THE INFLUENCE OF PEER FEEDBACK ON WRITING ACHIEVEMENT AND
INDIVIDUAL WRITING SELF-EFFICACY

By

Andrea Lynn Zellner

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Educational Psychology and Educational Technology—Doctor of Philosophy

2017

ABSTRACT

THE INFLUENCE OF PEER FEEDBACK ON WRITING ACHIEVEMENT AND
INDIVIDUAL WRITING SELF-EFFICACY

By

Andrea Lynn Zellner

This study examined the influence of peer feedback and review on individual writing achievement and self-efficacy. Undergraduate first-year composition students engaged in normal instructional activities used the Eli Review program in order to conduct peer feedback and review sessions. Using the data collected from surveys and through the web-based peer review system Eli Review, the influence of giving and receiving writing feedback in peer review groups on both individual writing achievement and individual self-efficacy was modeled using a social-network analysis methodology. The findings showed that students did not improve over the course of the semester in achievement or self-efficacy. Additionally, social network analysis suggested a negative relationship between the quality of feedback received on writing achievement, while no relationship was found between the quality of feedback given on writing achievement and self-efficacy. The findings suggest that practitioners should focus on modeling the feedback cycle, specifically ways to incorporate feedback into the revision process.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In writing courses, peer feedback is a pedagogical strategy employed by writing teachers with the objective to help a student improve his or her own writing skills by reading, reviewing, and offering feedback to another student on his or her piece of writing. This is generally done with students in groups of three or four individuals examining one another's writing, although implementation of peer feedback groups does vary in method of implementation: groups might be of different sizes, the frequency might be different from course to course, feedback might be given face to face or anonymously online. Feedback is both given and received by all members of the peer review group. It is widely accepted that this pedagogical strategy improves students' writing achievement. Various studies have shown large effect sizes for the use of peer feedback in writing classrooms when student writing achievement outcome measures in peer feedback conditions were compared to students writing alone or students receiving teacher correction of a text (Graham & Perrin, 2007).

The precise influence of both giving and receiving feedback on student writing development and achievement has been unclear, however. In one study, for example, students were divided into "givers" of feedback and "receivers" of feedback. The students in the "givers" group showed more writing growth over the course of the semester than did the receivers on external, timed writing assessments in post-tests (Lundstrom & Baker, 2009). Additionally, there have been few studies that examine both the influence of receiving peer feedback on in-process writing and giving peer feedback on fellow classmates' in-process writing as distinct processes occurring simultaneously, a situation in which most authentic uses of peer feedback would occur in a typical classroom. Finally, very few studies have included looking at changes in writing

achievement within the same group of students on writing pieces taken through the draft, peer feedback, revision cycle.

This study examined the influence of giving and receiving peer feedback as distinct but simultaneous processes on individual writing achievement and writing self-efficacy under online, asynchronous conditions. In addition, this study used social network analysis to examine these relationships. The study was based on the assumption, held in social network analysis, that the interactions among individuals and subgroups within a network are a significant factor contributing to individual change. Social network analysis then is ideal for studying peer feedback groups because it allows for the study of the influence of the individuals as they function in the peer feedback groups and classroom as a whole. Peer feedback groups in this study were conceptualized as units, or subgroups, within a larger classroom social network. Direct interactions among students within peer feedback groups facilitated the flow of expertise and influence among students. In this way, the specific influence of an individual student's peers on his or her writing achievement and individual self-efficacy over the course of semester was estimated.

REVIEW OF LITERATURE

Instructors seeking to influence writing achievement have a number of research-based pedagogical strategies from which to choose. Graham and Perrin's (2007) meta-analysis, for example, identified the highest impact strategies in order to highlight what works best for students and teachers when it comes to improving writing achievement, including peer-review approaches, process writing approaches, sentence combining, summarizing strategies, and specific goal setting. Of particular interest are the group of strategies that included collaboration with peers and other students in all aspects of the writing process from planning, to peer review, to revision.

In order to examine more carefully the strategy of peer review and its influence on writers, it is important first to understand the general approaches to teaching developing writers and the historical shifts in researchers' understandings of how best to teach writing. Approaches to the questions of how writers develop, and what strategies might best support that development, emerge from psychology, including motivational research and discussions of cognitive processes, to more theoretical understandings grounded in the study of rhetoric, argument, and the interactions between writer and audience. While this brief summary can do no justice to the depth and breadth of these multiple fields of inquiry, it instead focuses on general pedagogical trends and understandings reflected in both the academic and practitioner literature.

On-demand Writing versus Process Writing Approach

In writing pedagogy, there is a general distinction between on-demand writing and process writing. On-demand writing, in brief, is generally a timed writing assignment in which students are given a question, prompt, or theme about which to write within an allotted time. On-demand writing is generally reserved for measuring student performance (Gere, Christenbury, &

Sassi, 2005; Pajares & Johnson, 1994). Process writing, on the other hand, is often understood as the multiple steps a writer takes in order to produce a piece of writing, and takes its name from its focus on the process of writing rather than merely on the final product. While the canon on writing process defines each aspect of the process in multiple ways, in general the process writing approach involves more time for writers to develop a piece and includes a pre-writing stage, a drafting phase, a chance for response from outside readers, and a revision phase which marks the end of the process. In order for students to grow as writers and gain writing fluency, it is the field's general consensus that process writing is superior for improving developing writers, and many studies on writing development include practical interventions to test process writings' potential improvement on student writing achievement (Elbow & Belanoff, 1995; Flowers & Hayes, 1981; Pritchard & Honeycutt, 2006).

It is within the context of process writing that peer feedback is often included in the instructional process of writing development, although the use of peer feedback groups is not always utilized in studies of process writing approaches. When peer feedback is included in classroom practice, even more instruction is necessary to help writers know how best to revise in light of feedback.

Self-efficacy and Writing Achievement

The study of an individual writer's self-efficacy and its influence on writing achievement has a well-established connection in both the writing and psychological literature. Self-efficacy is a theory that has been well-described in the educational research literature, and focuses on students' expectations about the outcomes of goal directed behavior. Bandura defined self-efficacy as "People's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances (Bandura, 1986, p. 391). Self-efficacy is

considered domain and task-specific, and is often investigated within the context of a particular domain (writing, math) by asking students to give ratings about their confidence to complete discreet tasks within that domain (Pajares, 1996; Pintrich & Schunk, 1996; Shell, Murphy, & Bruning, 1989).

In studies within the writing domain, a student's self-efficacy has been shown to be independently predictive of his or her writing performance (Pajares, 1996). Pajares noted, "In general, results reveal that writing self-efficacy makes an independent contribution to the prediction of writing outcomes and plays the meditational role that social cognitive theorists hypothesize (p. 145)." This is thought to be in large part because students who have high writing self- efficacy are more likely to have more interest in writing, are more willing to engage in writing, and also are more willing to exhibit perseverance, rather than exhibiting doubt, when writing difficulties arise (Bandura, 1986; Hidi & Boscolo, 2006; Pajares, 1996; Pintrich & Schunk, 1996; Shell, Murphy, & Bruning, 1989).

Pedagogy and Effectiveness of Peer Review Groups

The implementation of peer feedback groups into writing classrooms has, in general, been shown to positively influence writing achievement. For example, a meta-analysis conducted by Graham and Perrin (2007) examined a variety of writing instruction strategies for improving adolescent writing quality outcomes. In this meta-analysis, only studies that reported an outcome measure of writing quality and those that compared a treatment group to a control or comparison condition were included. Of 123 studies that met these criteria, seven focused particularly on the peer feedback process. While there was some variation in the ways that peer feedback was implemented, in general each of the studies had in common that peers were involved in viewing an individual students' draft and helping them isolate areas for revision and plan a revision

strategy. Students from 4th grade through 12th grade were included across the studies. The weighted average experimental-control effect size of these interventions collectively was 0.75 and indicated that, as far as interventions go, peer feedback groups have a strong potential influence on increasing students' writing achievement.

DiPardo and Freedman (1988) spend a great deal of time in their review identifying the ways in which peer review groups are structured and implemented in the writing classroom. In their discussion, in which they use the term "peer response group" rather than "peer review" as I do in this study, they identify a number of issues with the use of this pedagogical strategy. They note two main issues with peer review. The first is the "(a) the degree of teacher control over groups and the effects of control structures and (b) the kinds of social interactions within group (p. 119)." To further complicate the discussion of peer response groups, the authors noted that the level of teacher control over what happens in peer response groups seems to have an influence on what happens within the groups. With so much variation in the ways teachers are implementing the strategy, it has thus far been very difficult to isolate the mechanisms at work in successful peer feedback groups. They noted,

>…progress in writing is difficult to measure and often occurs over extended periods of time. Even when no one-to-one relationship can be found between talk in groups and improvement on an individual piece of writing, learning might still be occurring in groups. Alternatively, even if a writer makes measurable improvement on a piece of writing that can be connected to talk in a group session, the writer may not have learned a concept that he or she can apply to a new writing situation. (DiPardo & Freedman, 1988, pp. 121-122).

This lack of ability to "see inside" peer review dyads or groups has been a stumbling block in understanding precisely how the demonstrated improvement from engaging in peer review occurs.

It has also been found that peer review groups and dyads are not working merely because students are tasked with giving feedback suggestions. Students tend to focus only on surface level errors and can be very lenient when it comes to giving feedback about areas of the writing that are incomplete or incoherent, and remained focused generally on grammatical and spelling issues rather than on higher-level rhetorical choices (Beason, 1993; Faigley & Witte, 1981). In most discussions of the strategy of peer review and feedback dyads and groups, the approach has been to implement peer review in a variety of ways in classrooms and then evaluate the writing gains of individuals, the writing attitude of individuals, or to try to capture the phenomenon of student talk in some way in addition to the textual feedback (DiPardo and Freedman, 1998; Graham & Perrin, 2007). Like most tests of educational practices, the focus is largely on individual gains rather than on the classroom as whole.

Peer review as a pedagogical strategy is frequently used in developmental writing courses and across the disciplines. In addition, it has been employed at both the K-12 level and within courses in higher education, both within writing courses and across the curriculum (Graham & Perrin, 2007; Lundstrom & Baker, 2009; Cho & Cho, 2011). It is the flexibility and adaptability of this strategy that is both its strength and its weakness. As Lundstrom and Baker (2009) noted, "The many choices available to teachers when setting up peer review can be daunting, especially since what method is best varies with the situation. Thus, the adaptability of peer review can actually create confusion for teachers as to what exactly peer review involves and the best way to utilize it (p. 30)." What is held in common when employing the strategy is the interaction of the

peers as they consider a piece of writing. Variation exists in terms of the training of students to respond to the writing; the substance of the feedback; number of students in a group; the types of pairings (random or expert-novice, for example); the length of time for the entire writing process cycle; anonymous reviewing; and whether the strategy will employ face-to-face versus asynchronous interaction or a combination of both. Largely, instructors and teachers decide on the activities, trainings, and other implementation factors in ways that seem most likely to meet the needs of their instructional goals and students. Despite this lack of consistency, the strategy of peer review has been consistently shown to yield improvements in student writing achievement when peers are asked to give and receive feedback to and from one another.

Defining Feedback and Revision

In examining feedback and its influence on writing, it is important to clearly define how feedback is understood within the writing classroom context. From the psychological sciences, feedback intervention studies have conceptualized feedback in terms of task-performance. Kruger and Denisi (1996) noted in their review that this variability in definition necessitated a clearer definition for inclusion in their own examination of the literature. They defined feedback as "actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one's task performance (p. 255)." There is a very clear focus in this conceptualization of feedback on the idea of quality of task performance. In the writing literature, Keh (1990) conceptualized the idea of feedback in this way.

"Feedback is a fundamental element of a process approach to writing. It can be defined as input from a reader to a writer with the effect of providing information to the writer for revision…. Through feedback, the writer learns where he or she has misled or confused the

reader by not supplying enough information, illogical organization, lack of development of ideas, or something like inappropriate word-choice or tense (p. 294-5)."

In this case, Keh is focused primarily on feedback not only as a response to the quality of task performance, but as actionable information with the express purpose of improvement upon revision. For Keh, feedback is intended as a way to improve task performance. Ideally, revision then is more than just fixing mechanical errors, but rather a return to the task performance with the explicit knowledge of how to improve the quality of the task, in this case, the piece of writing.

As Beach and Friedrich (2006) noted, "Teachers often do not have time to devote to extensive conferencing with each student, so they need to rely on trained peers to supplement their conferencing in pairs or small group conferences, online conferences, or "read-arounds" … (p. 229)." This idea that the peer review process is a pedagogical strategy in order to take the place of expert guidance is a common one in the literature. The assumption inherent in this conceptualization of the process is that students benefit largely from receiving feedback, ostensibly because they might locate weaknesses in the text and then revise accordingly. Recent research in this area, however, has indicated that the cause of the increases in student writing achievement is not attributable to receiving this feedback, but rather that the act of giving feedback itself is the mechanism by which writing is improved.

In Lundstrom and Baker's (2009) study, they developed an experiment to directly test the assumption that increases in writing achievement are directly related to the revision of a piece of writing based on comments and directions to improve the piece received from peers. This study sampled second language learners in an undergraduate institution. Students were enrolled in the regular writing courses, with half of the sections assigned as "givers" of feedback and half the

9

sections designated as "receivers." Each group received instruction on giving or receiving feedback on sample essays on particular aspects of the text over four instructional mini-lessons. The giver group provided feedback on sample student essays, while the receiver group received sample essays (not their own) with feedback marked, and were taught to use that feedback in order to revise the essay. Students were given a pre- and post-writing proficiency test that took the form of a timed essay. Students were then assessed on their gains in proficiency. The results of the study showed that the givers of feedback demonstrated greater writing improvement than the receivers of feedback over the course of the semester. Both groups showed gains in writing achievement; however, the giver group showed much larger gains than the receivers on nearly every aspect of writing fluency. The authors noted, "It is not just the added feedback students receive on their writing, nor the extra language interaction experience that helps improve student writing; the act of providing feedback may also improve student writing and may be the most beneficial aspect of peer review (p. 38)." The authors also noted that future research should focus more closely on the quality of student interactions in order to better understand the factors at work.

The weaknesses of Lundstrom and Baker's study, however, lie in the lack of authenticity of the tasks. While providing structured interventions to specifically look at how best to revise an introduction and then doing so based on feedback (receivers) or knowing what to comment on specifically (givers), the essays were not the students' own. In addition, while timed essays certainly provide one measure of students' writing ability, it remains a somewhat inauthentic measure of what students experience in most writing situations. For example, as previously noted, on-demand writing approaches are considered less authentic as compared to process writing approaches. Finally, teachers are unlikely to scaffold an intervention in this way, and, as

previously stated, it is unclear that it might generalize when peer feedback groups are implemented in a more authentic context where students are both giving and receiving feedback.

Another study to address the question of how giving feedback might influence student achievement was Cho and Cho's (2011) study. Cho and Cho mined the SWoRD system, a computerized writing software that facilitates peer review, to look at the quality of comments given and improvement in student writing achievement in an introductory physics course. Data were collected from 72 students, their comments analyzed, and the first and final drafts were rated by independent reviewers to establish both baseline and final writing achievement. Additionally, raters considered nearly 4000 comments traded over the course of the peer review process. There were a number of interesting findings from the study. First, even when controlling for initial writing achievement, the quality of feedback given to peers was shown to have a strong positive relationship with final writing achievement. When students gave high-quality feedback, their own achievement improved. Additionally, received comments had little influence on final writing achievement, and even in some cases were shown to have a negative relationship with final drafts. The authors stated, "…when writers received more praise from peer reviewers on surface features, their revised drafts tended to be of lower quality (p. 637)." Taken together, the findings suggest that the giving of feedback influences student writing achievement.

The limitations of the study, discussed by the authors at length, include the lack of investigation into the details of the giving of feedback and its influence on student achievement. For instance, it remains unclear whether the higher quality comments led in turn to similar types of revisions. The most efficacious reviewing activities remain unclear. Secondly, the writing ability was assessed narrowly and ignored physics domain-specific concerns. Finally, student

motivation throughout the process is unknown, and could be another factor influencing the strong correlations between quality of feedback and writing achievement.

These studies are highly suggestive that the strength of the peer feedback strategy lies in the individual act of giving, rather than receiving, feedback. Future research in this area is still needed, however, especially in terms of addressing the interaction of giving and receiving feedback, particularly in the context of authentic instructional situations, authentic writing assessment, and including examinations of student writing self-efficacy.

The field of writing pedagogy has considered the question of peer review and feedback in many circumstances and with varied implementation and methodologies. Consistently, the act of engaging students with peers around writing concepts has been shown to be a highly efficacious pedagogical strategy. Early exploratory studies suggest that the strength of the strategy is in the act of giving feedback to peers, and that the quality of that feedback is what leads to student writing gains. To date, many aspects of the strategy remain unclear and the need for more exploratory studies has been expressed by numerous researchers in the field. Additionally, the interactive aspects of peer feedback processes, namely understanding the social interactions and how they influence achievement and motivation, have not yet been examined. Social network analysis as a methodological choice to examine questions of these phenomena is an ideal approach to these remaining questions.

PURPOSE OF THE STUDY

This study employed a web-based peer review platform, Eli Review (http://www.elireview.com/). When students used Eli Review to facilitate peer feedback, the data collected in the process of using the program were essential for analyzing the influence of peer feedback as a pedagogical strategy. Eli Review efficiently facilitated the gathering of data in ways that were previously quite difficult, for example: tracking each incoming and outgoing comment, ratings of helpfulness, drafts over time, and revision plans.

The setting of the study was within sections of first-year writing courses at a large Midwestern university. The University provided Eli Review to instructors: a web-based software to facilitate the logistics of the peer review process. This online system helped instructors and students scaffold the peer review process through the draft, reviewing, and revision stages. The first-year writing course was required for all students regardless of major and taken by incoming freshman who have scored below a certain threshold on standardized college entrance exams. The population of students in this course represented the general population of college students.

The curricular goals of first-year writing and many of the core writing assignments/genres were held in common across courses. Nevertheless, some variation existed across sections. While some sections of first-year writing focused on composition in digital environments, creating digital storytelling products, blogging, and discussing visual rhetoric, another section focused on more traditional assignments. Nonetheless, the writing program held in common curricular goals and assignments that are standard across institutions of higher education, including methods of process writing, feedback giving, and revision. Section sizes ran

from 24-29 students, and some instructors taught multiple sections. Instructors included a mixture of current graduate students and adjunct instructors.

Using data captured in the peer feedback process using Eli Review and the use of student and instructor surveys, the research questions were as follows:

1. What is the influence of the feedback given and received by an individual on **writing achievement?**

2. What is the influence of feedback given and received on an individual **writing self-efficacy**?

METHOD

This study is a social network analysis of the influence of both giving and receiving peer feedback on writing achievement and writing self-efficacy.

Participants

A sample of 109 participants were drawn from first-year composition students across multiple sections of the required introductory writing course at a large Midwestern University. Data was collected from a sample of 23 sections (13 instructors) of the seventy-five total sections that were taught during the semester, which resulted in 633 participants who used the ELI review system to facilitate peer review in the section. The instructors volunteered to be a part of the study, and the students within those sections also volunteered. Of the 633 participants in the sample, 109 participants completed each of the data collection points: pre- and post-surveys, submitted an essay draft and final draft, as well as provided feedback to each of the group members and thus were included in the final analyses. For more details on data collection timelines, please refer to Appendix A.

Context of the study

Within each section, students were randomly assigned to a peer feedback group that was consistent throughout the duration of the semester. Each group had 2-4 students resulting in approximately 30 peer feedback groups. It was within these peer feedback groups that the majority of interactions around writing and feedback occurred. Indeed, these subgroups are conceptualized as the network subgroup of each individual student. Students received feedback only from the students in their peer feedback group and the instructor within the confines of the class. Students gave feedback only to the students within their peer feedback group for the duration of each assignment writing process. The writing assignment process began with a

15

student draft, was followed by the giving and receiving of feedback in the peer feedback groups, and was considered finished once the final draft was submitted.

In this study peer review groups are conceptualized as subgroups, and it is important to clarify what is meant exactly in terms of how these students are interacting. In Eli Review, feedback cycles are undertaken in dyads. Because students worked asynchronously in Eli Review, by necessity they engaged with only one other student's work at a time. During a feedback cycle, Student A will be giving feedback to Students B and C, for example, and receiving feedback from Students B and C in return. In this way, while students are exposed to one another's writing in a similar way to face-to-face writing groups might, the social context and synchronous interactions were missing. This impacted the nature of the interactions these students were having with one another as compared to more traditional implementation of the peer feedback strategy in classrooms that might rely on face-to-face conversations about a piece of writing.

In considering a network, it is important to examine both the impact of selection and influence. In social network analysis, it is recognized that, in general, individuals select one another not randomly, but based on certain criteria. This in turn can impact the influence those individuals have on one another. In the case of this study, however, selection is decoupled from influence because the individual students were randomly assigned to groups within the class and a proportion of their work for class is to engage in giving and receiving peer feedback as part of the peer feedback group. This means that issues of students choosing the "good writer" to partner with as a strategy to improve their own writing is controlled for in this study through random assignment.

By using Eli Review, the University also scaffolded the peer feedback process for students with a pedagogically sound design. Eli Review was developed by researchers in the Writing in Digital Environments group at Michigan State University for facilitating peer feedback in writing classrooms. Eli Review is a web-based interface that allows students to upload papers for peer feedback. Peers within classes are assigned a fellow student's essay on which they provide feedback. The student then can review the feedback he or she has received and rate the quality of that feedback. Instructors also have the option to rate the quality of the feedback. The group also rates the quality of the comments given by an individual commenter. All of this data is visible to students, creating a feedback loop that is designed to inform their future feedback quality: the idea that a student might adjust his or her behavior to improve commenting when given low scores on the quality of the feedback (Figure 1). Finally, Eli Review provided a scaffold that helps students to identify changes they have made in response to the feedback.

Student	Avg. Rating	# Comments	% of Comments Rated By Group
	4.57 out of 5	7	28.6%
	4.67 out of 5	3	33.3%
	4.20 out of 5	10	20.0%
	4.20 out of 5	5	20.0%
	4.17 out of 5	6	33.3%
	4.17 out of 5	6	33.3%
	4.00 out of 5	5	40.0%
	4.00 out of 5	2	50.0%
	4.00 out of 5	6	33.3%
	4.00 out of 5	4	50.0%
	4.00 out of 5	7	28.6%
	3.83 out of 5	6	33.3%
	3.67 out of 5	6	33.3%

Figure 1. Screenshot from Eli showing feedback quality, number of comments, and percentage of the comments that were rated.

Eli Review, while solving many of the logistical challenges of using peer feedback groups in the writing classroom, has taken a particular set of pedagogical stands in terms of operationalizing the use peer feedback groups (DiPardo & Freedman, 1988; Glenn, Goldthwaite, & Connors, 2003). By precisely guiding students through each step of the peer feedback process and facilitating the feedback online instead of face-to-face, the writing group process is largely prescribed and controlled by the ways the teacher uses the program.

Measures

Rubric Achievement. In the first-year writing courses, students were required to complete a number of writing tasks through the writing process. The writing process included generating an initial or rough draft, receiving feedback, developing a revision plan, and

submitting a revised draft. For this study, two writing tasks (see Appendix E) were used to generate data for analysis: the first task's writing prompt was assigned to students as their first writing assignment, and the second as the last task of the semester. Both tasks follow an iterative writing process of a student-generated draft 1, followed by peer feedback on draft 1, and then integrating that feedback into a draft 2 on the same prompt. The first writing task, Prompt A, was a personal literacy narrative or Learning Memoir. The second writing task, Prompt B, was writing about a cultural artifact. Additionally, the prompts were assigned to students in a counterbalanced manner: half of the sections wrote to Prompt A first, and Prompt B second; and the other half of the sections wrote to Prompt B first, and Prompt A second. Time 1 score was conceptualized as the initial time point in the data, and the revised draft score as the outcome variable and second time point (Achievement Time 2). This rubric was used to assess writing achievement (see Appendix D) that reflects both course instruction and commonly assessed aspects of writing such as sentence complexity and organization. Interrater reliability was calculated as weighted Cohen's $\kappa = 0.95 \ (CI: 0.83, 1.0)$.

Flesch-Kincaid Grade. Flesch-Kincaid grade level score was calculated as a separate measure of writing achievement. The Flesch-Kincaid grade level is a readability formula commonly used in studies to ascertain the approximate grade level of a piece of writing. It considers a ratio of syllables to words, and words to sentences to determine complexity (Kincaid et al., 1975). Flesch-Kincaid was measured using the R package Korpus (michalke, 2016).

Individual Writing Self-efficacy. Participants were surveyed at both the beginning and end of the semester. The student surveys focused on three main areas. The first was to establish basic demographic information and to gain knowledge of a student's previous achievement. The second was to gather data on the student's writing self-efficacy based on a validated instrument

of writing self-efficacy, and reliability of the instrument, assessed with Cronbach's alpha, was α= 0.92 for the task subscale which was used in the final analysis in this study. (Shell, Murphy, & Bruning, 1989). Finally, participants responded to a survey section that attempts to gauge their "openness to feedback" (See Appendices B and C for full pre- and post-surveys).

Quality of Feedback Given. During the course of peer review activities, students produced feedback on their subgroup members' writing. These comments were compiled and given a holistic rating based on the Eli Review feedback helpfulness scale as determined by the author (see Table 1). Interrater reliability was calculated as weighted Cohen's $\kappa = 0.95 \ (CI: 0.89, 1.0)$.

Influence of Feedback. In order to specify the quality of the incoming feedback for a particular student, an average of the quality of feedback given was calculated for each individual. This is consistent with social network analysis methodology and represents the exposure of the peer feedback group on the individual.

Gender and Ethnicity. Participants were surveyed at the beginning and end of the semester. The first portion of each survey (see Appendices B and C for full pre- and post-surveys) included questions about ethnicity and gender.

Table 1. *Rubric ratings of feedback quality*

5	4	3	2	1
The feedback exhibits all of the following elements: it clearly names what the writer has done, is specific with regard to the goals of the writing, speaks to quality of the writing, and is respectful in tone.	The feedback exhibits the majority of the following elements: it clearly names what the writer has done, is specific with regard to the goals of the writing, speaks to quality of the writing, and is respectful in tone.	The feedback exhibits the some of the following elements: it clearly names what the writer has done, is specific with regard to the goals of the writing, speaks to quality of the writing, and is respectful in tone.	The feedback exhibits few of the following elements: it clearly names what the writer has done, is specific with regard to the goals of the writing, speaks to quality of the writing, and is respectful in tone.	The feedback exhibits none of the following elements: it clearly names what the writer has done, is specific with regard to the goals of the writing, speaks to quality of the writing, and is respectful in tone.

Research Design

      This longitudinal study used social network analysis to follow participants throughout a full semester of the first-year writing course with data collection at four time points: two at the beginning of the semester and two at the end. Students were surveyed on a writing self-efficacy measure at the beginning and end of the semester, and additional demographic and behavioral data was gathered at those time points as well (see Appendix A). Students were randomly assigned to peer feedback groups for each writing assignment process, and initial and final drafts for each assignment were collected at the beginning and end of the semester. This design allowed for estimations of the effect of peer influence on individual writing achievement and self-efficacy as well as the influence of an individual giving quality feedback to peers.

Procedures and Data Collection

Data was collected over the course of the Fall 2014 semester. Course sections included in the sample followed a "business as usual" approach to instruction and course events, with some minor adjustments. Instructors held in common the following aspects of instruction: the essay prompts, the rubric used to evaluate students, the use of the Eli system, and the time allotted for the writing process from draft through finished essay. In addition, students were randomly assigned to peer feedback groups for the duration of the semester. While data was collected for two rounds of the writing-feedback-revision cycle, the analysis of writing achievement was reserved only for the first cycle for this study. In future studies, the second round of writing-feedback-revision cycle could be compared to the first. For the self-efficacy analysis, both the September and December data was used, as is discussed in more detail in the next section.

Analytic approach

The analyses in this study are correlational and used the method of multilevel modeling to estimate the relationships among variables. Each model included an influence term to estimate the relationship of exposure to peers' feedback on the outcome variables of individual writing achievement (both rubric and Flesch-Kincaid scores) as well as the outcome of individual self-efficacy at Time 2.

Because variance within peer feedback groups was expected to be different than the variance between peer feedback groups, the design of the study's analysis reflected the grouped nature of the data. As a result, multilevel models were used to pool the effects of the group level variances in the regression. This method of analysis reflects the structure of the data within the real-world context of students nested within groups. While students were also nested within classrooms as well, the variation between classrooms was minimal in the rubric achievement and

22

higher with the Flesch-Kincaid model (Rubric: ICC=*0.148*, Flesch-Kincaid :ICC=*0.37, and Self-Efficacy: ICC=0)*. Course differences were controlled as a fixed effect on Level 1 (see Appendix E).

The within-group model used in the analysis considered the outcome of individual writing achievement as it related to student-level factors including the quality of the feedback given, the mean quality of the feedback received. The social network terms is included as mean of the quality of incoming feedback on the individual level, specified as Influence of Feedback in the model.

$$RubricAchievementTime2_{ij} = \beta_{0j} + \beta_{1j}RubricAchievementTime1_{ij} +$$
$$\beta_{2j}QualityofFeedbackGiven_{ij} + \beta_{3j}Gender_{ij} + \beta_{4j}Ethnicity_{ij}$$
$$+ \beta_{5j}InfluenceofFeedback_{ij} + \varepsilon_{ij}$$

In order to capture the group level variance, the random intercept model was then estimated. The level-2 model was specified in this way for student in peer feedback group j:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} + \mu_{0j}$$

$$\beta_{pj} = \gamma_{p0k(p=2-5)}$$

These analyses were repeated with each of the Time 1 and Time 2 variables, substituting the measure of rubric achievement with the Flesch-Kincaid grade, as well as the self-efficacy measure.

RESULTS

Results are organized by participant flow and then by research question.

Participant Flow and Characteristics

There were n = 299 participants who did not complete either one of the surveys and/or participants who were not in an intact peer feedback group despite having data collected at each of the time points. Chi-squared analysis was used to compare the participants with partial data (n=408) and participants with intact data and intact peer feedback groups (n = 109) for both ethnicity and gender. There were no significant relationships between these two groups based upon ethnicity, $\chi2(4)$= 4.1263 (p = 0.39), or gender $\chi2(1)$ =0.66673 (p = .41), suggesting that membership in intact groups was independent of the demographic factors. Therefore, the null hypothesis was rejected and the inclusion based on intact groups is independent of the demographic factors.

Dependent t-tests on both the self-efficacy data and the achievement data indicated that there was no overall statistically significant difference in the pre- and post-measures for these student attributes.

Data collection in this study was negatively impacted in a number of ways, leading to a smaller subset of data for the final analysis. An attempt was made to include a control group of students who had not yet taken the course in order to compare their self-efficacy scores over the same time period to the students in the courses. Despite repeated attempts to recruit from this pool and an offer of incentives, I was unable to recruit a sample large enough for comparison. In terms of recruitment from writing courses, students who were absent or underage at Time 1 were unable to consent to the study or take the survey at Time 1. At Time 2, an additional set of students opted into the study, especially students who had turned 18 over the course of the

semester. While an offer was made to students to join the study via a parental signature, none of the participants engaged with that option and rather waited until I returned for the Time 2 survey collection to consent to participate. This allowed for access to their writing data over the course of the semester, but resulted in missing Time 1 survey data. Additionally, some students dropped the course between Time 1 and Time 2, resulting in an incomplete data set. Another factor that led to attrition was students who did not provide feedback or did not complete writing samples for Time 1 and/or Time 2. For every one participant with an incomplete data set, an entire sub-group was impacted by that missing data and thus were excluded from the final analysis.

Table 2 summarizes descriptive statistics for participant writing achievement and writing self-efficacy, as well as descriptive statistics for each of the key predictor variables and key co-variates of the final participant group used in all analyses. In addition, Table 3 is the correlation table among variables.

Table 2: Demographics and Descriptive Statistics

	n	M*(SD)*
Rubric Achievement (Time 1)	109	3.14 *(0.62)*
Rubric Achievement (Time 2)	109	3.68 *(0.74)*
Individual Self-efficacy (Time 1)	109	65.89*(15.39)*
Individual Self-efficacy (Time 2)	109	69.99*(16.62)*
Quality of Feedback Given	109	2.61 *(1.34)*
Influence of Feedback	109	2.60 *(1.12)*

Table 2 (cont'd)

Flesch Kincaid Grade (Time 1)   84   7.98 *(1.78)*

Flesch Kincaid Grade (Time 2)   92   7.96 *(1.90)*

Table 3: Intercorrelations Between Variables

	1	2	3	4	5	6	7
1. Rubric Achievement (Time 1)	----	0.22*	0.11	0.15	0.02	0.08	0.12
2. Rubric Achievement (Time2)		-----	-0.05	-0.01	-0.14	-0.08	-0.02
3. Flesch Kincaid Achievement (Time 1)			----	0.93**	0.18	-0.03	-0.09
4.Flesch Kincaid Achievement (Time 2)				----	0.16	0.08	0.00
5. Quality of Feedback Given					----	0.06	-0.08
6. Self-efficacy (Time 1)						---	0.67**
7. Self-efficacy (Time 2)							---

*p <.05. **p <.01.

The two measures of achievement, the rubric and Flesch-Kincaid, did not showed very low correlation with each other. This is likely due to the limitations of the Flesch-Kincaid score which focuses more on the length and complexity of words and sentences rather than reflecting the quality of a piece. In addition, there was high correlation between Time 1 and Time 2 on the Flesch-Kincaid measure, indicating that the text complexity was very similar at both time points. In effect, these two instruments, the rubric and the Flesch-Kincaid instrument, measure two very different approaches to understanding a text. The Flesch-Kincaid measure was included to add additional valid and objective measurements of achievement and should be interpreted with the limitations in mind. Text complexity is only one indicator of a text's quality and, in this case, was not correlated with the overall quality of text.

Influence of Feedback on Achievement

In order to analyze the influence of feedback on achievement, multilevel models were used to calculate the relationships between achievement and key variables.

The results indicate that, across groups, the quality of the feedback received has a negative correlation with the quality of the writing as assessed at Time 2. Thus, the higher quality of feedback a student received inversely correlated with the rubric achievement at Time 2. The results demonstrate that the demographic variables, the quality of the feedback given, and the quality of the writing within the peer group did not have a significant outcome. There was no consistent relation between achievement and any of the individual student predictors. The one exception was the achievement at time one, which significantly predicted the achievement at time 2. The results of this analysis are summarized in Table 3.

Table 4: Parameter Estimations of the Final Multilevel Model with Rubric Achievement as

Outcome Variable

	b	SE b	95% CI
Intercept (β_0)	3.23**	0.46	2.34, 4.12
Rubric Achievement (Time 1)(β_{1j})	0.26*	0.12	0.02, 0.49
Quality of Feedback Given (β_{2j})	-0.01	0.06	-0.12, 0.11
Gender (β_{3j})	0.00	0.15	-0.28, 0.28
Ethnicity (β_{4j})	0.02	0.05	-0.08, 0.12
Influence of Feedback (β_{5j})	-0.07*	0.06	-2.79, 0.00

*p <.05, **p<0.01

Adding interaction terms, random slopes, and exposure to quality of the group's writing did not improve the model and are thus not reported here (see Appendix E). The variation of the other factors across the peer feedback groups for the were consistent between groups. Outcome: Achievement (Flesch-Kincaid Grade). When the models were specified with the measure of achievement as the Flesch-Kincaid grades substitute for the rubric scores, the results were very similar (Table 4). The Time 1 and Time 2 Flesch-Kincaid Grades were highly correlated (.93, p < .01). In the multilevel model, the Time 1 score accounted for a great deal of the variance in the Time 1 score as well, consistent with the correlation matrix. Thus, the other

fixed effects were not significant. Adding interaction terms, random slopes, and exposure to quality of the group's writing did not improve the model and are thus not reported here (see Appendix F).

Table 5: Parameter Estimations of the Final Multilevel Model with Flesch-Kincaid Grade Time 2 as Outcome Variable

	b	SE b	95% CI
Intercept (β_0)	0.82	0.45	-0.05, 1.69
Flesch-Kincaid Grade (Time 1)β_{1j})	0.93**	0.05	0.83, 1.03
Quality of Feedback Given (β_{2j})	-0.05	0.07	-0.19, 0.09
Gender (β_{3j})	-0.11	0.19	-0.48, 0.25
Ethnicity(β_{4j})	-0.09	0.07	-0.22, 0.03
Influence of Feedback (β_{5j})	0.04	0.08	-0.17, 0.31

*p <.05, **p<0.01

Influence of Feedback on Self-efficacy

To examine the influence of feedback on self-efficacy, once again multilevel models were specified in the same way as in the achievement models above by substituting self-efficacy for the achievement variables. Self-efficacy at Time 2 was used as the outcome variable. In the models for self-efficacy, the Time 1 variable was the only significant predictor for self-efficacy at Time 2. Once again, Time 1 and Time 2 results were highly correlated, and t-tests on the scores showed no significant difference in the scores (Table 5).

Table 6: Parameter Estimations of the Final Multilevel Model with Self-Efficacy as Outcome Variable

	b	SE b	95% CI
Intercept (β_0)	19.2**	7.48	10.77, 37.60
Self-efficacy (Time 1) (β_1)	0.74**	0.08	0.57, 0.90
Quality of Feedback Given (β_2)	- 0.91	1.07	-3.21, 2.49
Gender (β_4)	1.3	2.56	-4.21, 5.84
Ethnicity (β_5)	1.42	0.93	-0.41. 3.26
Influence of Feedback (β_5)	-0.71	-1.28	-5.31, 2.08

*p <.05, **p<0.01.

DISCUSSION

In this study, I examined the influence of peer review on student self-efficacy and writing achievement. The first focus of this study was to investigate the relationship of both giving and receiving feedback on writing achievement. The findings on achievement, somewhat surprisingly, showed no impact on writing achievement for giving feedback on two different outcome measures of achievement: one human rated (Rubric Achievement) and the other computer rated (Flesch-Kincaid). When writing achievement was measured on Flesch-Kincaid grade level scores, the correlation between pre- and post-measures of writing was an almost perfect correlation (.93, $p < .01$). Additionally, the rubric pre- and post-measures also showed no statistically significant improvement. This research was designed with the assumption that changes in writing achievement would result over the course of the writing task. Therefore, changes that might be attributable to the peer feedback process are even less likely to be detected.

What might account for this lack of change increase in achievement? One possibility is that the time period from pre- to post-writing task was not long enough to detect a change in writing achievement. For some of the students, the pre- and post-writing represented only a few days, for others a few weeks. Despite the best efforts by instructors, it is possible that students required more time with the intervention.

A second possibility is related to prior research into the abilities of students to revise. Previous studies have, in general, utilized on-demand, rather than process, writing tasks to assess for writing growth. This leaves out of the equation the revision step. Interrogating the process of moving from feedback to revision was beyond the scope of this study, but it is possible that

students were not clear what do with the feedback they received and thus only made surface-level changes.

Future research would benefit from observing student growth over a longer period of time or over additional writing tasks in order to detect changes in writing achievement. Quality of feedback in this study focused on the naming of what the writer had done, included comments on the specific goals of the writing, spoke to the quality of the writing, and was respectful in tone. Similarly, to the Cho and Cho (2011) findings, a slight negative impact on writing achievement was detected (only as measured by the rubric) as the quality of the feedback increased. The results indicated that the higher the quality of feedback, the lower the achievement. Additionally, the giving feedback had no significant relationship to writing achievement. As has been found in earlier research, revision tends to be a very difficult skill for students to master. In scaffolding the writing process, participants may have understood how best to give feedback, but perhaps not how to then incorporate that feedback into meaningful revision, nor how to translate the moves they were suggesting to peers into their own writing. Thus, the participants were able to give valuable feedback and received valuable feedback, but were unable to turn that into a revision plan to improve the piece of writing. Although the results were not significant, even in the Flesch-Kincaid models, the suggestion of a negative relationship between quality of feedback given and received held true. Future research should focus on the process of revision as well, focusing on the degree to which drafts changed over time as a possible factor in explaining changes in writing achievement or lack thereof.

Finally, there is the possibility that the asychrnonous and online nature of the giving and receiving of feedback contributed to the negative association between feedback and achievement. Just as in the Cho and Cho (2011) findings, this study also relied on an online, asynchronous

environment for the facilitation of the feedback process. Indeed, previous research has suggested that online asynchronous environments can impact processes that are positive interventions in face-to-face environments, but have negative or null impacts when moved online (Roseth, Saltarelli, & Glass, 2011).

Writing self-efficacy is also often closely tied to writing achievement (Pajares & Johnson, 1994), and so the influence of feedback quality received and given on the self-reported self-efficacy scores at Time 2 was also investigated. Once again, the results suggested a negative relationship between quality of feedback both given and received on a student's self-efficacy, although the results were not statistically significant. Self-efficacy may be impacted by higher quality feedback: before beginning on a college writing journey, students may be filled with the confidence that their earlier schooling years afforded them. The decline in self-efficacy could be related to the increased quality of feedback--by highlighting areas of improvement within the writing, the writers themselves felt the limitations of their own abilities, even when the feedback was specific and constructive.

In previous literature on the use of peer review as a pedagogical strategy, the results have indicated that quality feedback that focused on the higher-level aspects of writing, including the rhetorical choices that the writer has made, is the most impactful. In addition, there have been some indications that the more the feedback received focused on revision plans, the more impact that feedback has on improving the revised writing.

In summary, peer feedback can be an impactful strategy for improving student writing, but previous research has identified that peer review process be scaffolded in very specific ways in order to be efficacious: the feedback should avoid praise, the feedback should focus on higher-level rhetorical strategies present and missing in the writing, and suggest a clear pathway to

revision. These caveats held true even when giving and receiving feedback were uncoupled, as in Lundstrom and Baker's (2009) study, in which one group only gave feedback, and the other only received. In the Lundstrom and Baker study, however, the students were assessed for writing achievement in on-demand writing situations, unlike studies which focus on process writing gains.

However, it also been shown that when students focus solely on surface level errors and engage in general praise of the writing, as a result the writing does not improve and may also decline in quality (Cho & Cho, 2011; Cho & MacArthur, 2010;). Nonetheless, on the whole, peer feedback as a pedagogical strategy has been respected as one that will improve students' writing.

The results found in this study diverge from previous scholarship, showing no impact whatsoever at best, and a negative impact on writing achievement at worst. Some of the main differences in this study are its focus on writing courses in a business as usual situation, with randomly assigned peer feedback groups. Additionally, the use of revised essays as the post-measure, rather than an on-demand writing task, is authentic to the real-world, but not used in the studies in Graham and Perrin's (2007) meta-analysis, for example. The design of this study in classrooms on typical writing tasks resulted in findings that suggest that the giving and receiving of feedback, while it may have its place in writing classrooms, does not always achieve the gains seen in more controlled studies.

Indeed, in trying to integrate the process of reviewing and giving feedback, students may not have established the writing behaviors to know how to incorporate the feedback into their revisions, thus resulting in a nearly identical draft in the revision. Achievement, as reflected as changes and improvement in the writing, was then divorced from the writing process. One way to improve on this study in future research would be to articulate more explicitly to students,

through pedagogical moves in the classroom, the ways in which feedback is most efficacious. Additionally, a more precise check of differences in pre- and post-essays on a variety of factors would allow for more clarity in terms of how much revision was done. Finally, examination of the revision plans would also account for the revision factor and its relationship to feedback and achievement. In this study, while the measure of quality of feedback was reliable and valid, additional factors that might be beneficial for inclusion would include examining how students articulate explicit guidance on revision and higher-level rhetorical choices, in line with findings from previous studies. While both of these factors were implied in the holistic quality of feedback rubric, stating them more explicitly would help to tease out the impact of such feedback more concretely.

In terms of changes in self-efficacy, again the data suggested same negative relationship between exposure to feedback and self-efficacy, although the effects were not significant. Additionally, while the students on average increased their self-efficacy, the individual differences were not statistically significant either. In general, the data indicated that there was little to no impact on students' perceptions of themselves as writers over the course of the semester.

While previous studies have shown large effect sizes on achievement as a result of incorporating opportunities for students to review one another's writing, feedback alone is not a magic bullet. Review and the giving of feedback may work better when students are writing outside of the writing process, if Lundstrom and Baker's study is a guide (2009). Revision based on feedback received on a particular piece of writing is a skill set all its own. It may be possible that, absent explicit instruction on revision, students were unclear on how to revise for improvement, despite the suggestions offered by their peers. Additional analyses might include

determining how much variation exists between pre- and post-writing in order to test the degree to which the participants revised their writing.

Limitations

Limitations of this study include threats to external validity exist in part because of the sample and the context (first-year college students and the lack of consistency in implementation across courses). There also exist threats to internal validity due to the lack of power which impacted the ability to detect an effect: the sample size was limited due to incomplete data collected for all members of a peer feedback group and missing data meant the originally calculated sample size requirement to detect an effect was not met. Other limitations include the high rate of attrition within the sample and the students who elected not to participate.

Additionally, there existed the possibility of sampling bias due to the convenience sample employed in the method. The use of statistical controls through the multilevel model alleviated some of that bias, however. Furthermore, the recruitment of the sample of students is due to their instructors opting into the study. There is no evidence that there are any factors related to writing achievement or writing self-efficacy that would possible influence the choice of a first-year writing section, so the bias from a convenience sample was small. Finally, this study is correlational and can make no claims to causality among the factors.

Ethical Issues

Because all of the activities within this study are part of the natural instructional experiences of students in first-year writing, there were very few ethical concerns. Student information was linked through ID numbers in order to track individuals while maintaining de-individuated data. All the student work and student quality ratings required for the study was pulled from the Eli Review system and de-individuated prior to analysis, consistent with

38

Institutional Review Board policies. In addition, a version of the student survey used in this proposed study has already been folded into the larger University study of the use of Eli Review as part of a pilot study. This survey has been approved according to IRBs FS13: #i044775 and SS14: #i044777.

Implications

This study focused by design on authentic classrooms, with teachers and students engaging in learning about writing in a variety of ways, with different approaches to scaffolding the writing process, peer review, and revision. By tracking achievement and self-efficacy through each phase of the writing process, it was not only truer to how writing is taught in live classrooms, but also reflects the real messiness of the writing classroom. Despite this variety of approaches, no significant difference across classrooms and peer feedback groups was detected. Peer feedback groups produced a range of quality in the feedback, and yet the positive impact on achievement seen in prior studies were not detected in this larger-scale, real-world investigation. The implications of the findings for practitioners is to be thoughtful in implementing peer review. Being explicit in the goals for peer review, identifying for writers how to implement quality feedback and use it for revision are important aspects of writing pedagogy. The work of giving and receiving feedback, in the end, showed no significant impact on achievement, nor on self-efficacy. Nonetheless, the giving and receiving of feedback is still its own skill that is valued by writers everywhere, and the expanse of literature on its positive impact must be considered. Teachers of writing, then, should consider implementing peer review with an eye towards revision, modeling for students how to use comments to improve their own writing, and teaching them ways in which peer review and revision are linked in order to produce increases in achievement on writing.

APPENDICES

Appendix A: Timeline

Time points	Instruments Measures
Time 1A Prior to instruction and peer review (September)	Student pre-survey Demographic information Initial student writing self-efficacy Student prior writing achievement Initial student openness to feedback Instructor experience teaching and teaching with Eli Review
Time 1B Initial draft (September)	Pre-draft writing prompt (A) Student submit initial draft of writing assignment to Eli Review Random assignment to peer feedback groups of 3-4 individuals
Time 1C Peer review (September/October)	Feedback cycle in Eli Review Students give feedback comments on peers' initial drafts Students rate helpfulness of feedback comments Students establish a revision plan
Time 1D After peer review, instruction, and revision (September/October)	Post-draft writing prompt (A) Students submit final draft of essay for Writing Prompt (A) in Eli Review Student drafts 1 and 2 rated on Rubric and Flesch-Kincaid Grade

Time 2A Initial draft (November/December)	Pre-draft writing prompt (B) Student submit initial draft of writing assignment to Eli Review Maintain random assignment to same peer feedback groups of 3-4 individuals
Time 2B Peer review (November/December)	Feedback cycle in Eli Review Students give feedback comments on peers' initial drafts Students rate helpfulness of feedback comments Students establish a revision plan
Time 2C After instruction, peer review, and revision (November/December	Post-draft writing prompt (B) Students submit final draft of essay for Writing Prompt (B) in Eli Review Student drafts 1 and 2 rated on Rubric and Flesch-Kincaid Grade
Time 2D Final Survey (December)	Student post-survey Student writing self-efficacy Perceptions of Eli Review Help received outside of class/peer feedback groups

Appendix B: Student pre-survey and post-survey

Name:

What was your High School GPA?

What was your ACT English/SAT Verbal score?

What was your ACT composite/SAT Composite score?

Demographic questions:

Major/School:

Ethnicity

Gender

Age

Writing Self-Efficacy (Shell, Murphy, & Bruning, 1989)

Rate the probability from 0 being no chance to 100 being completely certain that you would be

able to complete the following tasks:

1. Write a letter to a friend or family member.

2. List instructions for how to play a card game.

3. Compose a will or other legal document.

4. Fill out an insurance application.

5. Write an instruction manual for operation an office machine.

6. Prepare a resume describing your employment history and skills.

7. Write a one or two sentence answer to a specific test question.

8. Compose a one or two-page essay in answer to a test question.

9. Write a term paper of 15 to 20 pages.

10. Author a scholarly article for publication in a professional journal in your field.

11. Write a letter to the editor of the daily newspaper.

12. Compose an article for a popular magazine such as Time.

13. Author a short fiction story.

14. Author a 400-page novel.

15. Compose a poem on the topic of your choice.

16. Write useful class notes.

Component Skill Subscale

1. Correctly spell all words in a one-page passage.

2. Correctly punctuate a one page passage.

3. Correctly use parts of speech (i.e. nouns, verbs, adjectives, etc.)

4. Write a simple sentence with proper punctuation and grammatical structure.

5. Correctly use plurals, verb tenses, prefixes, and suffixes.

6. Write compound and complex sentences with proper punctuation and grammatical structure.

7. Organize sentences into paragraph so as to clearly express a theme.

8. Write a paper with good overall organization (for example: ideas in order, effective

transitions, etc.)

Openness to Feedback (not included in the final analysis)

       4 point Likert-scale: Strongly Agree, Agree, Disagree, Strongly Disagree

Think of times when you received suggestions or feedback on a piece of writing.

1. I find suggestions to improve my writing helpful.

2. I have incorporated suggestions to improve my writing when revising.

3. When others read my writing, they rarely find anything wrong with it.

4. Feedback from others on my writing has been useful to me.

5. I tend to seek help with my writing from friends/family/tutors.

6. I find suggestions to improve my writing to be generally useful.

7. I am never sure how to use suggestions to improve my writing.

8. I am open to suggestions of how to improve my writing.

The Post-Survey added the following questions:

How often did you visit the Writing Center to get help with writing this semester?

How often did you ask a friend or family member for help with writing this semester?

How often did you receive extra help from your instructor or visit office hours this semester?

Appendix C: Writing Achievement Rubric

The rubric was originally modeled from one given to me by Dr. Casey McArdle, a first-year writing instructor. It is modeled after the program's outcomes for students and, in its originally conception, was a holistic rubric. Because I am interested in questions beyond general writing achievement, including whether or not peer feedback might influence some domains more than others, this study uses an analytic rubric for analysis. Reliability was calculated on a trial set of essays that were rated on the rubric by myself and another expert rater with many years teaching writing to a similar population.

	Superior: convincingly communicates a noteworthy idea to an audience through sophisticated use of rhetorical strategies	Strong: effectively conveys an insightful idea to an audience through consistent and controlled use of rhetorical strategies	Competent: communicates an idea, but does not consistently address the needs of the audience	2 Inadequate: ineffectively communicates its idea to its audience	Incompetent Fails to present its ideas to the audience and does not meet some or all of the criteria for the assignment.

Thesis/focus	Demonstrates an awareness of audience, is sophisticated, and is clearly established and maintained throughout.	Is intelligent, clearly established, and consistently addressed throughout.	Has a central idea that is conventional or general?	Is superficial and inconsistently addressed; reveals limited awareness of audience.	Lacks a central idea; has no awareness, or limited awareness of its audience and purpose.
Organization	Has a clear sense of logical order appropriate to the content and the thesis	Is logical, clear, and controlled	The essay's organization is choppy and may, at times, be difficult to follow.	Reveals no apparent strategy and lapses in focus and logic.	Is random and without focus or logic.

47

Syntax & diction	Uses sophisticated language that engages the reader; manipulates sentence length to enhance the total effect of the essay; uses precise language that expresses complex ideas clearly.	Demonstrates knowledge of and skill with complex and varied sentence constructions and vocabulary.	Demonstrates competence with language use, but sentence constructions and vocabulary may be limited or repetitive.	Contains repetitive, incorrect, or ineffective sentence structure' displays limited vocabulary.	Fails to demonstrate competency with language use; sentence constructions and vocabulary may be inappropriate, simplistic, or incoherent.
Mechanics	Contains very few errors of spelling, grammar, paragraphing or manuscript format	May contain errors, but these errors do not interfere with the essay's overall effectiveness	Contains multiple errors that hinder the essay's readability	Contains many errors that garble the meaning or intent.	Contains serious and multiple errors that seriously hinder the reading of the paper.

48

Appendix D: Common Writing Prompts

Writing Prompt, A: Personal Literacy Narrative

A personal literacy narrative is the personal story of learning to read and write. In general, personal literacy narratives asks a writer to reflect on his or her learning journey, focusing particularly on a crucial moment in this journey in order to demonstrate a larger truth about the journey as a whole. From the Norton Field Guide to Writing:

"In general, it's a good idea to focus on a single event that took place during a relatively brief period of time. For example:

any early memory about writing or reading that you recall vividly

someone who taught you to read or write

a book or other text that has been significant for you in some way

an event at school that was interesting, humorous, or embarrassing

a writing or reading task that you found (or still find) difficult or challenging

a memento that represents an important moment in your literacy development (perhaps the start of a LITERACY PORTFOLIO)

the origins of your current attitudes about writing or reading

perhaps more recent challenges: learning to write instant messages, learning to write email appropriately, learning to construct a Web page

Retrieved from:

https://www.wwnorton.com/college/english/write/fieldguide/writing_guides.asp#BLUE04

Writing Prompt B: Cultural Artifact Narrative

(adapted with permission from Dr. Casey McArdle)

A cultural artifact narrative focuses on one object and tells the story of that object and its significance to your own personal understanding of your own culture. Your job for this assignment is to write an artifact or profile analysis of a digital or multimedia "object" of some kind. The object could be a film clip, television episode, song, recording of a speech or poem, commercial, YouTube video, blog, or a website. If you have another idea, feel free to suggest it. Think about what the artifact or profile says about the culture that created it. For example, consider:

• What does this artifact or profile tell us about its creator or author?

• How does the author portray him or herself through the artifact?

• In other words, who is this person?

• What are they saying about themselves?

• What argument is this artifact or profile making?

• What assumptions are we making?

• How does this relate to you?

Appendix E: Technical Appendix

This technical appendix includes links to data files, R code, and analytic results.

```
View(Eli_Study_full)
names(Eli_Study_full)

##  [1] "course_id"              "group"
##  [3] "ID"                     "pre_overall"
##  [5] "mean_exposure"          "mean_overall"
##  [7] "pre_org"                "pre_mech"
##  [9] "pre_thesis"             "pre_syn"
## [11] "gpa"                    "ethincity"
## [13] "gender"                 "age"
## [15] "t1_SE_Score"            "t2_SE_Score"
## [17] "SEScore_diff"           "t1_CS_Score"
## [19] "t2_CS_Score"            "CS_Score_diff"
## [21] "feedback"               "mean_feedback"
## [23] "flesch_kincaid_grade"   "mean_exposureFK"
## [25] "mean_flesch_kincaid"    "post_overall"
## [27] "post_org"               "post_mech"
## [29] "post_thesis"            "post_syn"
## [31] "post_flesch_kincaid_grade" "X32"
## [33] "X33"

library(psych)
```

Libraries used in this analysis:

```
library(nlme);library(reshape);library(psych);library(Hmisc)

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##    %+%, alpha

##
## Attaching package: 'Hmisc'
```

```
## The following object is masked from 'package:psych':
##
##     describe

## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

Descriptive Statistics

```
library(psych)
Eli_subset <- Eli_Study_full[c(4:6,12,13,15,16,21:26,31)]
names(Eli_subset)

##  [1] "pre_overall"           "mean_exposure"
##  [3] "mean_overall"          "ethincity"
##  [5] "gender"                "t1_SE_Score"
##  [7] "t2_SE_Score"           "feedback"
##  [9] "mean_feedback"         "flesch_kincaid_grade"
## [11] "mean_exposureFK"       "mean_flesch_kincaid"
## [13] "post_overall"          "post_flesch_kincaid_grade"

sapply(Eli_subset, mean,na.rm=TRUE)

##            pre_overall          mean_exposure
##              3.1376147              3.1421101
##           mean_overall              ethincity
##              3.1376147              1.7211538
##                 gender            t1_SE_Score
##              0.4220183             65.8902752
##            t2_SE_Score               feedback
##             69.9877821              2.6055046
##          mean_feedback   flesch_kincaid_grade
##              2.6001529              7.9807143
##        mean_exposureFK    mean_flesch_kincaid
##              7.8886190              7.8421560
##           post_overall post_flesch_kincaid_grade
##              3.6788991              7.9577174

sapply(Eli_subset, sd, na.rm=TRUE)

##            pre_overall          mean_exposure
##              0.6156582              0.4436801
##           mean_overall              ethincity
##              0.3613561              1.3542936
##                 gender            t1_SE_Score
##              0.4961626             15.3901879
##            t2_SE_Score               feedback
```

```
##       16.6239994           1.3403865
##      mean_feedback      flesch_kincaid_grade
##       1.1208933           1.7750430
##     mean_exposureFK     mean_flesch_kincaid
##       1.4860061           1.3764621
##      post_overall post_flesch_kincaid_grade
##       0.7438729           1.8994048
```

correlation table (all variables):

```
# select variables v1, v2, v3
myvars<- c("pre_overall","mean_exposure", "post_overall","flesch_kincaid_grade","post_flesch
_kincaid_grade","feedback","ethincity","gender","age", "mean_exposureFK", "t1_SE_Score", "t
2_SE_Score")
library(Hmisc)
CorTableVariables<-Eli_Study_full[myvars]
MyCorTable <- rcorr(as.matrix(CorTableVariables))
MyCorTable

##                          pre_overall mean_exposure post_overall
## pre_overall                   1.00        0.07        0.22
## mean_exposure                 0.07        1.00       -0.01
## post_overall                  0.22       -0.01        1.00
## flesch_kincaid_grade          0.11        0.03       -0.05
## post_flesch_kincaid_grade     0.15       -0.03       -0.01
## feedback                      0.02       -0.02       -0.14
## ethincity                     0.06        0.07        0.11
## gender                       -0.01        0.03       -0.03
## age                          -0.13        0.04        0.12
## mean_exposureFK              -0.01        0.08       -0.06
## t1_SE_Score                   0.08       -0.01       -0.08
## t2_SE_Score                   0.12        0.00       -0.02
##                          flesch_kincaid_grade post_flesch_kincaid_grade
## pre_overall                   0.11                  0.15
## mean_exposure                 0.03                 -0.03
## post_overall                 -0.05                 -0.01
## flesch_kincaid_grade          1.00                  0.93
## post_flesch_kincaid_grade     0.93                  1.00
## feedback                      0.18                  0.16
## ethincity                    -0.03                 -0.10
## gender                        0.36                  0.28
## age                          -0.16                 -0.16
## mean_exposureFK               0.56                  0.51
## t1_SE_Score                  -0.03                  0.08
## t2_SE_Score                  -0.09                  0.00
##                          feedback ethincity gender   age mean_exposureFK
```

```
## pre_overall               0.02      0.06 -0.01 -0.13        -0.01
## mean_exposure            -0.02      0.07  0.03  0.04         0.08
## post_overall             -0.14      0.11 -0.03  0.12        -0.06
## flesch_kincaid_grade      0.18     -0.03  0.36 -0.16         0.56
## post_flesch_kincaid_grade 0.16     -0.10  0.28 -0.16         0.51
## feedback                  1.00     -0.18  0.07 -0.21         0.17
## ethincity                -0.18      1.00 -0.02  0.19         0.05
## gender                    0.07     -0.02  1.00  0.06         0.18
## age                      -0.21      0.19  0.06  1.00        -0.03
## mean_exposureFK           0.17      0.05  0.18 -0.03         1.00
## t1_SE_Score               0.06      0.00 -0.14 -0.05        -0.08
## t2_SE_Score              -0.08      0.14 -0.09  0.03        -0.13
##               t1_SE_Score t2_SE_Score
## pre_overall               0.08      0.12
## mean_exposure            -0.01      0.00
## post_overall             -0.08     -0.02
## flesch_kincaid_grade     -0.03     -0.09
## post_flesch_kincaid_grade 0.08      0.00
## feedback                  0.06     -0.08
## ethincity                 0.00      0.14
## gender                   -0.14     -0.09
## age                      -0.05      0.03
## mean_exposureFK          -0.08     -0.13
## t1_SE_Score               1.00      0.67
## t2_SE_Score               0.67      1.00
##
## n
##               pre_overall mean_exposure post_overall
## pre_overall           109           109          109
## mean_exposure         109           109          109
## post_overall          109           109          109
## flesch_kincaid_grade   84            84           84
## post_flesch_kincaid_grade 92         92           92
## feedback              109           109          109
## ethincity             104           104          104
## gender                109           109          109
## age                   109           109          109
## mean_exposureFK       105           105          105
## t1_SE_Score           109           109          109
## t2_SE_Score           109           109          109
##               flesch_kincaid_grade post_flesch_kincaid_grade
## pre_overall                     84                        92
## mean_exposure                   84                        92
## post_overall                    84                        92
## flesch_kincaid_grade            84                        76
## post_flesch_kincaid_grade       76                        92
```

```
## feedback                           84                92
## ethincity                          80                87
## gender                             84                92
## age                               84                92
## mean_exposureFK                    80                88
## t1_SE_Score                        84                92
## t2_SE_Score                        84                92
##                 feedback ethincity gender age mean_exposureFK
## pre_overall          109      104   109 109           105
## mean_exposure        109      104   109 109           105
## post_overall         109      104   109 109           105
## flesch_kincaid_grade      84      80    84 84            80
## post_flesch_kincaid_grade   92      87    92 92            88
## feedback             109      104   109 109           105
## ethincity            104      104   104 104           101
## gender               109      104   109 109           105
## age                  109      104   109 109           105
## mean_exposureFK          105      101   105 105           105
## t1_SE_Score          109      104   109 109           105
## t2_SE_Score          109      104   109 109           105
##               t1_SE_Score t2_SE_Score
## pre_overall          109       109
## mean_exposure        109       109
## post_overall         109       109
## flesch_kincaid_grade      84        84
## post_flesch_kincaid_grade   92        92
## feedback             109       109
## ethincity            104       104
## gender               109       109
## age                  109       109
## mean_exposureFK          105       105
## t1_SE_Score          109       109
## t2_SE_Score          109       109
##
## P
##                 pre_overall mean_exposure post_overall
## pre_overall                 0.4392      0.0223
## mean_exposure        0.4392                0.8812
## post_overall         0.0223      0.8812
## flesch_kincaid_grade    0.3144    0.7546      0.6428
## post_flesch_kincaid_grade 0.1602    0.7816      0.9431
## feedback             0.8243      0.8674      0.1539
## ethincity            0.5634      0.4935      0.2697
## gender               0.9177      0.7868      0.7502
## age                  0.1702      0.6683      0.2119
## mean_exposureFK          0.9132      0.4063      0.5686
```

```
## t1_SE_Score           0.4260    0.8975      0.4043
## t2_SE_Score           0.1998    0.9790      0.8671
##                  flesch_kincaid_grade post_flesch_kincaid_grade
## pre_overall            0.3144          0.1602
## mean_exposure          0.7546          0.7816
## post_overall           0.6428          0.9431
## flesch_kincaid_grade                   0.0000
## post_flesch_kincaid_grade 0.0000
## feedback               0.1102          0.1399
## ethincity              0.8200          0.3399
## gender                 0.0007          0.0064
## age                    0.1563          0.1232
## mean_exposureFK        0.0000          0.0000
## t1_SE_Score            0.7630          0.4468
## t2_SE_Score            0.3939          0.9627
##                  feedback ethincity gender age    mean_exposureFK
## pre_overall        0.8243   0.5634   0.9177 0.1702 0.9132
## mean_exposure      0.8674   0.4935   0.7868 0.6683 0.4063
## post_overall       0.1539   0.2697   0.7502 0.2119 0.5686
## flesch_kincaid_grade  0.1102   0.8200   0.0007 0.1563 0.0000
## post_flesch_kincaid_grade 0.1399   0.3399   0.0064 0.1232 0.0000
## feedback                    0.0745   0.4590 0.0323 0.0789
## ethincity          0.0745            0.8503 0.0477 0.5931
## gender             0.4590   0.8503          0.5600 0.0725
## age                0.0323   0.0477   0.5600        0.7509
## mean_exposureFK       0.0789   0.5931   0.0725 0.7509
## t1_SE_Score        0.5037   0.9828   0.1579 0.6083 0.4270
## t2_SE_Score        0.4114   0.1616   0.3532 0.7624 0.1711
##                  t1_SE_Score t2_SE_Score
## pre_overall        0.4260      0.1998
## mean_exposure      0.8975      0.9790
## post_overall       0.4043      0.8671
## flesch_kincaid_grade  0.7630      0.3939
## post_flesch_kincaid_grade 0.4468      0.9627
## feedback           0.5037    0.4114
## ethincity          0.9828    0.1616
## gender             0.1579    0.3532
## age                0.6083    0.7624
## mean_exposureFK       0.4270    0.1711
## t1_SE_Score                0.0000
## t2_SE_Score        0.0000
```

ACHIEVEMENT DATA

Run GLS model on the model with outcome variable (Post essay)achievement data

Intercept only model--this allows to then compare to the addition of the random intercept. Random intercept helps to improve the model. We do this by looking at the AIC/BIC. In the following cacluations I also tested the chage in -2LL and I also used ANOVA to test the differences. If there is a significant change in the -2LL, which has a chi-square distribiution, we can conclude then that the intercepts vary significantly across the peer feedback groups.

intercept0nly<-**gls**(post_overall ~ 1, data=Eli_Study_full,method= "ML")
summary(intercept0nly)

```
## Generalized least squares fit by maximum likelihood
##   Model: post_overall ~ 1
##   Data: Eli_Study_full
##      AIC     BIC   logLik
##   247.821 253.2037 -121.9105
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept) 3.678899 0.0712501 51.6336      0
##
## Standardized residuals:
##      Min        Q1       Med       Q3       Max
## -3.6179201 -0.9168702  0.4336548  0.4336548  1.7841798
##
## Residual standard error: 0.7404528
## Degrees of freedom: 109 total; 108 residual
```

courseintercept<-**gls**(post_overall ~1, data=Eli_Study_full, method="ML",correlation=**corComp Symm**(form = ~1|course_id))
summary(courseintercept)

```
## Generalized least squares fit by maximum likelihood
##   Model: post_overall ~ 1
##   Data: Eli_Study_full
##      AIC     BIC   logLik
##   242.0284 250.1025 -118.0142
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | course_id
##  Parameter estimate(s):
##      Rho
## 0.1483225
##
```

```
## Coefficients:
##              Value  Std.Error  t-value  p-value
## (Intercept) 3.583025 0.09536127 37.57316      0
##
## Standardized residuals:
##       Min        Q1        Med        Q3        Max
## -3.5322185 -0.7972712  0.5702024  0.5702024  1.9376761
##
## Residual standard error: 0.7312755
## Degrees of freedom: 109 total; 108 residual
```

CourserandomIntercept0nly<-**lme**(post_overall ~ 1, data=Eli_Study_full, random=~1|course_id, method="ML")
summary(CourserandomIntercept0nly)

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##       AIC       BIC    logLik
##   242.0284 250.1025 -118.0142
##
## Random effects:
##  Formula: ~1 | course_id
##         (Intercept)  Residual
## StdDev:   0.2816337 0.6748677
##
## Fixed effects: post_overall ~ 1
##              Value  Std.Error DF  t-value  p-value
## (Intercept) 3.583025 0.09536128 90 37.57316      0
##
## Standardized Within-Group Residuals:
##       Min        Q1        Med        Q3        Max
## -3.5309874 -0.8010468  0.3298210  0.6018837  2.1624972
##
## Number of Observations: 109
## Number of Groups: 19
```

Run model with random intercept:

Achievement data

randomIntercept0nly<-**lme**(post_overall ~ 1, data=Eli_Study_full, random=~1|group,method="ML")
summary(randomIntercept0nly)

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##       AIC       BIC    logLik
```

```
##   246.1112 254.1852 -120.0556
##
## Random effects:
##  Formula: ~1 | group
##        (Intercept)  Residual
## StdDev:  0.3207079 0.6681705
##
## Fixed effects: post_overall ~ 1
##             Value  Std.Error DF  t-value p-value
## (Intercept) 3.684596 0.08519459 75 43.24918      0
##
## Standardized Within-Group Residuals:
##     Min       Q1       Med       Q3       Max
## -3.5991018 -0.7126524  0.2456600  0.6045344  2.1011584
##
## Number of Observations: 109
## Number of Groups: 34
```

Compare log likelihood to determine whether mlm is appropriate: (testing for variability across groups)

logLik(intercept0nly)*-2

'log Lik.' 243.821 (df=2)

logLik(randomIntercept0nly)*-2

'log Lik.' 240.1112 (df=3)

anova(intercept0nly, randomIntercept0nly)

```
##                 Model df    AIC      BIC   logLik   Test  L.Ratio
## intercept0nly        1  2 247.8210 253.2037 -121.9105
## randomIntercept0nly  2  3 246.1112 254.1852 -120.0556 1 vs 2 3.709876
##                 p-value
## intercept0nly
## randomIntercept0nly  0.0541
```

logLik(CourserandomIntercept0nly)*-2

'log Lik.' 236.0284 (df=3)

logLik(intercept0nly)*-2

'log Lik.' 243.821 (df=2)

anova(intercept0nly, CourserandomIntercept0nly)

```
##                 Model df    AIC     BIC    logLik   Test
## intercept0nly            1  2 247.8210 253.2037 -121.9105
## CourserandomIntercept0nly     2  3 242.0284 250.1025 -118.0142 1 vs 2
##                 L.Ratio p-value
## intercept0nly
## CourserandomIntercept0nly 7.792615  0.0052
```

In order to account for variance between courses, adding dummy variable on student level for easier interpretation.

threelevels<-**lme**(post_overall ~ 1, data=Eli_Study_full, random=~1|course_id/group,method="ML")
summary(threelevels)

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##      AIC     BIC    logLik
##   243.9244 254.6898 -117.9622
##
## Random effects:
##  Formula: ~1 | course_id
##      (Intercept)
## StdDev:   0.2749048
##
##  Formula: ~1 | group %in% course_id
##      (Intercept)  Residual
## StdDev:   0.1276446 0.6666015
##
## Fixed effects: post_overall ~ 1
##           Value  Std.Error DF  t-value p-value
## (Intercept) 3.588798 0.09654712 75 37.17147      0
##
## Standardized Within-Group Residuals:
##     Min       Q1      Med       Q3      Max
## -3.5454580 -0.7382692  0.2868053  0.6132141  2.1773060
##
## Number of Observations: 109
## Number of Groups:
##        course_id group %in% course_id
##            19             34
```

Continue with course ID as dummy variable

Step 2 Testing fixed effects: (achievement data) and feedback

This includes the pre-achievement score and the random intercept. Notice we haven't done random slopes yet. Keeping an eye on the BIC/AIC--going up means the model has a worse fit. Going down means the model has a better fit.

Step 1: specify random intercept with first fixed effect--quality of feedback received by an

individual

```
randomInterceptFeedback<-lme(post_overall~mean_feedback, random= ~1|group, method="ML
",data=Eli_Study_full)
summary(randomInterceptFeedback)

## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##      AIC    BIC   logLik
##   243.2756 254.041 -117.6378
##
## Random effects:
##  Formula: ~1 | group
##        (Intercept)  Residual
## StdDev:  0.2677888 0.6674917
##
## Fixed effects: post_overall ~ mean_feedback
##             Value  Std.Error DF   t-value p-value
## (Intercept)   4.088005 0.19642429 74 20.812115  0.0000
## mean_feedback -0.155266 0.06891201 74 -2.253104  0.0272
##  Correlation:
##          (Intr)
## mean_feedback -0.914
##
## Standardized Within-Group Residuals:
##      Min       Q1       Med       Q3       Max
## -3.50861084 -0.74492980  0.05802084  0.58905263  2.16023431
##
## Number of Observations: 109
## Number of Groups: 34
```

Testing fixed effects: feedback received and pre-achievement

Now adding a second fixed effect to the model.

```
randomInterceptPre<-lme(post_overall~pre_overall + mean_feedback , data=Eli_Study_full, ran
dom=~1|group, method="ML",na.action=na.exclude)
summary(randomInterceptPre)
```

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
239.3865 252.8433 -114.6933
##
Random effects:
Formula: ~1 | group
(Intercept) Residual
StdDev: 0.2723272 0.6462402
##
Fixed effects: post_overall ~ pre_overall + mean_feedback
Value Std.Error DF   t-value p-value
(Intercept) 3.248078 0.3962127 73 8.197812 0.0000
pre_overall   0.265593 0.1093611 73 2.428587 0.0176
mean_feedback -0.152670 0.0679690 73 -2.246165 0.0277
Correlation:
(Intr) pr_vrl
pre_overall   -0.872
mean_feedback -0.458 0.013
##
Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-3.56916854 -0.70934859 -0.01707894 0.61837401 2.31914890
##
Number of Observations: 109
Number of Groups: 34

Notes: First fixed effects with just feedback specified: AIC BIC logLik 247.4881 258.2535 -
119.7441

Second with feedback and pre-achievement score:

AIC BIC logLik 243.5269 256.9836 -116.7634

IMPROVED the AIC/BIC.

Note that pre-overall is significant in this model. Also, can square the SD of the intercept to get
the variance across groups.

More fixed effects: add gender

```
addGender<-update(randomInterceptPre,.~.+gender)
summary(addGender)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##      AIC     BIC    logLik
##   241.3806 257.5287 -114.6903
##
## Random effects:
##  Formula: ~1 | group
##       (Intercept)  Residual
## StdDev:  0.2729187 0.6460432
##
## Fixed effects: post_overall ~ pre_overall + mean_feedback + gender
##              Value Std.Error DF  t-value p-value
## (Intercept)   3.244601 0.4006163 72  8.099024  0.0000
## pre_overall   0.265710 0.1098795 72  2.418195  0.0181
## mean_feedback -0.153134 0.0686277 72 -2.231375  0.0288
## gender        0.010319 0.1366654 72  0.075507  0.9400
##  Correlation:
##           (Intr) pr_vrl mn_fdb
## pre_overall   -0.868
## mean_feedback -0.443  0.012
## gender        -0.112  0.012 -0.095
##
## Standardized Within-Group Residuals:
##       Min        Q1        Med        Q3        Max
## -3.56246532 -0.70171453 -0.01113686  0.62404465  2.32250199
##
## Number of Observations: 109
## Number of Groups: 34
```

The AIC/BIC increased.

FIXED EFFECTS: add Ethnicity

addEthnicity<-**update**(addGender,.~.+ ethincity,na.action=na.exclude)
summary(addEthnicity)

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##      AIC     BIC    logLik
##   236.6171 255.1279 -111.3086
##
## Random effects:
##  Formula: ~1 | group
##       (Intercept)  Residual
## StdDev:  0.2614636 0.6623594
##
```

```
## Fixed effects: post_overall ~ pre_overall + mean_feedback + gender + ethincity
##              Value Std.Error DF   t-value p-value
## (Intercept)   3.225012 0.4485578 66  7.189735  0.0000
## pre_overall   0.256884 0.1208935 66  2.124881  0.0373
## mean_feedback -0.153097 0.0708385 66 -2.161217  0.0343
## gender        0.001821 0.1450229 66  0.012553  0.9900
## ethincity     0.024224 0.0513964 66  0.471324  0.6390
##  Correlation:
##            (Intr) pr_vrl mn_fdb gender
## pre_overall   -0.857
## mean_feedback -0.460  0.047
## gender        -0.035 -0.052 -0.124
## ethincity     -0.208 -0.044  0.117  0.004
##
## Standardized Within-Group Residuals:
##      Min         Q1        Med        Q3        Max
## -3.48039672 -0.68729329  0.04425127  0.65811681  2.29823824
##
## Number of Observations: 104
## Number of Groups: 34

intervals(addEthnicity, 0.95)

## Approximate 95% confidence intervals
##
##  Fixed effects:
##                 lower       est.       upper
## (Intercept)    2.35123079  3.225012011  4.09879323
## pre_overall    0.02138627  0.256884305  0.49238234
## mean_feedback -0.29108920 -0.153097335 -0.01510547
## gender        -0.28068109  0.001820511  0.28432212
## ethincity     -0.07589481  0.024224375  0.12434356
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: group
##                 lower      est.      upper
## sd((Intercept)) 0.1127271 0.2614636 0.6064486
##
##  Within-group standard error:
##    lower     est.      upper
## 0.5605887 0.6623594 0.7826058
```

Does a better job of explaining the model, however ethnicity and gender are not significant.

Add quality of feedback given

addfeedback<-**update**(addEthnicity,.~.+ feedback,na.action=na.exclude)
summary(addfeedback)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
238.6004 259.7555 -111.3002
##
Random effects:
Formula: ~1 | group
(Intercept) Residual
StdDev: 0.2613396 0.6623364
##
Fixed effects: post_overall ~ pre_overall + mean_feedback + gender + ethincity +   feedbac
k
Value Std.Error DF   t-value p-value
(Intercept)   3.234686 0.4572957 65 7.073512 0.0000
pre_overall   0.257569 0.1216242 65 2.117742 0.0380
mean_feedback -0.149591 0.0764672 65 -1.956280 0.0547
gender   0.002094 0.1457681 65 0.014365 0.9886
ethincity   0.023478 0.0520007 65 0.451502 0.6531
feedback   -0.007589 0.0603862 65 -0.125681 0.9004
Correlation:
(Intr) pr_vrl mn_fdb gender ethnct
pre_overall   -0.837
mean_feedback -0.361 0.061
gender   -0.031 -0.052 -0.110
ethincity   -0.223 -0.049 0.066 0.002
feedback   -0.168 -0.045 -0.365 -0.015 0.115
##
Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-3.49343556 -0.68990884 0.04126493 0.64637716 2.28891847
##
Number of Observations: 104
Number of Groups: 34

intervals(addfeedback, 0.95)

Approximate 95% confidence intervals
##
Fixed effects:
lower   est.   upper
(Intercept) 2.34813992 3.234686430 4.121232944
pre_overall 0.02177924 0.257568714 0.493358187

```
## mean_feedback -0.29783609 -0.149591243 -0.001346393
## gender      -0.28050253  0.002093996  0.284690521
## ethincity    -0.07733389  0.023478424  0.124290737
## feedback      -0.12465847 -0.007589387  0.109479692
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: group
##              lower     est.    upper
## sd((Intercept)) 0.1126618 0.2613396 0.6062248
##
##  Within-group standard error:
##   lower    est.    upper
## 0.5605900 0.6623364 0.7825497
```

Now we add RANDOM SLOPES, too :)

```
addRandomSlope<-lme(post_overall~feedback+mean_feedback + +ethincity+gender+pre_overa
ll+mean_overall, data=Eli_Study_full, random= ~mean_feedback|group, method="ML",na.actio
n=na.exclude)
summary(addRandomSlope)

## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##      AIC     BIC    logLik
##   243.7844 272.8727 -110.8922
##
## Random effects:
##  Formula: ~mean_feedback | group
##  Structure: General positive-definite, Log-Cholesky parametrization
##            StdDev       Corr
## (Intercept)  5.997748e-07 (Intr)
## mean_feedback 1.084905e-01 -0.415
## Residual     6.494750e-01
##
## Fixed effects: post_overall ~ feedback + mean_feedback + +ethincity + gender +     pre_over
all + mean_overall
##             Value Std.Error DF   t-value p-value
## (Intercept)   3.190132 0.7377650 65  4.324049  0.0001
## feedback     -0.005386 0.0598332 65 -0.090022  0.9285
## mean_feedback -0.149415 0.0795499 65 -1.878255  0.0648
## ethincity    0.024933 0.0512905 65  0.486123  0.6285
## gender       -0.005195 0.1453204 65 -0.035750  0.9716
## pre_overall   0.284065 0.1394176 65  2.037514  0.0457
## mean_overall  -0.012958 0.2554951 32 -0.050718  0.9599
```

```
##   Correlation:
##            (Intr) fedbck mn_fdb ethnct gender pr_vrl
## feedback      -0.142
## mean_feedback -0.211 -0.336
## ethincity     -0.085  0.118  0.069
## gender        -0.016 -0.018 -0.111  0.011
## pre_overall   -0.060 -0.061  0.062 -0.011 -0.030
## mean_overall  -0.790  0.050 -0.007 -0.067 -0.011 -0.496
##
## Standardized Within-Group Residuals:
##       Min        Q1        Med        Q3        Max
## -3.30918969 -0.63939352  0.01713599  0.59873425  1.94393848
##
## Number of Observations: 104
## Number of Groups: 34
```

Adding the impact of the exposure to the group's feedback on the 2nd level as a random slope does not improve the model fit. This means we can conclude that there is no evidence that impact of feedback had a different impact across groups, the relationship stays the same even if the magnitude differs across groups

checking for an interaction between mean_feedback and feedback

```
addInteraction<-update(addEthnicity,.~.+ mean_feedback:feedback,na.action=na.exclude)
summary(addInteraction)

## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##      AIC      BIC    logLik
##   238.6153 259.7704 -111.3076
##
## Random effects:
##  Formula: ~1 | group
##       (Intercept)  Residual
## StdDev:  0.2614569 0.6623549
##
## Fixed effects: post_overall ~ pre_overall + mean_feedback + gender + ethincity +     mean_fe
edback:feedback
##                  Value Std.Error DF   t-value p-value
## (Intercept)          3.222320 0.4554703 65  7.074708  0.0000
## pre_overall          0.257148 0.1216736 65  2.113425  0.0384
## mean_feedback       -0.149689 0.1086034 65 -1.378310  0.1728
## gender               0.002072 0.1458861 65  0.014206  0.9887
## ethincity            0.023964 0.0520356 65  0.460540  0.6467
## mean_feedback:feedback -0.000879 0.0211616 65 -0.041560  0.9670
##   Correlation:
```

```
##                 (Intr) pr_vrl mn_fdb gender ethnct
## pre_overall          -0.855
## mean_feedback         -0.406  0.070
## gender              -0.040 -0.050 -0.050
## ethincity            -0.188 -0.050 -0.015 -0.001
## mean_feedback:feedback  0.142 -0.052 -0.755 -0.042  0.120
##
## Standardized Within-Group Residuals:
##      Min        Q1        Med        Q3        Max
## -3.48534455 -0.68733684  0.04487929  0.65690849  2.29413773
##
## Number of Observations: 104
## Number of Groups: 34
```

intervals(addInteraction, 0.95)

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                    lower        est.      upper
## (Intercept)       2.33931181   3.2223195779 4.10532735
## pre_overall       0.02126273  0.2571479176 0.49303310
## mean_feedback     -0.36023555 -0.1496891364 0.06085728
## gender            -0.28075297  0.0020724008 0.28489777
## ethincity         -0.07691549  0.0239644485 0.12484439
## mean_feedback:feedback -0.04190497 -0.0008794798 0.04014601
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: group
##               lower      est.      upper
## sd((Intercept)) 0.1127625 0.2614569 0.6062272
##
##  Within-group standard error:
##    lower     est.      upper
## 0.5606057 0.6623549 0.7825714
```

No significant improvement in the model.

Achievement with Flesch-Kincaid--same procedure as rubric achievement

intercept0nlyFK<-**gls**(post_flesch_kincaid_grade ~1, data=Eli_Study_full, method="ML", na.act
ion=na.exclude)
summary(intercept0nlyFK)

```
## Generalized least squares fit by maximum likelihood
##   Model: post_flesch_kincaid_grade ~ 1
##   Data: Eli_Study_full
##      AIC     BIC   logLik
##   382.1227 387.1663 -189.0613
##
## Coefficients:
##             Value Std.Error  t-value p-value
## (Intercept) 7.957717 0.1980266 40.18508      0
##
## Standardized residuals:
##      Min        Q1       Med       Q3       Max
## -1.6980551 -0.7663717 -0.1523077  0.4802842  2.8280204
##
## Residual standard error: 1.889054
## Degrees of freedom: 92 total; 91 residual
```

randomintercept0nlyFK<-**lme**(post_flesch_kincaid_grade ~1, data=Eli_Study_full,random=~1|group, method="ML", na.action=na.exclude)

courseinterceptFK<-**gls**(post_flesch_kincaid_grade ~1, data=Eli_Study_full, method="ML",na.action=na.exclude,correlation=**corCompSymm**(form = ~1|course_id))
summary(courseinterceptFK)

```
## Generalized least squares fit by maximum likelihood
##   Model: post_flesch_kincaid_grade ~ 1
##   Data: Eli_Study_full
##      AIC     BIC   logLik
##   365.7035 373.2689 -179.8518
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | course_id
##  Parameter estimate(s):
##      Rho
## 0.3699583
##
## Coefficients:
##             Value Std.Error t-value p-value
## (Intercept) 8.048234 0.3216682 25.0203      0
##
## Standardized residuals:
##      Min        Q1       Med       Q3       Max
## -1.7414710 -0.8121893 -0.1997082  0.4312530  2.7729373
##
## Residual standard error: 1.893936
## Degrees of freedom: 92 total; 91 residual
```

summary(randomintercept0nlyFK)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
372.0619 379.6272 -183.0309
##
Random effects:
Formula: ~1 | group
(Intercept) Residual
StdDev: 1.118629 1.498129
##
Fixed effects: post_flesch_kincaid_grade ~ 1
Value Std.Error DF t-value p-value
(Intercept) 7.845229 0.2522874 58 31.0964   0
##
Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-1.7105043 -0.6001819 -0.1459556 0.4917721 2.7424230
##
Number of Observations: 92
Number of Groups: 34

logLik(intercept0nlyFK)*-2

'log Lik.' 378.1227 (df=2)

logLik(randomintercept0nlyFK)*-2

'log Lik.' 366.0619 (df=3)

anova(intercept0nlyFK,courseinterceptFK)

Model df   AIC   BIC   logLik   Test L.Ratio
intercept0nlyFK   1 2 382.1227 387.1663 -189.0613
courseinterceptFK   2 3 365.7035 373.2689 -179.8518 1 vs 2 18.41914
p-value
intercept0nlyFK
courseinterceptFK <.0001

randomInterceptFKFeedback<-**lme**(post_flesch_kincaid_grade~feedback, data=Eli_Study_full,random=~1|group, method="ML", na.action=na.exclude)
summary(randomInterceptFKFeedback)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
373.6377 383.7249 -182.8189
##

```
## Random effects:
##  Formula: ~1 | group
##        (Intercept) Residual
## StdDev:    1.078215 1.508634
##
## Fixed effects: post_flesch_kincaid_grade ~ feedback
##             Value Std.Error DF   t-value p-value
## (Intercept) 7.581939 0.4703567 57 16.119551  0.0000
## feedback   0.106410 0.1591253 57  0.668719  0.5064
##  Correlation:
##        (Intr)
## feedback -0.848
##
## Standardized Within-Group Residuals:
##      Min       Q1       Med       Q3       Max
## -1.7176471 -0.5807427 -0.1082005  0.4890403  2.7995099
##
## Number of Observations: 92
## Number of Groups: 34
```

anova(intercept0nlyFK, randomintercept0nlyFK)

```
##                 Model df    AIC     BIC   logLik   Test L.Ratio
## intercept0nlyFK       1  2 382.1227 387.1663 -189.0613
## randomintercept0nlyFK   2  3 372.0619 379.6272 -183.0309 1 vs 2 12.06082
##                 p-value
## intercept0nlyFK
## randomintercept0nlyFK   5e-04
```

Significant differences in the groups--multilevel models are justified

testing feedback controlling for pre-score

randomInterceptPreFK<-**lme**(post_flesch_kincaid_grade~flesch_kincaid_grade + feedback +course_id, data=Eli_Study_full, random=~1|group, method="ML",na.action=na.exclude)
summary(randomInterceptPreFK)

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##      AIC     BIC   logLik
##   169.1793 183.1637 -78.58963
##
## Random effects:
##  Formula: ~1 | group
##        (Intercept)  Residual
```

71

```
## StdDev:    0.178416 0.6579935
##
## Fixed effects: post_flesch_kincaid_grade ~ flesch_kincaid_grade + feedback +     course_id
##                Value Std.Error DF   t-value p-value
## (Intercept)          0.6767000 0.4150589 40  1.630371  0.1109
## flesch_kincaid_grade  0.9256551 0.0461161 40 20.072286  0.0000
## feedback            -0.0196209 0.0647071 40 -0.303227  0.7633
## course_id           -0.0047638 0.0167411 32 -0.284555  0.7778
##  Correlation:
##              (Intr) flsc__ fedbck
## flesch_kincaid_grade -0.811
## feedback          -0.263 -0.173
## course_id         -0.393 -0.002  0.074
##
## Standardized Within-Group Residuals:
##      Min        Q1       Med        Q3        Max
## -2.43387958 -0.34484465 -0.04350798  0.35007121  2.97373262
##
## Number of Observations: 76
## Number of Groups: 34
```

Step 2 Testing fixed effects: (achievement data) and feedback

This includes the pre-achievement score and the random intercept. Notice we haven't done random slopes yet. Keeping an eye on the BIC/AIC--going up means the model has a worse fit. Going down means the model has a better fit. Remember in the Bickel where he actually includes tables highlighting these factors which correspond to the various models,which might be an idea for an appendix at least.

fixed effects: mean_feedback

randomInterceptFeedbackPreFK<-**lme**(post_flesch_kincaid_grade~feedback +flesch_kincaid_grade+course_id, data=Eli_Study_full, random=~1|group, method="ML", na.action=na.exclude)
summary(randomInterceptFeedbackPreFK)

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##       AIC     BIC    logLik
##    169.1793 183.1637 -78.58963
##
## Random effects:
##  Formula: ~1 | group
##        (Intercept)  Residual
## StdDev:    0.178416 0.6579935
##
```

Fixed effects: post_flesch_kincaid_grade ~ feedback + flesch_kincaid_grade +   course_id
Value Std.Error DF   t-value p-value
(Intercept)   0.6767000 0.4150589 40 1.630371 0.1109
feedback   -0.0196209 0.0647071 40 -0.303227 0.7633
flesch_kincaid_grade 0.9256551 0.0461161 40 20.072286 0.0000
course_id   -0.0047638 0.0167411 32 -0.284555 0.7778
Correlation:
(Intr) fedbck flsc__
feedback   -0.263
flesch_kincaid_grade -0.811 -0.173
course_id   -0.393 0.074 -0.002
##
Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-2.43387958 -0.34484465 -0.04350798 0.35007121 2.97373262
##
Number of Observations: 76
Number of Groups: 34

addmeanfeedbackFK<-**update**(randomInterceptFeedbackPreFK,.~.+mean_feedback)
summary(addmeanfeedbackFK)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
170.4823 186.7974 -78.24115
##
Random effects:
Formula: ~1 | group
(Intercept) Residual
StdDev: 0.1917282 0.6514248
##
Fixed effects: post_flesch_kincaid_grade ~ feedback + flesch_kincaid_grade +   course_id +
mean_feedback
Value Std.Error DF   t-value p-value
(Intercept)   0.5846828 0.4331750 39 1.349761 0.1849
feedback   -0.0479661 0.0732383 39 -0.654931 0.5164
flesch_kincaid_grade 0.9222839 0.0466504 39 19.770097 0.0000
course_id   -0.0040488 0.0169624 32 -0.238696 0.8129
mean_feedback   0.0702974 0.0865646 39 0.812080 0.4217
Correlation:
(Intr) fedbck flsc__ cors_d
feedback   -0.106
flesch_kincaid_grade -0.752 -0.107
course_id   -0.395 0.041 -0.009
mean_feedback   -0.263 -0.458 -0.098 0.057
##

Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-2.41333043 -0.43775748 -0.07510287 0.32838622 2.94523986
##
Number of Observations: 76
Number of Groups: 34

adding mean feedback makes the model slightly worse

Bestmodel<-**lme**(post_flesch_kincaid_grade~feedback+mean_feedback + flesch_kincaid_grade +ethincity +gender+course_id, data=Eli_Study_full, random= ~1|group, method="ML",na.action =na.exclude)
summary(Bestmodel)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
163.1513 183.6413 -72.57563
##
Random effects:
Formula: ~1 | group
(Intercept) Residual
StdDev: 0.1471644 0.6470963
##
Fixed effects: post_flesch_kincaid_grade ~ feedback + mean_feedback + flesch_kincaid_grad
e +   ethincity + gender + course_id
Value Std.Error DF   t-value p-value
(Intercept)   0.8209289 0.4549691 34 1.804362 0.0800
feedback   -0.0519561 0.0747062 34 -0.695472 0.4915
mean_feedback   0.0442984 0.0868799 34 0.509881 0.6134
flesch_kincaid_grade 0.9272544 0.0513235 34 18.066844 0.0000
ethincity   -0.0943905 0.0653477 34 -1.444435 0.1578
gender   -0.1145452 0.1914205 34 -0.598395 0.5535
course_id   -0.0007045 0.0174247 31 -0.040433 0.9680
Correlation:
(Intr) fedbck mn_fdb flsc__ ethnct gender
feedback   -0.127
mean_feedback   -0.238 -0.468
flesch_kincaid_grade -0.704 -0.099 -0.112
ethincity   -0.307 0.090 0.021 0.049
gender   0.107 0.005 0.054 -0.422 -0.022
course_id   -0.307 0.055 0.053 -0.127 -0.048 0.253
##
Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-2.65466447 -0.37951355 -0.06805221 0.41575408 2.85221343

74

```
##
## Number of Observations: 72
## Number of Groups: 33

intervals(Bestmodel,0.95)

## Approximate 95% confidence intervals
##
## Fixed effects:
##                        lower        est.        upper
## (Intercept)         -0.05758430  0.8209288522  1.69944200
## feedback            -0.19620853 -0.0519560813  0.09229636
## mean_feedback       -0.12346051  0.0442984478  0.21205741
## flesch_kincaid_grade 0.82815225  0.9272543843  1.02635652
## ethincity           -0.22057241 -0.0943905429  0.03179132
## gender              -0.48416473 -0.1145451845  0.25507436
## course_id           -0.03447083 -0.0007045359  0.03306175
## attr(,"label")
## [1] "Fixed effects:"
##
## Random Effects:
##   Level: group
##                 lower        est.      upper
## sd((Intercept)) 0.008603161  0.1471644  2.517373
##
## Within-group standard error:
##    lower      est.      upper
## 0.5231676 0.6470963 0.8003815
```

SELF-EFFICACY DATA

Run GLS model on the model with outcome variable

Intercept only model--this allows to then compare to the addition of the random intercept. Random intercept helps to improve the model. We do this by looking at the AIC/BIC. In the following cacluations I also tested the chage in -2LL and I also used ANOVA to test the differences. If there is a significant change in the -2LL, which has a chi-square distribiution, we can conclude then that the intercepts vary significantly across the peer feedback groups.

```
intercept0nlySE<-gls(t2_SE_Score ~ 1, data=Eli_Study_full,method= "ML")
summary(intercept0nlySE)

## Generalized least squares fit by maximum likelihood
##   Model: t2_SE_Score ~ 1
##   Data: Eli_Study_full
```

```
##       AIC      BIC    logLik
##   925.0887 930.4714 -460.5444
##
## Coefficients:
##             Value Std.Error  t-value p-value
## (Intercept) 69.98778   1.59229 43.95416       0
##
## Standardized residuals:
##      Min       Q1       Med       Q3       Max
## -2.5676151 -0.5658102 -0.0181466  0.8127913  1.6814991
##
## Residual standard error: 16.54757
## Degrees of freedom: 109 total; 108 residual
```

Run model with random intercept:

self-efficacy

```
randomIntercept0nlySE<-lme(t2_SE_Score ~ 1, data=Eli_Study_full, random=~1|group,method
="ML")
summary(randomIntercept0nlySE)

## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##       AIC      BIC    logLik
##   927.0887 935.1628 -460.5444
##
## Random effects:
##  Formula: ~1 | group
##         (Intercept) Residual
## StdDev: 0.001463306 16.54757
##
## Fixed effects: t2_SE_Score ~ 1
##             Value Std.Error DF  t-value p-value
## (Intercept) 69.98778   1.59229 75 43.95416       0
##
## Standardized Within-Group Residuals:
##      Min       Q1       Med       Q3       Max
## -2.5676151 -0.5658102 -0.0181466  0.8127913  1.6814990
##
## Number of Observations: 109
## Number of Groups: 34
```

```
courseinterceptSE<-gls(t2_SE_Score ~1, data=Eli_Study_full, method="ML",correlation=corCo
mpSymm(form = ~1|course_id))
summary(courseinterceptSE)
```

Generalized least squares fit by maximum likelihood
Model: t2_SE_Score ~ 1
Data: Eli_Study_full
AIC   BIC   logLik
927.0861 935.1601 -460.543
##
Correlation Structure: Compound symmetry
Formula: ~1 | course_id
Parameter estimate(s):
Rho
0.002758099
##
Coefficients:
Value Std.Error t-value p-value
(Intercept) 69.98974 1.607294 43.54506   0
##
Standardized residuals:
Min   Q1   Med   Q3   Max
-2.56773105 -0.56592777 -0.01826461 0.81267260 1.68137968
##
Residual standard error: 16.54758
Degrees of freedom: 109 total; 108 residual

Compare log likelihood to determine whether mlm is appropriate: (testing for variability across groups)

```
logLik(intercept0nlySE)*-2
```

'log Lik.' 921.0887 (df=2)

```
logLik(randomIntercept0nlySE)*-2
```

'log Lik.' 921.0887 (df=3)

Significance for 1df


```
anova(intercept0nlySE, randomIntercept0nlySE)
```

Model df   AIC   BIC   logLik Test
intercept0nlySE   1 2 925.0887 930.4714 -460.5444
randomIntercept0nlySE   2 3 927.0887 935.1628 -460.5444 1 vs 2
L.Ratio p-value

intercept0nlySE
randomIntercept0nlySE 9.111545e-08 0.9998

looks like MLM Madness is not necessarily warranted HOWEVER:

p. 247 in Gelman and Hill "[There is] little risk in setting up the multi-level model becuase it reduces to classical regression."

testing feedback controlling for pre-score

randomInterceptPreSE<-**lme**(t2_SE_Score~t1_SE_Score +course_id, data=Eli_Study_full, random=~1|group, method="ML",na.action=na.exclude)
summary(randomInterceptPreSE)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
863.1153 876.5721 -426.5577
##
Random effects:
Formula: ~1 | group
(Intercept) Residual
StdDev: 0.00143538 12.11488
##
Fixed effects: t2_SE_Score ~ t1_SE_Score + course_id
Value Std.Error DF t-value p-value
(Intercept) 18.607340 5.733269 74 3.245502 0.0018
t1_SE_Score 0.733179 0.076909 74 9.533099 0.0000
course_id   0.339843 0.239865 32 1.416814 0.1662
Correlation:
(Intr) t1_SE_
t1_SE_Score -0.903
course_id   -0.423 0.050
##
Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-2.62825151 -0.72474828 0.08053093 0.64109394 2.11462583
##
Number of Observations: 109
Number of Groups: 34

testing feedback controlling for pre-score

randomInterceptFDBKPreSE<-**lme**(t2_SE_Score~t1_SE_Score + feedback+course_id, data=Eli
_Study_full, random=~1|group, method="ML",na.action=na.exclude)
summary(randomInterceptFDBKPreSE)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
862.3745 878.5225 -425.1872
##
Random effects:
Formula: ~1 | group
(Intercept) Residual
StdDev: 0.001034185 11.96352
##
Fixed effects: t2_SE_Score ~ t1_SE_Score + feedback + course_id
Value Std.Error DF   t-value p-value
(Intercept) 22.149123 6.086957 73 3.638784 0.0005
t1_SE_Score 0.740767 0.076450 73 9.689607 0.0000
feedback   -1.438828 0.879931 73 -1.635161 0.1063
course_id   0.307432 0.238817 32 1.287310 0.2072
Correlation:
(Intr) t1_SE_ fedbck
t1_SE_Score -0.821
feedback   -0.356 -0.061
course_id   -0.423 0.045 0.083
##
Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-2.5641213 -0.7706880 0.1210987 0.6438724 2.1838034
##
Number of Observations: 109
Number of Groups: 34

adding mean feedback as fixed effect

addmeanfeedbackSE<-**update**(randomInterceptFDBKPreSE,.~.+mean_feedback)
summary(addmeanfeedbackSE)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
863.9961 882.8356 -424.9981
##

```
## Random effects:
##  Formula: ~1 | group
##       (Intercept) Residual
## StdDev: 0.001016127 11.94277
##
## Fixed effects: t2_SE_Score ~ t1_SE_Score + feedback + course_id + mean_feedback
##               Value Std.Error DF   t-value p-value
## (Intercept)   23.169172  6.336802 72  3.656288  0.0005
## t1_SE_Score    0.743034  0.076776 72  9.677981  0.0000
## feedback      -1.113786  1.034980 72 -1.076143  0.2855
## course_id      0.298867  0.239969 32  1.245440  0.2220
## mean_feedback -0.745691  1.240042 72 -0.601344  0.5495
##  Correlation:
##            (Intr) t1_SE_ fedbck cors_d
## t1_SE_Score   -0.777
## feedback      -0.153 -0.026
## course_id     -0.423  0.042  0.040
## mean_feedback -0.268 -0.049 -0.522  0.059
##
## Standardized Within-Group Residuals:
##      Min        Q1        Med        Q3        Max
## -2.59753806 -0.74765014  0.07581988  0.67997679  2.24204583
##
## Number of Observations: 109
## Number of Groups: 34
```

adding gender, then ethnicity

```
addgenderSE<-update(addmeanfeedbackSE,.~.+gender)
summary(addgenderSE)

## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##       AIC      BIC    logLik
##   865.7951 887.3258 -424.8975
##
## Random effects:
##  Formula: ~1 | group
##       (Intercept) Residual
## StdDev: 0.001002995 11.93176
##
## Fixed effects: t2_SE_Score ~ t1_SE_Score + feedback + course_id + mean_feedback +     ge
nder
##               Value Std.Error DF   t-value p-value
## (Intercept)   22.341266  6.638867 71  3.365222  0.0012
## t1_SE_Score    0.748292  0.078014 71  9.591796  0.0000
```

feedback   -1.122756 1.039237 71 -1.080366 0.2836
course_id   0.316360 0.244226 32 1.295359 0.2045
mean_feedback -0.785910 1.248310 71 -0.629579 0.5310
gender   1.069470 2.452457 71 0.436081 0.6641
Correlation:
(Intr) t1_SE_ fedbck cors_d mn_fdb
t1_SE_Score -0.779
feedback   -0.141 -0.029
course_id   -0.447 0.066 0.036
mean_feedback -0.235 -0.060 -0.519 0.046
gender   -0.286 0.155 -0.020 0.164 -0.074
##
Standardized Within-Group Residuals:
Min   Q1   Med   Q3   Max
-2.5704515 -0.7522537 0.1003310 0.6927138 2.2858146
##
Number of Observations: 109
Number of Groups: 34

--Worse!

addethnicitySE<-**update**(addgenderSE,.~.+ethincity)
summary(addethnicitySE)

Linear mixed-effects model fit by maximum likelihood
Data: Eli_Study_full
AIC   BIC   logLik
830.5567 854.3562 -406.2783
##
Random effects:
Formula: ~1 | group
(Intercept) Residual
StdDev: 0.00103364 12.03217
##
Fixed effects: t2_SE_Score ~ t1_SE_Score + feedback + course_id + mean_feedback +   ge
nder + ethincity
Value Std.Error DF   t-value p-value
(Intercept) 19.274048 7.479157 65 2.577035 0.0122
t1_SE_Score   0.740160 0.083373 65 8.877657 0.0000
feedback   -0.908245 1.072432 65 -0.846902 0.4002
course_id   0.345856 0.255222 32 1.355119 0.1849
mean_feedback -0.711872 1.282941 65 -0.554875 0.5809
gender   1.342036 2.562213 65 0.523780 0.6022
ethincity   1.467985 0.923853 65 1.588982 0.1169
Correlation:
(Intr) t1_SE_ fedbck cors_d mn_fdb gender
t1_SE_Score -0.762

81

```
## feedback      -0.185 -0.007
## course_id     -0.436  0.083  0.051
## mean_feedback -0.242 -0.046 -0.501  0.041
## gender        -0.235  0.121 -0.002  0.169 -0.100
## ethincity     -0.295 -0.005  0.113  0.026  0.078  0.001
##
## Standardized Within-Group Residuals:
##     Min        Q1        Med        Q3        Max
## -2.4628430 -0.7208902  0.1304265  0.6975294  2.3508857
##
## Number of Observations: 104
## Number of Groups: 34
```

EThnicity helps a LOT!

Now we add RANDOM SLOPES, too :)

addRandomSlopeSE<-**lme**(t2_SE_Score~feedback+mean_feedback + t1_SE_Score, data=Eli_St
udy_full, random= ~mean_feedback|group, method="ML",na.action=na.exclude)
summary(addRandomSlopeSE)

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: Eli_Study_full
##       AIC      BIC    logLik
##   867.6098 889.1406 -425.8049
##
## Random effects:
##  Formula: ~mean_feedback | group
##  Structure: General positive-definite, Log-Cholesky parametrization
##            StdDev       Corr
## (Intercept)   1.778834e-03 (Intr)
## mean_feedback 2.592313e-05 0
## Residual     1.203150e+01
##
## Fixed effects: t2_SE_Score ~ feedback + mean_feedback + t1_SE_Score
##               Value Std.Error DF   t-value p-value
## (Intercept)  26.506427 5.757430 72  4.603865   0.000
## feedback     -1.164910 1.036876 72 -1.123480   0.265
## mean_feedback -0.837349 1.241100 72 -0.674683   0.502
## t1_SE_Score   0.739013 0.076909 72  9.608944   0.000
##  Correlation:
##           (Intr) fedbck mn_fdb
## feedback      -0.150
## mean_feedback -0.268 -0.526
## t1_SE_Score   -0.838 -0.028 -0.052
##
```

```
## Standardized Within-Group Residuals:
##      Min        Q1        Med        Q3        Max
## -2.6891153 -0.7239731  0.1082547  0.6986978  2.2107264
##
## Number of Observations: 109
## Number of Groups: 34
```

anova(randomInterceptFDBKPreSE,addRandomSlopeSE)

```
##                         Model df    AIC      BIC    logLik   Test
## randomInterceptFDBKPreSE    1  6 862.3745 878.5225 -425.1872
## addRandomSlopeSE            2  8 867.6098 889.1406 -425.8049 1 vs 2
##                      L.Ratio p-value
## randomInterceptFDBKPreSE
## addRandomSlopeSE     1.235347  0.5392
```

Model is significantly worse for adding the random slope

WORKS CITED

WORKS CITED

Abbott, R. D., Amtmann, D., & Munson, Je. (2006). Statistical analysis for field experiments and lognitudinal data in writing research. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *The Handbook of Writing Research* (1st ed., pp. 374–385). New York: Guillford Press.

Arum, R., & Roksa, J. (2011). *Academically adrift : limited learning on college campuses*. University of Chicago Press.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, N.J: Prentice-Hall, 1986. xiii, 617 pp.* Englewood Cliffs, NJ: Prentice Hall.

Beach, R., & Friedrich, T. (2006). Response to Writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *The Handbook of Writing Research* (1st ed., pp. 222–234). New York: Guillford Press.

Beason, L. (1993). Feedback and Revision in Writing across the Curriculum Classes. *Research in the Teaching of English*, *27*(4), 395–422.

Cho, K., & MaArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, *20*, 328–338. http://doi.org/10.1016/j.learninstruc.2009.08.006

Cho, Y. H., & Cho, K. (2010). Peer reviewers learn from giving comments. *Instructional Science*, *39*(5), 629–643. http://doi.org/10.1007/s11251-010-9146-1

DiPardo, A., & Freedman, S. W. (1988). Peer Response Groups in the Writing Classroom: Theoretic Foundations and New Directions. *Review of Educational Research*, *58*(2), 119–149. http://doi.org/10.2307/1170332

Elbow, P., & Belanoff, P. (2000). *A community of writers : a workshop course in writing*. New York: McGraw-Hill.

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, *32*, 365–387. http://doi.org/10.2307/356600

Frank, K. A., & Fahrbach, K. (1999). Organization Culture as a Complex System: Balance and Information in Models of Influence and Selection. *Organization Science*, *10*(3), 253–277. http://doi.org/10.1287/orsc.10.3.253

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social Capital and the Diffusion of Innovations Within Organizations: The Case of Computer Technology in Schools. *Sociology of Education*, *77*(2), 148–171. http://doi.org/10.1177/003804070407700203

Gere, A. R., Christenbury, L., & Sassi, K. (2003). *Writing on demand* (1st ed.). Portsmouth, NH:

Heinemann.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445–476. http://doi.org/10.1037/0022-0663.99.3.445

Hidi, S., & Boscolo, P. (2006). Motivation and writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *The Handbook of Writing Research* (1st ed., pp. 144–157). New York, NY: The Guillford Press.

Keh, C. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, *41*(1), 294–304. Retrieved from http://eltj.oxfordjournals.org/content/44/4/294.short

Kincaid, J. P., Fishburne, Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Retrieved from http://www.dtic.mil/docs/citations/ADA006655

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284. http://doi.org/10.1037/0033-2909.119.2.254

Kreft, I. G. G., & Leeuw, J. de. (1998). *Introducing multilevel modeling*. Sage.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive : The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*, 30–43. http://doi.org/10.1016/j.jslw.2008.06.002

National Commision on Writing: Writing and school reform. (2006). Retrieved from http://www.writingcommission.org/prod_downloads/writingcom/writing-school-reform-natl-comm-writing.pdf

Pajares, F., & Johnson, M. J. (1994). Confidence and competence in writing: The role of self - efficacy, outcome expectancy, and apprehension. *Research in the Teaching of English*, *28*(3), 313–331.

Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How Robust Are the Findings of Academically Adrift? *Change: The Magazine of Higher Learning*, *43*(3), 20–24. http://doi.org/10.1080/00091383.2011.568898

Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education:Theory, research, and applications*. Englewood Cliffs, NJ: Merrill.

Pritchard, R. J., & Honeycutt, R. L. (2006). The process writing approach to writing instruction:Examining its effectiveness. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *The Handbook of Writing Research* (pp. 275–290). New York: Guillford Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data anaylsis methods* (2nd ed.). Newbury Park, CA: SAGE.

Roseth, C. J., Saltarelli, A. J., & Glass, C. R. (2011). Effects of face-to-face and computer-mediated constructive controversy on social interdependence, motivation, and achievement. *Journal of Educational Psychology*, *103*(4), 804–820. http://doi.org/10.1037/a0024213

Schunk, D., & Zimmerman, B. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly*, *23*, 7–25. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/10573560600837578

Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, *81*(1), 91–100. http://doi.org/10.1037//0022-0663.81.1.91

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*, 339–343. http://doi.org/10.1016/j.learninstruc.2009.08.003

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York, NY: University of Cambridge.

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology*. http://doi.org/10.1037/0022-0663.94.4.660