

Supplemental Material:

1. Description of Educational Programs:

We recruited students from three types of educational programs for this study. Group A students were enrolled in the Rainforest Expedition and Laboratory Course. This program involved developing individual research questions relevant to diverse microbial samples students isolated from materials collected during a two-week field research trip.

Ownership is an explicitly designed element of this undergraduate research program, since students develop their own research questions and direct the project from field collection to laboratory bench. Group B students were sampled from those doing independent research in various faculty laboratories throughout the university. An email invitation was sent to students in this category, and we interviewed the students who responded. Ownership is an implicitly designed element of this type of undergraduate research experience and varies among students. Group C students were drawn from two standard laboratory courses, one in organic chemistry and one in biochemistry. Students in these lab courses were invited by their instructors to participate in our research study. Ownership is not an element of this experience, since students follow a provided protocol on a topic selected by the instructor.

2. IRB Oversight: The student interview protocol was approved by the Human Subjects Committee at Yale University (IRB Protocol # 0807004013).

3. Computational Linguistic Analysis: Several different categories of information are relevant to project ownership. Project ownership involves a sense of personal responsibility, agency, control and connection. If a student is an active agent, in control and personally connected to a project, it is a fair assumption that this will be reflected in increased usage of first person personal pronouns (I, me, my, and mine). Thus in analyzing degrees of project ownership, the analysis of frequencies of first person, personal pronouns should allow the distinction between situations in which there is increased personal responsibility, agency, control and connection (in other words, project ownership) and those where this is not the case. The assumption is that higher levels of first person personal pronouns reflect increased project ownership.

Beyond personal connection and agency, additional features of project ownership are emotional engagement, identification and commitment. In linguistic terms these features should be reflected in the usage of the semantic category of words which denote an emotional response. If a student is emotionally engaged with a program, it is reasonable to expect that he/she will express a heightened sense of emotion through the increased usage of emotive words. Accordingly, in analyzing degrees of project ownership (in other words, levels of emotional engagement) the analysis of frequencies of emotional words reveals differences in levels of emotional engagement. The assumption is that higher frequencies of emotional words reflect higher levels of emotional engagement with a project.

There is also a content aspect to the ownership question. Project ownership is being discussed in

relation to a scientific inquiry process and an educational program. Accordingly, it is assumed that the actual content of discussion – scientific inquiry – will also be reflected in word usage. Specifically the LIWC program can count words which designate cognitive thinking and insight. It is a fair assumption that if students speak about a scientific inquiry process they will do so using cognitive words dealing with their thought processes and may have some moments of insight in relation to their findings. However, frequencies of cognitive and insight words are not expected to differentiate between degrees of project ownership but rather just specify that it is ownership in relation to a research project.

Our experimental approach incorporates features of project ownership and recent developments in analytical techniques of language processing to measure project ownership. If students involved in different educational programs (which differ in relation to the degree of choice, control, decision making, personal agency and responsibility) are asked to describe their research projects, a systematic analysis of the language produced by these students using the LIWC program in relation to the categories of first person, personal pronouns, emotive words, cognitive lexicon and insight words should be able to measure degrees of project ownership. Even more importantly, a quantitative hypothesis can be proposed in relation to the measurement of project ownership. High degrees of project ownership should be reflected in high degrees of first person pronoun and emotive word usage.

Based on this explication of the concept of scientific inquiry project ownership, four categories of linguistic information are important to the analysis of scientific inquiry project ownership:

1. *Usage of First Person Personal Pronouns:* Personal pronouns are words that reference a participant in an event. As such personal pronouns indicate the presence of human interaction in described events or experiences. Personal pronouns in English are divided into distinctions of person (first, second and third), gender (male, female) and number (singular and plural). First person personal pronouns reference personal action and involvement in a described situation or process.
2. *Usage of Emotional Lexicon:* Emotions are linguistically manifest through the usage of emotionally laden words such as happy, sad, cried, loved...etc. Pennebaker and colleagues have created a dictionary list of words that reflect the presence of affective processes (1,2). This dictionary of affective words can be used to judge the degree to which these words appear in any given corpus. Differences in the levels of usage of this emotional lexicon in different corpuses reflect differences in the degree to which a given corpus involves affective processes.
3. *Usage of Cognitive Lexicon:* Cognitive processes are linguistically manifest through the usage of words that address cognitive functions such as cause, know, effect...etc. Pennebaker and colleagues have created a dictionary list of words that reflect the presence of cognitive processes (1,2). This dictionary of cognitive words can be used to judge the degree to which these words appear in any given corpus. Differences in the levels of usage of this cognitive lexicon in different corpuses reflect differences in the degree to which a given corpus involves cognitive processes.

4. *Usage of Insight Lexicon*: The cognitive function of insight is reflected in the usage of specific words such as think and consider. Pennebaker and colleagues have created a dictionary list of words that reflect the presence of insight in a corpus (1,2). This dictionary of insight words can be used to judge the degree to which these words appear in any given corpus. Differences in the levels of usage of this insight lexicon in different corpuses reflect differences in the degree to which a given corpus involves insight.

References

1. J.W. Pennebaker, C.K. Chung, M. Ireland, A. Gonzales, R. J. Booth, *The Development and Psychometric Properties of LIWC2007*. (LIWC Inc, Austin, 2007).
2. J.W. Pennebaker, R.J. Booth, F.E. Martha, *Linguistic Inquiry and Word Count: LIWC2007*. (LIWC Inc, Austin, 2007).