JCA

Journal of College Access

Volume 5 | Issue 2 Article 5

12-19-2020

Inside the Black Box of Text-Message College Advising

Karen D. Arnold
Boston College, arnoldkc@bc.edu

Laura Owen
San Diego State University, lowen@sdsu.edu

Jonathan Lewis
uAspire, jonathanl@uaspire.org

Inside the Black Box of Text-Message College Advising

The study reported in this article was funded by the Institute of Education Sciences, U.S. Department of Education (Grant #R305A1410121).

Authored by
Karen D. Arnold (*Boston College*)
Laura Owen (*San Diego State University*)
Jonathan Lewis (*uAspire*)

ABSTRACT

Making college access and success more equitable at a national scale requires alternatives to intensive in-person modes of pre-college advising. Text-message advising campaigns are a promising intervention model for delivering college application and financial aid assistance affordably to large populations of college-intending, low-income students. College outcome results from a recent series of very large text-message programs have been disappointing however. Going inside the black box of text-message advising to understand why and how students engage in text-messaging programs can help explain program effects and inform the design of future virtual-advising programs. This study uses text mining techniques to investigate the content of 342,000 student text messages from a national text-message advising program. In the program under study, over 30,000 college-intending students from 745 high schools received two-way college advising for 15 months via text messaging with professional advisors. Data mining of the student text messages indicated that students needed substantial individualized assistance and that they used texting primarily for navigating discrete tasks related to testing, applications, and financial aid. In addition to providing substantive findings about college access advising, the study method illustrates how big data tools can extract meaning from large bodies of unstructured text like those generated by the growing number of text-message educational interventions.

Keywords: college access, low-income students, text mining, text-message advising, virtual advising.

Introduction and Background

Family socioeconomic status is a powerful indicator of whether high school students will begin college, what kind of college they will attend, and whether they will earn a college degree (Bailey & Dynarski, 2011; Chetty et al., 2017; Reardon, 2013; Shapiro et al., 2019). A large research literature identifies barriers to college attainment among students who come from low-income families, have non-college-educated parents, or are from minoritized racial groups. In a review of this body of research, Page and Scott-Clayton (2016) group college access barriers into financial, informational, and academic constraints. These constraints are particularly likely to be present among students from schools with high concentrations of students from low-income backgrounds and low college-going rates (Engberg & Wolniak, 2010, 2014; Palardy, 2013; Perna & Jones, 2013; Roderick et al., 2008; Willms, 2010). In-person assistance to overcome these constraints, such as school counseling and out-of-school college access programs, is frequently insufficient for students in high-poverty schools (Avery et al., 2014; Carrell & Sacerdote, 2017; Hyman, 2019; McKillip et al., 2012; Perna et al., 2008; Swail & Perna, 2002). Increasing face-to-face support is labor-intensive and expensive; even exemplary programs can serve only limited numbers of students. Making college access and success more equitable at a national scale therefore requires new modes of advising that can be delivered affordably to large populations of students.

Technology-Enabled College Advising Interventions

Technology-based college access interventions have begun to emerge over the past decade as a possible means for reducing socioeconomic gaps in the transition to college (Arnold et al., 2015; Bettinger et al., 2019; Bird et al., 2019; Castleman, 2015; Castleman & Page, 2015, 2016; Fesler, 2020; Fesler et al., 2019; Oreopoulos et al., 2020; Page et al., 2020; Phillips & Reber, 2019). Pre-college admissions and financial aid advising that is delivered by text message is a particularly promising way to engage high school students because of the ubiquity of text messaging among American youth (Lenhart, 2015). Compared with school-based and other face-to-face approaches, texting programs are affordable and can be delivered at scale.

Economists and education policy scholars measure the effectiveness of text-message advising programs through quantitative analysis of treatment effects in randomized controlled trials (RCTs). A recent review of results from large-scale RCTs in education in the U.K. and the U.S. concluded that these trials have been generally "uninformative," however (Lortie-Forgues & Inglis, 2019, p. 158). Recent randomized controlled trials of text-message advising programs, in particular, have shown some positive outcomes in localized settings but disappointing results when delivered at large scale (Bird et al., 2019; Bergman et al., 2019; Bettinger et al., 2019; Gurantz et al., 2019; Gurantz et al., in press; Hyman, 2019; Page et al., 2019, 2020; Phillips & Reber, 2019). This emerging body of results includes investigations of the impact of large-scale campaigns to encourage students to apply for college (Gurantz et al., in press; Phillips & Reber, 2019), to apply to selective colleges (Gurantz et al., 2019); to apply or reapply for financial aid (Bird et al., 2019; Page et al., 2019), and to understand tax benefits for college (Bergman et al., 2019). Researchers typically measure the effects of interventions on overall college enrollment. Other indicators of treatment effects include the type of postsecondary institution where students enroll, student re-enrollment rates, and financial aid outcomes (Bergman et al., 2019; Bird et al., 2019; Page et al., 2019, 2020). As a group, the remote large-scale advising interventions reported to date have not demonstrated statistically significant differences between treatment and control groups in overall measures of college enrollment. Some researchers have found modest positive effects for specific student subgroups, quality of college enrollments, or likelihood of filing financial aid applications (Page et al., 2019, 2020; Gurantz, 2019; Gurantz et al., in press; Hyman, 2019; Philips & Reber, 2019).

Despite the discouraging findings from text-message campaigns to date, it is too soon to give up on the search for large-scale virtual advising programs that have the potential to move the needle on college access. Instead it is vital to study the reasons why recent text-message interventions that enroll large numbers of students have failed to show treatment effects. Understanding RCT results requires investigating what occurs inside text-message advising and how students are using this increasingly popular advising medium. Researchers can access the data to answer these questions by capturing and analyzing the text messages that students and their advisors exchange. In going inside the black box of what actually happens within text-message campaigns, in sum, "researchers can use the content of the text messages to understand how and why a program worked in the way that it did" (Fesler et al., 2019, p. 708).

Researching the content of text messages in large-sample texting campaigns is challenging methodologically because such studies require analysis of huge amounts of unstructured data in the form of the unstandardized prose that text-message interventions produce. Extracting meaning from tens of thousands—or even millions—of text segments cannot be accomplished with standard qualitative analysis procedures that involve human inspection and manual coding of an entire collection of texts (Saldaña, 2015). From a methodological standpoint, education researchers will need to employ "big data" analytic tools that can extract meaning from large bodies of unstructured text like those captured in text-message interventions (Fesler et al., 2019; Fischer et al., 2020). The use of big data is relatively new in education and few graduate programs train students in data mining techniques such as text mining (Fischer et al., 2020).

In sum, research on text-message interventions enables new insights about virtual advising programs that call for the investigation of the content of prose messages through new methodological approaches. This paper takes up these substantive and methodological issues by reporting on a data mining study of a text-message advising program intended to improve the college enrollment rates of U.S. high school students.

Purpose of the Study

The research reported here is an investigation into the content of text messages from a large-scale, randomized controlled trial of a national college advising intervention involving 75,000 college-intending students: Digital Messaging to Improve College Enrollment and Success (DIMES). DIMES was intended to influence college application behavior and enrollment outcomes as measured by the difference between treatment and control groups at the end of the trial. However, quantitative impact measures are insufficient to understand what is happening within the intervention program. What kinds of topics do students raise with advisors through text message? Is it possible for students to establish relationships with advisors via text message? Are student needs and concerns sufficiently similar in content and timing that advising could be automated?

Given these questions, the goals of this study are to describe the type of content of text-message advising interactions and to illustrate the use of text mining as an analytic method for large bodies of text data such as DIMES. Specific research questions are:

1. What is the content of student text messages in text-message college advising?

2. What is the nature of the student/advisor relationship in text-message advising?

3. What kinds of variation in topics and timing appear in student texts?

Study Intervention

"Digital Messaging to Improve College Enrollment and Success" (DIMES) was a 15-month college access advising program that was conducted entirely by text message. In the DIMES program, professional advisors used two-way text-message advising to provide individualized assistance to a treatment group of 31,408 college-intending students from 745 high schools in 15 states. Funded by the U.S. Department of Education Institute for Education Sciences (IES), the RCT was conducted by university-based education researchers. As a partner in the study, the College Board recruited students during PSAT test-taking and provided administrative data about DIMES participants. The advising itself was delivered by uAspire, a national nonprofit organization that specializes in college and financial aid advising for low-income students. A text-messaging platform provider, Signal Vine, was the final partner, providing the technology for sending, receiving, and storing automated and personalized text messages. (See Avery et al., 2020, for a full description of the sample and intervention design.)

Through the partnership with the College Board, 75,000 students from 15 states signed up to participate in DIMES in Spring 2015 at the point of taking the PSAT as high school juniors. The 745 DIMES high schools had substantial proportions of students who were eligible for free and reduced lunch (Mean=63%) and low two- and four-year college-going rates (Mean=26% and 30%, respectively). The College Board included an invitation to receive text-message advising in PSAT registration materials for all of the PSAT test-takers in these schools. Students who signed up to participate were randomly assigned to either a treatment condition, consisting of 15 months of two-way advising, or a control group that received only automated messages over the program period.

The findings reported here come from analysis of the national study's treatment group, in which 31,408 students were assigned to a specific professional advisor from uAspire. Advising began in April (2015) of students' junior year of high school and ended at the end of August (2016) after their senior year. During this period, students received text messages on their cell phone marked with their individually-assigned advisor's name as the sender. Advisees

Table 1.
April 2015 to August 2016 DIMES Message Topics, Sample, and Timing

Message # and n	Start Date	Message Topic
1 n=6012	4/9/15	Introduction of program and individual advisor
2 n=8631	4/23	SAT (spring) – registration, preparation & resources
Extra	5/12	I am a human; counter possible misconception that texts are fully automated
3 n=7364	6/1	College search guidance; start on initial list of possible colleges under consideration
4 n=6560	6/30	College affordability; understanding how to pay for college
5 n=5784	8/4	SAT (fall) – registration; taking or retaking
6 n=6458	8/27	Applications overview and planning; fee waivers
7 n=9313	9/21	Deadlines and finalizing college lists of where to apply
8 n=6839	10/14	Application assistance
9 n=3810	11/5	Paying for college; information and normalizing concerns about affordability
10 n=3608	12/ 2	FASFA and financial aid preparation
11 n=4705	1/7/16	FASFA aid and state aid application assistance
12 n=5764	2/2	Financial aid deadlines
13 n=6497	3/1	Finishing aid applications
14 n=7012	3/29	Financial award letter: interpretation; comparisons
15 n=6584	4/26	Pre-enrollment decisions and tasks
16 n=4388	5/24	College accounts and summer tasks
17 n=4380	6/21	College bills; financial aid and initial bill concerns

Note: n indicates the number of students who texted to their advisor at least once during specified message flow period.

answered or initiated communication with their advisor using the same number and advisor name, as in typical two-way text messaging.

Texting was organized into 18 "message flows" of roughly a month each. Each message flow began with one to two automated "broadcast" messages, sent by the uAspire advisor, that focused on a particular topic such as taking college entrance examinations, deciding where to apply, completing applications, applying for financial aid, choosing where to attend, and carrying out matriculation tasks. Broadcast message topics were pegged to the calendar of college and financial aid decisions and tasks. Following the standardized broadcast message, students and advisors exchanged individualized two-way text messaging. A message flow consisted of the initial outgoing broadcast message and any student or advisor text messages exchanged before the next one. Altogether, the DIMES program yielded close to a million automated and non-automated messages.

Table 1 shows the message flow number, beginning date, and focal topic for each month's initial outgoing text message.

In order to encourage two-way discussions on the focal topic for the month, the majority of the automated DIMES program messages were phrased as questions that asked students to reply via text message. For example, students received the following broadcast message, marked with their advisor's name as sender, that asked them to text back the word *Yes* or *No* (Message 8):

> Hi [advisee first name], I wanted to see how your college applications are going. Have you started working on them? Reply Yes or No

In this example, one more round of automation followed a student response:

> *Response to Yes:* That's awesome. What questions can I answer about the application process?
>
> *Response to No:* That's ok, this is a good time to start working on them. Do you know which college application you want to start with?

After each initial outgoing message and any automated follow-up replies, uAspire advisors and advisees exchanged personalized text messages. Students were invited to initiate contact with their advisor and had the option to ask questions or request help by texting on any topic at any time. Advisors encouraged student responses by posing questions, following up on previous conversations, and checking in with advisees on their progress in completing tasks, solving problems, or making decisions.

Method

This study is among the first to pioneer the use of data mining methods to examine the content of a large body of unstructured college advising texts: approximately 342,000 text messages that students sent to their advisors over the 15 month period of DIMES. Text mining is a form of data mining that can "turn text into numbers" (Miner et al., 2012, p. 30) by employing algorithms to uncover themes and identify relationships among themes in participant responses. Text mining is especially useful when working with large-scale data sets where it is impractical to follow conventional qualitative methods that require manual inspection and coding of text segments (Zilvinskis & Michalski, 2016). The need to conduct this kind of large-scale, quasi-qualitative analysis will become increasingly common in education as virtual interventions and internet-enabled digital content produce big data in the form of unstructured text (Fesler et al., 2019; Fischer et al., 2020).

Crucially, text mining permits a researcher to preserve and analyze participants' perspective in their own words, embracing a constructivist epistemology typically absent from large scale data analysis (Lewis, 2020). In addition to feasibility and access to the student voice, text mining was ideal for analyzing the DIMES student message dataset because it allowed for the identification of themes and topics in our large body of unstructured text. This paper features results from a lexical (deductive, researcher-supervised) analysis of student DIMES texts based on a categorization dictionary (Miner et al., 2012) that we validated with the results of a machine-specified (inductive, researcher-unsupervised) algorithmic analysis of the same data. (See Fesler, 2020) for an example of an additional text-mining strategy, supervised machine learning, that we did not employ in this study.)

Because data mining is just beginning to appear in education studies of college access and completion, we assume that most readers are unfamiliar with text mining. For this reason, as well as to document our analytic process, we therefore introduce text mining concepts and describe the procedures we followed in some detail.

Analytical Framework for Text Mining

Data analysis followed procedures in the Cross-Industry Standard Process for Data Mining, or CRISP-DM (Miner et al., 2012). The CRISP-DM covers all activities related to data mining and was therefore an appropriate choice for this study. As Miner et al. (2012) recommend, we conceptualized the process of text mining as three sequential sets of activities: first, establish the corpus; second, preprocess the data; and third, extract the knowledge from the data. Figure 1 summarizes the steps we followed.[1]

Establishing the Corpus

A corpus refers to a collection of documents. The full set of DIMES program text messages included approximately 875,000 automated broadcast messages, manual advisor texts to

[1] We employed a commercially-available data mining software program, WordStat, to conduct the analysis (Provalis, 2016). Many text mining researchers use natural language processing (NLP) packages in R.

students, and student texts to advisors. For the study reported here, the corpus consisted of the 342,192 individual text messages that students sent from their cell phone to their virtual advisors. All of the texts were captured and saved in the Signal Vine online texting platform.

Preprocessing the Data

Preprocessing refers to a host of activities that happen behind the scenes of the text mining software that prepare the data for knowledge extraction (Miner et al., 2012). During tokenization, for example, the software recognizes distinct words (or tokens) among all characters included in the corpus, usually by identifying punctuation marks and space between words. We employed lemmatization to identify and modify words that are related to one another but appear in different grammatical forms (Ignatow & Mihalcea, 2017; Miner et al., 2012). For example, the text-mining program automatically reduced the words *applies*, *applied*, or *applying* to the simpler form *apply*. This reduction in the number of distinct terms increased the frequency that some words appeared across the corpus, allowing for more intelligible and comprehensive coverage of the student text-message corpus. We also employed a stop-word list, which removed from analysis words commonly found in natural language (e.g., articles, conjunctions, prepositions, and pronouns) or words like DIMES advisors' names that had little substantive relevance to the research questions.

Figure 1

Text-Mining Procedures

STEP 1 **Data corpus establishment**	Download 873,192 text messages from Signal Vine Separate text into three categories: automated program broadcast messages, student responses, and manual advisor responses Import 342,192 student text message responses, sorted by program message number, into WordStat text mining software
STEP 2 **Data pre-processing**	Identify distinct words (tokenize) Shorten related words to common root form (lemmatize) Revise stop-word list (exclusion dictionary)
STEP 3 **Lexical analysis (deductive)**	Identify most-frequent words and phrases (univariate frequency analysis) Create include-word list by category (categorization dictionary) Confirm correct classification (keywords-in-context inspection) Specify rules for words with multiple meanings (disambiguation)
STEP 4 **Unsupervised analysis (inductive)**	Extract topics via principal components analysis; determine significance of topics using scree plots and inspection of keywords-in-context (feature extraction) Examine relationships among words and phrases with dendrograms and network graphs (cluster analysis)

Deductive Lexical Analysis

This final phase of text mining included the full range of our deductive and inductive analytic procedures. We carried out the deductive lexical analysis by establishing and refining a categorization dictionary in which we grouped topically-related words and phrases iteratively into folders, and then grouped folders with similar content together into larger categories. Once we assigned a particular word or phrase to a category, the software program automatically assigned all subsequent appearances of that word or phrase to the same category. As in traditional qualitative coding, a given piece of text could be assigned to multiple categories. For instance, a text message about difficulty accessing tax information for the FAFSA federal financial aid application from a non-custodial parent would likely contain words or phrases assigned to the categories of "financial aid," "parents," and "problems."

We defined preliminary categories on the basis of DIMES advising topics and themes that uAspire advisors identified in a series of focus groups. In the focus groups, we asked advisors to describe the questions, issues, and problems that their students were bringing up in advising, along with any typical language that students used to express these topics. The themes and associated words and phrases that advisors identified included, for example, students writing about funding college with text like "*cost, pay for, afford*." Such words and phrases became part of a "Financial" dictionary category. In a more nuanced example, advisors shared examples of language that they had come to understand was signaling about whether or not a student understood an explanation or piece of advice. Based on advisors' input about how students communicated their degree of understanding, we created the dictionary category "Clarity" that indicated the degree of vagueness or certainty in a student text, "*I think so, OK…, kinda; OH! Got it.*"

We compiled advisors' observations into an initial set of categories by hand-coding the focus group transcripts into interview themes and associated keywords and phrases. The uAspire research director and project manager reviewed the initial categories and associated words for face validity and suggested minor changes and additions. At that point, we created a preliminary categorization dictionary of text-message topics and associated words and phrases.

Next, we inspected unsupervised (machine-generated) frequency lists in which the text-mining software automatically created tables with the words and phrases used most frequently by participants.[2] Retaining the initial classification from the uAspire focus groups, we then began grouping the student text language from the frequency table of words and phrases into our pre-existing

[2] In recognition of the fact that raw frequency of a word or phrase does not alone constitute its importance, we followed the conventional weighting method known as term frequency-inverse document frequency (TF-IDF). Specifically, TF-IDF "operates under the assumption that words that appear frequently should receive higher weight unless they also appear frequently across all documents" (Lewis, 2020, p. 236).

initial categories and subcategories. This process was analogous to typical qualitative coding; here the text being coded consisted of all instances of an intelligible word or phrase. For example, references to specific colleges were assigned to the single category "college names"; terms having to do with feelings or stress were grouped under "emotions"; and designations representing examinations (ACT, SAT, scores, placement testing) were labeled as "testing." Categories that were more complex required subcategories. For instance, the large "financial" category included the subcategories of financial aid, cost, aid deadlines, scholarships, and fee waivers, each containing its own set of words and phrases.

As the process continued, we refined the content of category folders and occasionally revised dictionary folder names and subcategory locations. To verify that we were interpreting a word or phrase correctly, we referred to keywords-in-context tables that showed the word or phrase embedded in its surrounding text. This enabled us to disambiguate words with multiple meanings and words whose meaning shifted depending on context (Ignatow & Mihalcea, 2017). Once the range of word usage was established in such cases, we crafted rules using Boolean characters and phrases in order to classify the word or phrase correctly. For instance, we specified that a phrase be classified as a question when the word *can* occurred directly before *I*, *You*, *U*, or *We*, or within a specified number of characters away from a question mark.

Revisions to the Categorization Dictionary

When the software succeeded in classifying approximately 60% of the students' words in our initial categorization scheme, we began relabeling some categories and combining them into overarching themes. This process is similar to axial coding in standard qualitative analysis. At this point, two college access scholars—a school counseling researcher and a higher education researcher—conducted an in-depth review of the draft dictionary based on:

1) degree of coherence and independence of each axial category and associated subcategories;

2) correspondence of the categories with the research and theoretical literature on college access; and

3) representation in the categories of the objectives of each of the DIMES program messages as well as any unanticipated content.

Additional movement of subcategories and associated words and phrases occurred in this expert review. Finally, members of the DIMES quantitative research team and the uAspire project leaders reviewed the final dictionary for face validity.

Table 2 presents the final categorization table themes and sub-themes for the messages. (The full categorization dictionary is available by request to the authors for inspection and to use in replication studies.) The final version of

Table 2.
Categorization Dictionary by Category with Sample Words and Phrases

Navigating Process
- •Deadlines/timing (Deadline_is, Waiting, Soon, Right_now)
- •Application process/tasks (Started, Common-App*, Fill*)
- •College list
  - -Decision criteria (First_choice, Size, Close_to_home)
  - -Institution type (HBCU, For_Profit)
- •Admission cycle (Early_Action, Binding, Rolling)
- •Eligibility (Qualify, Acceptance_rate, Class_rank)
- •Essay (Essay*, Personal_Statement)
- •Interview (Interview*)
- •Online processes (Website, Login, Password, Portal)
- •Recommendations (Recommendation*, Recs, Reqs)

Financial
- •Aid (FAFSA, CSS, SAR, Finan*)
- •Cost (Bill, Free, Pay, Price)
- •Deadlines (FAFSA_Deadline)
- •Scholarships (Any_Scholarsh*, Apply_for_Scholarsh*)
- •Waivers (Fee_waiv*)
- •Taxes and IRS (IRS, Tax*, Tax_return)
- *Grants (Pell, Pell_Grant, Cal_Grant)

Personalization
- •Counseling/personal issues (Options, Transfer, Career)
- •Explanation for situation (Because, Meant, Wrong)
- •Fit (Fit, Good_fit)
- •Judgments re better/best (Best, Better)
- •Questions (Question*, Ask, Should_I, Whats, Wondering)
- •Special Status (DACA, TPS, IEP, Dream_Act, No_SSN)
- •Wants and interests (Really_want, What_I want, Don't_want)

Clarity
- •Certainty (Decided, Sure, Definit*)
- •Uncertainty (Don't_know, IDK, Not_decided)
- •Understanding or realization (Oh!, Oh_OK)
- •Vagueness (Guess, Not_really_sure, Kinda)

Influencers
- •Community
  - -Organizations (College_Board, Questbridge)
  - -Non-DIMES Advisor (Advisor)
- •Family (Mom, Parent*, My_famil*)
- •High School (Teacher, Counselor*, Guidance_counselor)

Problems and Concerns
- •Barriers (Can't, Bad, Cannot, Forgot)
- •Can't find (Trouble_finding, Couldn*_find)
- •Confusion (Mak*_sure, Don't_understand)

Relational
- •Advising process (Remind, Contact_you)
- •Affirmed/encouraged (Awesome, Cool, Sounds_great)
- •Appreciating help (Thank*, Thanks_so_much)
- •Humor and informality (Haha, LOL, Woohoo, XDC)

College Names

College Programs (Math, Major, Want_to_Major)

Emotions (Feel, Stress*, Anxi*, Hope)

Testing (ACT, SAT, Retake, Subject_test, Test_scores)

Military (AFA, ROTC, Airforce)

Note: Asterisks show truncation, enabling retrieval of different forms of a word.

the categorization dictionary that emerged from the process described here successfully classified 80% of the non-excluded words and phrases in the body of student text messages for the October 2015 message flow (Message 8). This ratio represents a satisfactory metric identified in content analysis work using similar methods (Bengston & Xu, 1995). When we applied the same dictionary to the student texts from later DIMES message flows, some additional words and phrases (such as *taxes* or *college bill*) had appeared in response to new admission tasks. After adding these words and phrases to existing categories, we finalized the categorization dictionary and applied it to each DIMES message after removing the irrelevant ("stop-list") words. The final dictionary performed in all 18 messages at or within a few percentage points of the desired threshold of 80% classification of all student words and phrases.

Inductive analysis

In order to validate the categorization dictionary, we conducted a separate, unsupervised analysis and compared the extracted topics with the categorization dictionary topics and sub-topics. We began this analysis with an automatic calculation of the frequency of words and phrases for each message with the categorization dictionary disabled. Using this tool, we established baseline descriptive statistics, including the raw frequency of each word and the percent of student cases in which the most frequent words appeared.

Next, we used feature extraction, in which unsupervised algorithms extract topics through a principal components analysis of clusters of words and phrases. Feature extraction uses a computer-generated matrix of all unique words in rows and all participants in columns to "extract underlying or 'latent' dimensions that capture most information contained in the full data matrix" (Miner et al., 2012, p. 942). In this study, principal components analysis resulted in the extraction of 60 linear combinations of words with a factor loading of at least 0.4 in each message flow. Multiple factor loading is acceptable in text analysis because words are used in a variety of contexts, each of which may constitute a valid theme. For example, the words and phrases that suggested a student was feeling certain or uncertain (e.g., *decided*, *don't know*) might appear in conversations covering many substantive topics, such as whether to retake standardized tests, consider additional colleges, apply for specific scholarships, or submit an early decision application.

Following recommendations by Cattell (1966), we examined a scree plot and retained components above the inflection point for analysis. In cases where the primary factors were of little substantive interest to the research questions (e.g., college names), we employed Kaiser's (1960) criterion, examining all factors with an eigenvalue of at least 1.0 and retaining those with practical significance for analysis.

Our final inductive strategy was cluster analysis, also referred to as concept extraction or topic modeling. This technique produces machine-generated hierarchical grouping of words and phrases found near one another most frequently (Miner et al., 2012). In this study, the WordStat software performed a hierarchical clustering algorithm using Jaccard's coefficient to group words iteratively based on their similarity. The program depicted the clustered words in the form of a tree graph, or dendrogram, and produced network graphs indicating the strength of association among words in any unique cluster.

Validation of the Categorization Dictionary

Access to complementary deductive and inductive analysis strategies is one of the strengths of text mining. In this study, we used our set of inductive analysis results to determine the validity of the categorization dictionary by comparing the cluster of keywords that appeared in the machine-generated topics extraction function with the researcher-defined groupings in the categorization dictionary. The group of co-occurring keywords that made up each principal component topic was considered to match the corresponding category in the categorization dictionary when the actual keywords were identical or when the combined keyword set was a clear conceptual fit with the categorization dictionary top-level category or sub-category.

Using both eigenvalue and frequency ordering, we used this method of comparing principal components topics and categorization dictionary topics for each of the 18 DIMES messages. The machine-generated topics and researcher-generated categories matched for 95% of the words in the student text corpus. This high level of correspondence between the principal components analysis derived by unsupervised algorithms and the researcher-generated categorization dictionary led us to conclude that the categorization dictionary is a valid coding representation of the student text content. As described, we derived the categorization dictionary logic empirically from advisor focus group results and conceptually from the literature on college access. It is more parsimonious and more readily interpretable than topics extraction results. For these reasons, we present our findings according to the themes and topics in the categorization dictionary.

Findings

Following our research questions, we begin the presentation of results from the text mining analysis by classifying student text content by topic. Next we consider evidence from the text messages bearing on the advisor/advisee relationship. We then report on individualization and temporal variability in advisees' text-message topics. We conclude the section with an example of a text-message advising conversation that illustrates and summarizes the findings.

Topics in Student Texts

The results of our deductive text mining analysis using the categorization dictionary appear in Table 3 and Table 4. These tables show the percentage of student cases that included text material in each of the top-level topics in the categorization dictionary. As described earlier, each of the 18 DIMES program message flows consisted of an outgoing broadcast message from the uAspire advisor along with any text messages that the student or advisor sent during the period before the next broadcast message. In interpreting the values in the tables, it is important to reiterate that the case percentages within categories represent the categories of message content among the subgroup of students who responded within that specific message flow.

Values in Table 3 represent the percentage of students who referred at least once to a particular category in any of their text messages within the specified message flow. For example, only 3.5% of advisees included the name of a specific college in at least one text message they sent to their advisor during the first message flow in which advisors introduced themselves and the texting program. In contrast, 53% of advisees mentioned a specific college by name in at least one text during the third message flow, which began with a broadcast message about choosing where to apply.

Viewing the data by semester helps reveal patterns in advisee texts over the course of the academic calendar and admission cycle. Table 4 summarizes the mean percentage of student cases with text material by topic categories in each of the four different time periods of the DIMES program: the spring semester of students' junior year of high school, the first and second semesters of their senior year, and the post-high school summer. Again illustrating with the category of "College Names," Table 4 shows that an average of 30% of student text messages across the entire DIMES program included the name of at least one college.

Text Content

The categorization dictionary results in Tables 3 and 4 indicate that participating students' text messages to their advisors were dominated by instrumental issues related to understanding and carrying out college and financial aid tasks. Looking across the message flows, it is clear that discussions of specific colleges and issues related to navigating processes comprised the most frequent substantive topics in the student texts. College names were a top content category in 12 of the 18 message flows, reflecting individualized advising content. References to specific colleges were most heavily represented in student texts during the senior year of high school. Texts in this category fell off over the summer, after the typical college admission cycle was complete.

Texts about "navigation" made up the second-most prominent category throughout the students' senior year and constituted the most frequent topic in the post-graduation summer. As shown in the dictionary (Table 2),

Table 3.
Percent of Student Cases with Text-Message Content by Categorization Dictionary Topic

Message # and Topic	College Names	Navigation	Relational	College majors	Financial	Influencer	Personalization	Clarity	Problems	Testing	Emotions
1. Intro	3.54	11.62	18.83	2.7	5.13	6.33	8.13	4.54	3.63	15.04	1.06
2. SAT-Spring	14.84	13.47	14.74	6.99	2.93	6.69	8.43	5.81	3.69	24.61	1.13
3. College search	53.17	21.71	25.92	22.44	6.75	16.18	17.69	14	7.25	2.12	4.45
4. Affordability	47.13	18.1	24.52	4.67	11.16	11.65	12.95	9.56	7.4	6.08	2.23
5. SAT-Fall	13.84	23.67	24.2	5.88	9.4	11.62	13.86	11.34	9.76	11.84	3.15
6. How apply	44.63	22.85	19.94	4.51	8.7	11.69	10.93	11.73	9.71	4.41	1.76
7. Applicatn list	52.3	21.83	19.2	3.14	5.17	9.12	9.73	8.36	5.7	2.2	0.81
8. Applicatn help	32.78	31.37	23.98	4.52	13.21	13.62	15.65	11.57	9.87	4.15	1.67
9. How pay	16.46	33.78	34.19	4.09	35.46	20.33	20.68	15.69	12.37	1.99	2.24
10. Fin. aid prep	11.53	34.15	23.76	2.04	29.58	17.47	17.64	11.62	10.71	1.37	1.22
11. FAFSA tasks	24.44	29.76	20.54	1.51	22.59	16.97	13.5	10.96	11.58	0.8	1.51
12. Aid forms/ tasks	30.54	25.44	17.19	1.3	19.82	15.07	11.03	9.95	8.97	0.86	1.19
13. Aid deadlines	23.65	25.35	16.73	1.49	19.47	12.22	10.93	11.41	11.08	1.0	1.5
14. Aid offers	44.34	25.54	17.66	3.54	16.94	13.16	12.67	12.81	10.79	1.94	1.07
15. Coll. decision	45.03	27.35	19.64	4.6	17.59	14.37	12.85	13.67	12.18	2.58	1.35
16. Pre-enroll	22.18	26.07	17.9	4.77	17.38	14.13	12.08	14.61	12.08	3.06	1.07
17. College bill	21.55	35.35	25.22	8.63	32.31	22.99	20.59	18.12	17.04	5.08	1.94
18: End and eval	14.09	32.99	13.19	-	31.97	23.62	16.73	3.4	2.98	5.37	3.82

Note: "Military" category not shown (% cases with military category 0 to .44%) Dates of message flows: Spring 2015-high school junior (#1-5); Fall 2015-first semester senior year (#6-10); Spring 2016-second semester senior year (#11-16); Summer 2016 (#17-18)

Table 4.
Mean Percentage of Responding Students' Message Content by Topic and Time

Topic	Total mean % of cases for all messages	Mean % of cases for Spring/ Summer 2015 messages	Mean % of cases for Fall 2015 messages	Mean % of cases for Spring 2016 messages	Mean % of cases for Summer 2016 messages
College Names	28.7	26.5	31.5	31.7	17.8
Navigation	25.6	16.2	28.8	26.6	34.2
Relational	21.0	21.6	24.2	18.3	19.2
College Majors	17.6	8.5	3.7	2.9	0.0
Financial	17.0	7.1	18.4	19.0	32.1
Influencer	14.3	10.5	14.4	14.3	23.3
Personalization	13.7	12.2	14.9	12.2	18.7
Clarity	11.1	9.1	11.8	12.2	10.8
Problems	9.3	6.3	9.7	11.1	10.0
Testing	5.3	11.9	2.8	1.7	5.2
Emotions	3.7	2.4	1.5	1.3	2.9
Military	0.12	0.17	0.14	0.06	0.10

Note: Dates of message flows: Spring 2015-high school junior (#1-5); Fall 2015-first semester senior year (#6-10); Spring 2016-second semester senior year (#11-16); Summer 2016 (#17-18).

the navigating processes category included student text-message content referring to deadlines and timing, applications, college lists, admission cycles, eligibility, essays, interviews, recommendations, and online actions. Student texts to their advisors referred to deadlines, eligibility, and online actions across multiple areas such as testing, college applications, financial aid, and enrollment tasks.

Financial issues were another prominent category in student texts. Although appearing throughout the advising program, explicit text language about college costs and financial aid became one of the most frequent student issues for texters beginning in November of students' senior year of high school and continuing through the remainder of DIMES.

Inspection of the extent and longitudinal patterns of less-prominent categories captures additional information about the content of text-message advising. For instance, the personalization category was well represented in texts to advisors across the DIMES period, with up to 20% of student texts including content about the advisee's personal circumstances. Personal situations and questions were highest in texts regarding the college search, paying for college, filling out the FAFSA, and paying the first bill. Parents (a subgroup in the category of "influencers") were highly represented in the texts during November, when paying for college was the topic, and again in June when students were facing financial decisions and processes related to enrollment. The share of student cases texting about problems and concerns was highest during the last half of the advising program, peaking in Message 17 with content from responding students about dealing with insufficient financial aid and paying the first college bill.

Advising Relationships

About one in five advisees (21%) across all messages included expressions of warmth, humor, and appreciation in their text messages that point to at least some degree of interpersonal connection with their advisors (Table 4). In responding to the introductory message from their DIMES advisor (Message 1), 19% of students included at least one instance of a relational word or phrase in a text message, constituting the most frequent category in this introductory message flow. Interestingly, text messages from the subgroup of students who became frequent DIMES texters included a higher percentage of relational content in the very first message, in comparison to texts from students who sent rare or occasional texts over the program period.

Relational material was most highly represented in the texts that students sent their advisors in the first semester of their senior year (Fall 2015, 24% of cases) as they worked with advisors to decide where to apply, complete college applications, and begin financial aid processes. Relational content in texts declined in the final semester of high school, perhaps because the advising relationship was already established. Alternatively it is possible that students were more exclusively task-oriented at this point in the year as they faced school, admissions, and financial aid deadlines. Supporting this premise, relational language became more prominent in the first message of the summer (Message 17, Table 3) in which responding students who should presumably have been locked into their college choice by the national May 1 response date were still deciding which college to attend and struggling to pay the first college bill.

Needs and Problems in Student Texts

Review of keywords-in-context for the most frequent topics illustrates the kinds of questions and concerns that students communicated to advisors. The texts show relatively little advising that rises to the level of counseling, although there were some instances of students asking about what kind of college or program or career might be best for them. In keeping with the emphasis on navigating processes, transactional, task-oriented discussions predominated. There was a preponderance of messages including phrases like "when is the deadline?" and "How do I/Can I/Can I still…?" "and "What do I do about…?" and "How do I find out about…?" Students asked questions about the meaning of terms and concepts. These ranged from very basic questions, "How does financial aid work?" to specific questions about their own situation, "what if my sister is going to college this fall will my financial aid be less?" Some of the conversation was around texting logistics, figuring out when the student could text back a response or get the answer to an advisor's question. "I'm about 10 min away to finishing the application i just need my parents to sit down with me and help me finish it."

Overall, the set of DIMES texts showed that the students who chose to participate found college and financial aid processes to be opaque, complicated, and difficult. Texts show students learning about, discussing, and frequently misunderstanding the steps and processes in testing, college search, applications, financial aid, and enrollment**.** DIMES advisees rarely mentioned school counselors in their texts, and then almost exclusively in connection with getting fee waivers and transcripts.

Advisees continued to ask substantive questions about college application, financial aid, and enrollment tasks through the final message in the summer after high school graduation. Fully one-third of students responding after the typical college admission cycle ended (Message 17) had text content in the "financial" category, which was the highest incidence of that category across any message. Keywords-in-context inspection showed financial concerns at that point in time were divided among students who were struggling to locate or interpret their financial aid packages and those with insufficient aid to cover their costs.

In sum, students showed considerable confusion about the college topics covered in DIMES advising. Their messages detailed obstacles and problems that students experienced in all major parts of the college and financial aid application process.

Variability in Student Texts
Categorization results (Table 3) and subsequent inspection of keywords-in-context tables suggest that the students who texted their advisors in a given message flow were generally responsive to the intended focal issue of each message flow. Message 3 about college search, for instance, elicited student discussion about possible majors and yielded the highest percentage of students mentioning the name of particular colleges. The most frequent responses to the college affordability message flow (Message 4) were about starting at community college and transferring in order to save money. Messages devoted to FAFSA completion (Messages 10-12) prompted student texts about how to fill out the FAFSA and reports of having submitted it. In this sense, student texts can be characterized by common topics and timing as well as shared gaps in knowledge, misunderstandings, and problems.

As Table 3 shows, however, student texts in a given message flow covered a wide variety of issues. Within each category, students posed off-topic or off-time questions, communicated about multiple issues within given categories, and demonstrated different levels of understanding and sophistication. In the initial message introducing the study for example, 5% of the student texts referenced financial issues and 6% mentioned the College Board (the dominant keyword-in-context for the "influencer" category in that message flow). At the enrollment stage, text messages were spread out around many issues, including decisions about what college to attend, specific family issues, and problems accessing online materials. In the final message flow, some students were seeking advice about matriculation issues like orientation and class registration at the same time that other students were messaging their advisors with questions about how to begin applying for financial aid.

In addition to varied topics and timing, students demonstrated different levels of understanding and sophistication in their messages. In the same message, for instance, one student used the term "super scoring" to ask about how colleges handle multiple SAT scores while another student asked, "what is the benefit of taking the SAT and what is the score range of passing and not passing?"

Illustrating DIMES: An Example

An example of an advising conversation illustrates the key findings about the nature and content of DIMES text-message exchanges. The series of back-and-forth texts shown below took place over a week between a highly-engaged advisee and her uAspire advisor in mid-September of the student's senior year of high school. The following exchange, quoted with verbatim spelling and capitalization, occurred during the message flow that focused on putting together a college application list (Message 7).

Advising Conversation

Advisor (Broadcast message, delivered Thursday): Some students qualify for college application fee waivers, which makes it free to apply. have you talked to your counselor to see if you qualify?

Student: not yet

Advisor: ok np [no problem] application fee waivers can save you money (most colleges charge a fee to submit your application); do you know if you qualify for an sat or act waiver?

Student: nope i have no idea

Advisor: ok, then i would recommend meeting with your guidance counselor soon and asking them if you do qualify for an sat waiver, because if you do, you can use that waiver to "buy" college application fee waivers to apply to colleges for free. when do you think you can meet with your guidance counselor?

Student: i'm not sure whenever i get time things are already stressful haha

Advisor: i hear you, [student name]. the beginning of the year is a stressful time for a lot of students. and counselors, for that matter. do you think you'll be able to meet with them by the end of next week?

Student: I can try

Advisor: ok, sounds good. i'll follow up with you next week to see if you've been able to meet with your guidance counselor. by the way, how's your college search going?

Friday:

Student: i know i'm going to try and shoot for [open access local community college name]

Advisor: hey [student name]. ok, gotcha. the good news is that [community college name] accepts most students who apply and it has a very low sticker price, so i think you'll have no trouble getting in, and it could be pretty affordable too. is [community college name] the only school you're considering?

Student: yes for now

Advisor: ok. it's a good idea to apply to at least 4 colleges so you have some options to choose from when the time comes to make a final decision about where to go. what do you like about [community college name]?

Student: i'm not sure of any other community colleges

Advisor: gotcha. how did you decide only to apply to community colleges?

Student: it's cheaper plus i can get rid of my basic classes then transfer

Advisor: it's true that community colleges have a lower sticker price, but they usually don't have the same kind of resources to give a lot of financial aid to give to students that many 4 year schools have, so if you don't receive a lot of financial aid from the state or federal government through fafsa, you could end up paying more out of pocket than if you went to a 4 year school. does that make sense?

Student: yes that makes sense. so it's not always the best option?

Advisor: exactly, not always. and if you only apply to community colleges, you won't be able to compare the aid that you could get from 4 year schools. it's definitely still a good idea to apply to them as financial safety schools (meaning you can definitely get in and afford them), but i think you should at least [add] 3 or 4 4-year schools to your list sound good?

Student: yeah i'll start seriously looking at colleges. do you know any around [hometown] that have good psychology classes?

Advisor: [4-year college name] could be a good choice. psychology is one of the most popular majors there. here's their departments website: [URL]. i also used this site to find some more colleges near [student town], but it doesn't say anything about their psych programs [URL]

Student: [emoji] oh wow haha that sounds good! i'll check it out. i haven't really looked for colleges because i was so set on [community college name] but financially if i could [emoji] have a better price in the long run that would be good

Advisor: how about you give the site a shot tonight and i can connect with you tomorrow to see how it went!

Student: i don't have internet at my house yet. :(

Advisor: i'm sorry to hear that! can u check it out in school?

Student: [emoji] yeah i'll check it out at school haha yeah we are working on getting internet so it makes my senior year that much easier plus it's better for college to have internet haha

Advisor: very true! have a good weekend and i'll check back in monday!

Monday:

Advisor: hey [student name], how'd that link work out for you?

Student: i haven't had any time :(i haven't forgotten though

Advisor: no worries! do you think you'd be able to check it out today?

Student: tomorrow the latest

Advisor: ok, cool. i'll check in with you tomorrow afternoon to see if you were able to check out the link. here it is again: [URL]

Student: okay thank you!

Tuesday:

Advisor: hey [student name], did you get a chance to check out the link i sent you and add some schools to your list?

Student: yeah i have i don't really remember which ones but i made an account and added some and some colleges i thought i would be interested in

Wednesday:

Advisor: hey [student name], did you get a chance to meet with your guidance counselor yet this week?

Student: no not at all.. what do i need to ask them?

Advisor: the plan was to ask them about whether or not you qualify for an sat fee waiver so you can use it to "buy" college application fee waivers. also, that's great that you were able to make an account and add some schools on there. could you log in and text me the names of the schools you want to apply to?

Student: yes i'll go to guidance tomorrow morning and ask and yes i'll tell you the colleges tomorrow and i'll try to add some more

Advisor: ok, great. looking forward to hearing what you find!

Thursday:

Student: [emoji] so i do qualify for free lunches so i can get the sat free waver [*sic]* i just need to go back at lunch and get the form. [emoji] i have [3 college names] for my colleges so far

Advisor: this is great [student name]! thanks for getting back to me with this info. are you pretty confident that you'll be able to get into each of these schools? if you're in school, please don't text back unless you're allowed to be on your phone. thanks!

Student: not each of them but hopefully a good amount plus i'm going to the college fair on the 22nd

Advisor: gotcha. honestly, i think you shouldn't have too much trouble getting into [college name]. it might be worth checking how your gpa compares to the gpas of other students who were admitted to those schools and see how they stack up. you can do that on big future [URL]: you can search the name of your colleges in the search bar. click on their page, then click the applying tab on the left. there should be info on gpa ranges in the subsection in the middle. does that all make sense?

Student: yeah i just don't know my gpa

Advisor: gotcha. do you mean you just don't know your gpa on a 4.0 scale, or you don't know your hs average on a 100 scale?

Student: both haha

Advisor: gotcha. is there a way you can find out? for example, would it be possible to meet with your guidance counselor to find out your average

Student: I can ask tomorrow maybe

Advisor: ok, sounds good. i'll text you tomorrow to follow up and see if you were able to meet with your guidance counselor

Student: okay sounds good

Friday:

Student: so i have the form for the free waver thing for the sat and i have my gpa now too

Advisor: ok great! What's your gpa? Also, do you plan to use your waiver to register for the sat, or just "buy" college application fee waivers?

Student: i don't know what to do with the waver

Advisor: ok, so the waiver has a code on it. You can enter that during registration for the sat to be free, or, if you don't plan to take the sat, you can call the college board at [800 number] to request application fee waivers using your sat fee waiver. Does that make sense?

Student: it's all kind of confusing

Advisor: i hear you, [student name]. let's start at the very basic: are you planning on registering for the sat?

Student: i don't want to i need to though

Advisor: i get that, [student name]. standardized tests like the sat aren't fun, colleges like [college name] require them, so if you want to apply there, you will need to take it. The next sat is [date], and since you have the waiver, if you register by [day and date] registration is free. You can register online here: [URL]

This conversation illustrates many of the central themes in the text mining analysis. First, the student appears to be using the advising to become aware of, understand, and carry out specific processes, tasks, and decisions: obtaining and using fee waivers, finding out her GPA, deciding whether a college entrance examination is necessary, registering for the SAT, and coming up with a college list. The function of text-message advising for this student is primarily instrumental and transactional. However, the student also reveals some personal information, like the lack of internet at home. She shares her feelings about feeling stressed and not wanting to take the SAT. In another indication of emotional content, she appears to include emojis and exclamation points in the text messages when she has accomplished a task and achieved a goal like obtaining fee waivers from her counselor or choosing additional colleges for her application list. She uses some relational language (*sounds good, haha*). In another indication of a building relationship, she responds with longer texts as the exchanges progress. She expresses degrees of uncertainty and understanding (*don't know; kind of confusing; yes, that makes sense*). Although the explicit focus of this particular text exchange was choosing a college list, the student still needs personalized help with the "off-topic" issues of testing and fee waivers.

The advisee clearly needed very basic information about the net cost of different kinds of colleges, her eligibility for admission, SAT testing requirements and procedures, and fee waivers. To address these needs, the advisor conveyed specific information, checked for understanding, prompted the student for particular actions, and followed up on whether the student had carried out these actions. The advisor remained focused on the broader topic of choosing where to apply but individualized the text-message advising according to the advisee's particular situation and stage in the application process.

For instance, the advisor gently debunked the student's assumption that community college would be the only feasible financial option. The advisor also picked up that the student needed to know whether and how to sign up for the SAT before being able to act on information about using fee waivers.

Finally, it is worth noting that this student was highly engaged, remaining in contact with the advisor and carrying out the actions that the advisor suggested. Less-engaged students might not have shared enough with their advisor to receive appropriately customized advice, failed to carry out the advisor's suggested actions, or ignored follow-up texts.

Discussion

Text mining analysis of the nearly 350,000 student text messages provides a methodologically rigorous look at the nature, content, and variability of text-message advising conversations between advisors and students who attended high schools with high percentages of low-income students and low college-going rates. Results carry implications in three areas: technologically-delivered advising modes; student college access needs; and data mining methodology.

Text-Message Mode of Advising

What is the Nature of Advising in this Mode?
Deep conversations and counseling interactions are infrequent in this medium. Instead, data mining results clearly indicate that students use text-message advising to address concrete, practical issues. Information, logistics, troubleshooting, and responses to nudges for action (Thaler & Sunstein, 2008) comprise the bulk of the content in this form of advising. This finding suggests that text-message advising is particularly well suited to helping students understand and carry out specific college enrollment choices, tasks, and decisions. Combining this mode with in-person advising is a potentially promising model. More research is needed that studies such a blended approach or investigates direct comparisons of in-person and remote advising. Virtual advising could also be expanded beyond text messages with the addition of phone calls, screen sharing conversations, in-person events, workshops, webinars, and videos. These complementary remote modalities would presumably be particularly useful for addressing complicated advising issues as well as helping students fill out forms and interpret documents.

Can Students Establish a Relationship With a College Advisor via Text Message?
Text messages that reach students on their cell phones seem to be a feasible way to deliver personalized college advising for students. A significant group of text-message advisees ask questions, raise individual issues, and use language that indicates a relationship with their advisor. Establishing a relationship over text messages can clearly occur for students who engage with their advisors. Relational

content is plentiful but not ubiquitous in student texts, however. It is important to note that the majority of students who participated in DIMES exchanged texts with their advisors occasionally or rarely.[3] Nearly a quarter of the students who had not opted out of DIMES never texted with their assigned advisor. Increasing the engagement of students who sign up for text-message advising programs is vital to tap the potential of virtual advising for moving the needle on college outcomes more broadly. It is likely that students will be more apt to participate in virtual advising when the advisor is someone they already know. Alternatively, engagement might be improved when someone that students know and trust provides a warm hand-off to a text-message advisor from an outside organization (Bird et al., 2019).

Can Text-Message Advising be Automated?
DIMES text-message advising introduced specific topics that were timed according to the college application and financial aid calendar. Students were responsive to these topics; however, text mining results show variability in advising conversations that indicate the presence of substantial individualization. Relational language and conversations about students' particular situations indicate that engaged students were using advising to get tailored assistance. According to their questions and comments, students' advising needs ranged widely. Students consistently brought up topics that were unrelated to the focus of a particular advising program message flow. These off-time topics indicate what immediate questions and concerns the student had at a given time and underscore nuances in their individual circumstances that can make advising less generic. In sum, because data mining indicates that participating students use text-message advising for individualized situations and timing, DIMES study results support two-way advising delivered by a human advisor.

The promising results of a fully-automated intervention at Georgia State University in reducing summer melt would seem to contradict this conclusion. The Georgia State POUNCE program uses artificial intelligence "chatbots" that draw from other data sources to provide tailored answers and referrals to would-be incoming students in the summer before beginning college (Page & Gehlbach, 2017). It seems unlikely, however, that algorithms can produce the kind of advising that is required in a longitudinal, comprehensive college intervention that attempts to provide assistance with all of the steps in the college-going process (Klasik, 2012).

[3] All 31,408 students were divided into engagement groups through k-means cluster solution on the basis of the number of message flows to which students responded and the number of characters they texted back to their advisors over the course of the entire advising program. Averages are within message flows for all messages where student sent at least one text to their advisor: High engagers (3% of all students, sending their advisor an average of 9.6 texts and 475 text-message characters); Medium engagers (21%, 3.6 texts and 97 characters); Low engagers (52%, 1.9 texts and 31 characters); and Never-engaged (24%).

Text Mining

This study is one of the very first to use text mining methods to examine the content of college advising for a large, nationally representative sample of students from majority low-income high schools. Text mining procedures, presented in detail earlier, can be used by evaluators and researchers to investigate the needs and characteristics of groups of advisees. In contrast to the typical researcher coding of unstructured qualitative texts (Saldaña, 2015), the approach offers a rigorous method for analysis of text messages for large samples of students. Text mining offers the scale and replicability of positivist statistical methods while including respondent voices as in qualitative methods (Lewis, 2020). For these reasons, this method is ideal for informing funding and policy decisions. Researchers have access to an increasing number of software programs for conducting text mining. As Fischer et al. (2020) suggest, taking advantage of big data in education requires incorporating data mining training for texts and other types of information in graduate curricula and collaborating with computer science and other campus data scientists.

In this study, text mining was used to describe the content of text messages and to investigate variability in the timing of student topics. Specialized categorization dictionaries are available for focused examinations, such as linguistic, opinion, or sentiment analysis (Redhu et al., 2018). Text mining can also be used for probabilistic analysis, such as predicting college enrollment or success based on student text content. Fesler (2020), for instance, used supervised machine learning techniques to predict which text-message advising interactions led to productive student engagement in the form of student responses and reported action. Data mining has significant potential for the analysis of student social media and other large bodies of unstructured text. In particular, text mining studies of college access can draw from bodies of data such as publicly-available social media communications related to college and financial aid, college essays, or new or reanalyzed sets of interviews and other written accounts.

Conclusion

Remote advising delivered through technology offers new modalities to reach large numbers of students who are likely to need assistance in choosing, applying to, and paying for college. With this goal, over 30,000 college-intending high school students from schools with high percentages of low-income students were offered DIMES advising in the form of two-way text messaging with a trained advisor. The resulting data set of advising messages enabled the use of data mining methods to examine the content of text-message advising for a large sample of college-intending students.

The study demonstrates that it is possible for students to use individualized text-message advising to build a relationship with a counselor, learn about colleges, and receive help with college choice, application, and financial aid tasks and decisions. If virtual advising intervention engagement rates remain low, however, stand-alone text-messaging programs like DIMES are unlikely to move the needle in eliminating socioeconomic gaps in college-going.

The DIMES message analysis method and results present a definitive picture of the challenges faced by students who engage with text-message advisors. Text mining is a relatively new methodology for education that offers a rigorous, replicable method for analyzing the kind of large, unstructured bodies of words that text-message advising can generate and preserve. The detailed picture of student needs for information and assistance that resulted from mining DIMES advisee texts can inform in-person and blended virtual/face-to-face advising models. In brief, students' text messages to their advisors showed extensive needs for assistance in understanding and carrying out tasks related to college admission and financial aid. Advisees showed considerable confusion about the processes related to college applications, financial aid processes, accepting admissions and aid offers, and preparing to matriculate. They encountered obstacles and problems in multiple aspects of these processes, completed tasks in a non-linear fashion, and faced challenging individual circumstances. For all of these reasons, students require individualized, two-way advising. Regardless of the advising mode, study findings clearly suggest that virtual advisors need to do more than present information and advice. Instead, advisors need repeatedly and proactively to explain, translate, and unpack terms and concepts. They need to double-check students' comprehension, provide repeated prompts for action, and confirm that students have followed up on tasks in an appropriate and timely manner.

"...study findings clearly suggest that virtual advisors need to do more than present information and advice. Instead, advisors need repeatedly and proactively to explain, translate, and unpack terms and concepts."

DIMES contributes to knowledge about what students need and what best practices in college advising might entail. Analysis of the "black box" actual content of virtual advising, such as described in this article, can be used to improve the design of future interventions, thereby increasing the likelihood that redesigned advising campaigns will produce positive treatment effects at scale.

However, the study also shines a light on systemic problems that are at the heart of the continuing socioeconomic gap in college access. Getting to college, this study shows, is complicated, opaque, and difficult for students in high schools with a concentration of low-income students and low college-going rates. It is unlikely that any intervention can fully overcome these structural barriers

(Deil-Amen & Rios-Aguillar, 2014). Both improved advising and policy changes are needed to help this large population of students succeed in negotiating the college-going process.

References

Arnold, K.D., Chewning, A., Castleman, B., & Page, L. (2015). Advisor and student experiences of summer support for college-intending, low-income high school graduates. *Journal of College Access, 1*, 4-28.

Avery, C., Castleman, B.L., Hurwitz, M., Long, B.T., & Page, L.C. (2020). *Digital messaging to improve college enrollment and success* (NBER Working Paper No. 27897). National Bureau of Economic Research.

Avery, C., Howell, J. S., & Page, L. (2014). A review of the role of college counseling, coaching, and mentoring on students' postsecondary outcomes. *College Board*, 1–15. https://files.eric.ed.gov/fulltext/ED556468.pdf

Bailey, M. J. & Dynarski, S. M. (2011). Inequality in postsecondary education. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 117-132). Russell Sage Foundation.

Bengston, D. N., & Xu, Z. (1995). *Changing national forest values: A content analysis*. (Research Paper NC-323). St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station.

Bergman, P., Denning, J. T., & Manoli, D. (2019). Is information enough? The effect of information about education tax benefits on student outcomes. *Journal of Policy Analysis and Management*, *38*(3), 706–731. https://doi.org/10.1002/pam.22131

Bettinger, E. P., Castleman, B. L., & Mabel, Z. (2019). *Finishing the last lap: Experimental evidence on strategies to increase college completion for students at Risk of late departure* (working paper) https://scholar.harvard.edu/files/zmabel/files/finishing_the_last_lap_-_working_paper_2-27-19.pdf

Bird, K. A., Castleman, B. L., Denning, J. T., Goodman, J., Lamberton, C., & Rosinger, K. O. (2019). *Nudging at scale: Experimental evidence from FAFSA completion campaigns* (NBER Working Paper No. 26158). National Bureau of Economic Research. https://www.nber.org/papers/w26158

Carrell, S., & Sacerdote, B. (2017). Why do college-going interventions work? *American Economic Journal: Applied Economics*, *9*(3), 124–151. https://doi.org/10.1257/app.20150530

Castleman, B. L. (2015). *The 160-character solution: How text messaging and other behavioral strategies can improve education*. Baltimore, MD: Johns Hopkins Press.

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, *115*, 144-160.

Castleman, B. L, & Page, L. C. (2016). Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence. *Journal of Human Resources, 31*(51), 389-415.

Cattell, R. B. (1966). The scree test for number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Chetty, R., Friedman, J.N., Saez, E., Turner, N., & Yagan, D. (2017). *Mobility report cards: The role of colleges in intergenerational mobility*. (Working Paper No. 23618). National Bureau of Economic Research.

Deil-Amen, R., & Rios-Aguilar, C. (2014). From FAFSA to Facebook: The role of technology in navigating the financial aid process. In A. P. Kelly & S. Goldrick-Rab (Eds.), *Reinventing financial aid: Charting a new course to college affordability* (pp. 75–100). Harvard Education Press.

Engberg, M. E., & Wolniak, G. C. (2010). Examining the effects of high school contexts on postsecondary enrollment. *Research in Higher Education, 51*, 132–153

Engberg, M. E., & Wolniak, G. C. (2014). An examination of the moderating effects of the high school socioeconomic context on college enrollment. *The High School Journal*, *97* (4), 240-263.

Fesler, L. (2020). *Opening the black box of college counseling* (CEPA Working Paper No. No.20-03). Retrieved from http://cepa.stanford.edu/wp20-03

Fesler, L., Dee, T., Baker, R., & Evans, B. (2019). Text as data methods for education research. *Journal of Research on Educational Effectiveness*, *12*(4), 707-727.

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130-160.

Gurantz, O., Howell, J., Hurwitz, M., Larson, C., Pender, M., White, B. (in press). A national-level informational experiment to promote enrollment in selective colleges. *Journal of Policy Analysis and Management*.

Gurantz, O., Pender, M., Mabel, Z., Larson, C., & Bettinger, E. (2019). *Virtual advising for high-achieving high school students*. EdWorkingPaper No. 19-126. Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-126

Hyman, J. (2019). Can light-touch college-going interventions make a difference? Evidence from a statewide experiment in Michigan. *Journal of Policy Analysis and Management*, *39*(1), 159-190. https://doi.org/10.1002/pam.22155

Ignatow, G. & Mihalcea, R. (2017). *Text mining: A guidebook for the social sciences*. Thousand Oaks, CA: SAGE.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Klasik, D. (2012). The college application gauntlet: A systematic analysis of the steps to four-year college enrollment. *Research in Higher Education*, *53*, 506–549. https://doi.org/10.1007/s11162-011-9242-3

Lenhart, A. (2015). Teens, social media & technology overview 2015. Pew Research Center. http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/fn-13190-2

Lewis, J.S. (2020). Amplifying participant voices through text mining. In E. Kennedy (Ed.) *Advancing Educational Research with Emerging Technology*. IGI-Global.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, *48*(3), 158–166. https://doi.org/10.3102/0013189X19832850

McKillip, M. E. M., Rawls, A., & Barry, C. (2012). Improving college access: A review of research on the role of high school counselors. *Professional School Counseling*, *16*(1). https://doi.org/10.1177/2156759x1201600106

Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., & Nisbet, R. A. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Waltham, MA: Academic Press.

Oreopoulos, P., Petronijevic, U., Logel, C., & Beattie, G. (2020). Improving non-academic student outcomes using online and text-message coaching. *Journal of Economic Behavior and Organization*, *171*, 342–360. https://doi.org/10.1016/j.jebo.2020.01.009

Page, L. C., Castleman, B. L., & Meyer, K. (2020). Customized nudging to improve FAFSA completion and income verification. *Educational Evaluation and Policy Analysis 42* (1), 3–21. https://doi.org/10.3102/0162373719876916

Page, L. C., & Gehlbach, H. (2017). How an artificially intelligent virtual assistant helps students navigate the road to college. *AERA Open*, *3*(4), 1-12. https://doi.org/10.1177/2332858417749220

Page, L.C., Sacerdote, B., Goldrick-Rab, S., & Castleman, B. (2019). *Financial aid nudges: A national experiment with informational interventions* (Working Paper No. 3). The Hope Center for College, Community, and Justice. https://hope4college.com/wp-content/uploads/2019/09/NPSAS-Nudging-Working-Paper-FINAL.pdf

Page, L. C., & Scott-Clayton, J. (2016). Improving college access in the United States: Barriers and policy responses. *Economics of Education Review*, *51*, 4–22. https://doi.org/10.1016/j.econedurev.2016.02.009

Palardy, G. J. (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal*, *50*(4), 714-754).

Perna, L.W., & Jones. A. (Eds.). (2013). *The state of college access and completion: Improving college success for students from underrepresented groups*. Routledge.

Perna, L. W., Rowan-Kenyon, H. T., Thomas, S. L., Bell, A., Anderson, R., & Li, C. (2008). The role of college counseling in shaping college opportunity: Variations across high schools. *Review of Higher Education*, *31*(2), 131–159. https://doi.org/10.1353/rhe.2007.0073

Phillips, M., & Reber, S. J. (2019). *Does virtual advising increase college enrollment? Evidence from a random assignment college access field experiment* (NBER Working Paper No. 26509). National Bureau of Economic Research.

Provalis Research (2016). *Wordstat 7 user's guide*. Montreal: Author.

Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership*, *70*(8), 10-16.

Redhu, S., Srivastava, S., Bansal, B., & Gupta, G. (2018). Sentiment analysis using text mining: A review. *International Journal on Data Science and Technology, 4*(2), 49-53. doi: 10.11648/j.ijdst.20180402.12

Roderick, M., Nagaoka, J., Coca, V., & Moeller, E. (2008). *From High School to the Future: Potholes on the Road to College. Research Report*. Consortium on Chicago School Research.

Saldaña, J. (2015). *The coding manual for qualitative researchers* (3rd ed.). Sage.

Shapiro, D., Ryu, M., Huie, F., Liu, Q., & Zheng, Y. (2019). *Completing college 2019 national report* (Signature Report No. 18). National Student Clearinghouse Research Center. https://nscresearchcenter.org/wp-content/uploads/Completions_Report_2019.pdf

Swail, W. S., & Perna, L. W. (2002). Pre-college outreach programs: A national perspective. In W. G. Tierney & L. S. Hagedorn (Eds.) *Increasing access to college: Extending possibilities for all students* (pp. 15-34). State University of New York Press.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press.

Willms, J. D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record, 112,* 1008–1037.

Zilvinskis, J. & Michalski, G. (2016). Mining text data: Making sense of what students tell us. *Association for Institutional Research Professional File, Fall* 2016(2), 15-24.