Attendees:

- **RAC members:** Gary Henry, Linda Espinosa, Steve Barnett, Marie Kanne Poulsen, Mark Appelbaum, Vera Gutiérrez-Clellan, Peg Burchinal, Eboni Howard, Floraline Ingram Stevens, Faith Polk, Barbara Goodson, David Elkind, Alice Kuo, Carrie Rothstein-Fisch

- **First 5 LA:** Katie Fallin, Patricia Lozano, Christine Ong, Mike Lopez, Karen Blackeney

- **LAUP:** Gary Mangiofico, Celia Ayala, Relda Robertson-Beckley, Kimberly Hall, Jessica Brauner, Ama Atiedu, Julia Love, Rocele Estanislao, Lisa Shimmel, Thalia Polychronis, Alex Zepeda

- **Mathematica Policy Research:** Cheri Vogel, Sally Atkins-Burnett (by phone)

**Introductory remarks - Katie Fallin and Mike Lopez (First 5 LA):**

- **Purpose of this RAC meeting:**
  - Share the work that has been accomplished since our meeting in July of last year, including the initial data from the Fall 2007 baseline data.
  - Solicit RAC members’ input on potential future directions for preschool-related research and evaluation efforts that build upon and expand the work of the UPCOS study.

- **RAC role (reminder):** provide us with advice, input and best ideas; RAC is not a decision-making body.

- **First 5 LA’s contextual issues:**
  - Mentioned state budget deficit and legislative proposals to use Proposition 10 (First 5) funding
  - LAUP funding decision – longer term investment but will require other sources of funding in order to sustain.
  - Pressure to have evidence of impact (sustainability of LAUP and First 5 LA as a whole) The context in which the study was designed called for a more formative approach that should yield information of importance to First 5 LA and LAUP moving forward.

- **Review of last RAC meeting:**
  - Outgrowths from the pilot study work and our last meeting:
    - Measurement feasibility and measurement development work
      - Language routing methodology (head start research conference posters)
      - Respect for differences scale
      - LISn development
    - Further investigate the psychometric properties of the DRDP-R
      - Hired a 3rd party contractor (Westat) to provide additional expertise and input on this component
      - LAUP enrollment form update
  - LAUP Enrollment Form and Child Level Tracking System Update
**Introductory Remarks - Gary Mangiofico (LAUP):**  
Impact of recent funding decisions based on negotiations between First 5 LA and LAUP:  
- LAUP moved to a shorter school year system (180 days of funding over 10 months) vs. a year-round funding approach (carry over funds to fund summer program for this year only).  
- CDE: try to get unused money from State Pre-K program to be a “super grantee”  
- “Grant funding” for special initiatives or enhanced programs.  
- Exploring potential options to create “for profit” program operations (i.e., establish some preschools that make profits)  
- Systemic Improvement System –IF coaching  
- Decreased enhancement and subsidy rates  
- Talked about “enrollment committee” – focusing more on helping providers improve their outreach, recruitment and enrollment efforts

**Fall Study Preliminary Findings – Presentation by Cheri Vogel (MPR, Inc.)**  
RAC members expressed the following questions and/or concerns:  

**Language of the assessment:** Having only the English version for the ROWPVT in the spring data collection, in combination with the EOWPVT-Bilingual version. They suggested including a rationale within report about why we are using the ROWPVT English for the Spring and what this means for the related analytic strategy. We should also explain the trade off of giving up the option of looking at Fall to Spring gains in one of the measures (the ROWPVT). It was pointed out that the decision to not continue the conceptually scored receptive language measure this spring was based on RAC input in July 2007. At that time, there was impetus for the study to place greater emphasis on learning about children’s acquisition of English and preparation for the English language instruction they will receive in kindergarten.  

- Follow-up: Check correlation between EOWPVT and the ROWPVT.  
- Within the report, will also want to report on whether/how many assessors are native speakers, number of children requiring Spanish track in Fall, Spring.  
- Also will be important to place the UPCOS findings in the context of LAUP not requiring either a specific language of instruction or any specific curricula.  
- **Household definition:** there may be different number of adults living in a “household”. What does that mean for different families? The MPR team clarified that it takes many questions to get a more detailed understanding of these issues.
RAC suggestion:
- Add a question about broader extended family (and relevant social) networks for families in future studies.

Enrollment: It’s known from previous studies that the most advantaged kids will participate in high quality preschool programs. Disadvantaged families are least likely to enroll when they are not preferred or targeted specifically in some way.

AGN /NON-AGN: Some confusion with the term “AGN”. There is a tendency for people to equate the term AGN with income-based need at the individual family level versus the intended use of the term to represent the level of early education/care service need at the community-level. Thus, we may need to find a term that better captures the intent.

RAC questions/suggestions:
- Where are the AGNs and how do the kids in those areas compare to children who are not in LAUP. (Perhaps use Census data?)
- Since AGN was based on service availability, the non AGN areas potentially include areas that offer more means tested programs (such as State Prek, and Head Start).
- Do we know if kids in AGN programs live in these AGNs?
- Is it possible to compare demographics of families served by different types of programs (e.g., programs under different auspices, such as Head Start vs. CDE, etc.) within same AGN?
- AGNs may not be poorest of poor areas, since the AGN represents relative service need and not average family income. Is it possible that these Tier 2 areas are LAUP’s real niche?
- Do we know if some families were reluctant to participate in either the study or LAUP program due to their legal status? If so, what are the implications for the present study?
- How are waiting lists used in LAUP programs? (Since LAUP isn’t a means-tested program, they are not given priority on waitlist?)

Low rates of parental reported depression: Mexican-Americans/Latinos are less likely to report depression.

Health insurance: The percentage of children with insurance is higher than anticipated.
RAC comments/suggestions:
• Are the 4% of kids w/ Healthy Kids insurance from higher income families (versus lower-income and/or undocumented families that make up most of Healthy Kids’ client base)?

• **Add a question in the future:** When there is an illness, where do you go (to get sense of access to different types of health facilities, especially since percentage of children reporting health insurance is so high)?

• Performance of other minorities: add more data about the performance of African American children on the Pre LAS.

Discussion:

Power Analyses:
• Degree to which we met the power analysis assumptions. The more important question is what power we have now, regardless of what our assumptions were. Cheri responded to Gary H that we were selecting subgroups in part to ensure we have large enough Ns in each cell. We will also do the power analysis based on the sample we obtained.
• Power might be very low across some of the subgroups.
• Interclass correlations may be higher than what we initially assumed, contributing to lower power.

Fall to Spring
• Are fall to spring gains better than what we expect from maturation alone?
• MPR might use Covariate Adjusted Model, against a norm (with the exception of the Executive Function measures that don’t have published norms).
• RAC raised concerns about doing HLM modeling. Concerns about self-selecting and the amount of “noise” in the data.
• Are we interested in looking at the comparison of effect sizes across measures? As a function of child, program, and family characteristics?
• We should characterize whether different children benefit more or less from being in the program
• What program characteristics seem to make a difference?
• What is the time lag between fall & spring data collection?

Two levels of descriptive questions:
- Are certain subgroups benefiting more or less?
- Correlations between program characteristics and child measures over time
- Determine in what domains and/or on which measures are kids progressing/advancing relative to the norms?
- Descriptions of programs also helpful
- Also need to make it clear that the design of the study does not allow us to make inferences about causality, but only associations. Note: We should be cautious about comparing to norms (First 5 LA will double check this comment to be sure that we captured it correctly)

Another option for data analysis:
- **First Level:**
  - What was the learning and development of the kids overall and by subgroups (What happened to them?, no inferences about causal relations)

- **Second Level:**
  - Relationship between program characteristics and kids’ learning.
    - There’s a concern about these analyses. The ideal would be to do both a well defined comparison with propensity scores matching analyses, as well as regression adjusted analyses.
    - An example for LAUP would be: Star Ratings above and below match kid’s outcomes in above and below.
    - There is an example of this combination in work by Glazerman, Myer and Levy.
    - Using just regression adjustments is not strong.
    - It’s important to delineate a set of questions in advance for program/research. STAR ratings or quality alone may not yield much (based on more recent research).

Areas to look at:
- **Language of instruction:** We should consider ELL kinds of questions like what characteristics of the pre-k environment support ELL children. Using language interactions and language of instruction.

- Are there enough Spanish Primarily Programs? There are some concerns about the variability of the data (Observational data from the LiSn vs. teacher self-reported data).

- **Curriculum/quality of instruction:** We might think about using the LiSn to cross validate the teachers’ reports since the LiSn data is only available on a subsample versus teacher reported data from all teachers. Also consider comparing the LiSn with the CLASS.

**RAC members suggested the following:**
- Create a subcommittee once we have some data on the variability within the sample.
• Look at means and distributions of the different variables of interest first to help guide the analytic discussion.

• Look at measures of quality and variability (such as in language instruction) and do the best job at taming selection bias.

• Plan qualitative analysis about children. Identify programs now for these analyses (i.e., those that “look different” based on data).

• Look at language facilitation or language that the teacher uses in the classroom and analyze its relationship to gain during preschool year.

• Be cautious about “diluting” data by conducting analyses without the power to detect differences.

• Include confidence intervals and standard errors.

RAC members also suggested to:
• Think about topics that people care about like school readiness and prioritize the information that allows us to weave a story about data that is linked to later outcomes.

• Prioritize the following:
  ▪ Outcome Measures questions
  ▪ Program characteristics questions
  ▪ Variability
  ▪ Analytic questions: how to proceed – focus on the most strategically important results.

• Share an updated analysis plan with the RAC before a conference call (use WEB-X)

Gary Henry and Bill Gormley’s Presentations:

Georgia’s Pre-K Evaluation (Gary Henry):

There were 4 purposes of the evaluation:
• Assessment of merit and worth
• Program improvement
• Accountability and oversight
• Knowledge and development

• They wanted to see how much program has grown over time (not fully implemented so not appropriate to evaluate for impact).
• 4 instruments were used overtime (norm referenced): PPVT, WJ LW-ID, OWIS, WJ-AP. They wanted to follow kids to third grade to look at predictive validity but funder was not willing to do it.

• Sample: - All able to be assessed in English
  - Small probability sample
  - Programs didn’t have an inclusion component

• Assumption was that kids would be worse off in Pre-K than in Head Start.

• Pre-K comparison to Head Start: A subset of children from Pre-K who were eligible for Head Start were selected and matched on 72 variables (they did one to one matches, propensity scores and/or weighted analysis).

• Kindergarten stratified first and then Elementary Schools (stratified by County)

• Got more out of using the norm-referenced measures than rescaling measures to generate connections for policy audiences.

Considerations for LA:

Design and Approach:

• It’s better to use longitudinal data or wait until stable program (cohorts) because program keeps changing.
• Counterfactual problem: there is no comparison group so we cannot do causal attributions.

Measurement and Method:

• Implementation variation (curriculum)
• Implementation fidelity
• We should consider having some consistent measures across all years rather than adding better measures later.
• Measures should be consistent and age appropriate.

Logistical/Practical:

• Establish intellectual and property rights for publication and dissemination
• Consider the nature of the evaluation and sponsor relationship (the relationship needs flexibility as project evolves)

Oklahoma (Bill Gormley)

There were three waves of data collection: Fall 2001, Fall 2003 and Fall 2006.
Characteristics of the study:

- Home grown measure (created by Tulsa Board of Ed.) used for the first two waves of data collection
- Regression Discontinuity Design using data from schools – the advantage of using this design is that it eliminates selection bias. The conditions were ideal in Tulsa for implementing such a design (based on the strict adherence to the age eligibility cut-off date and the existence of an established assessment process).
- Always tried to complete the assessments before the academic year started

During Fall 2001:
- Focus on impact from the beginning: Impact on school readiness was bottom line.
- Used Home Grown Measure
- There were concerns about ceiling effects for middle class kids (in the home grown measure)

During Fall 2003:
- They used Woodcock Johnson Letter Word Identification, Spelling and Applied Problems (only English), the home grown survey and a parent survey.
- Results showed that all groups benefited substantially (all incomes, all ethnic groups)
- Tried to use Head Start Data but could not use it in the end.

During Fall 2006:
- Added Woodcock Munoz to test Spanish speaking children
- Visited every Pre-K classroom
- Used CLASS and the Emerging Academic Snapshot
- Socio-emotional measure from University of Pennsylvania -ASPI (developed by John Fantuzzo and colleagues).
- Captured 80-85% of the universe of eligible children.
- They found that teachers with no BA were more lenient than those with a BA.
- Added Head Start data
- 5000-6000 children participated

Three Analysis Techniques used:
1. Regression Discontinuity Design: the best quasi-experimental design to use. This method was used in year three of program implementation.
2. Propensity Score Matching:
3. Ordinary Least Squares
**The last two methods underestimated the impact when compared to Regression Discontinuity.**

Ethical/other Issues to consider:
- Whether to release findings at all given problems with the Head Start data and concerns about the potential to misuse the findings
- Must first “do no harm” due to concerns that even when findings are presented with many disclaimers, they may still be used to make policy decisions
- How much to criticize your own findings
- How to explain idea of “regression discontinuity” design to public
- Where to make cut-off for regression discontinuity (how far out from the specific cutoff age to go).

Lessons Learned:
- Prioritize merit and worth and net impact.
- Gary H. would have used a different design in Georgia to identify net effects
- More children are being served in Georgia and costs are lower because they found that an Associate Degree is good enough.
- In Tulsa:
  - Would have liked better measures like the ones we have in UPCOS
  - More long term results
  - Advocates and public officials have used findings to fight for UPK and for additional funding.
  - Full day programs provided greater benefits than half day but full day children were also more disadvantaged. Tulsa’s school system interpreted this to advocate for full day kindergarten.

Advice for UPCOS
- Most important elements to look at should be things that are modifiable and manipulated by program administrators.
- We should also look at teacher characteristics and their relationship with pedagogical choices (there is no consistent relationship between teacher qualifications and classroom quality in the most recent literature – see meta-analysis from FPG, GA)
- Focus on what the UPCOS data can and cannot address.
- Remember the primary focus of the study.
- Try to identify what’s most important from the quality data.
- Peg recommended looking at monthly learning gains controlling for age – taking into account that it might require some strong assumptions.
• Explore the possibility of cross-validating LAUP’s Quality Improvement Plan with UPCOS data - always considering the issues of having coaches as reporters and how this could put the coach in an awkward position. It will be important to first look at what data is available from the QIPs and whether it is systematically available across all programs.

• Look at the 20 classrooms that were observed using the LISn and analyze the language of instruction changes. Also, can explore the use of the LISn data from the subgroup to validate the teacher-reported data from the complete sample.

• Further investigate what is happening in the classroom around language – considering variability within classrooms.

• Include confidence intervals on all estimates.

• Collect school records two and three years down the road to identify how kids are doing in school – might be difficult to track this data without children’s Social Security numbers. However, it was noted that SSN for this group may be less informative than for others, given that SSN is likely to be correlated with other family characteristics.

• Use the data to understand how kids are doing overall and by subgroups (taking into account that power might go down for the subgroup analyses). We can then ask “how kids are doing” across domains & by characteristics, where are we happy with their progress, where do we feel more should be done, etc.

• Keep the option of a longitudinal study open, or at least one follow-up wave of data collection.

• Be very careful not to just correlate program characteristics with outcomes. (Such investigations may be premature since the current LAUP model may not be well understood/characterized).

• Look at the UPCOS sample and compare it with others served in the same conditions.

• Create a document for external audiences answering some very basic questions like: What is preschool in LA County? Who is being served? What is happening in these LAUP classrooms?

• Be sure of taking into consideration family needs and support.

• Think more about who is being served by the LAUP programs? How does this relate to the original master plan and whether or not the intent was for LAUP to serve those who are considered most needy (and what definition of need is)? What do we know about the characteristics of children already
being served by other available programs? Do we know where other kids not served by LAUP (but who are eligible) are?