

# Persuasive Subgenre Analysis of University Coursework Writing Data

## University Coursework Writing Data Collection:

In order to examine the relationship among NLP writing analytics and broader academic achievement, we collected authentic coursework writing data (N=1,058), HEIghten assessment data, survey data, and outcomes data from university students enrolled at one of six different four-year postsecondary institutions. A representative from each institution acted as the site liaison and supported recruitment efforts by providing a list of courses across different disciplines that targeted first and second year students and involved one or more substantive writing assignments (i.e., assignments with approximately 300 words or more). Students enrolled in these courses were contacted via email with an invitation to participate in the study. Participants were asked to complete a writing survey, upload electronic copies of all of the writing assignments they submitted for the target course to a web portal, and complete two HEIghten exams (Written Communication and Critical Thinking).

## Writing Genre Analysis:

After generating Automated Writing Evaluation (AWE) features from the dataset and conducting preliminary analyses, we decided to investigate how NLP writing analytics might vary by genre, and examine relationships between genre and students' outcomes and success indicators. As a first step, we completed a genre annotation of the authentic coursework writing data (N=1,058) using a taxonomy developed by Dr. Norbert Elliot, who has an expertise in writing studies. Three research assistants examined the writing assignments uploaded by students (and, when available, the corresponding writing assignment descriptions provided by instructors) and classified the papers into one of three broad assignment types ("bins"). Drawing inspiration from Melzer (2014) and Gere, Limlamai, Wilson, MacDougall Saylor, and Pugh (2019), these classifications were based on an analysis of aim (writing purpose) and primary audience. Aims under consideration included to persuade, to inform, to explore, and to reflect, while audiences were broken down into the following categories: (1) the instructor, (2) a hypothetical audience outside of the classroom (e.g., an imaginary employer), and (3) the self. This analysis allowed us to classify the course writing data into one of three "bins," which we have rebranded for the purpose of this study as: (1) persuasive, (2) expository, and (3) reflective.

## Persuasive Subgenre Analysis:

While the original taxonomy was deliberately broad in an effort to optimize our sample sizes for each genre, preliminary analyses revealed unexpected results in terms of the absence of relationships among the features as well as the absence of relationships between the features and criterion variables. Consequently, we selected one of the three broad genres (persuasive) and created a finer-grained classification system to account for the different types of persuasive essays observed in the dataset. Dr. Norbert Elliot, with whom we developed the original taxonomy for genre annotation, created a new taxonomy for the persuasive subgenre analysis. This new taxonomy was based on a review of the 397 papers identified as "Persuasive" in the original genre annotation completed by the RAs. This taxonomy

consisted of six sub-categories of persuasive writing that varied in terms of argument value (on a continuum of low to high), source use and integration, and support (see table 1).

Subgenre		Characteristics
1	High Value Source-Based Argument--High Value Source Integration with Bibliography	<ul style="list-style-type: none"> <li>• High value argument</li> <li>• Effective source integration</li> <li>• Claims supported with evidence from sources</li> <li>• Bibliography with two or more sources</li> </ul>
2	Moderate Value Argument--Some Source Integration, Claims Supported, but no Bibliography or Single Citation Bib	<ul style="list-style-type: none"> <li>• Moderate value argument</li> <li>• Some source integration</li> <li>• Claims supported with evidence from sources</li> <li>• No bibliography or a bibliography with a single citation</li> </ul>
3	Moderate Value Argument--No Source Integration but Claims Supported	<ul style="list-style-type: none"> <li>• Moderate value argument</li> <li>• Some source integration</li> <li>• Claims supported with evidence</li> <li>• No source integration</li> </ul>
4	Low Value Argument--No Source Integration with Claims Unsupported	<ul style="list-style-type: none"> <li>• Low value argument</li> <li>• No source integration</li> <li>• Claims unsupported</li> </ul>
5	Sentence Level Outline, with Thesis	<ul style="list-style-type: none"> <li>• No attempt at organization beyond outline structure.</li> </ul>
6	Annotated Bibliography	<ul style="list-style-type: none"> <li>• No attempt at organization beyond basic factual exposition.</li> </ul>

*Table 1 Persuasive Subgenre Taxonomy*

The first subgenre was labeled “High Value Source-Based Argument--High Value Source Integration with Bibliography.” These papers consisted of high quality arguments that leveraged multiple reputable sources (e.g., a peer-reviewed journal article) to support their claims. Not only were these sources effectively integrated into the paper, but they also had to be formally cited in a bibliography. The second subgenre was labeled “Moderate Value Argument--Some Source Integration, Claims Supported, but no Bibliography or Single Citation Bib.” The persuasive Subgenre 2 papers also supported claims with evidence from outside sources, yet the argumentation and source integration were somewhat less sophisticated in terms of integration. Markers of lower sophistication may include the absence of a bibliography (or the inclusion of a bibliography with a single citation), exclusive use of non-academic sources (e.g., blog post, dictionary), a heavy reliance on verbatim quotes as opposed to synthesized, integrated evidence from outside sources, or parenthetical citations included with no argumentation. Subgenre 3 papers were also described as “Moderate Value Argument--No Source Integration but Claims Supported.” Like Subgenre 2 papers, Subgenre 3 papers include claims that are supported with evidence. The distinguishing trait, however, is the evidence for Subgenre 3 papers do not come from outside sources. Instead, they may come from the text under discussion, the author’s own knowledge, observations, and life experiences. Subgenre 4 papers were labeled “Low Value Argument--No Source Integration with Claims Unsupported.” These papers contained unsubstantiated claims, and did not use any outside sources. Subgenres 5 and 6 were used to house uploads that contained limited prose, but were building towards a persuasive argument. More specifically, uploads containing a thesis statement

accompanied by a sentence level outline were labeled Subgenre 5 (“Sentence Level Outline, with Thesis”), while annotated bibliographies were labeled as Subgenre 6 (“Annotated Bibliography”).

Once the persuasive subgenre categories were established, Dr. Elliot annotated the writing upload data according to the persuasive subgenre taxonomy. He began this process using the 397 uploads that were classified as “persuasive” in the original genre annotation. If he agreed that a paper was persuasive, he would assign it a persuasive subgenre category. Then, he would search for additional uploads submitted by that same student to see if a second writing sample contained sufficient argumentation to warrant its inclusion in the persuasive subgenre analysis. If a participant submitted more than two persuasive uploads, the remainder were excluded from the study. This process yielded a sample of 324 uploads with persuasive subgenre annotations. The results of this review were captured in a table mapping participant identifiers to their 1-2 persuasive uploads and corresponding subgenre classifications.

After Dr. Elliot completed the subgenre annotation, a second annotator, an RA who had also been involved in the original genre annotation, reviewed the findings from Dr. Elliot’s subgenre analysis to validate the results. During this phase, annotator 2 reviewed the corpus of persuasive uploads identified by Dr. Elliot (N=324 uploads from 200 participants) alongside his annotations and decided whether (a) she agreed with the subgenre classifications, and (b) any uploads should be removed from the persuasive subgenre analysis. During this review, a decision was made to remove any papers that were previously classified as expository or reflective (Bins 2 or 3) in the original genre annotation. Common examples include literary and rhetorical analyses, which contain persuasive elements but are powerfully expository. With this new criteria, 50 uploads were removed from the persuasive subgenre analysis. An additional 12 uploads were removed due to the presence of duplicates in the dataset. In some cases, these identical or nearly identical uploads were uploaded by different students (which may be explained by collaborative writing assignments and/or instances of plagiarism). In other cases, the same student may have uploaded drafts or components of the same assignment (in which case, the final or complete version was retained and earlier drafts or components were removed). Finally, both annotators recommended the removal of one anomalous data file which contained multiple essays copied and pasted into one text file (presumably corresponding to all of the writing assignments for that course). Each time Annotator 2 would remove an upload, she would check to see if the upload could be replaced with another submission from that same student. Out of the remaining uploads in the persuasive dataset that Dr. Elliot identified, annotator 2 reclassified seven uploads. In the end, 263 persuasive papers from 196 participants (71 of which had two uploads) were selected for the subgenre analysis. The final distributions by persuasive subgenre can be viewed in table 2.

Category	Count	%
Subgenre 1 High Value Source-Based Argument--High Value Source Integration with Bibliography	104	39.54%
Subgenre 2 Moderate Value Argument--Some Source Integration, Claims Supported, but no Bibliography or Single Citation Bib	47	17.87%
Subgenre 3 Moderate Value Argument--No Source Integration but Claims Supported	76	28.90%

Subgenre 4 Low Value Argument--No Source Integration with Claims Unsupported	34	12.93%
Subgenre 5 Sentence Level Outline, with Thesis	1	0.38%
Subgenre 6 Annotated Bibliography	1	0.38%
Total	263	100%

*Table 2 Distribution of Persuasive Subgenres (N=263 persuasive essays)*