

Weighted log-odds-ratio, informative Dirichlet prior method to compare peer review feedback for top and bottom quartile college students in a first-year writing program

Valerie Ross

University of Pennsylvania
3808 Walnut Street
Philadelphia, PA 19104
215-573-2729

vross@writing.upenn.edu

Mark Liberman

University of Pennsylvania
Williams Hall, Room 619
Philadelphia, PA 19104
215-898-0363

myl@cis.upenn.edu

Lan Ngo

University of Pennsylvania
3808 Walnut Street
Philadelphia, PA 19104
215-573-2729

lngo@writing.upenn.edu

Rodger LeGrand

University of Pennsylvania
3808 Walnut Street
Philadelphia, PA 19104
215-573-2729

legrand@writing.upenn.edu

u

ABSTRACT

The purpose of this paper is to use a weighted log-odds-ratio, informative Dirichlet prior method (“bag of words” approach) to analyze student comments and scores posed to MyReviewers, a web-based tool designed for collecting student writing as well as their peers’ comments and scores of their colleagues’ drafts. Our preliminary findings suggest that students who receive lower scores may also be receiving significantly different kinds of feedback that some in the field of writing studies have suggested may have a negative impact on student learning and motivation. Findings point to the possibility of identifying the effectiveness of different kinds of feedback on lower and higher performing student writers; evaluating the impact of feedback on student revision and grading practices; and identifying and analyzing symmetries and asymmetries in teacher and student feedback commentary and scores.

1. INTRODUCTION

Asking students to give writing feedback to their peers is a common practice and may be introduced to students as early as kindergarten and continue into graduate school based on the understanding that evaluating and sharing feedback with peers may accelerate the learning of the writer and peer reviewer alike. However, while much attention has been paid to the quality and nature of feedback given to student writers by teachers and writing tutors, very little exploration has been done of the impact of the feedback given by students to their peers’ work (Kelly 2016, Poe 2016, Bouzidi, L. & Jaillet A 2009, Stanley 1992).

This paper represents a preliminary investigation of data gathered as part of a larger, cross-institutional study of peer review of writing assignments in undergraduate courses. Employing the web-based tool *MyReviewers* we collected collected student writing, commentary, and scores from one semester of a first-year writing-in-the-disciplines program at the University of Pennsylvania. We used a simple “bag of words” approach to explore whether the type

of comments students receive correlates with the score they give to the students’ assignment. Results suggest that students in the lower quartile receive significantly different commentary--more prescriptive and negative--from that given to students in the upper quartile. This has considerable implications, for it suggests that good writers receive the kind of positive reinforcement from peers that many in the field of writing studies consider most effective for advancing writing skills, while underperforming writers receive the sort of commentary from peers that are generally regarded as the sort that hinders development of writing skills.

Writing feedback is typically divided into two categories: direct and indirect instruction. Direct instruction includes telling, suggesting, explaining, and exemplifying (Mackiewicz & Thompson) and is often contrasted with open problem solving or discovery learning (see, for example, Kirschner, Sweller, & Clark, 2006). In discussing their analysis of discourse between writing center tutors and writing students, Mackiewicz and Thompson (2015) describe directiveness as worrisome but admittedly necessary. Though directiveness provides students with essential knowledge, it may curtail opportunities for learners to generate solutions on their own and may not foster motivation and curiosity (Glogger-Frey, Fleischer, Gruny, Kappich, & Renkl, 2015). However, Mackiewicz and Riley’s (2003) analysis of a technical editor’s role in providing feedback to writers shows that indirect strategies are less clear in communicating the hearer’s obligation to implement the implied directive, thereby potentially creating ambiguity. Research on directiveness in various K-12 settings has also highlighted the important role of direct instruction in student learning. Glogger-Frey et al. (2015) studied the effect of an open problem solving approach (i.e. indirect approach) to teaching physics to eighth grade students. Contrary to their expectation that open problem solving would cultivate knowledge transfer, they found that the students were less prepared for learning and transfer in physics than students who had received direct instruction. Their findings echo Kirschner et al.’s observation that there is little research to support the effectiveness of solely using minimal guidance when teaching. As such, Mackiewicz and Thompson (2015) suggest a balance between directness and indirectness in providing writing support to students.

In the context of writing center tutoring, Mackiewicz and Thompson (2015) suggest that motivation plays a key role in the amount of time and effort that students devote to writing tasks. They explain the importance of motivational scaffolding strategies in encouraging students through “praise, assurances of

caring, and statements reinforcing student writers' ownership of their work" (Mackiewicz and Thompson, 2015, p. 5). Drawing on studies by Harris (1992) and Lunsford (1991), Mackiewicz and Thompson (2015) particularly emphasize the effectiveness of helping students to maintain control of their own writing, and their research underscores tutors' use of motivational scaffolding in fostering students' ability to monitor their own learning. However, the effect of motivational scaffolding may depend on various factors, including a student's self-efficacy level. For example, Boyer, Phillips, and Wallis (2008) examined tutorial dialogue in the context of computer science learning and found that direct encouragement appeared to aid students with low self-efficacy, though it may not have been helpful for high self-efficacy learners. As Boyer et al. (2008) suggest, balancing motivational scaffolding and cognitive scaffolding--which encourages students to reflect on their own thinking and reasoning (Mackiewicz & Thompson, 2015)--remains an issue to be studied.

Though students may like to receive positive feedback, including praise, research highlights the complexity of feedback in terms of circumstance and effect. Straub (1997) surveyed 142 first-year college writing students to investigate their perceptions about teacher comments on their writing and found that students preferred praise, even when it was merely in the form of the word "good" adjacent to criticism. However, students most preferred praise that was accompanied by reasons for the positive evaluation (Straub 1997). Hattie and Timperley (2007) reviewed research, including meta-analyses, on feedback and consider it to be one of the strongest influences on learning and achievement. According to their review, positive feedback may increase a student's persistence and, for high self-efficacy students, positive feedback builds their ability to cope with future negative feedback (Deci, Koestner, & Ryan, 1999; Hattie & Timperley, 2007; Swann, Pelham, & Chidester, 1988). On the other hand, students with a low level of self-efficacy may react to positive feedback by avoiding tasks to limit the risk of receiving future negative feedback (Hattie & Timperley, 2007). Research on second language learning indicates that low-performing students may continue to underperform if they are consistently given positive feedback rather than information on how to move forward (Hiver, 2014).

Negative feedback, which is equally as complicated as positive feedback, may either hinder or bolster learning, depending on the student and context. High self-efficacy learners view their performance optimistically, and therefore, may seek negative feedback to outperform on tasks (Hattie & Timperley, 2007). For low self-efficacy students, disconfirmation of their performance may adversely impact their motivation and future performance (Brockner, Derr, & Laing, 1987; Hattie & Timperley, 2007; Moreland & Sweeney, 1984). Negative feedback from either a teacher or peer may hurt a student's confidence. After surveying 200 students of English as a foreign language, Kaivanpanah, Alavi, and Sephehrinia (2015) note that negative feedback from classmates may be confusing and harmful to a student's confidence. However, disconfirmation of performance may be welcomed if presented in terms of guidance (Straub, 1997). A study by Muis, Ranellucci, Trevors, and Duffy (2015) emphasizes the complex nature of negative feedback and its impact. In examining the attitudes, emotions, engagement, and learning outcomes of kindergarten students who received immediate feedback from a literacy learning app on iPad, the researchers had expected negative evaluative feedback to decrease enjoyment and increase boredom (Pekrun, 2006). Ultimately, results from their study demonstrated that the impact of negative feedback on the students was mixed.

2. DATA SET

The Critical Writing Program has over the past decade developed and refined a genre-, discourse- and discipline-based shared curriculum for introducing students to what is called "authentic" writing situations that involve real genres, audiences, motives, and subject knowledge, as well as introduce and provide students with practice in using a shared vocabulary and concepts about writing, from knowledge domains such as genre and process to more specific aspects of rhetorical analysis and production. The curriculum emphasizes peer review and reflection throughout.

The data set consists of the work of 1,183 undergraduates, mostly first-year, who completed a writing seminar at the University of Pennsylvania in Spring 2016. This data set includes up to five drafts of a literature review as well as the peer reviews each draft received. Peer reviews consist of rubric-guided comments and numeric scores. Most drafts are accompanied by the student writer's pre- or post-outline that provides a rhetorical analysis as well as line of reasoning for the draft. All drafts receive from one to six peer reviews (comments and scores) as well as comments and scores from their instructor. In addition, most students will provide a revision plan that responds to the feedback they have received from their peers and instructor.

All undergraduate students at Penn across the four undergraduate schools--College of Arts & Sciences, Wharton, Engineering, and Nursing--are required to complete a writing seminar. Penn's writing seminars are administered by the Critical Writing Program; most seminars within the program are situated within a specific discipline, bounded by a particular, discipline-based inquiry, and taught by a Ph.D. in that discipline who frequently is engaged in the line of inquiry focused upon in the seminar. Thus the topics and disciplines vary and along these lines while all students are assigned the genre of the literature review, we recognize that literature reviews written in the bench sciences are often substantially different from those written in, say, the humanities. Some instructors teach two sections of the same topic. In addition to discipline-based seminars, we have two sets of seminars that share a single topic across disciplines: Craft of Prose (14 sections representing 161 students) and Upper Division seminars (8 sections representing 104 students). Students, sometimes with the help of their advisors, choose the type of seminar that best suits their needs, including their self-assessment of their competence and confidence as writers. Craft of Prose seminars are designed for students who may have less preparation in writing or have confidence issues or other concerns about writing. Upper Division seminars are designed for upperclassmen and transfer students who were uncomfortable in seminars designed mainly for freshmen. We also have single topic Global English classes for international students who are concerned with learning how to write for an American academic audience. For more information on our directed self-placement criteria, visit: http://writing.upenn.edu/critical/seminars/choosing_the_right_seminar.php

The specific data in this set includes peer review scores and comments produced by 1,183 undergraduates enrolled in 90 writing seminars during the spring 2016 semester. The Excel file housing the data is organized into 19 columns and 14,010 rows. The column headings include:

Column Heading	Definition
Class Code	Signifies the discipline for the writing seminar topic and the course title. The data set includes writing seminars in 21 disciplines.
Section Number	Unique numerical identifier for each writing seminar.
Instructor	Full name of the course instructor.
Date	Date on which a peer review was completed.
Project	The writing assignment and draft number. Draft 1 of the literature review was a one-on-one review. Draft 2 was a multiple-reciprocal peer review with 1-6 peers. Draft three of the literature review was a multiple reciprocal review. Drafts 4 and 5 were optional. The first draft of the public argument peer review was one-on-one, and the second draft was multiple-reciprocal. The students also completed a multiple-reciprocal peer review for their final portfolio drafts.
Student Name	Full name of the student writer (anonymized)
Grader Name	Full name of the student peer reviewer (anonymized)
Rubric Score	Represents the total score for across the 4 scoring categories identified in our rubric of Cognitive and Heuristic Processes, Invention, Reasoning, and Presentation (see below).
Final Grade	The letter grade conversion of the cumulative rubric score.
Grade Cognitive and Heuristic Processes	Peer reviewer's numerical assessment of student writer's knowledge of writing and rhetorical awareness.
Grade Invention	Peer reviewer's numerical assessment of student writer's novelty and persuasiveness for a targeted audience.
Grade Reasoning	Peer reviewer's numerical assessment of student writer's reasonableness and logical coherence.
Grade Presentation	Peer reviewer's numerical assessment of student writer's ability to produce voice, vocabulary, syntax, sentence structure, punctuation, and tone appropriate to the genre and audience.

Comment on Cognitive and Heuristic Processes	Peer reviewer's written assessment of student writer's knowledge of writing and rhetorical awareness.
Comment on Invention	Peer reviewer's written assessment of student writer's creativity, novelty, and inventiveness in what they select to persuade their target audience.
Comment on Reasoning	Peer reviewer's written assessment of student writer's reasonableness and logical coherence.
Comment on Presentation	Peer reviewer's written assessment of student writer's ability to produce voice, vocabulary, syntax, sentence structure, punctuation, tone, source handling, etc., appropriate to the genre and audience.
General Comments	Peer reviewer's final comments, insights, and observations of student writer's writing.
Combined	Compiles the peer reviewer's written assessments of Comment on Cognitive and Heuristic Processes, Comments on Invention, Comments on Reasoning, Comments on Presentation, and General Comments into one field.

2.1 Rubric

Students are given a detailed rubric, the same one used by instructors in our program to assess individual students in their classes as well as to assess mid-term and final portfolios. The rubric acts as a guide for formative as well as, at semester's end, summative assessment.

Cognition/Metacognition: Knowledge of Writing

- Recognizes the purpose of the assignment
- Conceives of a procedure for fulfilling it
- Perceives the problem(s) to be solved in the assignment
- Follows directions through all stages of the assignment
- Able to detect flaws in reasoning in one's own or other's reasoning (outlines and peer reviews)
- Able to identify and evaluate (in plan, outlines, peer reviews, cover letters, other artifacts):
 - Rhetorical Strategies
 - Audience
 - Purpose
 - Genre
 - Plan/Arrangement
 - Complex Synthesis
 - Presentation

Invention: Idea/Audience (test of novelty, creativity, persuasion)

- Selection of an appropriate and engaging subject within the topic
- Ability to select and work successful within a genre
- Selection of an appropriate proposition and reasons to support it, attuned to the audience and purpose

- Selection of the appropriate amount and type of evidence and materials to support the proposition, attuned to the audience and purpose
- Arrangement and style attuned to the audience, purpose, and genre, including ability to evaluate the strength of reasons and evidence
- Identification of shared premises to enable an effective introduction and conclusion
- Ability to grasp feedback or detect problems with invention and revise accordingly
- Ability to vary voice and style to accommodate different audiences and genres

Reasoning: Development/Coherence (test of reasonableness)

- Creation or selection of an appropriately justificatory or explanatory proposition
- Creation or selection of reasons that directly support the proposition
- Selection of evidence that confirms, illuminates or otherwise develops the reasons
- Ability to test argument through strategies of counterargument
- Demonstration of logical coherence: all reasons support the proposition, all evidence supports the reasons, and to the extent possible, reasons do not contradict each other
- Demonstration of semantic coherence: sentences and paragraphs stick together

Presentation

- Ability to produce a voice and style appropriate to the genre and audience
- Control of vocabulary, syntax, sentence structure, punctuation, tone
- Ability to integrate rhetorical strategies and sources so that they create a consistency of style appropriate to the genre and audience
- Demonstrated ability to proofread and polish work for an outside reader
- Creation and use of grammar checklist to identify context and patterns of error in mechanics and usage, as well as to correct them
- Appropriate formatting, citation, documentation of sources

3. METHOD

We exclude peer reviews where the numerical Rubric Score is listed as 0, since many of these appear to be cases where the score is simply missing, although the reviewer evaluated the draft positively. We also exclude reviews where the score was greater than 0 but less than 2 out of 4 -- these represent evaluations with failing grades, and in such cases the reviewers often didn't bother to supply detailed comments. Among the remaining examples, the bottom quartile includes 3046 reviews with scores between 2 and 3.3 out of 4, and the top quartile includes 3054 reviews with scores above 3.78. The combined comments in the bottom-quartile reviews comprise 1,022,709 words, and the combined comments in the top-quartile reviews comprise 759,637 words.

In order to evaluate the degree of association between individual words* [FN: here "word" simply means a string of alphabetic letters set off by space, punctuation, or other symbols] and score

quartiles, we used the "algorithm from section 3.5.1 of Monroe et al. 2008.

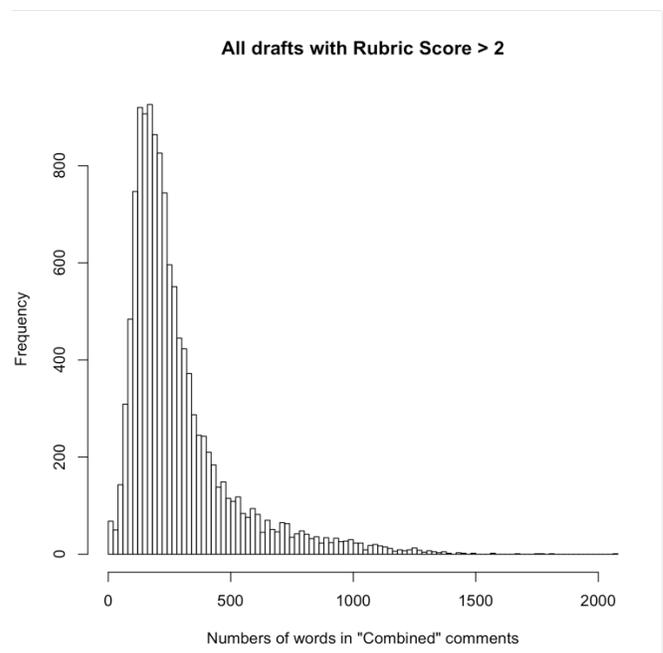
This method, originally developed for a study of political writing, starts with a simple ratio of estimated word frequencies in two collections of text. The problem is that when the overall frequency of a word is low, so is our confidence that the ratio is not a statistical accident -- and so is the value of that word as a predictive marker of the distinction under study. Thus in this collection, the word *judged* occurs five times in the bottom-quartile reviews, and only once in the top quartile; given the different overall word counts in the two groups, the maximum-likelihood estimate is that *judged* is about 4 times as common in bottom-quartile reviews as in top-quartile reviews. But we can't be very confident that in the next batch of reviews, this ration might not be quite different, or even reversed. And in any case, *judged* doesn't occur often enough to be a very strong indicator of a reviewer's sentiment.

In contrast, the word *should* occurs 3,780 times in the bottom-quartile comments, and 1,914 times in the top-quartile comments. Allowing for the groups' overall word counts, this tells us that the frequency of *should* is about 1.5 times greater in the bottom-quartile comments than in the top-quartile comments. But in this case, the overall frequency is high enough that we can be fairly confident that *should* will also be about 50% more frequent in the low-quartile comments in next semester's sample -- and *should* is common enough to be a useful indicator of overall review sentiment.

In order to deal with these issues, the cited method shrinks the odds ratio for each word based on a factor derived from a simple statistical estimate of the process generating the counts, along with an estimate of that word's overall frequency in a relevant more general source. The result is a number, the "weighted log-odds ratio", that we can use to rank words according to their apparent affinity for one text sample or the other.

4. PRELIMINARY RESULTS

The bottom quartile has more words per (combined) comment than the top quartile: 336 vs. 249.



The words most reliably associated with the bottom quartile include:

Word	Low Q Count	Low Q Freq Per Million	High Q Count	High Q Freq Per Million	Weighted Log Odds Ratio
be	10219	9992.09	5710	7516.75	7.261
sentence	9278	9071.98	5092	6703.2	7.221
more	7664	7493.82	3505	4614.05	10.11
paragraph	7001	6845.54	3443	4532.43	8.445
not	6424	6281.36	3516	4628.53	6.309
but	5123	5009.25	2742	3609.62	5.949
should	3780	3696.07	1914	2519.62	5.687
some	2984	2917.74	1529	2012.8	4.925
however	2701	2641.02	1255	1652.1	5.617
than	1938	1894.97	945	1244.02	4.536
seems	1719	1680.83	720	947.821	5.626
sure	1268	1239.84	549	722.714	4.78
rather	1052	1028.64	425	559.478	4.708
try	888	868.282	316	415.988	4.962
needs	793	775.392	253	333.054	5.439
media	731	714.768	208	273.815	5.235
pass	300	293.339	35	46.0746	5.514
chaplin	205	200.448	19	25.0119	4.765
mid	168	164.27	8	10.5313	4.835

The words most reliably associated with the top quartile include:

Word	Low Q Count	Low Q Freq Per Million	High Q Count	High Q Freq Per Million	Weighted Log Odds Ratio
the	71418	69832.2	56903	74908.1	-5.51

and	26778	26183.4	23808	31341.3	-8.526
is	20103	19656.6	16680	21957.9	-4.52
very	3391	3315.7	5269	6936.21	-14.319
well	3474	3396.86	4763	6270.1	-11.643
was	3185	3114.28	3842	5057.68	-8.508
good	3222	3150.46	3169	4171.73	-4.742
topic	2738	2677.2	2751	3621.47	-4.619
piece	2647	2588.22	2591	3410.84	-4.236
clear	2206	2157.02	2211	2910.6	-4.13
all	1773	1733.63	2083	2682.86	-5.637
job	1440	1408.03	1984	2611.77	-7.544
great	1149	1123.49	1811	2384.03	-8.504
really	1330	1300.47	1682	2214.22	-6.091
interesting	1447	1414.87	1653	2176.04	-4.985
easy	676	660.99	1096	1442.79	-6.835
strong	824	805.703	1091	1436.21	-5.281
read	906	885.882	1059	1394.09	-4.177
written	593	579.833	770	1013.64	-4.303
liked	130	127.113	324	426.52	-5.147
enjoyed	99	96.8017	257	338.319	-4.64
picasso	27	26.4005	112	147.439	-4.294
twins	1	0.977795	111	146.122	-5.658
identical	5	4.88898	76	100.048	-4.47
poincare	0	0	63	82.9343	-4.81

There's obviously some mixture of content with commentary here, but overall it makes sense -- most if not all of the content admixture could be removed by limiting the list to words that occur at least at a rate of 100 per million in both sets as well as removing words that are related to the specific content of a particular seminar, such as "Picasso."

"But" is the 6th most negative word:
but 5123 (5009.25) 2742 (3609.62) 5.949

The evaluations in the top quartile by rubric score are more heavily weighted with evaluative words of all types. The

evaluations in the lower quartile are more heavily weighted with prescriptive terms such as “should,” “try,” “more,” “needs,” “unnecessary.” One symptom: on the negative side, there's just one evaluative word that's both reasonably common overall (frequency > 15/100k) and more than twice as common in the lower quartile of evaluations: "unclear". The rest of that end of the list of the lower quartile includes:

WORD	RATIO Q4/Q1
unclear	2.004
incorrect	1.969
unnecessary	1.825
needs	1.729
clearer	1.688

At the upper end of the quartile, more than 20 positive terms are employed:

WORD	RATIO Q4/Q1
easy	2.939
great	2.857
very	2.816
nice	2.716
flows	2.553
logically	2.547
organized	2.500
job	2.497
well	2.485
supported	2.456

fits	2.419
strong	2.400
really	2.292
nicely	2.251
convincing	2.211
presentation	2.155
persuasive	2.122
coherent	2.118
engaging	2.111
interesting	2.071
consistent	1.983
supports	1.949
clearly	1.932
helps	1.927
appropriate	1.925

This may be because the Q4 (lower-scoring) evaluations are full of specific complaints/suggestions such that even the negative evaluative words are diluted in frequency.

4. FUTURE DIRECTIONS

Our next steps will include identifying cases where the commentary and grade are asymmetrical. We will also analyze instructor commentary and scores to see if similar patterns obtain. Of particular interest will be to explore correlations between the types and length of commentary as well as scores from peers in relation to improvement in drafts and final scores. We anticipate a corpus of several thousand papers from our own program as well as from the other institutions participating in a broader NSF-funded study of peer review: USF, MIT, UCNS, and Dartmouth, and will engage in a range of corpus-based approaches to text analysis. We hope to

contribute to research on the effectiveness of different kinds of feedback, particularly as concerns struggling writers.

5. ACKNOWLEDGMENTS

This research is supported by the National Science Foundation under Award #1544239, "Collaborative Research: The Role of Instructor and Peer Feedback in Improving the Cognitive, Interpersonal, and Interpersonal Competencies of Student Writers in STEM courses. Our thanks for the vital contributions of Joe Moxley, Principal Investigator, Alex Rudniy and the USF team Rajeev Reddy Rachamalla, Dat Ba Le, Brooke Downey, and Natalie Kass; Norbert Elliot, Professor Emeritus of English at New Jersey Institute of Technology and our NSF grant consultant; fellow investigators Chris Anson, Suzanne Lane, and Christiane Donahue; and Brighid Kelly, who solves all problems, big and small.

6. REFERENCES

- [1] Boyer, K. E., Phillips, R., Wallis, M., Vouk, M., & Lester, J. (2008). Balancing cognitive and motivational scaffolding in tutorial dialogue. *Intelligent Tutoring Systems, 5091*, 239-249. DOI= <https://link.springer.com>
- [2] Bouzidi, L., & Jaillet, A. (2009). Can Online Peer Assessment be Trusted?. *Educational Technology & Society* 12 (4), 257-268.
- [3] Brockner, J., Derr, W. R., & Laing, W. N. (1987). Self-esteem and reactions to negative feedback: Towards greater generalizability. *Journal of Research in Personality, 21*(3), 318-334. DOI= [10.1016/0092-6566\(87\)90014-6](https://doi.org/10.1016/0092-6566(87)90014-6)
- [4] Deci, E. L., Koestner, R., & Ryan, M. R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627-668. DOI= [10.1037/0033-2909.125.6.627](https://doi.org/10.1037/0033-2909.125.6.627)
- [5] Glogger-Frey, Fleischer, C, Gruny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction, 39*, 72-87. DOI=[10.1016/j.learninstruc.2015.05.001](https://doi.org/10.1016/j.learninstruc.2015.05.001)
- [6] Harris, M. (1992). Collaboration is not collaboration is not collaboration: Writing center tutorials vs. peer-response groups. *College Composition and Communication, 43*(3), 369-383. DOI= <https://www.jstor.org>
- [7] Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. DOI= [10.3102/003465430298487](https://doi.org/10.3102/003465430298487)
- [8] Hiver, P. (2014). Attractor states. In Z. Dornyei, A. Henry, & P.D. MacIntyre (Eds.), *Motivational dynamics in language learning* (pp. 20-28). Multilingual Matters: Bristol, UK.
- [9] Kaivanpanah, S., Alavi, S. M., & Sepehrinia, S. (2015). *The Language Learning Journal, 43*(1), 74-93. DOI= [10.1080/09571736.2012.705571](https://doi.org/10.1080/09571736.2012.705571)
- [10] Kelly, L. Effectiveness of Guided Peer Review of Student Essays in a Large Undergraduate Biology Course. *International Journal of Teaching and Learning in Higher Education* 2015, Volume 27, Number 1, 56-68 DOI= <http://www.isetl.org/ijtlhe/> ISSN 1812-9129
- [11] Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86. http://dx.doi.org/10.1207/s15326985ep4102_1
- [12] Lunsford, A. (1991). Collaboration, control and the idea of a writing center. *The Writing Center Journal, 12*(1), 3-10. <https://www.ebscohost.com/academic/education-source>
- [13] Mackiewicz, J., & Thompson, I. K. (2015). *Talk about writing: The tutoring strategies of experienced writing center tutors*. New York, NY and London, UK: Routledge.
- [14] Moreland, R. L., & Sweeney, P. D. (1984). Self-expectancies and reactions to evaluations of personal performance. *Journal of Personality, 52*(2), 156-176. DOI=[10.1111/j.1467-6494.1984.tb00350.x](https://doi.org/10.1111/j.1467-6494.1984.tb00350.x)
- [15] Muis, K. R., Ranellucci, J., Trevors, G., Duffy, M. C. (2015). The effects of technology-mediated immediate feedback on kindergarten students' attitudes, emotions, engagement and learning outcomes during literacy skills development. *Learning and Instruction, 38*, 1-13. DOI=<http://dx.doi.org/10.1016/j.learninstruc.2015.02.001>
- [16] Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315-341. DOI= [10.1007/s10648-006-9029-9](https://doi.org/10.1007/s10648-006-9029-9)
- [17] Poe, S. and Emily O. Gravett (2016) Acknowledging Students' Collaborations through Peer Review: A Footnoting Practice, *College Teaching, 64*:2, 73-83, DOI= [10.1080/87567555.2015.1094441](https://doi.org/10.1080/87567555.2015.1094441) <http://dx.doi.org/10.1080/87567555.2015.1094441> Published online: 08 Mar 2016.
- [18] Stanley, Coaching Student Writers to Be Effective Peer Evaluators, *JOURNAL OF SECOND LANGUAGE WRITING, 1*(3), 217-233 (1992)
- [19] Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English, 31*(1), 91-119. DOI= <http://www.jstor.org/stable/40171265>
- [20] Swann, W. B., Pelham, B. W., & Chidester, T. (1988). Change through paradox: Using self-verification to alter beliefs. *Journal of Personality and Social Psychology, 54*(2), 268-273. DOI= [10.1037/0022-3514.54.2.268](https://doi.org/10.1037/0022-3514.54.2.268)