTEST evaluate analytic results based on boundary values
 - Similar to PROC MIANALYZE synthesizes the multiple results



Results

ę	3	/ 🛄 🔼 48 🖂 💱 🔨 +z +a 🔚 🕶 🛍 🖽 🔛 😎										
			Index1	Estimate	Standard Error	Variable	_Scale_	_Stage_	Nobs			
		1	BPAS	0.3372	1.8021	TxG	MLE	1	25			
		2	BPAS	-0.02598	1.1469	TxG	MLE	2	54			
		3	BPAS	-0.4052	0.9013	TxG	MLE	3	80			
		4	BPAS	-0.3599	0.8802	TxG	MLE	4	90			





Adaptive Skills: Parent vs. Teacher Reports





Externalizing Behaviors: Parent vs. Teacher Reports





Parent-Teacher Relationship: Parent vs. Teacher Reports





Social Skills: Parent vs. Teacher Reports





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)	· ·	21	(N _{teach} =25)	(N _{teach} =54)	(Nteach=80)	(N teach=90)	
PASC	Est.	0.34	-0.03	-0.41	-0.36	BASC – Adaptive Skills (Teacher)	Est.	3.63	2.59	2.39	2.20	
DAGU -	SE	1.80	1.15	0.90	0.88		SE	1.41	1.12	0.91	0.89	
Adaptive	t	0.19	-0.02	-0.45	-0.41		t	2.57	2.32	2.63	2.46	
Skills	P	0.43	0.49	0.33	0.34		P	0.01	0.01	0.00	0.01	
(Pareni)	Dec.	Continue	Accept H ₀				Dec.	Reject Ho		2274 Providen		
PASC	Est.	-3.11	-0.55	0.58	0.39	BASC – Externalizing Behavior (Teacher)	Est.	-0.82	-1.02	-2.23	-2.43	
Externalizing	SE	2.31	1.43	1.27	1.26		SE	2.11	1.59	1.29	1.27	
Externalizing	t	-1.35	-0.39	0.46	0.31		t	-0.39	-0.64	-1.73	-1.91	
(Dament)	P	0.09	0.35	0.32	0.38		P	0.35	0.26	0.04	0.03	
(Parent)	Dec.	Continue	Accept Ho				Dec.	Continue	Continue	Continue	Reject Ha	
0000	Est.	-1.05	0.33	1.73	0.94	SSRS – Social Skills Score (Teacher)	Est.	3.13	3.16	3.90	3.72	
SSKS -	SE	3.22	2.05	1.91	1.89		SE	2.55	1.98	1.51	1.49	
Social Skills	t	-0.33	0.16	0.90	0.50		t	1.22	1.60	2.58	2.50	
Score (Darant)	P	0.37	0.44	0.18	0.31		P	0.11	0.06	0.01	0.01	
(Parent)	Dec.	Continue	Accept Ho				Dec.	Continue	Continue	Reject Ha		
Derent	Est.	0.03	0.05	0.05	0.03	Parent- Teacher Relationship (Teacher)	Est.	0.13	0.23	0.20	0.19	
Parent-	SE	0.24	0.14	0.11	0.11		SE	0.15	0.09	0.08	0.07	
Peletienship	t	0.13	0.34	0.45	0.26		t	0.82	2.69	2.67	2.57	
Relationship	p	0.45	0.37	0.33	0.40		P	0.21	0.00	0.00	0.01	
(Parent)	Dec.	Accept Ho	Bright Servers				Dec.	Continue	Reject Ho			
			-							•		



Results

- Fixed design results indicate significant treatment effects on some outcomes, but not all.
- Same conclusion met for all outcomes under sequential analysis.
 Average sample size under sequential design: 55 classrooms/teachers
- Capacity for early termination can be beneficial prevents unnecessary exposure or withholding & can save both time and resources
 - 39% reduction in the # of teachers



Discussion

- Limitations
 - Sample must be *sequentially recruited* & the experiment must be short
 - Increased design complexity & computational burdens
 - Threat to the validity of the study due to early termination
 - May be some statistical instability due to smaller sample sizes (largesample assumptions of asymptotic principles in MLM)
 - Requires consistency across both primary and secondary outcomes, risk groups, etc.
- <u>Future Directions</u>
 - Consider classroom/teacher as group rather than cohort



Thank you!

jbovaird2@unl.edu http://mapacademy.unl.edu https://cehs.unl.edu/edpsych/faculty/jamesbovaird



