

1-2021

Exploring the Relationship Between Clout and Cognitive Processing in MOOC Discussion Forums

Robert L. Moore
Old Dominion University, robmoorephd@gmail.com

Cherng-Jyh Yen
Old Dominion University, cyen@odu.edu

F. Eamonn Powers
Old Dominion University, eamonnpowers@gmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/stemps_fac_pubs



Part of the [Curriculum and Instruction Commons](#), and the [Online and Distance Education Commons](#)

Original Publication Citation

Moore, R.L., Yen, C-J., & Powers, F.E. (2020). Exploring the relationship between clout and cognitive processing in MOOC discussion forums. *British Journal of Educational Technology*, 52(1), 482-497. <https://doi.org/10.1111/bjet.13033>

This Article is brought to you for free and open access by the STEM Education & Professional Studies at ODU Digital Commons. It has been accepted for inclusion in STEMPS Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Exploring the relationship between clout and cognitive processing in MOOC discussion forums

1 2 3 Robert L. Moore , Cherng-Jyh Yen and F. Eamonn Powers

Robert L. Moore is an assistant professor of instructional design and technology at Old Dominion University. He graduated with his PhD in Instructional Technology from North Carolina State University. His research interests include using learning analytics to explore and understand student engagement behaviors in Massive Open Online Courses. Cherng-Jyh Yen is an associate professor of educational research and statistics at Old Dominion University. He received his PhD in Educational Research from the University of Virginia. He specializes in quantitative research design and data analysis. F. Eamonn Powers is a doctoral student in the Instructional Design and Technology program at Old Dominion University. His research interests are centered around emergent technology use in complex technical training environments. Address for correspondence: Robert L. Moore, Darden College of Education and Professional Studies, Old Dominion University, 4301 Hampton Blvd, Suite 4128, Norfolk, VA 23529, USA. Email: robmoorephd@gmail.com

Abstract

4 Moore *et al.* (2019)'s exploratory study of the predictors of cognitive processing within massive open online course (MOOC) discussion forums identified a negative relationship between clout and cognitive processing. Cognitive processing, a category variable generated by the automated text analysis tool, Linguistic Inquiry Word Count (LIWC), is made up of six sub-scores (insight, causation, discrepancy, tentativeness, certainty and differentiation). Clout is a summary variable in LIWC used to understand the level of confidence in writing (Abe, 2020; Jordan *et al.*, 2019). However, clout is a nontransparent variable, meaning we do not know the algorithm used to calculate its value. To better understand clout and increase the specificity of Moore *et al.*'s findings, this study examines this negative relationship. A series of linear mixed models were fitted to evaluate, after controlling for gender, degree and type of post, if the focal relationship between each sub-score of cognitive processing as the predictor variables and clout as the dependent variable changed across courses with different pacing (self-paced or instructor-paced). Next, the focal relationship between each predictor and clout was examined with or without regard to pacing. Findings showed words classified as showing discrepancy, certainty or differentiation were negatively associated with clout scores.

We use words to express our thoughts and feelings, and in an educational setting, words are how we demonstrate our learning. In online learning environments, students spend a significant amount of time engaging and writing in asynchronous discussion forums (Moore, 2016). It is in these forums where students have a space to share ideas and interact with each other while discussing the course content and areas of confusion (Beckmann & Weber, 2016; Kent, Laslo, & Rafaeli, 2016; Sharif & Magrill, 2015). Examining the transcripts of these asynchronous discussions can identify the patterns of engagement as well as demonstrations of critical thinking by students (Hayati, Chanaa, Khalidi Idrissi, & Bennani, 2019).

A technique to examine the types of words being used and identify patterns is known as text mining (Hayati *et al.*, 2019; Moore, 2019). With text mining approaches, a corpus is examined to

Practitioner Notes

What is already known about this topic?

- While a majority of student-to-student interaction happens in MOOC discussion forums, the volume of text makes it a challenge to analyze and study these interactions.
- The use of computational linguistics analysis can provide insight into a student's comfort level with the content and confidence in their own knowledge.
- Clout can show the perspective that the author is speaking from—whether it is from high expertise or from a more uncertain or less confident position.

What this paper adds?

- Words classified as either discrepancy, certainty or differentiation negatively impact the demonstration of clout.
- Higher or lower clout scores are not indicative of student performance or ability but are part of the overall insight into a student's learning progression and development.
- Because clout measures the level of certainty, posts that are higher in clout may in turn limit continued engagement within MOOC discussion forums.

Implications for practice and/or policy.

- Understanding the relationship between clout and cognitive processing can lead to discussion prompts that foster more engagement and interaction.
- Identifying specific types of words to either avoid or encourage within discussion forums can influence the crafting of discussion prompts.
- Instructors and course designers can examine posts after a course and determine if the prompts are leading to the type of interaction that is most appropriate for the course.

find specific words, themes or other patterns that can help to examine the context and content of that corpus (Hayati *et al.*, 2019; Moore, 2019). While this type of analysis can be an effective way to measure and understand student engagement and learning within discussion forums, manual analysis is tedious and time consuming (Darabi, Arrastia, Nelson, Cornille, & Liang, 2011; Dowell & Graesser, 2014; Kovanović, Gašević, Hatala, & Siemens, 2017). This is because content analysis is a two-step process that first involves reading the information and then, converting that information into shared meaning (Krippendorff, 2019; Pilny, McAninch, Slone, & Moore, 2019). While this can be manageable with smaller classes that do not generate many discussion forum posts, it is untenable for large corpus, such as those generated in a massive open online course (MOOC) that may have thousands of learners engaging within the discussion forum. This challenge of manual text analysis can be mitigated by leveraging an automated content analysis tool (Krippendorff, 2019).

Research aim

In a study on pacing within MOOC discussion forums, clout was negatively associated with cognitive processing, perhaps due to the showing of more confidence not fostering as much openness for discussion (Moore, Oliver, & Wang, 2019). However, one of the limitations of that study is that clout is a nontransparent summary variable of LIWC. This means that while we can understand what the scores mean, we have limited information regarding the components (ie, words) that go into the calculation of the scores. One way to understand clout is by analyzing it alongside a more

transparent variable (Newman, Jones, & Ritter, 2016), in this case cognitive processing. Unlike the summary variable clout, the cognitive process dimension is made up of specific target words which can be readily identified by researchers during analysis. In this study, we use the sub-scores of cognitive processing to help better illustrate the specific meaning of clout and its potential usefulness in practice.

To expand our understanding of the relationship between clout and cognitive processing, we designed a new study using the same dataset from the study by Moore *et al.* (2019) to examine what role the pacing condition and other predictor variables had on the relationship between cognitive processing and clout. Considering the previously found relationships between the cognitive processing dimension and clout, this study provides a more focused approach to delve deeper in the specifics of those relationships. Our focal dependent variable is clout, as we are interested in learning more about this nontransparent summary variable. Clout has been shown to be an indicator of confidence and comfort level, so we wanted to better understand how clout relates to specific words in different cognitive processing categories. Thus, with the use of the sub-category scores of cognitive processing (eg, insight and causation) as the independent variables, a more nuanced examination of the interplay between the transparent and nontransparent variables of cognitive processing and clout is rendered possible. Our study answered the following questions:

1. How will course pacing moderate the focal predictive relationship between each dimension of cognitive processing (insight, causation, discrepancy, tentative, certainty & differentiation) and clout of post words after controlling for gender, degree and type of posts?
2. Given the moderator effect of course pacing, how will each dimension of cognitive processing predict clout in two differently paced courses respectively, after controlling for gender, degree and type of posts?
3. Given no moderator effect of course pacing, how will each dimension of cognitive processing predict clout regardless of course pacing, after controlling for gender, degree and type of posts?

Literature review

Massive open online courses (MOOCs) allow global learners to participate in an online learning environment. Shah (2018) estimated that as of 2017, nearly 81 million students had registered for a MOOC. These courses are either self-paced or instructor-paced (Moore *et al.*, 2019) and are primarily classified as either a cMOOC or an xMOOC (Bonk, Lee, Reeves, & Reynolds, 2017). Most MOOCs are xMOOCs and these have the familiar course structure that focuses on instructor-generated content, discussion forums and self-assessments (Adams *et al.*, 2014; Bonk *et al.*, 2017). One reason that the majority of MOOCs are xMOOCs is because these are typically developed and taught by faculty from traditional higher-education institutions such as MIT, Harvard and University of Texas at Arlington (Al-Imarah & Shields, 2019; Veletsianos & Shepherdson, 2016).

In MOOCs, much of the course content engagement and student-to-student engagement happens within the discussion forums (Gardner & Brooks, 2018; Martin & Ndoeye, 2016; Moore *et al.*, 2019). This creates a rich data source that can provide unique opportunities to study learner engagement and experience within these massive courses (Gardner & Brooks, 2018). However, these forums also present challenges due to the amount of data that is generated. Being able to efficiently mine this data for context, understanding and patterns is a challenge that can be addressed through the use of text mining, specifically automated text analysis techniques and tools (Dowell, Graesser, & Cai, 2016; Gardner & Brooks, 2018; Kovanović *et al.*, 2017; Krippendorff, 2019; Lin, Yu, & Dowell, 2020; Moore *et al.*, 2019).

Linguistic inquiry word count (LIWC)

The Linguistic Inquiry Word Count (LIWC; Pennebaker, Booth, Boyd & Francis, 2015) is a validated tool that searches for target words or word stems from an extensive dictionary, categorizes them into linguistic dimensions and then, converts the raw counts to percentages of total words (Abe, 2020; Khazaei, Lu, & Mercer, 2017; Simms *et al.*, 2017; Smith-Keiling & Hyun, 2019). LIWC uses predetermined categories and will analyze and classify words within these preassigned categories (Pilny *et al.*, 2019). This dictionary approach allows for a consistent measurement of words resulting in concurrent validity (Humphreys & Wang, 2018). LIWC has been used to look at social media, online forums, written notes, reflections, self-introductions and admissions essays (Oliver, Houchins, Moore, & Wang, 2020; Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014; Robinson, Navea, & Ickes, 2013). LIWC has also been used for final course grades (Robinson *et al.*, 2013), group dynamics and their interactions (Van Swol & Kane, 2018), video reflections of makerspace projects (Oliver *et al.*, 2020), deception (Newman, Pennebaker, Berry & Richards, 2003) and interactions between people (Kacewicz, Pennebaker, Davis, Jeon, & Graesser, 2014).

Clout

Clout is one of four summary variables in LIWC meant to measure the level of confidence and certainty in the text (Kacewicz *et al.*, 2014; Pennebaker, Booth, *et al.*, 2015). The proprietary algorithm which calculates the clout score was developed based on results of studies focusing on personal interactions (Kacewicz *et al.*, 2014). This focus on personal interactions makes the clout score particularly relevant in an analysis of online discussion forums. The analysis of a text using the clout score helps to determine whether a writer is writing from a perspective of confidence and certainty or using words more indicative of doubt and uncertainty (Jordan, Sterling, Pennebaker, & Boyd, 2019). A high clout score suggests the author is demonstrating confidence and high expertise while a low score suggests a more tentative manner of speaking (Jordan *et al.*, 2019; Pennebaker, Booth, *et al.*, 2015). It is important to note that the use of “confidence” or “tentative” is not meant to assign value propositions to what the students are saying. The clout score is not insinuating that a tentative response is less accurate or valuable. Instead, it is identifying how comfortable and confident the speaker is about the subject they are speaking on; in a learning environment unease is not necessarily a bad thing. In fact, researchers have found that clout scores have improved over a course as students became more confident in their expertise and knowledge (Zhu, Herring, & Bonk, 2019). Students who are referring to others’ work or how they have connected to others’ work, may have a higher clout score as they would use more “we” and “you” pronouns, compared with those whose work was not social using “I” first-person singular pronouns (Adaji & Olakanmi, 2019; Duncan, Chohan, & Ferreira, 2019; Oliver *et al.*, 2020). Using more outward-facing words, the poster is demonstrating that they are more focused on the people that they are interacting with versus themselves (Adaji & Olakanmi, 2019). Understandably so, this type of interaction would be ideal in a discussion forum and foster the type of interaction and engagement that is necessary for knowledge construction. Clout has also been described as an indication of expertise in professional settings (Newman *et al.*, 2016), a barometer for relative social status or leadership (O’Dea, Larsen, Batterham, Calear, & Christensen, 2017) and even a bellwether for bias (Hagiwara, Slatcher, Eggly, & Penner, 2017). The role of clout has been examined in a variety of contexts including education (Moore *et al.*, 2019; Oliver *et al.*, 2020; Smith-Keiling & Hyun, 2019; Smith-Keiling, Swanson, & Dehnostel, 2018), mental health (O’Dea *et al.*, 2017; Tadesse, Lin, Xu, & Yang, 2019), dreams (Bulkeley & Graves, 2018), social hierarchies (Kacewicz *et al.*, 2014) and bias (Hagiwara *et al.*, 2017).

Cognitive processing

Unlike the summary variable clout, cognitive processing is a dimension made up of specific target words. In total, there are 797 words that are indicative of cognitive processing that can be further

divided into sub-scores of insight, causation, discrepancy, tentativeness, certainty and differentiation (Figure 1; Pennebaker, Booth, *et al.*, 2015; Pennebaker, Boyd, Jordan, & Blackburn, 2015). Causation words (eg, “because,” “effect” and “hence”) and insight words (eg, “think,” “know” and “consider”) have been related to an individual’s active processing of events and have been present at higher levels within emotional and traumatic writing, as the individual is creating explanations for difficult cognitive processes (Tausczik & Pennebaker, 2010). Furthermore, studies have found that tentative words (eg, “maybe,” “perhaps” and “guess”) and filler words (eg, “I mean,” “you know” and “blah”) are used more often by individuals who are not sure about a topic, including those who are disclosing an event for the first time (Tausczik & Pennebaker, 2010). The use of tentative words (eg, “maybe” and “perhaps”) was found to be greater among those lower in the social hierarchy (Kacewicz *et al.*, 2014), which seemingly has direct parallels to the clout score.

Methods

We used the same data from Moore *et al.* (2019) and built our models with research variables different from the previous study due to the changes in research questions. A brief overview of the participants and context of the data are provided below.

Participants

This study uses data from the 671 students that made discussion forum posts in either the self-paced or instructor-paced offerings of the *Visualizing Japan (1850s–1930s): Westernization, Protest Modernity* MOOC co-offered by HarvardX and MITx. A data use agreement with the Harvard VPAL Research Team provided the identifiable discussion forum transcripts and the self-reported level of education and gender of participants. Students that contributed to the discussion forum were predominantly female (53%) and highly educated, with 41% holding either a bachelor’s or associate’s degree, 30% with a master’s and 5% with doctoral degrees.

Context

Visualizing Japan (1850s–1930s): Westernization, Protest Modernity, is a MOOC offered as a collaboration between MITx and HarvardX and was offered in both a self-paced and instructor-paced

COLOUR online, B&W in print

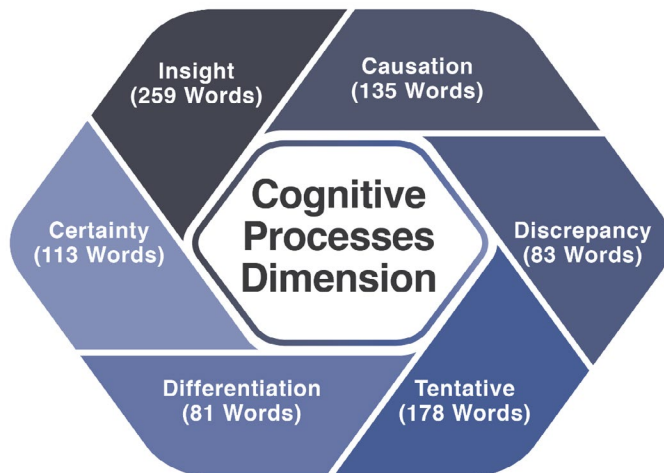


Figure 1: Cognitive processing dimension and sub-scores

1 method, each method sharing identical activities and objectives. The versions used in this study
2 were 09/01/16 to 09/31/2017 (self-paced) and 09/15/15 to 11/3/15 (instructor-paced). This
3 was a post hoc analysis and the researchers had no direct interaction with either the course in-
4 structors or the students.

5 *Procedures*

6 Through a data cleaning process, instructor posts, self-introduction posts and posts from stu-
7 dents missing demographic (gender or education) information were removed. The final dataset
8 had 2423 discussion forum posts created by 671 students. This study used clout as the dependent
9 variable and the six sub-category scores of cognitive processing as the focal predictor variables.
10 More details are provided in the following sections regarding what roles other research variables
11 play in various linear mixed models.

12 *Research variables*

13 *Dependent variable*

14 Clout served as the dependent variable in each linear mixed model. Clout is one of four LIWC
15 summary variables and is defined as the level of confidence and certainty in the post (Pennebaker,
16 Boyd, *et al.*, 2015). Clout scores are calculated by LIWC and are standardized as percentile scores
17 (0–100) of confidence levels exhibited by the words in the posts. Using clout scores allows for an
18 understanding of confidence or doubt within a specific post (Jordan *et al.*, 2019; Kacewicz *et al.*,
19 2014).

20 *Predictor variables*

21 Each of the six dimensions of cognitive processing in posts were the level-1 focal predictor for
22 clout in various linear mixed models. The specific dimensions of cognitive processing are insight,
23 causation, discrepancy, tentative, certainty and differentiation (Pennebaker, Boyd, *et al.*, 2015).
24 The data on the dimensions of cognitive processing represented the percent of words indicating
25 various cognitive processing dimensions in the posts.

26 *Moderator variable*

27 Course pacing was the level-2 moderator. Types of pacing in various courses were self-paced
28 or instructor-paced. With the use of pacing as the level-2 moderator, the focal relationship be-
29 tween each dimension of cognitive processing and clout was examined to see if it changed across
30 courses with different types of pacing or not.

31 *Control variables*

32 Gender, educational degree and type of posts were the control variables in linear mixed models.
33 The categories in degree consisted of high school, college, master's and doctorate. The types of
34 posts were initial, response and comment.

35 Means and standard deviations of posting features used in the model appear in Table 1.

36 *Data analysis*

Linear mixed modeling

37 Due to the nested nature of the data, the two-level random-intercept linear mixed models were
38 estimated with SPSS (Heck, Thomas, & Tabata, 2014; West, Welch, & Galecki, 2015). Specifically,
39 individual students contributed multiple posts so that the post data were nested within individual
40 students. In comparison with the general linear models, the linear mixed models have the ad-
41 vantages of flexibly including explanatory variables at different levels (eg, the post-level and the
42 student level) and explicitly addressing the issues of the negatively biased standard errors due to
43 the potential independence assumption violation in nested data (Heck *et al.*, 2014).

Table 1: Descriptive statistics of posting characteristics

	Clout	Insight	Causation	Discrepancy	Tentative	Certainty	Differentiation
<i>Pacing</i>							
Self (<i>n</i> = 1349)	52.08 (24.11)	3.63 (3.59)	1.95 (2.52)	1.22 (2.00)	2.83 (3.34)	1.40 (3.86)	2.89 (3.24)
Instructor (<i>n</i> = 1078)	53.80 (23.46)	3.84 (4.41)	2.20 (2.97)	1.29 (1.97)	2.91 (3.25)	1.26 (2.34)	2.90 (3.05)
<i>Gender</i>							
Male (<i>n</i> = 1136)	53.63 (24.18)	3.68 (4.28)	2.16 (3.05)	1.25 (1.94)	2.84 (3.36)	1.34 (2.99)	2.65 (3.02)
Female (<i>n</i> = 1291)	52.16 (23.52)	3.76 (3.69)	1.97 (2.42)	1.25 (2.02)	2.89 (3.24)	1.34 (3.50)	3.11 (3.25)
<i>Degree</i>							
High School (<i>n</i> = 435)	54.27 (23.78)	3.76 (4.07)	1.97 (2.37)	1.38 (2.01)	2.88 (3.30)	1.49 (5.09)	2.95 (3.69)
College (<i>n</i> = 1105)	52.10 (24.17)	3.73 (3.83)	2.11 (2.92)	1.19 (1.93)	2.89 (3.41)	1.44 (3.15)	2.91 (3.05)
Master (<i>n</i> = 803)	52.92 (23.56)	3.64 (3.84)	2.04 (2.65)	1.29 (2.05)	2.84 (3.17)	1.34 (2.05)	2.92 (3.04)
Doctorate (<i>n</i> = 84)	54.53 (22.34)	4.35 (6.14)	1.97 (2.85)	0.95 (1.94)	2.80 (2.99)	1.11 (1.88)	2.19 (2.48)
<i>Type of posts</i>							
Initial (<i>n</i> = 1138)	53.24 (23.65)	3.75 (3.89)	2.04 (2.82)	1.25 (1.93)	2.76 (3.20)	1.21 (2.13)	2.75 (3.13)
Response (<i>n</i> = 879)	53.06 (23.37)	3.57 (3.65)	2.09 (2.58)	1.18 (2.04)	2.89 (3.23)	1.51 (4.51)	2.87 (2.93)
Comment (<i>n</i> = 410)	51.34 (25.30)	4.00 (4.81)	2.05 (2.82)	1.39 (2.02)	3.10 (3.66)	1.34 (2.65)	3.34 (3.62)

Note: Numbers in parentheses are standard deviations.

Model specification

The level-1, post-level models contained each cognitive processing dimension of the posts, one at a time, as the focal predictor variable and type of post (TP) as the control variable. In the level-2, student-level models, gender and degree were specified as the explanatory variables for the level-1 intercept alongside the random intercept so that they could serve as the level-2 control variables. Furthermore, type of pacing was specified as the explanatory variable for the level-1 intercept and the regression coefficient of the focal predictor variable to examine if the focal relationship between a cognitive processing dimension and the dependent variable, clout, changed across types of course pacing. Accordingly, type of pacing (Pacing) played the role of the moderator variable in the full linear mixed models. The following are the level-1 and level-2 models with Insight as the focal predictor variable. Similar models were specified for the other five dimensions of cognitive processing respectively. Depending on whether there was the moderator effect of type of pacing (Pacing), similar linear models without the moderator variable were specified in sub-group analyses or complete-data analyses to evaluate the focal relationship between a cognitive processing dimension and clout.

Research Question 1 (Moderator effect analyses).

Level-1 model: $Clout_{ij} = \pi_{0i} + \pi_{1i}Insight_{ij} + \pi_{2i}TP_{ij} + \epsilon_{ij}$

Level-2 model: $\pi_{0i} = \gamma_{00} + \gamma_{01}Pacing_i + \gamma_{02}Gender_i + \gamma_{03}Degree_i + \zeta_{0i}$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}Pacing_i$$

Research Question 2 (Sub-group analyses by pacing).

Level-1 model: $Clout_{ij} = \pi_{0i} + \pi_{1i}Insight_{ij} + \pi_{2i}TP_{ij} + \epsilon_{ij}$

Level-2 model: $\pi_{0i} = \gamma_{00} + \gamma_{01}Gender_i + \gamma_{02}Degree_i + \zeta_{0i}$

$$\pi_{1i} = \gamma_{10}$$

Research Question 3 (Complete-data analyses).

Level-1 model: $Clout_{ij} = \pi_{0i} + \pi_{1i}Insight_{ij} + \pi_{2i}TP_{ij} + \epsilon_{ij}$

Level-2 model: $\pi_{0i} = \gamma_{00} + \gamma_{01}Gender_i + \gamma_{02}Degree_i + \zeta_{0i}$

$$\pi_{1i} = \gamma_{10}$$

Model estimation

The Restricted Maximum Likelihood (RML) estimation method (Singer & Willett, 2003) was adopted to obtain the unbiased estimates of covariance parameters (Heck *et al.*, 2014).

Tests of model parameters

The statistical significance of fixed-effect parameter estimates was determined by the *t*-test (Heck *et al.*, 2014) with the alpha level of .05.

Results

The purpose of the study was to further explore the negative relationship between cognitive processing and clout first reported by Moore *et al.* (2019). The results of the linear mixed models are summarized in Table 2.

Insight as the focal predictor

The results suggested the moderator effect of pacing on the focal relationship between insight and the dependent variable (Clout) after controlling for gender, degree and TP, $t(2416.46) = 2.45$, $p = .014$ and $b = 0.60$. Accordingly, the focal relationship between insight and clout would differ between the self-paced course and the instructor-paced course.

Table 2: : The focal predictive relationship between each cognitive processing and clout with or without regard to course pacing

Predictor	Moderator effect of pacing	Focal relationship
Insight	Yes	Self-paced: Negative Instructor-paced: Negative
Causation	Yes	Self-paced: Positive Instructor-paced: Zero
Tentative	Yes	Self-paced: Negative Instructor-paced: Zero
Certainty	No	Negative*
Differentiation	No	Negative*

¹*Indicates statistically significant.

The subgroup analysis by pacing was implemented to get a closer look at how the focal relationship changed across courses of different pacing after controlling for gender, degree and TP. Insight was negatively related to clout in the self-paced course, $t(1337.32) = -5.20, p < .001$ and $b = -0.95$, and in the instructor-paced courses, $t(1337.32) = -2.27, p = .024$ and $b = -0.35$. Therefore, the posts with a higher percent of insight words were less likely to exhibit more clout in both the self-paced and instructor-paced courses. However, the above negative relationship was stronger in the self-paced course given the larger fixed-effect estimate in absolute value (ie, -0.95 vs. $-.035$).

Causation as the focal predictor

The moderator effect of pacing on the focal relationship between causation and clout, after controlling for gender, degree and TP, was supported, $t(2415.69) = -2.76, p = .006$ and $b = -0.98$. Therefore, the focal relationship between causation and clout of posts was further examined in the self-paced course and the instructor-paced course respectively.

In the subsequent subgroup analysis by pacing, for the same TPs by students of the same gender and degree level, the relationship between causation and clout was present and positive in the self-paced course, $t(1338.40) = 3.20, p = .001$ and $b = 0.83$, but not in the instructor-paced course, $t(1069.69) = -0.70, p = .482$ and $b = -0.17$. Accordingly, in the self-paced course, the posts with a higher percent of causation words tended to show more clout as a whole.

Discrepancy as the focal predictor

Pacing did not moderate the focal relationship between discrepancy and clout after controlling for gender, degree and TP, $t(2408.87) = 1.59, p = .112$ and $b = 0.78$. As for the focal relationship between discrepancy and clout, it was statistically significant and negative, $t(2407.06) = -3.27, p = .001$ and $b = -1.06$. As a result, the increased percent of discrepancy words was accompanied by the weakened overall clout in the posts regardless of course pacing while gender, degree and TPs were held constant.

Tentative as the focal predictor

The results supported the moderator effect of pacing on the focal relationship between tentative and clout, after controlling for gender, degree and TP, $t(2416.02) = 2.03, p = .043$ and $b = 0.60$. In light of the varying focal relationship across course pacing, the focal relationship between tentative and clout was further examined in the self-paced course and the instructor-paced course respectively.

1 Based on the subgroup analysis by pacing, the relationship between tentative and clout was found
2 to be negative in the self-paced course, $t(1331.93) = -4.17, p < .001$ and $b = -0.82$, but nil in
3 the instructor-paced course, $t(1066.30) = -0.91, p = .366$ and $b = -0.20$, after controlling for
4 various control variables. Accordingly, the posts with a higher percent of causation words tended
5 to show stronger clout in the self-paced course.

6 *Certainty as the focal predictor*

7 Pacing did not moderate the focal relationship between certainty and clout after controlling for
8 gender, degree and TP, $t(2403.39) = 0.64, p = .524$ and $b = 0.22$. In addition, the focal relation-
9 ship between certainty and clout was statistically significant and negative, $t(2313.19) = -1.97,$
10 $p = .049$ and $b = -0.33$. Therefore, for the same TPs by students of the same gender and degree
11 level, the increased percent of certainty words tended to align with the decreased overall clout in
12 the posts within courses of both types of pacing.

13 *Differentiation focal predictor*

14 The focal relationship between differentiation and clout after controlling for gender, degree and
15 TP did not change across types of pacing, $t(2410.02) = -0.53, p = .598$ and $b = -0.15$. As to the
16 actual relationship between differentiation and clout, it was statistically significant and negative,
17 $t(2411.36) = -17.66, p < .001$ and $b = -3.21$. Specifically, the increased percent of differentia-
18 tion words are likely to be accompanied by decreased clout in posts for the courses of both types
19 of pacing with gender, degree and TPs being held constant.

20 **Discussion**

21 This study used the sub-scores of cognitive processing as predictor values for our variable of inter-
22 est, clout. We were particularly interested in this relationship for several reasons. First, clout is a
23 reflection of a poster's orientation—whether that is inward or outward facing—and their status
24 (both real and perceived) within the group (Kacewicz *et al.*, 2014; Oliver *et al.*, 2020). The use
25 of these sub-scores for cognitive processing allowed us to better understand the types of words
26 that were affecting the negative relationship between clout and cognitive processing (Moore *et al.*,
27 2019). Ultimately, we found that five of the six sub-scores of cognitive processing were negatively
28 associated with clout scores. Some of the negative associations were moderated by the pacing
29 condition, but for three sub-scores that was not the case. The three sub-scores—discrepancy, cer-
30 tainty and differentiation—had a statistically significant negative association not moderated by
31 pacing and are discussed below.

32 *Discrepancy*

33 We found that posts with an increased number of words classified as showing discrepancy, words
34 such as “should” or “would,” had lower clout scores. Understanding that it is impossible to thor-
35 oughly read through all the discussion transcripts, the discrepancy sub-score can be useful for
36 evaluating demonstration of student knowledge. The words in this category are typically associ-
ated with students that are uncertain of their answers and may be using vague language to mask
their lack of understanding (Williams & D'Mello, 2010). The fact that an increase in this score
would show a lower clout score makes sense because clout measures confidence and certainty
in a post. Additionally, scores that are high in clout will feature more “we” words, be outward
facing, refer and connect to others' work, and include fewer discrepancy words (Cohn, Mehl, &
Pennebaker, 2004; Jordan & Pennebaker, 2017; Kacewicz *et al.*, 2014; Oliver *et al.*, 2020). We
also found that higher frequency use of less confidence-inducing words found in the discrepancy
sub-score (eg, “hope” and “would”) resulted in lower clout scores.

Certainty

The negative association between the certainty sub-score and clout is confusing at first. The initial thought is that since clout is about how confident or certain a post is, that certainty words would increase this score. We found the opposite to be true—the more certainty words, the lower the clout score. The difference is likely attributable to the way the scores are calculated. The certainty score is looking for specific words, while clout is looking more at how the words are being used (eg, parts of speech and phrases). Words that demonstrate certainty, words such as “always,” “never” or “clearly,” have a certain level of finality and assertiveness. These types of assertions can be perceived as less analytical and more subjective, which results in the poster being perceived as not open to discussion (Momeni, Haslhofer, Tao, & Houben, 2015). Other researchers have also found that high use of certainty words are indicative of higher levels of irony or satire (Ravi & Ravi, 2017). The use of certainty words may not foster the type of discussion and engagement that one would find with the more inclusive language found in higher clout scores. Thus, it follows that in our analysis of discussion forum posts, we found that higher levels of certainty words resulted in lower clout scores.

Differentiation

Just as certainty words can contribute to less discussion and engagement, the same can be true for differentiation words. Words in this category include “exclude” and “despite” which are considered negative words in a discussion post. We found that posts using more words that indicated differentiation (eg, “exclude” and “despite”) resulted in lower clout scores. Our findings support those found by Pilny *et al.* (2019) where a relationship was established between a poster’s confidence in their answer and lower clout scores; in that the less certain (clout) a poster was the more differentiation words they used. It follows that the more confident or differentiating a post is, the less opportunities exist for engagement and interaction within the discussion forum.

Limitations

Likely the greatest limitation within this study is central to the very purpose of the study itself, which was to further explore the negative relationship between cognitive processing and clout first reported by Moore *et al.* (2019). Clout, being a summary variable, does not have a specific definition and thus the resultant data of the LIWC algorithm which generates the four LIWC summary variables must be taken at face value or interpreted in conjunction with other more transparent variables (Newman *et al.*, 2016). Another limitation of LIWC is that it uses a dictionary approach—meaning it matches words based on predefined categories instead of analyzing the text for contextual clues (Pilny *et al.*, 2019). While this approach is particularly helpful for efficiently analyzing a large corpus, such as we had in our study, it does present some analysis and interpretation challenges. This limitation leaves LIWC analysis open to criticism as concepts like deference and humility could be analyzed by LIWC as being low in a summary variable such as clout. For example, in our study, words within the tentative sub-category of cognitive processing showed a negative relationship to clout. But in practice, the expression of deference and humility are generally considered to be associated with greater confidence, as in higher clout. While the analysis of posts for clout scores can be a helpful indicator of student comfort level, it should not be used as a definitive measure.

Implications and recommendations

Despite the limitations of fully understanding the components that make up clout, our study has provided a deeper understanding of the relationship between cognitive processing and clout. The

1 clout score is of interest for instructors as it can show a level of confidence in student responses.
2 As previously noted, when students are hedging their answers and being vague (discrepancy
3 words), their clout scores go down. Ideally, an instructor should want to see both high clout and
4 high cognitive processing scores, as both are positive attributes within a learning environment.
5 As our study has identified, five of the six sub-scores of cognitive processing were shown to neg-
6 atively impact clout, but one category was positively associated: causation. Thus, a potential way
7 to increase clout among students is by developing discussion forum prompts that facilitate the
8 words found within the causation category. Examples of words in this category are “create,” “be-
9 cause” and “effect.” To foster use of those words, prompts could be framed in such a way as to
10 limit personal opinion while emphasizing consensus-building and teamwork, potentially counter-
11 acting some of the negative attributes associated with high-clout posts found in this study.

12 One way to do this is to focus on the application of content knowledge to a real-world example, as
13 this can change the focus from being inward facing to more outward facing. Additionally, instruc-
14 tors could consider using avatars in discussion forums, as several studies within the game-based
15 learning and online-based instruction space indicate that the use of avatars has novel effects on
16 user self-confidence (Falloon, 2010; Liu, Liu, Wang, Chen, & Su, 2012). The use of avatars could
17 allow students to contextualize situations in a way that, while personal to the poster, is abstracted
18 in the student’s representation of themselves within the discussion space. This, combined with
19 the explicit suggestion by prompts or teacher intervention to use words that would be found in the
20 causation category, could lead to replies and posts which foster further interaction as the focus
21 would be on more social words that invite further discussion such as first-person plural (eg, “we”)
22 or second-person versus first-person singular (eg, “I”) pronouns. These plural pronouns and the
23 use of second-person voice have been found routinely in posts with higher clout scores (Jordan &
24 Pennebaker, 2017; Kacewicz *et al.*, 2014), thus indicating that there exist potential activities and
25 instructional designs that can find the balance between high clout and high cognitive processing
26 within a robust discussion space.

27 **Conclusion**

28 The global pandemic COVID-19 has demonstrated that online learning will be increasingly turned
29 to in the future (Hill, 2020; Hodges, Moore, Lockee, Trust, & Bond, 2020; Moore, 2020). As more
30 learning happens in online contexts, identifying efficient and effective ways to analyze online
31 learner interactions will be important. This study adds to the empirical evidence for how text min-
32 ing, specifically automated text analysis, can be leveraged to better describe, understand and pre-
33 dict demonstration of student learning within MOOC discussion forums. We were interested in
34 identifying the ways to measure clout and cognitive processing and specifically identifying ways
35 to improve both scores. We knew from previous research that there was a negative relationship
36 between clout and cognitive processing (Moore *et al.*, 2019), but we did not know which specific
aspects of cognitive processing were negatively associated. This study identified that five of the
six sub-scores of cognitive processing were negatively associated. We further found that the use
of more words in the discrepancy, certainty and differentiation categories would result in lower
clout scores. Understanding the types of words that were lowering clout scores helps inform ways
to raise these scores. If students take these actions, their posts will show not only higher clout
scores but also higher cognitive processing scores. Future research should explore the parts of
speech, for example, personal pronouns and papers that are influencing the clout scores. Armed
with this knowledge, more specific recommendations can be made to help students improve their
posts and further demonstrate their mastery of the text while also creating more opportunities for
engagement within discussion forums.

Acknowledgements

This research used data provided through a data use agreement between the first author and the Harvard VPAL Research Team.

Statements on open data, ethics and conflict of interest

The data used in this study was identifiable and obtained through a data use agreement that included IRB approval. Because of the identifiable nature of the data, we cannot share the data publicly.

The authors declare they have no funding or conflict of interest in this work.

References

- Abe, J. A. A. (2020). Big five, linguistic styles, and successful online learning. *The Internet and Higher Education*, 45, 100724. <https://doi.org/10.1016/j.iheduc.2019.100724>
- Adaji, I., & Olakanmi, O. (2019). Evolution of emotions and sentiments in an online learning community. *Proceedings of Artificial Intelligence in Education (AIED), 2019*, 1–5.
- Adams, C., Yin, Y., Vargas Madriz, L. F., & Mullen, C. S. (2014). A phenomenology of learning large: the tutorial sphere of xMOOC video lectures. *Distance Education*, 35(2), 202–216. <https://doi.org/10.1080/01587919.2014.917701>
- Al-Imarah, A. A., & Shields, R. (2019). MOOCs, disruptive innovation and the future of higher education: A conceptual analysis. *Innovations in Education and Teaching International*, 56(3), 258–269. <https://doi.org/10.1080/14703297.2018.1443828>
- Beckmann, J., & Weber, P. (2016). Cognitive presence in virtual collaborative learning. *Interactive Technology and Smart Education*, 13(1), 52–70. <https://doi.org/10.1108/ITSE-12-2015-0034>
- Bonk, C. J., Lee, M. M., Reeves, T. C., & Reynolds, T. H. (2017). The emergence and design of massive open online courses (MOOCs). In R. A. Reiser & J. V. Demsey (Eds.), *Trends and issues in instructional design and technology* (4th ed., pp. 250–258). Pearson.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, 15(10), 687–693. <https://doi.org/10.1111/j.0956-7976.2004.00741.x>
- Darabi, A., Arrastia, M. C., Nelson, D. W., Cornille, T., & Liang, X. (2011). Cognitive presence in asynchronous online learning: A comparison of four discussion strategies. *Journal of Computer Assisted Learning*, 27(3), 216–227. <https://doi.org/10.1111/j.1365-2729.2010.00392.x>
- Dowell, N., & Graesser, A. C. (2014). Modeling learners' cognitive, affective, and social processes through language and discourse. *Journal of Learning Analytics*, 1(3), 183–186. <http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4203>.
- Dowell, N., Graesser, A. C., & Cai, Z. (2016). Language and discourse analysis with Coh-Matrix: Applications from educational material to learning environments at scale. *Journal of Learning Analytics*, 3(3), 72–95. <https://doi.org/10.18608/jla.2016.33.5>.
- Duncan, S. Y., Chohan, R., & Ferreira, J. J. (2019). What makes the difference? Employee social media brand engagement. *Journal of Business & Industrial Marketing*, 34(7), 1459–1467. <https://doi.org/10.1108/JBIM-09-2018-0279>
- Falloon, G. (2010). Using avatars and virtual environments in learning: What do they have to offer? *British Journal of Educational Technology*, 41(1), 108–122. <https://doi.org/10.1111/j.1467-8535.2009.00991.x>
- Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, 28(2), 127–203. <https://doi.org/10.1007/s11257-018-9203-z>
- Hagiwara, N., Slatcher, R. B., Egely, S., & Penner, L. A. (2017). Physician racial bias and word use during racially discordant medical interactions. *Health Communication*, 32(4), 401–408. <https://doi.org/10.1080/10410236.2016.1138389>

- 1 Hayati, H., Chanaa, A., Khalidi Idrissi, M., & Bennani, S. (2019). Doc2Vec & Naïve Bayes: Learners cognitive presence through asynchronous online discussion TQ transcripts. *International Journal of Emerging Technologies in Learning (IJET)*, 14(8), 70–81. <https://doi.org/10.3991/ijet.v14i08.9964>
- 2
- 3 Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS* (2nd ed.). Routledge.
- 4 **8** Hill, P. (2020). The COVID-Fueled hybridization of higher education. *Phil on EdTech*. <https://philonedtech.com/the-covid-fueled-hybridization-of-higher-ed/>.
- 5
- 6 Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *EDUCAUSE Review Online*., <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>.
- 7
- 8 Humphreys, A., & Wang, R. J.-H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, 44(6), 1274–306. <https://doi.org/10.1093/jcr/ucx104>
- 9
- 10 Jordan, K. N., & Pennebaker, J. W. (2017). Trump's first state of the union address. <https://wordwatchers.wordpress.com/2017/03/01/trumps-first-state-of-the-union-address>.
- 11
- 12 Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, 116(9), 3476–3481. <https://doi.org/10.1073/pnas.1811987116>
- 13
- 14 Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143. <https://doi.org/10.1177/0261927X13502654>
- 15
- 16 Kent, C., Laslo, E., & Rafaeli, S. (2016). Interactivity in online discussions and learning outcomes. *Computers & Education*, 97, 116–128. <https://doi.org/10.1016/j.compedu.2016.03.002>
- 17
- 18 Khazaei, T., Lu, X., & Mercer, R. (2017). Writing to persuade: Analysis and detection of persuasive discourse. **9** *In Proceedings of the IConference 2017* (pp. 203–215). <https://doi.org/10.9776/17022>.
- 19
- 20 Kovanović, V., Gašević, D., Hatala, M., & Siemens, G. (2017). A novel model of cognitive presence assessment using automated learning analytics methods. *In Analytics for Learning (A4L) Network* (Issue January).
- 21 **10** Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Dawson, S., & Gašević, D. (2017). Understanding the relationship between technology use and cognitive presence in MOOCs. *In Proceedings of the seventh international learning analytics & knowledge conference on - LAK '17, March 2017* (pp. 582–583). <https://doi.org/10.1145/3027385.3029471>
- 22 **11** Krippendorff, K. (2019). *Content analysis: An introduction to its methodology*. Sage.
- 23
- 24 Lin, Y., Yu, R., & Dowell, N. (2020). LIWCS the same, not the same: Gendered linguistic signals of performance and experience in online STEM courses. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Proceedings of international conference on artificial intelligence in education (AIED 2020)* (Vol. 12163, pp. 333–345). Springer International Publishing. https://doi.org/10.1007/978-3-030-52237-7_27.
- 25 **13** Liu, C.-C., Liu, K.-P., Wang, P.-H., Chen, G.-D., & Su, M.-C. (2012). Applying tangible story avatars to enhance children's collaborative storytelling. *British Journal of Educational Technology*, 43(1), 39–51. <https://doi.org/10.1111/j.1467-8535.2010.01146.x>
- 26
- 27 Martin, F., & Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice*, 13(3), 1–20. <https://doi.org/10.1177/0047239516656369>
- 28
- 29 Momeni, E., Haslhofer, B., Tao, K., & Houben, G.-J. (2015). Sifting useful comments from Flickr Commons and YouTube. *International Journal on Digital Libraries*, 16(2), 161–179. <https://doi.org/10.1007/s00799-014-0123-1>
- 30
- 31 Moore, R. L. (2016). Interacting at a distance: Creating engagement in online learning environments. In L. Kyei-Blankson, J. Blankson, E. Ntuli, & C. Agyeman (Eds.), *Handbook of research on strategic management of interaction, presence, and participation in online courses* (pp. 401–425). IGI Global. <https://doi.org/10.4018/978-1-4666-9582-5.ch016>.
- 32 **14** Moore, R. L. (2019). The role of data analytics in education: Possibilities and limitations. In B. Khan, J. R. Corbeil, & M. E. Corbeil (Eds.), *Responsive analytics and data mining in education: global perspectives on quality, support, and decision making* (pp. 101–118). Routledge. <https://doi.org/10.4324/9780203728703-8>.
- 33
- 34 Moore, R. L. (2020). Developing lifelong learning with heutagogy: Contexts, critiques and challenges. *Distance Education*, 41(3), 381–401. <https://doi.org/10.1080/01587919.2020.1766949>
- 35
- 36

- Moore, R. L., Oliver, K. M., & Wang, C. (2019). Setting the pace: examining cognitive processing in MOOC discussion forums with automatic text analysis. *Interactive Learning Environments*, 27(5–6), 655–669. <https://doi.org/10.1080/10494820.2019.1610453>
- Newman, D. S., Jones, J., & Ritter, C. (2016). Vertical peer supervision of consultation: A linguistic exploration of relational hierarchy. *The Clinical Supervisor*, 35(2), 287–304. <https://doi.org/10.1080/07325223.2016.1218811>
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675. <https://doi.org/10.1177/0146167203251529>
- O'Dea, B., Larsen, M. E., Batterham, P. J., Calear, A. L., & Christensen, H. (2017). A linguistic analysis of suicide-related Twitter posts. *Crisis*, 38(5), 319–329. <https://doi.org/10.1027/0227-5910/a000443>
- Oliver, K. M., Houchins, J. K., Moore, R. L., & Wang, C. (2020). Informing makerspace outcomes through a linguistic analysis of written and video-recorded project assessments. *International Journal of Science and Mathematics Education*. <https://doi.org/10.1007/s10763-020-10060-2>
- 15 Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015 operator's manual* (pp. 1–23). Pennebaker Conglomerates.
- 16 Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin. https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf.
- Pennebaker, J. W., Chung, C., Frazee, J., Lavergne, G., & Beaver, D. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12), 1–10. <https://doi.org/10.1371/journal.pone.0115844>
- Pilny, A., McAninch, K., Slone, A., & Moore, K. (2019). Using supervised machine learning in automated content analysis: An example using relational uncertainty. *Communication Methods and Measures*, 13(4), 287–304. <https://doi.org/10.1080/19312458.2019.1650166>
- Ravi, K., & Ravi, V. (2017). A novel automatic satire and irony detection using ensembled feature selection and data mining. *Knowledge-Based Systems*, 120, 15–33. <https://doi.org/10.1016/j.knosys.2016.12.018>
- Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology*, 32(4), 469–479. <https://doi.org/10.1177/0261927X13476869>
- Shah, D. (2018). By the numbers: MOOCs in 2017. *Class Central*. <https://www.class-central.com/report/mooc-stats-2017/>.
- Sharif, A., & Magrill, B. (2015). Discussion forums in MOOCs. *International Journal of Learnings, Teaching and Educational Research*, 12(1), 119–132.
- Simms, T., Ramstedt, C., Rich, M., Richards, M., Martinez, T., & Giraud-Carrier, C. (2017). Detecting cognitive distortions through machine learning text analytics. *IEEE International Conference on Healthcare Informatics (ICHI), 2017*, 508–512. <https://doi.org/10.1109/ICHI.2017.39>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.
- 17 Smith-Keiling, B. L., & Hyun, H. I. F. (2019). Applying a computer-assisted tool for semantic analysis of writing: Uses for STEM and ELL. *Journal of Microbiology & Biology Education*, 20(1), 1–6. <https://doi.org/10.1128/jmbe.v20i1.1709>
- Smith-Keiling, B. L., Swanson, L. K., & Dehnostel, J. M. (2018). Interventions for supporting and assessing Science writing communication: Cases of Asian English Language Learners. *Journal of Microbiology & Biology Education*, 19(1), 1–15. <https://doi.org/10.1128/jmbe.v19i1.1522>
- Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *IEEE Access*, 7, 44883–44893. <https://doi.org/10.1109/ACCESS.2019.2909180>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Van Swol, L. M., & Kane, A. A. (2018). Language and group processes: An integrative, interdisciplinary review. *Small Group Research*, 1–36. <https://doi.org/10.1177/1046496418785019>
- 18

- 1 Veletsianos, G., & Shepherdson, P. (2016). A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. *International Review of Research in Open and Distributed Learning*, 17(2), 198–221. <https://doi.org/10.19173/irrodl.v17i2.2448>.
- 2
- 3 West, B. T., Welch, K. B., & Galecki, A. T. (2015). *Linear mixed models: A practical guide using statistical software* (2nd. ed.). Taylor & Francis Group LLC.
- 4 **19** Williams, C., & D'Mello, S. (2010). Predicting student knowledge level from domain-independent functiona and content words. In: *Proceedings of 10th international conference on intelligent tutoring systems* (pp. 62–71). https://doi.org/10.1007/978-3-642-13437-1_7.
- 5
- 6 Zhu, M., Herring, S. C., & Bonk, C. J. (2019). Exploring presence in online learning through three forms of computer-mediated discourse analysis. *Distance Education*, 40(2), 205–225. <https://doi.org/10.1080/01587919.2019.1600365>
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36