

Text Simplification of College Admissions Instructions: A Professionally Simplified and Verified Corpus

Zachary W. Taylor

College of Education and Human Sciences
University of Southern Mississippi
z.w.taylor@usm.edu

Maximus H. Chu

Department of Computer Science
The University of Texas at Austin
maximuschu@utexas.edu

Junyi Jessy Li

Department of Linguistics
The University of Texas at Austin
jessy@utexas.edu

Abstract

Access to higher education is critical for minority populations and emergent bilingual students. However, the language used by higher education institutions to communicate with prospective students is often too complex; concretely, many institutions in the US publish admissions application instructions far *above* the average reading level of a typical high school graduate, often near the 13th or 14th grade level. This leads to an unnecessary barrier between students and access to higher education.

This work aims to tackle this challenge via text simplification. We present **PSAT** (**P**rofessionally **S**implified **A**dmissions **T**exts), a dataset with 112 admissions instructions randomly selected from higher education institutions across the US. These texts are then professionally simplified, and verified and accepted by subject-matter experts who are full-time employees in admissions offices at various institutions. Additionally, PSAT comes with manual alignments of 1,883 original-simplified sentence pairs. The result is a first-of-its-kind corpus for the evaluation and fine-tuning of text simplification systems in a high-stakes genre distinct from existing simplification resources. PSAT is available at <https://doi.org/10.5281/zenodo.7055024>.

1 Introduction

Access to the higher education system for minoritized populations, especially low-income students, first-generation in college students, students of Color, and students whose first spoken language is not English, is an important social challenge (Auerbach, 2004; Cook et al., 2012; Flores, 2010; Pérez and McDonough, 2008; Rosa, 2006; Taylor, 2018, 2020a). However, researchers have consistently explained that much of higher education’s

communication with prospective students is too complex, too lengthy, and requires a wealth of prior knowledge of the higher education system to successfully navigate (Auerbach, 2004; Pérez and McDonough, 2008); concretely many institutions in the US publish admissions application instructions at or above the 14th grade English reading comprehension level, far too high for the average prospective student or adult to read and comprehend (Taylor, 2018, 2019, 2020b). As a result, the verbose, difficult communication places an unnecessary barrier between students and access to higher education (Ardoin, 2013; Goff et al., 2004; Hartman, 1997; Kanno, 2018; Taylor, 2017).

With modern conditional text generation models, one plausible way to lower such barriers is automatic text simplification (Siddharthan, 2014; Alva-Manchego et al., 2020), i.e., simplify text such that it is more readable and accessible, while adhering to the texts’ original semantic content. Text simplification is known to benefit a range of readers, including children (Javourey-Drevet et al., 2022) and L2 learners of English (Yano et al., 1994). Yet, no studies have explored how to simplify this information, without losing important details, to render the admissions process more accessible for prospective students and their families. Crucially, existing work in text simplification largely rely on two established datasets in the News and Wiki domains (Woodsend and Lapata, 2011; Coster and Kauchak, 2011; Xu et al., 2015; Zhang and Lapata, 2017); unlike these two domains, college admission texts certainly falls under the category of being more specialized and difficult to generalize, where both the jargon and concepts encapsulated within these texts are not ones that one encounters every day. Consequently we lack understanding both in terms of the nature of

(Original) All conditionally accepted applicants must consent to, submit to and successfully complete a criminal background check through Certiphi Screening, Inc. Failure to do so will constitute failure to meet the pre-matriculation requirements established by SUNY Optometry and will result in the withdrawal of a conditionally accepted offer.

(Simplified) If you are conditionally accepted, you must consent to, submit to, and complete a criminal background check through Certiphi Screening, Inc. If you do not, we will withdraw your conditionally accepted offer.

(Original) You must complete the following steps before USF will consider your application complete and begin admission evaluation. Pay the non-refundable \$30 application fee or submit an application fee waiver.

(Simplified) You must submit an online application with a nonrefundable \$30 application fee. You can also submit an application fee waiver.

Table 1: Original vs our expert simplified version of college admissions texts.

simplification in college admissions texts, as well as model performance in this new domain. The lack of domain diversity has been identified as a critical issue of text simplification as a whole (Alva-Manchego et al., 2020).

This work introduces **PSAT (Professionally Simplified Admissions Texts)**, the first manually simplified and verified corpus of admissions instructions from 112 randomly sampled US post-secondary institutions. Our professional simplification process are guided by extensive existing literature on manual simplification, including lowering syntactic complexity, improving lexical cohesion, and elaborating jargon (Crossley et al., 2008; McNamara et al., 2014; Siddharthan, 2014). Given that the admissions process normally requires application fees, completion of an application, writing and submitting of essays, sending of transcripts, and possibly more processes, it is crucial to understand how to compose admissions application instructions to clearly but accurately explain these processes, especially to first-generation in college students who may not have support from their secondary school, family, or household (Taylor, 2020b). *To this end, every document in PSAT was manually verified by 2 subject-matter experts (among a total of 10 experts) who are employed as full-time admissions professionals in US institutions of higher education.*

In addition to full document simplification, we also create manual alignments (by the author who originally simplified all documents), aligning each sentence in the original text to its simplified version.

This entailed 1,883 sentence pairs total.

Our analyses showed that the simplified documents reduced the reading level of these texts from grade 13.3 to grade 9.8, making it much more accessible for minority and emergent bilingual students. For experiments on automatic simplification, in this paper we focus on *sentence* simplification utilizing the aligned sentence pairs. We start from zero-shot transfer from existing simplification models pre-trained on both Wikilarge (Zhang and Lapata, 2017) and Newsela (Xu et al., 2015), as well as fine-tuned models on PSAT. We showed that this is a challenging domain for simplification. Additionally, our professional manual inspection of model outputs points out domain-specific errors that impact the accuracy of the simplifications.

2 Background and related work

2.1 Readability in admissions text

Decades of research in higher education has demonstrated that higher education information is often unreadable by prospective student audiences who read at average, pre-enrollment reading comprehension levels, roughly the 12th grade (Taylor, 2018, 2019, 2020b). At a larger scale, literacy research has consistently found that the average United States adult reads and comprehends below the 9th grade level, suggesting that many adults many not be able to understand how to read college admissions information and successfully apply to an institution of higher education in the United States (Hauser et al., 2005; Mamedova and Pawlowski, 2019; Sum et al., 2004).

Pertinent to the study at hand, prior research has found that undergraduate admissions instructions (what a student needs to accomplish and submit to the institution to be considered for admission) are often written above the 14th grade English reading comprehension level (Taylor, 2020b). Moreover, qualitative research focused on student experiences during the application process has found that prospective students, many of them from low-income and first-generation in college backgrounds, often struggle understanding higher education jargon such as *undergraduate*, *FAFSA*, and *verification*, all of which are critical to comprehend during the admissions and enrollment management process (Ardoin, 2013; Taylor and Bicak, 2019). Here, many students may be academically prepared for the rigors of a college curriculum, yet they may not understand the language and processes of higher

education writ large. Such a misunderstanding may systematically exclude these students due to a lack of support during the application process, not a lack of preparedness for college itself.

However, little has changed in the decades during which higher education researchers have been investigating the complexity of the college admissions process. In the 1980s and early 1990s, researchers called for the admissions process to be simplified and for common admissions applications to be widely embraced by institutions across the country (Astin et al., 1982; Gandara, 1986; Post, 1990). Yet by the early 2000s, researchers were still reporting that many minoritized students and their families struggled to comprehend how to complete the college admissions process and enroll in higher education (McBrien, 2005; Rosa, 2006; Tornatzky et al., 2002; Ward, 2006).

Recently, given the considerable personal and financial hardships facilitated by the COVID-19 pandemic, students and families continue to report that the college admissions process is too difficult (McCulloh, 2022) and often requires dozens of hours to complete (Reilly, 2021), and that institutions of higher education have not made efforts to ease the admissions process by simplifying the instructions for how to apply and enroll into higher education (Hurtado et al., 2020; Morrison, 2021). Subsequently, this study attempts to analyze simplifications of college admissions instructions to create an automated model that may automatically simplify college admissions instructions, easing the information burden for prospective students and families, while also easing the workload of admissions practitioners working for institutions of higher education.

2.2 Text simplification

Modern text simplification systems are typically encoder-decoder models that performs monolingual translation from the original to simplified text (Wang et al., 2016; Zhang and Lapata, 2017; Kriz et al., 2019; Dong et al., 2019; Martin et al., 2020; Devaraj et al., 2021; Maddela et al., 2021). The majority of this research are at the sentence level, trained using two corpora: the Wikipedia-Simple Wikipedia aligned corpus (Woodsend and Lapata, 2011; Coster and Kauchak, 2011; Zhang and Lapata, 2017) and the Newsela simplification corpus (Xu et al., 2015). As a result, it is unclear whether models trained on these datasets can be

applied in other domains, and progress is hindered by the lack of diverse datasets (Alva-Manchego et al., 2020). This work is a step towards mitigating towards this issue.

The training of supervised models rely on automatically aligned sentence pairs (Jiang et al., 2020). However, Devaraj et al. (2022) demonstrated that noisy alignments can lead to inaccuracies during the simplification process. Instead, in this work, all documents are manually aligned by the author of the simplification.

3 Data collection and simplification

The main work of PSAT included performing manual text simplification, followed by acceptability judgements made by subject-matter experts with domain-specific professional work experience in either admissions or financial aid. Below, we detail the corpus development process, including the gathering of college admissions instructions (Section 3.1), professional simplification and its principles (Section 3.2) and verification (Section 3.3).

3.1 Raw text extraction

To gather the original texts (i.e., publicly available admissions instructions texts) for this study, we employed the National Center for Education Statistics' Integrated Postsecondary Education Data System (IPEDS)¹ to gather institutional URLs to each institution's admissions application instructions website. Moreover, we decided to gather data during the college search and exploration process, typically occurring between August and November of each year in the US (Hossler and Gallagher, 1987). Understanding both student exploration and institutional information practices, we gathered all website data for this study in October 2019. Initially, we gathered undergraduate admissions instructions from 335 institutions, which we then manually inspected and kept 112 that contain long, meaningful discourse other than metadata (e.g., addresses).

All instructions are then cleaned, by manually extracting the raw text.

3.2 Simplification

Personnel The texts, written in English (from US institutions), were manually simplified by an author of this paper who is a native English speaker. The author has a doctoral degree in education and has worked professionally in US postsecondary

¹<https://nces.ed.gov/ipeds/>

education for over a decade, including work in undergraduate admissions. Thus the author engaged with their professional insight to simplify without losing critical information necessary for its comprehension and understanding.

Principles Longitudinal research (Crossley et al., 2008; McNamara et al., 2014; Siddharthan, 2014) has suggested that text can often be simplified using a smaller lexicon than the original text, rewriting sentences so that adjoining sentences share syntactic features (e.g., punctuation marks, independent clauses followed by main clauses), and ensuring that previously used lexical items appearing earlier in a text are re-introduced later in the text to improve comprehension of the text. Thus our manual simplification process include *reducing syntactic complexity*, *increasing lexical cohesion* (word overlap and frequency), *the elaboration and explanation of jargon and acronyms*, and *domain-specific principles*.

(1) Reducing syntactic complexity. We adopted Coleman (1962)'s framework for sentence-level simplification: raising clause fragments to full sentences, dividing sentences joined by conjunctions (e.g., because, but, for, or), avoid dividing sentences joined by the conjunction "and", and shortening clauses by using periods where other forms of grammatical punctuation may be found (e.g., semicolons, colons, commas). While shortening, we ensured that we do not lose critical information which would cause a factual error (Devaraj et al., 2022).

Additionally, extant research has suggested that writing or speaking in active voice rather than passive voice can increase simplicity, and this rewrite can often lead to shorter sentences (DeVito, 1969; Ferreira, 1994). As a result, we identified instances of passive voice in each admissions text and re-wrote them into active voice, e.g.,

(Original) The application must be completed by the student.

(Simplified) The student must complete the application.

(2) Increasing lexical cohesion. Hulme et al. (1997) learned that increasing the word frequency in an informative text helped with the short-term memory recall of research participants regarding the content of the text, supporting the finding that increasing the word frequency in a text may lead to a better understanding of the text on behalf of the reader. These words included "you," "must,"

"official," "transcript," and many others. Therefore, we simplified in a way that promotes lexical overlap across documents.

Pertinent to this study, Monaghan et al. (2017) also found that individual differences across bilingual readers in terms of word frequency effects were due to exposure to word diversity, not an individual's vocabulary size (personal lexicon). This finding supported the use of increasing word frequency to increase a text's simplicity and possible readers' comprehension of the text. Thus, we also attempted to identify content words that could be repeated earlier or later in each text separately, including words like "you," "must," "submit," "transcripts," "contact," and many others.

(3) Elaboration and explanation of jargon and acronyms. Research on acronyms and initialisms has found that using these lexical items in potentially unfamiliar text can be confusing to readers, thus making the text more difficult to read (Cannon, 1989; Ibrahim, 1989; Rúa, 2002; Laszlo and Federmeier, 2007). Using acronyms often hinders clarity, as the reader may need to parse extra text or consult another text in order to decipher the acronym or initialism and fully comprehend the text (Rúa, 2002; Taghva and Gilbreth, 1999). We attempted to locate acronyms (e.g., ACT) and initialisms (e.g., GPA) within each text and ensure that these acronyms and initialisms were clear and commonly used, so that students would not be confused when reading the simplification. Common acronyms and initialisms in this dataset included "ACT," "SAT," "GED," "US," and several others. Subject-matter experts determined that although acronyms and initialisms may be confusing, including the acronyms and initialisms without lengthy definitions was best to assist with simplification, as subject-matter experts determined that the vast majority of students would understand these acronyms and initialisms without context clues or definitions.

(4) Domain-specific principles. As admission instructions are tied with direct consequences for postsecondary education, we work with 10 subject-matter experts (SME, their background detailed in Section 3.3) such that our simplification is accurate as possible within this particular domain.

Subject-matter experts unanimously determined that admissions instructions should only contain the instructions themselves, and that any extraneous text should be removed to ensure minimalism and simplicity. Consequently, text was removed

(from the original instructions) by SMEs if the text was not pertinent to the application process itself, including welcome statements, espoused institutional beliefs, and language that was determined to be marketing and/or branding and not admissions instructions. Examples of removed texts include, “*You don’t have any time to waste, so we made applying as simple as possible*”, “*The Office of Admissions assists prospective students in exploring the academic opportunities*”.

Moreover, there were several keywords within sentences that subject-matter experts insisted remain or be added from original to simplified texts. For example, the words “official” and “transcript” needed to be retained in the simplified versions; and if these words did not appear in the original, they were added into the simplified version for clarity for the prospective student. In addition, SMEs agreed that simple admissions text includes second-person pronoun usage and modal verbs to provoke action on behalf of the student, for example:

(Original) Application fee waivers are available for students with demonstrated financial need.

(Simplified) If you cannot pay the fee, you can apply for an application fee waiver.

Ultimately, subject-matter experts felt that this language rendered the text simpler because students would better understand that they were responsible for completing parts of the admissions process.

3.3 Acceptability Judgements by Subject-Matter Experts (SMEs)

To determine whether the simplification of admissions application instructions were acceptable—that is to say they did not lose critical information or accuracy between the pre- and post-simplification process—we engaged with ten subject-matter experts (SMEs). Each simplified text was verified by 2 SMEs independently; in total, we engaged with 10 SMEs, who volunteered their time.

Personnel All ten of the SMEs had professional backgrounds in U.S. postsecondary admissions, having worked at least five years full-time in college admissions offices in the United States. These SMEs were identified through professional networks and snowball methods, as several of our SMEs knew colleagues from different institutions or educational entities who would serve as high-quality, knowledgeable SMEs.

Moreover, we engaged with a diverse group of SMEs from different institution types (i.e., com-

munity colleges, public four-year institutions, private liberal arts colleges) and with various lengths of experience to capture the potential variability of admissions and financial aid parlance, jargon, and communication style. As the first study of its kind, identifying SMEs from diverse backgrounds provided more generalizability and reliability of findings, thus informing future research and practice regarding the communication of admissions application instructions to students and their support networks. Four subject-matter experts worked at public, four-year universities, four worked at private, four-year universities, and two worked at public, two-year community colleges.

Process To perform the acceptability judgement, the SME was presented with both pre- and post-simplification texts in real time over a Zoom video conference meeting. Then, we asked the SME to read the pre-simplified (original) text, followed by the post-simplification (simplified) text and determine whether the simplified text was acceptable. For example, changing the verb “submit” to “complete” is not acceptable because “submit” implies the documentation or information is being submitted by a submitter to a submittee, while “complete” only implies the documentation or information is completed and not directed to any educational stakeholder.

If a simplification was deemed unacceptable by one or more SMEs, we asked the SME what simplification would be acceptable through an iterative process in real time across all texts in this study. Once the SME provided their feedback and we integrated their feedback into the simplified text, the same SME was again asked to read both the both pre- and post-simplification texts in real time and render their acceptability judgement. If at any time there was an instance where a lexical item (e.g., single word, acronym, initialism, compound adjective), sentence, or paragraph could not be simplified, the pre-simplified section of that text was used.

3.4 Manual Alignments

In the PSAT corpus, the author of the simplified documents manually added sentence alignments across the original and simplified texts. Specifically, for each sentence in the original version, we provide the indice(s) of the corresponding simplified sentence. This comes to a total of 1,883 alignments, excluding sentences that are kept un-

(Simplified texts) just, development, following, prior, credit, sent, steps, recommended, applications, freshmen, payment, better, no, currently, once, chapman, plan, well, way, submitting, educational, aid, ensure, enroll, stonehill, options, completed, documents, code, create, credits, rolling, option, begin, appropriate, completing, now, earned, out, reviewed, checklist, items, submission, during, process, general, colleges, receive, please, standardized

(Original texts) codes, materials, how, andor, csu, non-native, paintings, prepare, invite, dualcredit, homeschool, cannot, using, words, admit, certiphi, can, must, native, responses, resume, you, selfreport, letter, georgia, closer, bonaventure, prompted, event, inclusion, syllabi, statements, dualdegree, significant, quebec, tb, former, thinking, filmmaking, carrolls, concise, testingservice-sucmoedu, baton, crucial, parker, coach, cresson, did, esl, autobiography

Table 2: Words most associated with simplified vs. original texts, calculated with log-odds ratios.

changed, metadata (e.g., addresses), and short (< 2-word) headers.

4 Analysis

Table 3 shows comparative statistics of the 112 original—simplified bitexts; we see that on average the simplified versions have shorter and fewer sentences. The original texts in PSAT are above the 12-th grade reading level (measured with Flesch-Kincaid (Kincaid et al., 1975)) of an average high school graduate, confirming the findings in Taylor (2020a,b). The simplified version lowered this reading barrier.

We also study lexical items that are most associated with the original texts vs. the simplified version. Concretely we calculate the log-odds ratio of each word w_i , comparing the conditional probability of w_i in the original instructions \mathbf{D}_o or the simplified ones \mathbf{D}_s : $\text{logodds}(w_i) = \log(P(w_i|\mathbf{D}_o)/P(w_i|\mathbf{D}_s))$ (Nye and Nenkova, 2015). Excluding punctuation and numbers, the words with the highest and lowest log-odds ratios with respect to simplified texts are shown in Table 2. We see that after simplification, the vocabulary becomes more standardized towards the application and admissions process itself.

Finally, we measure the abstractiveness of the simplified texts; namely, how much paraphrasing was used? For this we use the *aligned* portion of PSAT; the percentage of unique uni-, bi-, and tri-grams in the simplified texts are 34.4%, 52.4%, 58.6%, respectively. Thus the amount of new jargon introduced through simpli-

	# sents	# tokens	FK
Original	29.6	10.2	13.3
Simplified	17.9	7.1	9.8

Table 3: Comparison of original vs. simplified texts: average number of sentences, tokens, and the Flesch-Kincaid grade level.

fication is enough to be significant; this indicates that simpler vocabulary is being introduced, and significant paraphrasing happened during simplification.

5 Sentence Simplification Experiments

We establish baselines for the sentence simplification task in PSAT, using the aligned sentences. With these experiments we hope to evaluate existing models—which are trained in other domains—in a zero-shot manner, gauge the utility of fine-tuning on PSAT, and discuss the challenges of simplification in this domain.

5.1 Models

ACCESS Our baseline is the ACCESS model (Martin et al., 2020), trained on Wikilarge (Zhang and Lapata, 2017). ACCESS uses a transformer trained from scratch (i.e., randomly initialized).

T5 We also fine-tune the T5 model (Raffel et al., 2020) for the simplification task. T5 is a pre-trained large-scale encoder-decoder model optimized on conditional generation tasks (e.g., summarization and machine translation), in addition to unsupervised objectives. We fine-tune T5-base on both Wikilarge and Newsela (Xu et al., 2015), denoted as T5-wiki and T5-newsela. Following Devaraj et al. (2022), the models are fine-tuned with prefix “*summarize:*” for 5 epochs with a batch size of 6 and learning rate $3e-4$. The default greedy decoding was used.

Additionally, we further fine-tune these models on PSAT’s training data (Section 5.2) for 3 more epochs, denoted as T5-wiki-ft and T5-newsela-ft.

5.2 Experimental setup

Data We split PSAT documents into 50% train, 30% test, and 20% validation (the test and validation sets constitute 50% of the data, such that the number of test and validation documents

	Train	Test	Validate
# Files	56	33	23
# Sentence Pairs	955	559	369

Table 4: The number of files and sentence pairs for the train, test, validation splits in PSAT.

	BERTScore	BLEU	SARI
ACCESS	0.898	0.104	0.271
T5-newsela	0.906	0.132	0.226
T5-wiki	0.903	0.171	0.169
T5-newsela-ft	0.919	0.210	0.241
T5-wiki-ft	0.923	0.239	0.216

Table 5: Automatic evaluation results of simplification systems.

will be large enough). We then use the manual alignments in these document sets for training/testing/validation examples. Table 4 shows the number of documents and aligned sentences in each split.

Evaluation For automatic evaluation, we report the following metrics:

(1) **SARI** (Xu et al., 2016), a widely used metric especially designed according to the edit nature of simplification. We report the average n-gram F1 scores corresponding to keep, delete, and add operations.²

(2) **BLEU** (Papineni et al., 2002), a standard metric in machine translation and other conditional generation tasks. Xu et al. (2016) stated that while SARI is better at capturing simplicity, BLEU has a stronger correlation with grammar and meaning.

(3) **BERTScore** (Zhang et al., 2019), which is shown to correlate better with human references in generation tasks, and has been adopted as an evaluation metric for summarization (Ahuja et al., 2022).

5.3 Results

Results are shown in Table 5. First, notice that all scores are lower in terms of absolute scale compared to the datasets they are trained on; for instance, Devaraj et al. (2022) reported SARI scores over 36 for all models on Newsela and Wikilarge. Although ACCESS obtained the highest SARI score, its BLEU and BERTScores are the

²Note that SARI scores are more reliable if multiple reference simplifications are present, but we are limited to one reference due to personnel, time, and budget constraints.

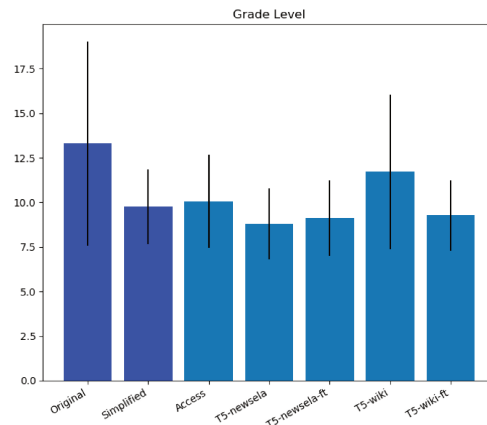


Figure 1: The Flesch-Kincaid Grade Level averages and standard deviation for the original and manually simplified texts, as well as the simplified texts generated by the five models.

lowest; our manual inspection (Section 6) confirmed that this model does not produce satisfactory simplification for admissions instructions. When comparing T5-newsela and T5-wiki, there is no clear superior model as T5-wiki has higher BLEU but lower SARI, and the two have very little difference in terms of BERTScore. This suggests that training on either Wikipedia articles or news articles as a starting point is acceptable.

The clearest trends across all metrics is that fine-tuning on PSAT substantially improves performance regardless of whether the model is pre-trained on Wikilarge or Newsela, demonstrating that in-domain fine-tuning is useful even with a moderately sized dataset like PSAT. This finding is consistent with recent discoveries in summarization (Yu et al., 2021).

Figure 1 depicts the Flesch-Kincaid grade level for the original sentences, their manually simplified versions, and the model outputs among the aligned data. Note that although we perform this analysis due to Flesch-Kincaid’s superior performance in assessing the readability of other technical texts in an unsupervised manner (Devaraj et al., 2021), we do not intend to use Flesch-Kincaid as one of our main automatic evaluation metrics for reasons pointed out by Alva-Manchego et al. (2021) and Tanprasert and Kauchak (2021). Comparing the five models, T5-wiki yielded a higher (and more varied reading level) than all other models and the manually simplified. On the contrary, T5-newsela’s output reading level is slightly lower than the manual simplification. Meanwhile,

	Fluency	Simplicity	Accuracy
ACCESS	1/25	1/25	1/25
T5-newsela	4/25	4/25	4/25
T5-wiki	4/25	4/25	5/25
T5-newsela-ft	2/25	2/25	1/25
T5-wiki-ft	5/25	5/25	5/25
None	9/25	9/25	9/25

Table 6: Number of times (out of 25) that each system received the highest rank, along with # of times that no system was selected.

both T5-newsela-ft and T5-wiki-ft produced desirable readability levels; notably fine-tuning was able to improve the readability of T5-wiki outputs.

6 Manual Inspection

Finally, because PSAT is a very new domain that comes with its own characteristics and requirements for accurate simplification, we perform a small scale manual inspection with the professional who authored the simplified documents; this professional has not seen model-generated outputs in NLP before. Because of time commitment required to inspect outputs from all 5 models, we were only able to study 25 examples. Nonetheless we hope the insights here will guide future work on domain-specific evaluation metrics.

Concretely, we randomly sample 25 sentences from the test set. We ask the professional to inspect outputs based on three metrics: **fluency, simplicity, and accuracy**. Under each metric, we ask the professional to choose the highest ranked simplification; an option of *none* is available if no model provided satisfactory outputs for any criteria.

Results in Table 6 confirmed that T5-wiki-ft performed the best, and that ACCESS outputs do not outperform other models. Similar to automatic evaluation results, T5-wiki vs T5-newsela. However, there is also a large portion of examples where “none” was chosen. Additionally, T5-newsela-ft performed much worse than T5-newsela. This suggests that automatic metrics do not capture some of the aspects that the professional deems important in their judgments. This echoes findings from Alva-Manchego et al. (2021) but emphasizes the importance of human evaluation for specific domains. We detail some of the insights qualitatively below.

First, several simplifications required adding

words to the simplified version that did not appear in the original, however the model did not learn to do so. The most frequent example was simplifications adding “official” to the phrase “Official Transcript(s)”. Here, subject-matter experts felt that adding “official” was necessary to make it clearer and more simple for students to understand that they needed to submit official transcripts instead of copies or screenshots of their transcripts.

Similarly, many simplifications added the subject and modal verb “you must” to phrases such as “submit your official transcript(s)”, “complete an application”, and “apply by the priority date”, as discussed in Section 3.2. The subject-matter experts felt it was necessary to clarify that a student “must” complete these tasks, otherwise they would not be considered for admission. However, the model could not generate such phrasing.

Finally, many models removed keywords from the original versions. This most often happened when an institution accepted multiple application types, such as the Coalition for College Application and the Common Application. For instance, one sentence original sentence read, “*We accept the Coalition for College Application and the Common Application*”. However, several models attempted to simplify the sentence and remove either the Coalition for College Application or Common Application, as they both contain the word “application”, and the model may have removed these “redundant” phrases. However, both application types needed to appear in the simplified version to gain subject-matter expert approval, as an accurate admissions instructions contain all of the different applications that a student may use to apply for admission, not just the simplest application. As a result, several models did not accurately keep keywords from original to simplified versions because the models did not have this domain-specific information that subject-matter experts had.

7 Discussions and Conclusions

This work presents PSAT, a text simplification corpus consisting of admission instructions texts from 112 US higher education institutions and their simplified versions. PSAT texts are professionally simplified and verified, rendering it the first-of-its-kind and most accurate dataset in this high-stake domain. We showed that this dataset is challenging for existing simplification models, especially due to domain mismatch, and domain-specific requirements for

the accuracy of information.

We recognize the limitations of this study. While there are over 6,000 institutions of higher education in the US, this dataset sampled a small number of these institutions given time constraints and the substantial work necessary to gather and simplify text and work with subject-matter experts to approve simplified texts.

Ultimately, we will perform future research—using this dataset—to specifically identify which simplifications were acceptable and unacceptable (at the lexical item-, sentence- and paragraph-level), thus informing admissions practitioners as to how text can or cannot be simplified for students and prospective students, broadly. We hope PSAT is a first step towards automatic simplification systems for fairer access to higher education.

Acknowledgements

The authors would like to acknowledge the subject-matter experts who volunteered their time to read the texts from this study and provide their professional judgement. This research is partially supported by NSF grants IIS-2145479, IIS-2107524. We acknowledge the Texas Advanced Computing Center (TACC)³ at UT Austin for many of the results within this paper.

References

- Ojas Ahuja, Jiacheng Xu, Akshay Gupta, Kevin Horecka, and Greg Durrett. 2022. ASPECTNEWS: Aspect-oriented summarization of news documents. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6494–6506.
- Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2020. Data-driven sentence simplification: Survey and benchmark. *Computational Linguistics*, 46(1):135–187.
- Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2021. The (un) suitability of automatic evaluation metrics for text simplification. *Computational Linguistics*, 47(4):861–889.
- Mary Sonja Ardoin. 2013. *Learning a different language: Rural students' comprehension of college knowledge and university jargon*. North Carolina State University.
- Alexander W Astin et al. 1982. *Minorities in American higher education. Recent trends, current prospects, and recommendations*. ERIC.
- Susan Auerbach. 2004. Engaging latino parents in supporting college pathways: Lessons from a college access program. *Journal of Hispanic Higher Education*, 3(2):125–145.
- Garland Cannon. 1989. Abbreviations and acronyms in english word-formation. *American Speech*, 64(2):99–127.
- Edmund Benedict Coleman. 1962. Improving comprehensibility by shortening sentences. *Journal of Applied Psychology*, 46(2):131.
- Amy Cook, Rachele Pérusse, and Eliana D Rojas. 2012. Increasing academic achievement and college-going rates for latina/o english language learners: A survey of school counselor interventions. *The Journal of Counselor Preparation and Supervision*, 4(2):2.
- William Coster and David Kauchak. 2011. Simple English Wikipedia: a new text simplification task. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 665–669.
- SA Crossley, J Greenfield, and DS Mcnamara. 2008. Text readability using assessing based indices. *TESOL Quarterly*, 42(3):475–493.
- Ashwin Devaraj, Iain Marshall, Byron C Wallace, and Junyi Jessy Li. 2021. Paragraph-level simplification of medical texts. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4972–4984.
- Ashwin Devaraj, William Sheffield, Byron Wallace, and Junyi Jessy Li. 2022. Evaluating factuality in text simplification. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7331–7345.
- Joseph A DeVito. 1969. Some psycholinguistic aspects of active and passive sentences. *Quarterly Journal of Speech*, 55(4):401–406.
- Yue Dong, Zichao Li, Mehdi Rezagholizadeh, and Jackie Chi Kit Cheung. 2019. EditNTS: An neural programmer-interpreter model for sentence simplification through explicit editing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3393–3402.
- Fernanda Ferreira. 1994. Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33(6):715–736.
- Stella M Flores. 2010. State dream acts: The effect of in-state resident tuition policies and undocumented latino students. *The Review of Higher Education*, 33(2):239–283.
- Patricia Gandara. 1986. Chicanos in higher education: The politics of self-interest. *American Journal of Education*, 95(1):256–272.

³<https://www.tacc.utexas.edu>

- Brent Goff, Vanessa Patino, and Gary Jackson. 2004. Preferred information sources of high school students for community colleges and universities. *Community College Journal of Research & Practice*, 28(10):795–803.
- Kenneth E Hartman. 1997. College selection and the internet: Advice for schools, a wake-up call for colleges. *Journal of College Admission*, 154:22–31.
- Robert M Hauser, Christopher F Edley Jr, Judith Anderson Koenig, and Stuart W Elliott. 2005. *Measuring Literacy: Performance Levels for Adults*. ERIC.
- Don Hossler and Karen Gallagher. 1987. Studying student college choice: A three-phase model and the implication...-supersearch powered by summon. *College and University*, 62:201–21.
- Charles Hulme, Steven Roodenrys, Richard Schweickert, Gordon DA Brown, Sarah Martin, and George Stuart. 1997. Word-frequency effects on short-term memory tasks: evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5):1217.
- Sylvia Hurtado, Hector Vicente Ramos, Edwin Perez, and Xochilth Lopez-Salgado. 2020. Latinx student assets, college readiness, and access: Are we making progress? *Education Sciences*, 10(4):100.
- AM Ibrahim. 1989. Acronyms observed. *IEEE Transactions on Professional Communication*, 32(1):27–28.
- Ludivine Javourey-Drevet, Stéphane Dufau, Thomas François, Núria Gala, Jacques Ginestíe, and Johannes C Ziegler. 2022. Simplification of literary and scientific texts to improve reading fluency and comprehension in beginning readers of french. *Applied Psycholinguistics*, 43(2):485–512.
- Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. 2020. Neural CRF model for sentence alignment in text simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7943–7960.
- Yasuko Kanno. 2018. Non-college-bound english learners as the underserved third: How students graduate from high school neither college-nor career-ready. *Journal of Education for Students Placed at Risk (JESPAR)*, 23(4):336–358.
- J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch.
- Reno Kriz, João Sedoc, Marianna Apidianaki, Carolina Zheng, Gaurav Kumar, Eleni Miltsakaki, and Chris Callison-Burch. 2019. Complexity-weighted loss and diverse reranking for sentence simplification. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3137–3147.
- Sarah Laszlo and Kara D Federmeier. 2007. Better the dvl you know: Acronyms reveal the contribution of familiarity to single-word reading. *Psychological Science*, 18(2):122–126.
- Mounica Maddela, Fernando Alva-Manchego, and Wei Xu. 2021. Controllable text simplification with explicit paraphrasing. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3536–3553.
- Saida Mamedova and Emily Pawlowski. 2019. Adult literacy in the united states. data point. nces 2019-179. *National Center for Education Statistics*.
- Louis Martin, Éric Villemonte De La Clergerie, Benoît Sagot, and Antoine Bordes. 2020. Controllable sentence simplification. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4689–4698.
- J Lynn McBrien. 2005. Educational needs and barriers for refugee students in the united states: A review of the literature. *Review of educational research*, 75(3):329–364.
- Edna McCulloh. 2022. An exploration of parental support in the retention of rural first-generation college students. *Journal of College Student Retention: Research, Theory & Practice*, 24(1):144–168.
- Danielle S McNamara, Arthur C Graesser, Philip M McCarthy, and Zhiqiang Cai. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- Padraic Monaghan, Ya-Ning Chang, Stephen Welbourne, and Marc Brysbaert. 2017. Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, 93:1–21.
- Joe Morrison. 2021. *4 ways to transform college admissions*. *eSchool News*.
- Benjamin Nye and Ani Nenkova. 2015. Identification and characterization of newsworthy verbs in world news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1440–1445.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- Patricia A Pérez and Patricia M McDonough. 2008. Journal of hispanic higher. *Journal of Hispanic Higher Education*, 7:249.

- David Post. 1990. College-going decisions by chicanos: The politics of misinformation. *Educational Evaluation and Policy Analysis*, 12(2):174–187.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21:1–67.
- Katie Reilly. 2021. [Applying to college was hard. covid-19 made it impossible.](#) *Time*.
- Mari Luna De La Rosa. 2006. Is opportunity knocking? low-income students’ perceptions of college and financial aid. *American behavioral scientist*, 49(12):1670–1686.
- Paula López Rúa. 2002. On the structure of acronyms and neighbouring categories: A prototype-based account. *English Language & Linguistics*, 6(1):31–60.
- Advaith Siddharthan. 2014. A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2):259–298.
- Andrew Sum, Irwin Kirsch, and Kentaro Yamamoto. 2004. Pathways to labor market success: The literacy proficiency of US adults. *Policy Information Report, Educational Testing Service*.
- Kazem Taghva and Jeff Gilbreth. 1999. Recognizing acronyms and their definitions. *International Journal on Document Analysis and Recognition*, 1(4):191–198.
- Teerapaun Tanprasert and David Kauchak. 2021. Flesch-Kincaid is not a text simplification evaluation metric. In *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, pages 1–14.
- Zachary W Taylor. 2017. Twenty-first-century communicators: How registrars can reject ‘university idiolect’. *The Successful Registrar*, 17(9):6–6.
- Zachary W Taylor. 2018. Intelligibility is equity: Can international students read undergraduate admissions materials? *Higher Education Quarterly*, 72(2):160–169.
- Zachary W Taylor. 2019. Lost in translation: Linguistic hurdles facing us language-minority students of color pursuing international higher education. . *Lost in Translation: Linguistic Hurdles Facing US Language-Minority Students of Color Pursuing International Higher Education. Emerging Voices in Education*, 1(1):45–60.
- Zachary W Taylor. 2020a. ¿comprenderán mis amigos y la familia? analyzing spanish translations of admission materials for latina/o students applying to 4-year institutions in the united states. *Journal of Hispanic Higher Education*, 19(2):195–209.
- Zachary W Taylor. 2020b. College admissions for 12 students: Comparing 11 and 12 readability of admissions materials for us higher education. *Journal of College Access* (2020).
- Zachary W Taylor and Ibrahim Bicak. 2019. What is the FAFSA? an adult learner knowledge survey of student financial aid jargon. *Journal of Adult and Continuing Education*, 25(1):94–112.
- Louis G Tornatzky, Richard Cutler, and Jongho Lee. 2002. College knowledge: What latino parents need to know and why they don’t know it. *Joyce Foundation*.
- Tong Wang, Ping Chen, John Rochford, and Jipeng Qiang. 2016. Text simplification using neural machine translation. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- Nadia L Ward. 2006. Improving equity and access for low-income and minority youth into institutions of higher education. *Urban Education*, 41(1):50–70.
- Kristian Woodsend and Mirella Lapata. 2011. Wikisimple: Automatic simplification of wikipedia articles. In *Twenty-fifth AAAI conference on artificial intelligence*.
- Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.
- Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.
- Yasukata Yano, Michael H Long, and Steven Ross. 1994. The effects of simplified and elaborated texts on foreign language reading comprehension. *Language learning*, 44(2):189–219.
- Tiezheng Yu, Zihan Liu, and Pascale Fung. 2021. Adaptsum: Towards low-resource domain adaptation for abstractive summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5892–5904.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. BERTscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.
- Xingxing Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594.