Comparing the Formality and Clarity of My Technical Writing from High School to My Technical Writing from University Coursework

Ashley Miller

Department of Technical Communication, University of North Texas

TECM 4500: Content Analysis

December 1, 2022

Table of Contents

Tables

Figures

Executive Summary

The purpose of this report is to detail the research, methodology, results, and conclusions of a content analysis I conducted on my own writing. My first research question was, “Is my high school writing significantly more formal than my college writing?” The second research question was, “Is my writing in college significantly higher in clarity than my writing in high school?” To answer these questions, I assessed 10 writing samples (17,009 words) of technical writing from my high school courses and 21 text documents (17,015 words) from my university technical communication coursework for formality and clarity.[1] Using TagAnt, AntConc, and Microsoft Excel, I recorded instances of self-mentions, unattended anaphoric pronouns, passive voice, and short sentences. My results yielded statistically significant differences in self-mentions, passive voice, and unattended anaphoric pronouns between high school and college. However, both of my research questions ultimately had inconclusive answers as only one criterion was fulfilled for each.

[1]My high school writing includes samples from 2016–2019, while my university writing samples are from 2021–2022.

Introduction

I conducted a content analysis on 31 of my technical writing texts to explore how the formality and clarity of my writing compare between high school and college.

I examined each writing sample for the following grammatical and syntactical features:

1. Self-mentions
2. Anaphoric demonstratives
3. Active and passive voice
4. Sentence length

More specifically, to test for formality, I recorded the frequency of first-person pronouns referring to the author(s) and the percentage of anaphoric demonstratives unattended. To quantify the clarity of my writing, I measured how long each sentence was and calculated the ratio of sentences written in passive versus active voice.

Formality

Few self-mentions

Many writers and scholars consider personal reference in academic research to be informal because it degrades the objectivity of empirical research (Hyland and Jiang, 2017). Conversely, others believe such self-mentions elevate authors' credibility by emphasizing an individual's contribution to research and establishing that author as a knowledgeable authority in the field (Hyland and Jiang, 2017). Most of my high school teachers preached that formal writing avoids self-mentions, whereas my technical communication professors favor self-mentions for credibility. Hence, while self-mentions in my writing could correlate with formality, a change in their quantity wouldn't exhibit a shift in professionalism.

Attended anaphoric determiners

When attended, the anaphoric demonstratives *this, that, these,* and *those* function as determiners; when unattended, they function as free-standing pronouns (Boettger & Wulff, 2014). Leaving these words unattended (i.e., not following them with a noun phrase) is often discouraged in academic and technical writing as there are rhetorical advantages to attending them. Specifically, attended *this* reduces ambiguity and supports a formal tone (hence why I'm using it to indicate formality) (Swales, 2005). Concerning the context of this study, unattended *this* is relatively common in student-authored texts; 43% of the instances of *this* in Wulff et al.'s sample and 33% in Boettger and Wulff's sample were unattended (Wulff, Römer, & Swales, 2012; Boettger & Wulff, 2014). Additionally, Swales found that in published papers across disciplines, around 35% of occurrences of *this* are unattended (Swales, 2005). These statistics provide a measure to evaluate how my use of demonstratives in high school and college compares to that of students and academics.

Clarity

Active voice

Active voice is essential to a document's clarity because it explicitly defines the subject of a sentence and communicates what the subject is doing. Contrarily, in a clause written in passive voice, an action is being performed, but it is unclear who is performing that action (Plain Language Action and Information Network, 2022). However, passive voice is often favored in scientific and academic research reports; passive voice can be acceptable in methods sections of scientific reports because it places emphasis on the experiment rather than the researcher, and the reader can infer who is performing the action (University of Toronto). Still, the Microsoft Word Grammar Expert Plus utility largely calculates readability based on percentage of passive sentences (Harvey, 2005). Thus, in my content analysis, the presence of active versus passive voice serves as a strong indicator of clarity.

Short sentences

Shorter sentences directly influence a text's coherence, so I recorded sentence length in my sample to assess clarity. Communicating technical information is complex, so breaking long sentences into smaller components aids reading comprehension (U.S. General Services Administration, 2011). Furthermore, the Flesch Reading Ease and Flesch-Kincaid Grade Level formulas—two of the most popular readability scores—evaluate text based on word and sentence length (Harvey, 2005). Most sources I've found recommend sentences be no longer than 20 to 25 words, so for this study, I defined short sentences as those shorter than 25 words (Vincent, 2014; Last, 2019).

Research Questions

1. **Is my high school writing significantly more formal than my college writing?**
   1.1. Are there significantly more unattended demonstratives in the high school sample?
   1.2. Are there significantly fewer personal pronouns in the high school sample?

2. **Is my writing in college significantly higher in clarity than my writing in high school?**
   2.1. Is there significantly less passive voice in the college sample?
   2.2. Are sentences in the college sample significantly shorter?

Methods

Gathering Materials

To assemble the samples, I first pasted all the content I have access to from my high school technical writing program into a Word document, noting which content came from which original file. The purpose of this initial step was to see how many words composed my sample of high school writing so I could gauge how much recent work to pull for the sample of university work. I subsequently compiled the text from several of my technical communication assignments at UNT to create a similar sample size. Ten high school texts totaled 17,009 words; 21 university assignments totaled 17,015 words.

Next, I cut and paste each of the 31 texts into a text converter to transform the samples into plain text. Then, I named each .txt file with tags according to Table 1 and uploaded the text files into TagAnt to assign each word a part of speech. (Table 2 displays the file names in chronological order of composition.) Finally, I created two corpuses in AntConc (one of to search the samples for each feature.

Indication	**Tags**
Whether the content is from high school or college, and where the file sits chronologically in relation to the other sample files	hs#, unt#
Semester (fall or spring) and year written (XX)	f16, f17, s18, f18, s19, f19, s21, f21, s22, f22
Type or context of assignment (technical report, interview report, design justification memo, or reflection memo)	report, interview, justmemo, refmemo, reflog[2]
Subject of material	culture, studying, etc.

Table 1: System of naming text files.

TXT ID	**File Name**	**Course Level**	**Context**
hs01	hs01_f16_report_3D.txt	High school	Report
hs02	hs02_f16_reflog_3D.txt	High school	Reflection
hs03	hs03_f16_report_balloons.txt	High school	Report
hs04	hs04_f16_interview_colorblind.txt	High school	Interview
hs05	hs05_f17_report_pendulum.txt	High school	Report
hs06	hs06_ s18_interview_pirouette.txt	High school	Interview
hs07	hs07_ s18_reflog_pirouette.txt	High school	Interview
hs08	hs08_ f18_interview_ginger.txt	High school	Reflection
hs09	hs09_s19_interview_google.txt	High school	Interview
hs10	hs10_f19_interview_737max.txt	High school	Interview
unt01	unt01_s21_report_culture.txt	College	Report
unt02	unt02_s21_interview_stud4180.txt	College	Interview
unt03	unt03_s21_interview_pro4180.txt	College	Interview
unt04	unt04_s21_justmemo_psycprop.txt	College	Justification
unt05	unt05_s21_report_studying.txt	College	Report
unt06	unt06_s21_refmemo_pdp.txt	College	Reflection
unt07	unt07_f21_report_heuristic.txt	College	Report
unt08	unt08_f21_justmemo_puzzles.txt	College	Justification
unt09	unt09_f21_report_cardsort.txt	College	Report

[2] My high school documents were comprehensive reports called “logs”; files tagged “reflog” are reflection sections pulled from logs. They have similar structure and purpose to the college reflection memos (refmemo).

unt10	unt10_f21_refmemo_3200_website.txt	College	Reflection
unt11	unt11_s22_justmemo_shortdoc.txt	College	Justification
unt12	unt12_s22_justmemo_logo.txt	College	Justification
unt13	unt13_s22_report_union.txt	College	Report
unt14	unt14_s22_justmemo_longdoc.txt	College	Justification
unt15	unt15_s22_justmemo_4400_contrib.txt	College	Justification
unt16	unt16_s22_refmemo_recipe.txt	College	Reflection
unt17	unt17_f22_interview_stakeholder.txt	College	Interview
unt18	unt18_f22_interview_users.txt	College	Interview
unt19	unt19_f22_justmemo_A9.txt	College	Justification
unt20	unt20_f22_report_3550_recs.txt	College	Report
unt21	unt21_f22_report_p1.txt	College	Report

Table 2: The file names and their corresponding contexts.

Collecting Data

Table 3 contains the codebook defining how I operationalized each grammatical feature; Table 4 shows how I searched for each. After recording all my data in an MS Excel workbook, I summarized my findings with tables and PivotTables.

Grammatical or Syntactical Feature	**Bins**	**Definition**
Self-mentions	Present Not present	Any instance of: I, we, my, me, myself, our, ours, us None of above words present
Anaphoric demonstratives	Attended (determiner) Unattended (pronoun)	Instance of *this*, *that*, *these*, or *those* followed by a noun phrase Stand-alone instance of above words not followed by a noun phrase
Voice	Active Passive	Not passive; subject performs action. Verbal phrase contains a “BE” verb, plus past participle. Action is performed on subject.
Sentence length	Short Long	25 words or fewer Longer than 25 words

Table 3: Codebook defining features.

Feature Searched for	**Tool Used**	**Search Entries**
Presence of self-mentions	AntConc	I We My Me Myself Our

		Ours Us
Attended anaphoric determiners	AntConc	All determiners: this_DT* that_DT* these_DT* those_DT* Attended determiners: this_DT* *NN* that_DT* *NN* these_DT* *NN* those_DT* *NN*
Active voice	AntConc	be_V* *_VBN am_V* *_VBN is_V* *_VBN are_V* *_VBN was_V* *_VBN were_V* *_VBN been_V* *_VBN being_V* *_VBN *To account for adverbs in between, I searched all of the following in the form* be_V* *_R* *_VBN *(1 adverb) and* be_V* *_R* *_R* *_VBN *(2 adverbs).*
Short sentences	MS Excel	Cell word count formula: =LEN(TRIM(G2))-LEN(SUBSTITUTE(G2," ",""))+1

Table 4: AntConc and MS Excel search entries to identify each feature.

Self-mentions

After locating the personal pronouns in each corpus in AntConc, I recorded the number of "hits" for each pronoun in my Excel workbook. I summed the number of hits for each corpus to determine the total count of self-mentions.

Unattended anaphoric pronouns

To identify unattended anaphoric pronouns, I searched the corpus for all instances of *this, that, those,* and *these* and for the attended instances (see Table 4). To calculate the percentage of demonstratives left unattended, I divided the count attended by the total count and subtracted that number from 1.

Passive voice

To determine how many sentences were written in active or passive voice, I searched for "be" verbs attached to past participles by searching variations of *be_V* *_VBN* in the KWIC in

AntConc. Additionally, since adverbs can split “be” verbs and past participles, I searched for instances of passive voice with one and two adverbs in between (see Table 4).

Sentence length

To investigate sentence length, I first isolated each sentence: I used a sentence splitter to separate the sentences, pasted them into Column G of an Excel sheet within my workbook, and deleted sheet rows of blank cells in between. Then, I entered a formula from ExcelJet (see Table 4) into Column E that auto-calculated how many words were in each sentence.

Statistical significance

Finally, to ultimately deduce how my past versus present writing is significantly more formal or clear, I measured each stylistic feature's frequencies and conducted one-way z-tests for significance, calculating p-values at $\alpha = 0.05$ (with the STAT function of a TI-84 calculator).

Results

R1: Is my high school writing significantly more formal than my college writing?

Self-mentions

Pronoun	# of hits		Pronoun	# of hits
I	158		I	444
We	21		We	65
My	52		My	162
Me	22		Me	32
Myself	3		Myself	11
Our	18		Our	72
Ours	1		Ours	0
Us	2		Us	10
	277			**796**
HS			**UNT**	
The value of z is -16.0941. The value of p is < .00001. The result is significant at p < .05.				

Figure 1: Results of self-mentions data collection.

Unattended Anaphoric Pronouns

Pronoun	*_DT*	*_DT* *_NN*	% are UNattended
This	159	79	50%
That	125	20	84%
These	38	24	37%
Those	11	3	73%
Total	333	126	**62%**
HS			
Pronoun	***_DT***	***_DT* *_NN***	**% are UNattended**
This	129	88	32%
That	81	21	74%
These	24	20	17%
Those	9	4	56%
Total	243	133	**45%**
UNT			
The value of z is 4.0488. The value of p is < .00001. The result is significant at p < .05.			

Figure 2: Results of unattended anaphoric pronouns data collection.

R2: Is my writing in college significantly higher in clarity than my writing in high school?

Passive Voice

"Be" + VBN	# of Hits	1 adverb	2 adverbs	"Be" + VBN	# of Hits	1 adverb	2 adverbs
be	65	2	0	be	25	0	0
am	0	1	0	am	6	0	0
is	58	8	4	is	23	2	0
are	31	9	0	are	16	2	0
was	37	7	0	was	10	2	0
were	15	0	1	were	3	2	0
being	12	0	0	being	1	2	0
been	9	0	0	been	6	0	0
Totals	**227**	**27**	**5**	**Totals**	**65**	**10**	**0**
Total	**259**			**Total**	**75**		
HS				**UNT**			
The value of z is 11.8025. The value of p is < .00001. The result is significant at p < .05.							

Figure 3: Results of passive voice data collection.

Sentence Length

	Shorter than 20 Words?					
	NO		YES		Total NUMBER OF SENTENCES	Total % of Sample Sentences
Sample	NUMBER OF SENTENCES	% of Sample Sentences	NUMBER OF SENTENCES	% of Sample Sentences		
College	276	**40%**	413	60%	689	100.00%
High School	292	**43%**	383	57%	675	100.00%
The value of z is -1.1243. The value of p is .13136. The result is *not* significant at p < .05.						

Figure 4: Results of sentence length data collection.

Discussion

R1: Is my high school writing significantly more formal than my college writing?

Since the results of my data synthesis for self-mentions and unattended anaphoric pronouns are statistically significant, I can validly assume that I used significantly fewer self-mentions and significantly more unattended *this* in high school. However, unattended *this* is an example of *in*formality; attended *this* supports formal tone and reduces ambiguity (Swales, 2005). Thus, based on this criteria of this study, my high school technical writing samples are not necessarily significantly more formal than my college writing.

R2: Is my writing in college significantly higher in clarity than my writing in high school?

As evident in Figure 3, there is statistically significantly less passive voice in the university corpus than the high school one; thus, I can conclude that I use significantly less passive voice in my technical writing today than I did in high school. On the other hand, there was not a significant difference in the proportion of short sentences in the samples, so I cannot claim to use shorter or longer sentences now than I have in the past. Therefore, based on these two criteria, my writing in college is not significantly more clear than my technical writing was in high school.

Conclusion

Even though I didn't emerge from this study with my research questions wholly answered, I still learned that I use significantly less passive voice and significantly fewer unattended anaphoric pronouns than I did in high school. I was particularly interested by the reduced frequency of unattended *this* since it's not a grammatical feature I've consciously focused on reducing in my writing.[3] Going forward, I will:

- Be more conscious of sentence length and shorten my sentences for clarity
- Be aware of unattended anaphoric demonstratives and attend them when possible
- Remember that my writing is improving, even when I don't notice a tangible difference

[3]In comparing my percentages of unattended *this* to previous studies' findings on student papers (see *Introduction*), I tended to use more unattended *this* in high school and college than other students. 62% of the demonstratives were unattended in high school, and 45% were in college; Wulff, Romer, Swales, and Boettger found percentages ranging from 33% to 45% in other students.

References

Boettger, R. K., & Wulff, S. (2014). The Naked Truth about the Naked This: Investigating Grammatical Prescriptivism in Technical Communication. *Technical Communication Quarterly, 23*, 115-140.

Harvey, V. S. (2005). Variables affecting the clarity of psychological reports. *Journal of Clinical Psychology, 62*(1), 5-18.

Last, S. (2019). 2.2 Communicating with Precision. In *Technical Writing Essentials.* University of Victoria.

Swales, J. M. (2005). Attended and Unattended "this" in Academic Writing: A Long and Unfinished Story. *ESP Malaysia, 11*.

U.S. General Services Administration. (2011). *Federal plain language guidelines.* Retrieved 2022, from plainlanguage.gov: https://www.plainlanguage.gov/media/FederalPLGuidelines.pdf

University of Toronto. (n.d.). *Passive Voice: When to Use It and When to Avoid It*. Retrieved 2022, from University of Toronto Writing Advice: https://advice.writing.utoronto.ca/revising/passive-voice/

Vincent, S. (2014). *Sentence length: why 25 words is our limit*. (Government Digital Service) Retrieved from GOV.UK: https://insidegovuk.blog.gov.uk/2014/08/04/sentence-length-why-25-words-is-our-limit/

Wulff, S., Römer, U., & Swales, J. (2012). Attended/unattended this in academic student writing: Quantitative and qualitative perspectives. *Corpus Linguistics and Linguistic Theory, 8*, 129-157.