Spring Semester 2006
SPED 3018
Observational Methods
201 Wyatt
Wednesdays 1:10 – 4pm

Course description:
This doctoral level course addresses what is known about quantitative, systematic observation of behavior to measure behavior that may or may not be used to infer status on psychological constructs in single subject and group research designs. The content emphasis is on providing students with the rationale for selecting among the many options at all stages of observational measurement. Among the topics covered are (a) classical measurement theory and Generalizability theory as they relate to observational measurement, (b) principles for selecting measurement procedures, selecting behavior sampling methods, designing coding systems, selecting appropriate metrics (including nonsequential and sequential variables), (c) sequential analysis of behavior, (d) the tension between ecological validity, representativeness, and construct validity, (e) interobserver reliability issues, and (f) other issues related to the direct observation of behavior.

Course objectives:
1. Describe quantitative, structured observational research is and how this differs from qualitative and unstructured observational research.
2. Describe the general steps to structured observational research.
3. Distinguish between descriptive questions and falsifiable research questions.
4. State your falsifiable research questions
5. Identify whether your variables are constructs or behaviors.
6. Indicate why you are or are not inferring past the physical attributes of the behavior.
7. Provide a rationale for selecting an observational measurement procedure to measure variables (and, if appropriate, underlying constructs).
8. Give rationale for selecting a standardized vs. naturalistic measurement procedure/context.
9. Indicate what is known about the relative validity of structured vs. unstructured procedures and lab vs. home settings?
10. Indicate whether you are attempting to measure individual differences between people that you assume exist outside of your measurement context.
11. Give the rationale for selecting among (a) behavioral counting, (b) checklist or (c) rating system.
12. Describe a decision study (D study) and when such a study is necessary.
13. Describe and give rationale for the type of behavior sampling method and recording medium (e.g. computer or paper pencil), and coding system you selected.
14. Describe the capabilities of Procoder DV.
15. Indicate what type of variable you plan to derive: sequential or nonsequential?
16. If nonsequential observational variables are planned, indicate which metric you

will use in your analysis (e.g., frequency, duration, rate, proportion, compound proportion, etc) and describe rational for decision.
17. If sequential analysis, describe your intended variable derivation process.
18. Describe the training of a second coder to reliably use your measurement system
19. Give your plan for achieving interobserver reliability.
20. Indicate how you will test the construct validity of your observational variable.
21. If you use a standardized procedure, how can a variable from a standardized procedure be construct valid if it is not "representative" of what the individual(s) usually do(es)?

Honor code

Violations of the Honor Code are cause for disciplinary actions imposed by the appropriate honor council.

The following are included as violations:

- Plagiarism on an assigned paper, theme, report, or other material submitted to meet course requirements. Plagiarism is defined as incorporating into one's own work the work or ideas of another without properly indicating that source.
- Failure to report a known or suspected violation of the Code in the manner prescribed.
- Any action designed to deceive a member of the faculty, a staff member, or a fellow student regarding principles contained in the Honor Code, such as securing an answer to a problem for one course from a faculty member in another course when such assistance has not been authorized.
- Any falsification of class records or other materials submitted to demonstrate compliance with course requirements or to obtain class credit, including falsifying records of class attendance.
- Submission of work prepared for another course without specific prior authorization of the instructors in both courses.
- Use of texts, papers, computer programs, or other classwork prepared by commercial or noncommercial agents and submitted as a student's own work.

Attendance policy

Students are expected to attend all class periods. Absences need to be cleared by the instructor ahead of time, if at all possible. Unavoidable events that require an absence without prior approval need to be addressed with documentation of the precipitating event.

Required readings.

Text: Bakeman, R. & Gottman, J. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University. Available at Rand bookstore.

Other required readings are available through Peabody library electronic reserve service. http://eres.library.vanderbilt.edu/cgi-bin/viewcopyright.pl?SPED3930_YODER
Password: yode3930

Baer, D. (1977). Reviewer's comments: Just because it's reliable doesn't mean that you can use it. *Journal of Applied Behavior Analysis, 10,* 117 – 119.

Cone, J. D. (1982). Validity of direct observation assessment procedures. In D. Hartmann (ed.), *Using observers to study behavior* (pp. 67 – 79).

Gardner, F. (2000). Methodological issues in the direct observation of parent-child interaction: Do observational findings reflect the natural behavior of the participants? *Clinical Child and Family Psychological Review, 3*, 185 – 198.

Jacobson, N. S. (1985). Uses versus abuses of observational measures. *Behavioral Assessment, 7*, 323 – 330.

Kerlinger, F. N. (1979). *Behavioral Research: A Conceptual Approach* (pp. 29 – 40). New York: Holt, Rinehart, & Winston.

Hartmann, D. (1977). Considerations in the choice of Interobserver reliability estimates. *Journal of Applied Behavior Analysis, 10,* 103 – 116.

McWilliam, R. & Ware, W. (1994). The reliability of observations of young children's engagement: An application of Generalizability theory. *Journal of Early Intervention, 18,* 34 – 48.

Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychological Bulletin, 86,* 376 – 390.

Nichols, D. P. (1998). *Choosing an intraclass correlation coefficient*. Tech support paper from the SPSS tech support web site (www.spss.com/tech/stat/articles/whichicc.txt).

Powell, get citation

Rosenblum, L. (1978). The creation of a behavioral taxonomy. In. G. Sackett's (ed.), *Observing behavior* (Vol. 2): *Data collection and analysis methods* (pp. 15 – 24). Baltimore: University Park Press.

Suen, H. K., & Ary, D. (1989). *Analyzing Quantitative Observation Data* (pp. 1-32; 65 – 98; 158 – 192). Hillsdale, NJ: Erlbaum.

Taylor, C., & Yoder, P. (in press, 2006). Interpreting kappa in observational research. *American Journal of Mental Retardation.*

Taylor, C., Yoder, P., McWilliam, R. (in press, 2006). Generalizability and decision studies: An example using conversational language samples. *Journal of Early Intervention.*

Whitley, B. (1996). *Principles of research in behavioral science* (pp. 98 – 115). Toronto: Mayfield.

Yoder, P. J. , Short-Meyerson, K., & Tapp, J. (2004). Measurement considerations in counting behaviours with special emphasis on sequential analysis of behaviour. In E. Emerson, C. Hatton, T. Thompson, & T. Parmenter & (Eds.)

International Handbook of Applied Research in Intellectual Disabilities (pp. 179 – 202).Wiley.

Yoder, P. J., & Tapp, J. (2004). Empirical guidance for time-window sequential analysis of single cases. *Journal of Behavioral Education, 13,* 227-246.

get the citations for the article critique models.

Course requirements:

Attendance, readings, 6 assignments, 1 oral presentation of article critique of study using quantitative, systematic observational measurement, final project.

Assignments:

1. Write the rationale for using an observational approach to measuring the traits or behaviors in your research questions. Type your falsifiable research questions that contain variables derived from observations. For your research question(s), indicate whether you are solely interested in within-context effects or whether you are wanting to measure individual differences that exist in the future in similar situations or across many observational contexts. If you have traits implied in your research questions, label these traits as precisely as you can. If you aren't measuring traits, indicate why this is not necessary for your research question. Describe the location, instructions, materials, activities, and other variables that influence the variable score. Indicate the rationale for your decisions. Indicate why using a coding system better addresses your research question than using a checklist or rating scale.

2. Decide whether you are going to measure your trait or behavior that is assumed to be stable across time or context and whether you are measuring it in a location that is frequently occurring for the population you are studying (i.e., a naturally occurring location) using an unstructured procedure, then decide if you need to average several sessions. If not, indicate why. If so, indicate how the lessons of D studies were used to help you decide whether to average several sessions together for each unit of analysis (i.e., data point in single subject studies, participant score at each measurement period in group design studies). If observing more one person at a time, how will you be able to get all of the data you need (i.e., what is your participant sampling method?)? Indicate whether you are using interval, event, or timed event coding and why. If interval, give rationale for which type of interval data you selected, what the interval width is, and why. If event, indicate why duration is not important? If timed event, indicate why is duration important? Indicate how and why you will collect these data (computer-aided or paper & pencil method of recording and whether you will observe from video-tape or live). Decide whether your coding system must be mutually exclusive and exhaustive and provide a rationale for your decision. Provide the operational definitions of the codable behaviors.

3. Write down whether your project variable is sequential or nonsequential and give

the rationale for your decision. If nonsequential observational variables are planned, which metric will you use to quantify your variables and why? For example, if a proportion metric is considered, what process will you conduct to determine whether to use a proportion metric. If you don't have a sequential level variable in your project use the one you selected in class for this section of the course to identify your antecedent and target behaviors of interest? Do you think the antecedent will inhibit or elicit the target behavior? What type of sequential analysis will you use (concurrent interval, concurrent time unit, lag1 event or interval, or time window analysis)? Why? What index of sequential association will you use (sequential frequency, transitional probability, or Yule's Q)? Why? Provide an example of the stream(s) of behavior or coded time units that could be a subset of coded data that will be sequentially analyzed. Provide the 2x2 table used to compute your index of sequential association (make sure you label your rows and columns). How will you determine whether you have "enough" instances of the antecedent or target events to see a reliable sequential dependency? What will you do if you do not?

4. During the observer training stage, on which variables will you attempt to achieve and maintain interobserver agreement? Why? Will you use a well-coded tape (i.e., standard) to help train observers to code difficult categories? Will the primary coders/raters know when they are being checked for reliability? If so, how will you address the tendency for people to be more accurate when they know they are observed than when they don't? If you can't address this issue, what are the possible negative effects on this regarding type I or type II errors in your analysis? How will you check for observer drift as primary data is being collected? If observer drift occurs, what will you do? During the reliability training period and code/rating scale development period, what unit of analysis (i.e., will sessions be pooled?) will you use? Defend your decision. Is agreement to be assessed on the summary (i.e., small over large) or point-by-point level? Why? What coefficient of agreement will you use? Why? If you are using point by point agreement, produce the confusion matrix you plan to use. If using a nonmutually exclusive system, will "disagreements" in the confusion matrix and agreement estimate include those in which one coder sees a relevant behavior and the other does not AND those in which there is disagreement on classifying behavior? Why or why not? How will you decide whether to change your coding system or whether to do further training on the coding/rating system you already have? What criteria will you use to determine whether you must recode your tapes? How will you decide which categories to collapse and drop? Will you use consensus coding and if so, under what conditions? Will you average multiple coders' scores and use the average in the analysis of your research question? If so, why?

5. When estimating reliability on the variables used to test the research questions, on which variables (include the metric for the variables) will you estimate interobserver reliability? Why? If different from the variables (and metric) you are estimating interobserver agreement on for coder training and monitoring, why is this reasonable? What proportion of your sample of people or sessions will you

include in your reliability sample? How will you select these people or sessions? Why? Assuming random measurement error, what are the consequences of attempting to address a research question with a coded or rated variable that is not reliable? What is nonindependence of behaviors or observers in the context of your project and how will you prevent it? What is correlated measurement error in the context of your project and how will you prevent it? What coefficient of reliability will you use to estimate reliability on the variables used to test the research questions? Why is it sometimes wise to have a different approach to reliability for code/rating scale development and reliability coding training than is used to determine whether the data used to address the research question is reliable? What option do you have if you fail to achieve a minimum level of reliability between coders? Which option will you choose and why?

6. What type of research are you conducting (specify the "why" and the "what")"? What type of validity is most relevant to your type of research? Who will your "experts" be for establishing content validity? What type of procedure might you use to ask these experts whether your behavioral categories are sufficient and good matches with what you want to measure? What are the pros and cons of your content validity procedure? If you need social validity, indicate who the experts will be, what you will be asking your experts to judge (video or past experience). If you are using video clips, what is the length and selection method of the video clip? What are the social validity items and rating scale you will use? Why did you select these items? What are pros and cons of your social validity procedure? If criterion-related validity is important to your research project, what will be your gold standard variable and why did you select it? Will you use the multi-method, multi-trait approach to testing criterion-related validity? If so, what will be the other methods and traits you will use? Produce an example of the expected correlation values you hope to get. What are the pros and cons of the criterion-related approach to validity? If construct validity is appropriate for your research question, will you the nomological network theory, diagnostic, or treatment validity approaches to construct validity? Indicate the nodes, diagnosis, or treatments you will use and what you expect to find. How can accruing evidence for construct validity help you address concerns about representativeness?

Article critique

This critique will be orally presented. The purpose of these are to give students experience in oral critiquing of the measurement procedure, coding system, variable metric, reliability estimation method, and validity of the observational variables used in published studies. Note this is assignment and template does not cover all points necessary when critiquing an article. It only covers those points most relevant to this course material. Use the questions and directives in the above-listed 5 assignments to address the relevant parts of the article.

Oral presentation grading rubric

I Content adequacy

Student describes and critiques the authors decisions for each of the following topics (10 pts each)

___ Measurement procedure for the observational variable

___ Coding system for the observational variable

___ Variable metric for the observational variable

___ Reliability estimation method for the observational variable

___ Was construct validity of the observational variable supported? What are the limitations on this support?

II Content quality

___ Presentation is approximately ½ hour long (10 pts)

___ Article used for critique is adequately described and referenced (10 pts.)

___ Concepts learned in class are used in presentation (30 pts.)

Article critique template.

A. Overview of purpose, RQ, Design

1. RQ or hypothesis
    a. Is it falsifiable?
        1) If not ask for that.
        2) If can't be stated as falsifiable question, ask for it to be framed as an exploratory question/study.
        3) If the authors are calling is a "descriptive study" note that small sample sizes prevent speaking to normative issues.
2. Reason for prediction
    a. Give theory behind the prediction.
    b. Do they intend to measure a context-specific behavior or state or a highly generalized behavior or signs of construct?
        1) If generalized, make note to check whether procedure is structured.
3. Importance of the study
4. Research design used:
    a. single subject (make note to give greater scrutiny to agreement issues)

b. group (make note to see if variables used to test RQ are used in reliability estimates)

B. Measurement system

1. Describe whether their procedure is structured or unstructured.

a. If generalized tendency to act is being measured,

1) Indicate whether the behaviors they are measuring are likely to be influenced by the uncontrolled aspects of the measurement procedure.

2. If unstructured and behavior likely to be influenced by contextual factors, are they averaging across several sessions?

a) If so, did they use a D study to decide how many to average across?

b) If no, this is a serious flaw in the study. Probable consequences are type II errors.

1) If there are many, many significance tests without family-wise error protection, then type I errors can occur (i.e., sample-specific results).

2. Describe whether the duration of the procedure from which observational variables are derived can have differing durations across participants.

a. If different, infer why the durations are different.

1. If different because the less able the participant the harder it is to complete the protocol for the procedure, then make that that negative association between session duration and frequency is likely and rate may be contra-indicated.

2. If not different, then pro-rating by time is unnecessary.

3. Describe the participant sampling method.

a. If more than 1 person observed at a time, is the selected method likely to be adequate? (minor issue)

4. Describe the behavior sampling method (interval, timed event, event).

a. If interval,

1) indicate the type of interval method they used

2) make a note regarding whether they used kappa as the agreement method and address this when get to agreement issues.

3) make note whether they used record-observe-record cycle or whether they had a computer program to aid interval coding.

a) If the former, did they acknowledge the loss of information through observe intervals in the discussion. Minor issue.

4) If not momentary interval, is it likely that they'll have mixed intervals?

a) If so, indicate the issues regarding over and/or underestimation.

b) The consequences are increased type II errors.

1. If predicted findings occur, those findings are despite this issue and thus becomes a minor issue.

2. If predicted findings don't occur, the over or under-estimation could have caused the null findings.

b. If event (no onset time),

1) make note whether point by point agreement used to train coders and tweak coding system.

a) If not but reliability was sufficient, call minor issue.

5. Indicate whether the data collection was from tapes or live

a. If live, were reliability checks independent and blind?

1. If not, was problem acknowledged in discussion section? Problem could invalidate the reliability checks as estimates of accuracy (agreement in single subject) or reliability (ICC in group).

6. Were the definitions for the coded behaviors sufficient? If not ask for this.

7. What is the metric for each variable?

a. Indicate whether the variable(s) are sequential or nonsequential.

1. If sequential,

a. What type of sequential analysis did they use in terms of direction of analysis (forward vs backward), coded unit (interval, time, event) and how immediacy was defined (concurrent, lag1 event or interval, or time window analysis)?

1) Did the type of analysis fit the theory motivating the theory. If not, ask for rationale.

b. Give the contingency table for each index of sequential association in the study,

1) Was contingency table correctly constructed (base rates correctly represented). If insufficient information, ask for it.

c. Is the sequential association predicted to be positive or negative?

d. Was sufficient information provided to determine if sufficient number of events was present?

1) Were there sufficient number of events for the index of sequential dependency to be stable (expected value of cells > 5) ?

a. If not, this is a serious flaw. Consequences are uninterpretable indices of sequential association, particularly if Q or other contingency indices are used.

e. Give sequential dependency index they used.

1) Was an appropriate index of sequential dependency selected?

a. If not, could differences in base rate of the target behavior offer an alternative explanation to the study findings.

1. If so, ask for different index.

2. If not, ask for authors to address issue in results and discussion section.

2. If nonsequential, indicate the metric used (frequency, duration, rate, relative duration, target/opportunity proportion, target/target actors relevant behavior proportion, compound proportion).

a. Does the metric fit the theory for why the predicted association or difference was made?

1. For example, if the theory behind why the predicted association or difference was made is really referring to a tendency to act (i.e., style), consistency to respond, or opportunities for the target behavior are easy to identify and vary across subjects, then a proportion may be a better fit to theory than frequency.

a. If not, ask for it to be changed or ask for clarification of rationale.

b. If frequency,

1) make sure you don't insist on rate just because the duration of the session can be different across participants.

2) if the duration of the observation session varies across participants, does it make sense that the duration of the observation session could be positively related to frequency (e.g., is the examiner allowed to continue as long as the child wants to?)

a. If so, ask them to check whether duration of session is positively related to frequency of target behavior. If so, rate may be a better choice than frequency.

c. If proportion is selected,

1) see if there is information to guess whether the denominator is likely to be positively related to the numerator.

a. If not, ask for this information.

b. If so, but you think it likely the association is negative, then a proportion is counter-indicated.

1. The consequences of using a proportion when numerator and denominator are negatively correlated is increased type II error. Minor issue.

C. Agreement or reliability study design

1. Is at least 20% of the data sampled for agreement/reliability estimation, are sessions to checked for agreement/reliability selected at random, and is the unit of analysis for the reliability/agreement estimate the same as used to test the RQ?

a. If not the estimate could be nonrepresentative of the larger study sample and the consequences could be lower (or higher) than shown agreement/reliability.

2. Are agreement/reliability checks conducted independently and without the

primary coder's knowledge??

a. If not the estimate could overestimate the agreement/reliability of the study sample.

3. Are agreement/reliability checks made throughout the study?

a. If not, observer drift could occur and the agreement/reliability estimates could overestimate the agreement/reliability of data used to test the RQ.

4. If consensus coding was used to generate the data to test the RQ, was agreement or reliability estimate prior to creating the consensus record?

a. If not, we have no estimate of agreement/reliability of the data to test the RQ.

D. If the study uses a single subject design, scrutinize agreement.

1. Was agreement assessed on the summary (i.e., small over large) or point-by-point level?
    a. If summary level, was limitation noted in the discussion section?
2. Are disagreements on unitizing and classifying considered in the same agreement estimate?
    a. If not, agreement is overestimated.
3. What coefficient of agreement was used?
    a. If not kappa, was not controlling for chance agreement acknowledged in the discussion section?
    b. If kappa used,
        i. were kappas computed for each category from 2 x 2 tables or was one used for the entire coding system?
            1. If the latter, we don't know agreement on particular variables.
        ii. If kappas were < .60, was the base rate of the target category much different from .5 or was estimated accuracy computed?
            1. If so, low base rate is a likely explanation for low kappas. Point this out for editor so they won't throw out the data needlessly.
4. If proportion metric is selected, is agreement on the numerator and denominator "sufficient". If each is right at threshold, consider asking for the reliability coder's proportion scores to be graphed with the primary coders' scores.
    a. If the primary and reliability scores don't show the same pattern of effects and correlated measurement error is a possibility, then coder bias offers an alternative explanation for the apparent effect.
5. Could there be systematic underestimation of the true score in the baseline and systematic overestimation of the true score in the treatment phase?
    a. If so, possible consequences of low agreement is type I error (confirming the RQ even though it is not really so). If agreement is overestimated, then it may appear that agreement issues could not cause type I errors, when they really can. Major issue.
    b. If not, probable consequence of low agreement is type II error. If

effect seen anyway it was despite the low agreement. If agreement is overestimated, then it is a minor issue.

E. If group design, scrutinize reliability issues,
1. Is the variable on which reliability is estimated the same as that used to test the RQ?
a. If not, there is not reliability estimate on the primary data. Possible consequence is low reliability of data when we think otherwise. Consequence of low reliability in group designs is almost always increased probability of type II error.
2. Is ICC used to estimate reliability. If agreement is used then between-participant variables not reflected in reliability estimate and it should be.

F. Validity
1. Are appropriate types of validity used given the type of observational research this article represents?
2. Are reasonable decisions made regarding conducting the part of the study that is relevant to testing the relevant type of validity.
a. Content validity: were appropriate experts selected
b. Social validity: were appropriate respondants selected, was length of the videoclip and the selection method of the videoclips likely to represent what occurred in general, was the instrument used to request ratings reasonable, were participants blind to the condition or phase?
c. Criterion-related validity: Were appropriate gold standard selected. Was MTMM approach used. Was expected pattern of correlations seen?
d. Construct validity: Was the theory sufficiently laid out to identify clearly falsifiable predictions. What proportion of the predicted associations were detected to be statistically significant, of moderate to large effect size, and in the predicted direction (i.e. algebraic sign). Do these associations correspond with extant literature on the construct validity of the variable and psychological reality of the construct?

Final Project

The final project will be a proposal, written in future tense, using APA format for a research study (without the literature review) that has (a) at least one observational variable, and (b) at least one falsifiable research question that will be addressed through your proposed plan.

(a) Describe a measurement procedure that specifies the following:
1. The number of sessions
2. The location of the sessions
3. Any instructions the examiner will give the participants [if any]
4. Any materials used during the observation session
5. Any activity imposed on the participants
6. Other aspects of the procedure that might influence the observational variable in a trivial manner

7. Which of these factors will be constant across all participants (or sessions within participant)
8. Which of the factors will be left to vary across participants (or sessions within participant)

(b) Provide the rationale for the described measurement procedure
(c) Describe the method of recording the behavior from the observation session and why this was chosen
(d) Describe the sampling method (continuous vs time sampling) used and the type of observational data (event, timed event, interval) you'll produce and why these were chosen
(e) Briefly describe the coding system and coding procedure (enough for me to understand what you consider relevant to measure)
(f) Indicate the metric (e.g., frequency, proportion, Yule's Q) you will use to quantify your trait of interest and why you selected this metric

1) If proportion metric is selected, indicate the numerator and denominator of the proportion.

(g) Describe your plan to check reliability for the purposes of training and monitoring observers (see corresponding assignment for required elements)
(h) Describe your plan to estimate reliability of the observational variable that you will use to test your falsifiable research question. (see corresponding assignment for required elements). Note that there are cases in which it is appropriate that (g) and (h) are identical, but if you are asserting yours is such a case, a statement of such is needed and a rationale for why you believe this is the case is necessary.

Grading:
Attendance (15%), oral participation in discussion (15%), assignments (graded after 1 post-feedback remediation; 5% each for 5 assignments for a total of 25%), oral presentation of article critique (10%), final project (35%). If the student chooses, an assignment may be revised after initial instructor feedback. The student must tell the instructor on the day they receive their initial feedback and provisional grade, if he or she intends to revise and resubmit for a final grade.

A+ = 100 – 97
A = 96 - 93
A- = 92 – 90
B+ = 89 – 87
B = 86 – 83
B- = 82 – 80
C+ = 79 – 77
C = 76 – 73
C- = 72 – 70

Schedule of lectures, readings, and evaluation activities

Dates of class and topics	Required readings	Assignments, presentation, and project due dates
1/11/06 Overview of course and Intro to systematic quantitative observational measurement		
1/18/06 "The tension between ecological validity, "representativeness", and construct validity as they relate to selecting measurement procedure; introduction to classical measurement theory; and rating scale vs coding systems	Kerlinger, 1979 (pp. 29 – 40); Suen & Ary (1 – 32), Jacobson, 1985	Falsifiable research question handed in for feedback
1/25/06 Introduction to Generalizability theory with special emphasis on Decision studies regarding how many sessions are needed to provide stable estimates of traits.	McWilliam & Ware, 1994; Taylor, Yoder, McWilliam, in press	Assignment 1 due
2/01/06 Developing an observational system: Behavior sampling and coding system development	Bakeman and Gottman, chapters 2, 3; Suen & Ary p. 65 – 98; Rosenblum, 1978 in Sackett, 1978 (pp. 15 – 24). Gardner, 2000	
2/08/06 Example recording software: Procoder DV; Emphasis on recording the data and the structure of the output. Meet in 304 Peabody library from 2 – 4, 1st hour is in Wyatt 201. UserID for Procoder: yoderclass@vanderbilt.edu, PW: 1039	ProcoderDV website URL: Mingus.kc.vanderbilt.edu/pcdv. Procoder manual	Revised assignment 1 due (if revision is chosen)
2/15/06: Observational variable metrics	Bakeman & Gottman, 1997 pp. 1 – 13;	Assignment 2 due.
2/22/06 Intro to sequential analysis	Yoder, Short-Meyerson, and Tapp, 2004	

3/01/06 Intermediate coverage of sequential analysis	Bakeman & Gottman, 1997, Chapters 5, 6,	Revised assignment 2 due (if revision is chosen)
3/8/06 Spring break		
3/15/06 Intermediate coverage of sequential analysis continued 304 Peabody library, 1 – 4pm	Yoder & Tapp, 2004	Assignment 3 due
3/22/06 Interobserver agreement for coder training, observer drift, and coding system refinement, AB single subject studies, single-subject tests of sequential associations 304 Peabody library, 1 – 4pm	Hartmann, 1977; Baer, 1977; Taylor and Yoder, in press	
3/29/06 Interobserver reliability for estimating the reliability of observational variables in group designs. Overview of when to use which type of reliability index.	Bakeman & Gottman, 1997 chapter 4; Mitchell, 1979; Nichols, 1998	Assignment 4 due (revised assignment 3 is due, if revision is chosen)
4/05/06 Construct validity of observational measures vs representativeness (revisited); Demonstration of MOOSES 304 Peabody library, 1 – 4pm	Suen & Ary, 158 – 192; Cone, 1982 in Hartmann, 1982 pp 67 – 79. Whitley, 1996	Assignment 5 due
4/12/06 Two examples of critiques of observational articles using concepts from course		(revised assignment 4 is due, if revision is chosen)
4/19/06 Graded oral presentations of article critiques (and any catch up that is needed)		oral presentations of article critique (revised assignment 5 is due, if revision is chosen)
4/25/06 student evaluations of class done on line		Final projects due and should be turned in by e-mail (paul.yoder@vanderbilt.edu)