Qualities of “Bad” Teachers: A Natural Language Analysis of 300,000+ Student Descriptions of Teachers Rated Very Low

Authors

Teachers have always been evaluated. The performance of their job is ordinarily public, and these evaluations are also often quite public, too. Legend has it that judgments about the quality and performance of the teachers of Athens were posted in the Agora for the people to see. Socrates was judged to be a bad teacher. Apparently, he spent too much time asking his students to think. A walk-through evaluation by his supervisor determined that “sometimes Socrates’s students meander through endless dialogues examining challenging questions that do not have one right answer.” This led Jacobs (2012) to think that Socrates might be replaced, or perhaps be required to take an intensive summer professional development program in Sparta. This shows both the flaws of high stakes evaluations and the importance of feedback from his students!

The public nature of teaching invites school administrators, peers, students, and parents to evaluate the quality of teachers, although across different metrics. These evaluations can be formal, such as those used by administration, while others can be deeply personal, communal, and not shared like those evaluations done daily by students. Certainly, informal evaluations focus on different criteria, record qualities or accomplishments which are more difficult to measure and do not carry as much weight. For example, interpersonal relationships are high in importance to students (Uttio, 2012), as are fairness and being treated respectfully. Formal evaluations often put far less of a premium on these. And, since these are often subjective interpretations done by adolescents and young adults, they are often dismissed and may be seen as unreliable. Indeed, as we have sometimes found in our work, often one student’s worst teacher was another’s best. Despite these challenges, student observations of their teachers do demonstrate correlation to other forms of teacher effectiveness (Chaplin et al., 2014) and are somewhat stable for a given teacher year-to-year (Polikof, 2014).

Millions of evaluations happen in schools across the country by those students closest to the performances of their teachers. Despite the inherent variation in these students’ experiences and judgements, there is value in understanding and using feedback from such evaluations as complementary to the more formal criteria used for professional teachers. They may paint a “bigger picture” of teacher effectiveness and contribute to a more complex--and more meaningful--portrait of a given teacher. Several decades of studies already indicate that students have insight into how teachers perform and how it impacts them (see, for example, Rodin & Rodin, 1972; Check 1986, 1999; Uitto, 2012; Hosgorur, 2015; Raufelder et al., 2016; Chang-Kredl and Colianno, 2017). Yet, research on what they uncover remains scant (Polikof, 2014), and what does exist remains somewhat small in scope, with surveys or interviews done on dozens or a few hundred respondents.

In this study, we seek to increase the scale--and perhaps the volume--of these voices. We asked the following research questions: How do students describe and define “bad” teachers they think judge as being bad? How do student’s descriptions of “bad” teachers inform us about the practices, attitudes and beliefs of those teachers? How can these descriptions of bad teachers inform good teaching? We began by collecting more than 4.8 million publicly available student comments and ratings on the RateMyTeacher.com website and used natural language processing to identify common descriptors of teacher performances. We narrowed our focus to the worst rated teachers on the website specifically to uncover those characteristics students deemed the worst that teachers may possess.

**Conceptual Framework**

*Good and Bad Teachers*

There is recognition for “good” teachers, if not actual awards, because they are certainly known by reputation in the school and often in the community. There is also recognition of “bad” teachers especially by reputation. “Bad” is a word with so many connotations that we hesitated, at first, to use it. It can easily be misconstrued and misused. But “bad” in the field of teaching, might be a small and vital category in need of being addressed forthrightly. Because two of the authors have spent a total of 100+ years learning to be good teachers, examining how to teach students to become great teachers, recognizing the excellence in teaching, researching about how to prepare and recognize good teachers and even being married to two terrific teachers, we think we have the credentials to credibly examine the reality and myth of a group we, like everyone else, heard about through the grapevine-bad teachers. In our experiences in schools, we have seen great teaching in PreK-20 and, also, bad teaching, but never wanted to affirm that often misleading and uninformed description of our colleagues. We have come to realize that in looking for bad teaching, as we observe and investigate, we will overwhelmingly inform the good to great teaching all around us. So, this study was born.

In digging deeper, we know that there is some question about how teacher quality can be defined (Goodwin, 2016). First we look, as a backdrop and point of comparison, how a “good” teacher is described in research. There is some consensus about the desired qualifications of a teacher. Starting with another study which looks at the qualities of a good teacher as described by students, 6th graders, Dorham (1987) says “Three distinct themes regarding the efficacy of teachers emerged from students' comments: (1) instruction; (2) personality; and (3) classroom management”(3). Good instruction means that teachers presented material in ways in which students could understand and they did it with patience and creativity. Personality was described mainly as being “nice,” “not yelling” and not looking as if they were bored. Classroom management again emphasizes, no yelling and also the responsibility of the teacher to intervene when necessary. Goodwin and Oyler (2008) are more specific as they talk about the most important quality any teacher has to have is content knowledge. Although, worthwhile content knowledge seems to be debatable and determined by teaching context. These researchers go on to list other required characteristics: language proficiency and fitness to teach. Hattie (2015) talks about teachers having to make the connection between student achievement and their teaching practices in addition and paying attention to what students are actually doing and that they are visibly learners themselves. Peneul & Shepard (2016) describe the Deweyin idea of essential skills as “seeing on the horizon the full mastery of disciplinary knowledge and practices and translating that into intermediate understandings and ways of participating connected to the experience of the learner” (787).

There are two specific studies about bad teachers which are more closely related to our study and which shed a light on our findings. In the first, Raufelder et al’s (2016), research had a similar goal to ours, but was based on a much smaller sample. These researchers questioned 86 German junior high school students about bad and good teachers. They then organized, as we do, the responses into themes and sub-themes. Although our analyses resulted in many more themes being identified, what we found does overlap with their research. Apparently bad teachers, as is true of good teachers, have common characteristics, and at least in the Western world, such characteristics are valued similarly.

From their interview data three prominent bad teacher themes were identified. The first was about the quality of the teacher student relationship. Three sub-themes were recognized. First was *relational aggression*. Teacher behaviors of this type included teacher yelling, and the teacher being insulting (e. g. calling students stupid). Vilification of the students was also categorized as relational aggression (e. g. showing little or no respect for what they had accomplished). Sub-theme two was labeled *injustice*. Bad teachers were seen as playing favorites in the classroom, or who had opaque and changing evaluative criteria. Sub-theme three was called *antipathy*. Antipathy by students toward their teachers developed out of a general dislike of the teacher. This theme developed because students found their teachers often to be incomprehensible in communicating subject matter content, or because their teachers really did not know the content.

Theme two also arose out of the coding of the interviews, and was based on students’ concerns about their teachers’ lack of expertise in teaching. One subtheme was the perception of students that their teachers were *disinterested or indifferent* to the materials being presented, which was disappointing to the students. A second subtheme was labeled *incomprehensible teaching*. It is quite easy to understand why students would be harsh judges of this perceived teacher characteristic. A third subtheme was labeled *teacher-centered instruction*. The students resented copying material, or seatwork, where in both cases, the teacher has it “easy” and the students do the instructional work.

Theme three consisted of comments about the *personal characteristics* of the teacher, for example, boring, repetitive, or disinterested. It included, as well, examples of teachers’ lack of assertiveness. Student examples of this negative trait included allowing whispering or talking among students that should have been stopped, newspaper reading in class, disrespect to the teacher that was uncriticized, students throwing things in class, etc.

In this study the students’ comments about their bad teachers were weighted more heavily by their teachers’ failings in terms of their interpersonal skills and abilities, more than by their academic skills. The analysis in our study supports that distinction, but, as will be seen, an unusually high number of criticisms of teachers in our data set were about the teachers’ subject matter content deficiencies and pedagogical abilities. In each study it is interesting to note that many students are judging their teachers to be “bad” on the basis of their inability to teach them as much as they want to learn.

In another venue, Chang-Kredl and Cloannino (2017) were interested in examining the image of teachers in the public sphere—movies, television, and on the web. Analyzed in this study were the descriptive comments made about the best and worst teachers encountered by subscribers to Reddit, a popular website for social news aggregation, web content ratings, and discussions. Characteristics of the sample are unknown, but consisted of individuals sufficiently motivated to post their comments on Reddit, often well after their personal encounters with “bad” teachers. Their analyses revealed many characteristics of bad teachers that mirror what we found in our data. Table 1 provides these data.

**INSERT TABLE 1 HERE**

Student comments like those in Table 1, and from our own analyses below, suggest that principals and other supervisors of classroom teachers are not paying close enough attention to life in the classrooms over which they have responsibility. These descriptions, comments, and complaints make clear that for some, or even many students in a class, “bad” teachers do exist, and in number, though certainly not in the numbers often touted.

*How many teachers are legitimately talked about as “bad?”*

So, a foundational question for this study, is, “How many teachers are we talking about?” In fact, the numbers of “bad” teachers of the type described in Table 1, and in our own work, actually appear to be rather low. Author (2014) estimates their numbers at about 3%. And in the well-respected Hechinger report, Butrymowicz (2014) says that states such as Tennessee, Michigan, Georgia, Florida, and Pennsylvania, particularly in Pittsburgh, all provided estimates of “bad” teachers in this same low range. Danielson (2016), who has visited and coded many hundreds of classrooms, estimates the “bad teacher” number to be around 6%. That seems to be the high end of estimates from those who are experienced classroom analysts. In our own study we found that, out of more than 4.8 million reviews, 55% gave a maximum rating of 100, 75% gave a rating of 80, and 89% gave a rating greater than 50, reflecting what those analysts found in an informal and voluntary review process.

*Student Evaluations of Teachers*

While informal and unscientific evaluations have been common throughout history, it is still difficult to find contemporary examples of teacher evaluation techniques that meet the standards for reliability and validity proffered by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (2014). For example, test based accountability systems, such as value-added models (VAMs) generally are unreliable year to year (Amrein-Beardsley, 2014), from subject matter to subject matter (Amrein-Beardsley & Collins, 2012), and even unreliable from class to class in the same subject and in the same school year (Newton, Darling-Hammond, Haertel, & Thomas, 2010; See also Konstantopoulos, 2014; and Author, 2014).

Consistent unreliability of this type strongly, and negatively, affects validity. Test-based accountability, especially VAMs-based accountability, ought to be avoided (Pivovarova, Amerein-Beardsley, and Broatch, 2016). Nevertheless, politicians and policy makers seem partial to test-based models of teacher evaluation, even when researchers point out that the amount of variance in student test scores that is attributable to their teachers is negligible (American Statistical Association, 2014). Demographic factors (family income, mothers’ level of education, abilities of the cohort that one goes to school with, etc.) are almost always the best predictors of performance on standardized tests of achievement: not teachers, nor schools (cf. Haertle, 2013).

Observational instruments to evaluate teachers have reliability problems as well, similarly limiting their validity. They almost always require more observers and more observation time than can be afforded by principals, peers, or the school systems that seek such data. Thus, their reliability is often questionable..Among their other drawbacks are the fact that observational instruments usually cover only a short period of teaching time, and cannot be trusted to be valid if consequential decisions about teachers are to be made on the basis of such limited observational data. Nevertheless, there are some that are commonly used and found by many educators to be useful in providing feedback to teachers (e.g. Danielson, 2008; Pianta, LaParo, & Harmre, 2008).

Furthermore, the results of test-based teacher accountability methods and observational methods of accountability are not substantially correlated. For example, in the multi-million dollar MET study, funded by the Bill and Melinda Gates Foundation (Kane, McCaffrey, Miller, & Staiger, 2013) four different observation instruments were correlated with the VAMs associated with math achievement test scores. Those correlations were .12, .18, .25, and .34., averaging about .22. With the VAMS derived from reading and language arts tests the observation instruments correlated .12, .11, and .09, averaging about .11 (Bill and Melinda Gates Foundation, 2012). A separate study using this data set found that the correlations between an observational measure of excellence in teaching, and two measures of excellence in teaching derived from VAMs, were trivial: .16 and .09, respectively (Grossman, Cohen, Ronfeldt, & Brown, 2014). Strunk, Weinstein, & Makkonen (2014) correlated observational data and VAMs for reading and math, over one year. They found correlations under .216. Similarly, Morgan, Hodge, Trepinski, and Anderson (2014), found correlations between observations of teachers and their pupil’s performance on tests that were roughly between .20 and .40, indicating, once again, that these two different measures of teacher competence have in common only between 4% and 16% of the variance observed. The latter investigators noted, additionally, that neither teacher performance in classrooms, nor teacher effectiveness as judged by test scores, were highly stable over multiple years of the study.

Since the variance in common between test-based accountability measures and observational measures is the square of the correlation coefficients just cited, these two methods of evaluating teachers are not measuring the same thing at all. They measure different constructs, or perhaps different aspects of what is sought. Each of these approaches to evaluation has problems: The test-based accountability systems do not look at teachers’ classroom behavior, and the observational systems do not assess learning outcomes. And neither has access to teachers’ thinking, which determines both their classroom behavior and, indirectly, their students’ likelihood of scoring well on tests. Although there is a public aspect to teaching, and students test scores provide important artifacts associated with classroom teaching, much of the most important part of the teachers’ job is cognitive. Thus, it is unobservable. Teachers make a number of decisions per day that cannot be easily captured from observational instruments or via student test scores. Borko, Livingston and Shavelson (1990) estimated that teachers make at least .7 consequential decisions per minute, 42 per hour, over 250 per day. Jackson (1990), writing at about the same time, believed that teachers engage in 200 to 300 consequential exchanges with students every hour (between 1,200-1,800 a day!). Most of these are unplanned and unpredictable and the thoughts that are behind them are unknowable. Choosing between these two most common and equally flawed evaluation systems (test-based and observational evaluations) is akin to being between Scylla and Charybdis (Author, 2018). Problem-free teacher evaluation systems do not exist.

But these are just the two most common ways to assess teachers. There are other methods, each with their own strengths and weaknesses. Scriven (1994), for example, has proposed that teachers be rated on the basis of their performance of the essential duties. This “duties based” assessment has much to offer. Users can learn to use the system reliably in a short period of time, and its face validity is quite high. But duties-based evaluation systems are infrequently employed. It is thought that this form of evaluation is too removed from the heart of the teachers’ job, namely, interactive classroom teaching. Instead, a duties-based evaluation system focuses on other important aspects of the teaching job, such as showing up to class on time, giving students back papers with useful comments on them, communicating regularly with parents, and a host of other “duties” expected to be adequately fulfilled by teachers. This is an assessment system of important aspects of the teachers’ job—related to what happens in classrooms and on tests—but not directly assessing those factors. The correlation of duties-based evaluation systems with test-based or observational systems is unknown at this time, but it is likely to be low.

There is one other method occasionally used for evaluating teachers. It is by means of student evaluations of their teachers. Such evaluations are most likely to be used at the college level, where raters are thought to be mature enough to engage in this activity. Students are less likely to be used as evaluators in the K-12 system because of their purported immaturity. Such student evaluations of their teachers are likely not to correlate well with the other three measures. We think that is because all four approaches to evaluating teachers (test-based, observation-based, duties-based, and student-based rating systems) only deal with a piece of the teachers’ job but cannot adequately describe the overall qualities of “teacher.” In a sense, this limitation provides a contemporary example of the parable of the blind men and the elephant in which we can only tell a part of the whole through any one (or even several) evaluations. This fourth method discussed, like the other three, has advocates for its use in the K-12 system.

For example, Scriven (1995), offers nine reasons to consider student evaluations in a positive light:

1. The positive and statistically significant correlation of student ratings with learning gains.
2. The unique position and qualifications of the students in rating their own increased knowledge and comprehension.
3. The special situation of the students in rating changed motivation (a) toward the subject taught; perhaps also (b) toward a career associated with that subject; and perhaps also (c) with respect to a changed general attitude toward further learning in the subject area, or more generally.
4. The singular ability of the students to be able to rate observable matters of fact relevant to competent teaching, such as the punctuality of the instructor and the legibility of writing on the board.
5. The peculiar circumstances of the students in identifying the regular presence of teaching style indicators. Is the teacher enthusiastic; does he or she ask many questions, encourage questions from students, etc.?
6. Relatedly, students are in a good position to judge—although it is not quite a matter of simple observation—such matters as whether tests covered all the material of the course.
7. Students as consumers are likely to be able to report quite reliably to their peers on such matters of interest to them as the cost of the texts, the extent to which attendance is taken and weighted, and whether a great deal of homework is required--considerations that have little or no known bearing on the quality of instruction.
8. Student ratings represent participation in a process often represented as "democratic decision-making."
9. Students may be the "best available alternative" for learning about what goes on in some classrooms.

Similarly, Ripley (2010) and Cuban (2012) both make persuasive arguments for using student ratings and evaluations of teachers. Others, however, especially those in higher education (Lawrence, 2018) argue that the information obtained from student evaluations is invalid, and does more harm than good.

But we are not using this information to make compensation or tenure decisions, so in this paper we join with Scriven, Ripley, and others. We use student evaluations and their accompanying commentaries in the investigation of a particular subset of teachers, those that students (and some parents) had previously judged to be “bad” teachers.

*Definition of bad teacher.*

The working definition of “bad teacher” we developed as we reviewed the ratings is:

*One who, for whatever reasons, either cannot teach the intended curriculum (e. g. because of recurring inadequate preparation for their classes, lack of knowledge of the subject area, problems of classroom control, extensive absences) or, is harmful cognitively, physically, or socially to students or their families (e.g. displays prejudice, punishes excessively, criticizes in demeaning ways, or is uncommunicative or disrespectful of parents or colleagues) (cf.* Authors, 2018).

We used RateMyTeacher.com data to identify these teachers, and categorized the comments made by students that accompanied the ratings.

## 

**Data and Methodology**

The data we used for this study were reviews of teachers submitted through RateMyTeacher.com, a website where students and/or parents can submit a review of a teacher along with a rating from 0 to 5 stars (½ stars are allowed). In 2018, the website changed ownership, and the data collected previously was removed from the site as was a rater’s ability to submit written reviews. We scraped the data in early 2018, when all of the reviews submitted since 2001 were still visible on the site. The last review in our dataset was submitted on January 16th, 2018.

Reviews were attached to specific teachers and schools and were collected from 6 English speaking countries (United States, Canada, United Kingdom, Australia, New Zealand, and Ireland). We limited our data set to only those originating in the U.S. We collected 4,884,479 reviews from the U.S. Each review was accompanied by a rating from 0 to 5 stars. Some were one quarter or half filled, so we scaled the ratings to account for the partial scores, and used a 0-100 scale. Then we rated levels that separated the reviews into large, similarly sized groups. We used natural language processing methods, described below, which benefit from large sample sizes. The data were heavily skewed, with almost half of all reviews being 5-star ratings. This distributional skew held even when we only considered teachers with at least 50 reviews (about 7% of the total dataset). This study focused on a filtered sample from the 359,387 worst reviews rated 0-35 in our dataset.

*Preparing the Data*

Our first step was processing the text data so it could be analyzed by statistical models. The “cleaning” process involved removing stop words (e.g. “a”, “is”), coded characters (e.g. “\n”, “\r”), and infrequent words such as personal names. We removed conjugations and pluralization using the word lemmatizer from Python's Natural Language Toolkit (Bird, Klein & Loper, 2009). We then used Gensim for Python to identify phrases by using n-grams to identify words that co-occurred often enough to warrant a unique meaning (Röder, Booth, & Hinneburg, 2015). For example, if “laid” appeared next to “back” enough times, every instance of both words appearing in that order would be replaced by “laid\_back” such that it made a new, unique word. We allowed phrases of up to four words (i.e. “as\_soon\_as\_possible”). Cleaning the data was an iterative process, and certain misspellings, names, or nonsense words were only detectable when reviewing the output of the language models. We experimented with the parameters of the cleaning process and reported the results in the online supplemental resources. One experimental parameter that dramatically improved our model was a lower bound on review length. We found that limiting each review to a minimum of 100 characters dramatically increased the probability that any given review in the sample contained non-trivial information. Of the 359,387 “worst” reviews, 211,224 met the 100-character threshold.

*Natural Language Processing*

Ever since natural language processing became popular in the 1980’s a seminal criterion used to evaluate algorithms has been text classification through information retrieval (Lewis, 1992). Information retrieval refers to a language model’s ability to identifyand retrieve words, sentences, or paragraphs that are alike. When one does not have reliable labels that can be used to characterize documents (in this case, reviews), these are analyzed using unsupervised learning. This refers to statistical methods that cluster documents together based on how close they are to one another in the metric space created by a language model (Mikolov, Chen, Corrado & Dean, 2013). The words that most frequently appear in the cluster can be used to characterize it and these can be interpreted as topics (Papadimitriou, Raghaven, Tamaki & Vempala, 2000).

Our goal in this step was to understand which key words appeared as descriptions of the different topics in reviews rated 0-35 that were over 100 characters in length. We used a latent dirichlet allocation (LDA) model to generate clusters to derive topics from. In LDA, clustering works by randomly sorting the documents into K groups and then iteratively moving them around until the members of each cluster are closest to each other and furthest from members of other clusters. One hurdle in this process was that we had no idea what the true value of K was. Some algorithms come prepackaged to help determine that number (Teh, Jordan, Beal, & Blei, 2005) but we chose to use the concept of coherence to estimate an optimal number of topics.

*Measuring Topic Coherence*

Coherence metrics measured the spread (or concentration) and orthogonality (or mutual exclusivity) of topics. The logic behind using these metrics was that 1) clusters that were more spread-out would be less informative than clusters that were more dense and compact, and 2) clusters that overlapped significantly would be less informative (more redundant) than clusters that were mutually exclusive (Stevens, Kegelmeyer, Andrzejewski & Buttler, 2012; Mimno, Talley, Leenders, Wallach & McCullum, 2011). In our analysis we iteratively tested different numbers of topics (10 to 100 in steps of 15). We used Gensim to estimate coherence scores for the topics generated in each test. Several coherence measures exist, we chose the “Cv” measure derived in Röder et al (2015), in which they compared this rating to others from the literature and found that it yielded the results that were most highly correlated with human ratings. With this we arrived at a suggested number of topics.

Estimating coherence yielded a suggested number of topics as well as descriptive terms for the topics themselves, so we proceeded to the task of interpretation. The key words we used were the words that best predicted whether a review belonged to a given topic. Every word had a predictive coefficient for each topic, making it possible to rank the words that were most relevant to each topic. While these were useful, more informative still were the set of example reviews we drew for each topic. The same predictive model described above allowed us to predict the probability that any given review belonged to a topic. This resulted in a dataset that assigned each review topic percentages that could be interpreted as the presence of any given topic in that review. Using these, we identified the 50 reviews that were most exclusively descriptive of (or dominated by) a given topic and read them searching for an obvious theme. We combined these sample reviews with topic key words to create comprehensive descriptions of each topic. Finally, we visualized the results using the LDAvis method which illustrated the topic distributions by plotting them in a 2-dimensional space that illustrated both their share of the corpus and their overlap with one another (Sievert & Shirley, 2014). This provided us a more intuitive view of the language used to describe the lowest rated teachers.

*Worked Example*

The results section contains descriptions of all topics we discovered. However, to afford the reader insight on how we arrived at those topics we provide this example of the process used to derive each one.

Upon identifying the LDA model with the highest coherence score, we proceeded to examine two characterizations of each topic. The first was the topic key words. These provided a general idea of what the reviews in each topic discussed. However, similar topics were often hard to differentiate on key words alone. The second characterization was more helpful in differentiating similar reviews, it consisted of examining sample reviews from each topic.

Using the language model, each review was assigned a probability that they discussed a given topic. For example, a review could be assigned 40% Topic 8, 35% Topic 18, and 25% Topic 5. For each topic, we examined the reviews with the highest probability of belonging to them. Considering Topic 1 (see Table 3), we examined the reviews that had more than a 60% chance of belonging to Topic 1. Our goal was to identify a common thread that resonated with the key words assigned to the topic. Most reviews were predicted to belong to (or discuss) more than one topic. This yielded a right skewed distribution where a minority of reviews were assigned a probability of over 60% for any given topic. These were the most valuable reviews in this process because our intuition was that these almost exclusively discussed the topic of interest. Table 2 provides 5 example reviews that were predicted to belong to Topic 1 with a probability of at least 60%.

**INSERT TABLE 2 HERE**

These and other sample reviews that were highly predictive of belonging to Topic 1 made it clear that the students writing these reviews took issue with specific teacher-related obstacles to getting good grades. We summarized the students’ motivation and the examples they presented by assigning Topic 1 the title “Teacher is a harsh grader, assigns too much work with unrealistic timetables.”

## Limitations of the Study

This study is limited by the data collection process. The reviews were voluntarily submitted by students or (much less frequently) by parents. We did not know whether the students or parents were incentivized in some way. For example, a teacher may have offered extra credit to students for submitting a review or a parent may have been trying to get a teacher removed from a school or district. We also did not know the degree of bias in each review. For example, we did not know if a student was lashing out against a teacher for a bad grade they may have deserved, or if a student was pressured by classmates into submitting a good review despite feeling differently. Furthermore, in articles about RateMyTeacher.com and teacher forums, opinions were split. Some took offense at negative comments or were concerned about personal information being posted. Others appreciated it as an open forum for students and argued that site administrators had guidelines for removing innapropriate comments.

These sorts of uncertainties would have rendered a smaller dataset useless. Fortunately, the size of our dataset let us scrutinize some of these concerns. Our dataset contained over 4 million reviews, covering over 27,000 schools. Of the 128,344 teachers reviewed in our final sample:

* 68.5% received only 1 bad review,
* 17.6% received 2 bad reviews
* 13.9% received 3 or more bad reviews.

To put this in context, the average number of bad and total reviews for teachers in our sample was 1.6 and 11.3, respectively. Put another way:

* 11.5% of teachers in our sample received 100% bad reviews.
* 10% received 50% bad reviews.
* 42.3% of the teachers in our sample received at least 20% bad reviews.

These distributions were reflected in the average ratings for teachers. The average rating for the entire dataset was 92.8, where the average ratings for teachers in our sample was 79.8. Such a high rating made it clear that many teachers receiving bad reviews must have simultaneously received a much better rating and better reviews by some other classmates. This appeared to be especially true of bad teachers rated that way for having classroom favorites.

In all, the assumption that the entire dataset was completely biased seemed less plausible than the assumption that the dataset represented a diverse array of motives, contexts, and incentives for submitting reviews about teachers. Most teachers received mixed reviews, while a small minority received mostly bad reviews. Furthermore, the reviews for teachers were submitted across long periods of time (the average time between the first and last review for each teacher was 5 years). Thus, the topics discussed by reviewers in our sample likely represented common “flaws” that could be found in classrooms across the country, rather than the views of any one type of mythical reviewer.

In fact, the results of the analysis increased our confidence that this was the case. The results detailed the variety of ways in which students, primarily, expressed the behaviors they observed in teachers, and gave, as well, their thoughts about the origins of those behaviors. Furthermore, one finding in this study was that filtering reviews by length (100 characters) was essential for identifying consistent topics among the reviews. This enforcement of minimum length increased the proportion of substantial reviews that were analyzed and decreased the likelihood of including reviews that required no effort or a lack of sincere thought to write.

**Results**

Clustering the reviews into 25 topics resulted in the highest coherence score, 0.502. As context for this score, in Röder’s 2015 paper, where they explore the performance of several coherence measures on many different datasets, the most difficult dataset to evaluate for all measures was a “Movie” dataset originally used by Rosner, 2014 in which the “Cv” coherence score was .544. Given the frequent typos and informal nature of our data we considered 0.502 a good stopping point. As noted, we restricted the corpus to reviews with a length of 100 characters or more. The model used the words in each review to predict the probability that a given review was related to a given topic. Reviews were assigned topics based on the topic with the highest probability, the logic being that this was the topic most relevant to the words and syntax of the review. The resemblance of these procedures using language, to factor analysis in psychometrics, is obvious.

In this case 7.9% of reviews in the sample were predicted to discuss Topic 1. This means that if we picked a random review in that group there was a higher probability that it would be concerned with something related to that topic’s description more explicitly than any other. For example, if one review was predicted to discuss Topic 1 with a probability of 20%, and another review was predicted to discuss the same topic with a probability of 90%, each was *dominated* by Topic 1 as long as no other topic had a higher score in either review. We referred to those reviews as being dominated by Topic 1. Table 2 shows all 25 topics along with their descriptive titles, sorted by the percentage of reviews they dominated.

The estimates of topic dominance reflected that the model allowed for reviews to discuss more than one topic at a time. The topic descriptions we assigned, and example reviews, made it clearer still that topics were often linked to one another in concurrent (“they’re mean and ignore questions”) and/or explanatory (“I failed because they’re mean”) ways. Table 2 in the *Worked Example* section above showed an example of key words and sample reviews used to assign the title to Topic 1. A full table with the key words and three sample reviews for each of the 25 topics is presented in a modified format in the Appendix.

**INSERT TABLE 3 HERE**

## *Result Verification*

One topic that stood out was Topic 8 which appeared to capture reviews submitted by parents. Reviews submitted after August, 2015 sometimes had labels identifying whether they had been submitted by parents or students. We compared reviews dominated by Topic 8 to the “Submitted By” label in the data and confirmed that our model was able to accurately identify reviews submitted by parents. There were 4,448 parent reviews that were longer than 100 characters and rated between 0 and 35. Of these 4,448 reviews, 42% of them were dominated by Topic 8. To highlight this further, another 15% of the parent reviews had Topic 8 as the second most dominant topic. No other topic came close. Topic 1 was a distant second and dominated only 7% of the reviews submitted by parents.

The fact that our language model was able to accurately identify reviews submitted by parents increased our confidence that the language model we trained and selected yielded consistent and coherent topics.

# Discussion

When analyzing all 25 topics, we saw that five themes could be distilled which could characterize bad teachers. The similarities between topics became much clearer once we plotted the themes in two-dimensional bubble plots. Figure 1 shows the topic distribution in 2-dimensional space where each topic is a bubble, topic frequency and bubble size are proportional, and themes are highlighted and labeled “T#.” Many of them seemed completely common sense (and seem to align with the kinds of topics identified by Raufelder et al., 2016 and Chang-Kredl & Cloannino, 2017), but it was surprising that students noticed them and could describe them with such nuance. Students really did not like “wasting” their time by sitting in classrooms with ineffective or disrespectful teachers. It seems obvious that there would be an expectation by students that consideration of students should be first. Not at all unreasonably, they wanted good communication and pleasant relationships.

As a guide to examining and labeling these themes, we asked: 1) how these data represent students concerns, 2) what they tell us about their teachers, and 3) what they tell us about students, in general. In addition, we went into more detail about the 25 topics and how they shaped the 5 themes. We wanted to be sure that the connection between them was clear and that our logic in creating them was made apparent.

The following subsections present the analysis of each theme. The three criteria above organized the descriptions that fill the remainder of this section.

**INSERT TABLE 4 HERE**

Table 4 shows the topics that make up Themes 1-5 and several “Residual Topics.” The column on the left shows the topic number, the middle column displays the description of each topic and the column on the right shows the percentage of reviews dominated by that topic. The row highlighted in grey shows the theme label as well as the total percentage of reviews dominated by the topics that make up the theme. It is worth repeating that the “closeness” that determined which topics were grouped together into a single theme represented how often those topics appeared together in the same review.

## Theme 1 - Unable to Create a Productive Learning Environment

## The students who submitted these reviews were concerned about the lack of a productive learning environment. The topics covered two specific aspects of the learning environment. Topics 3, 9 & 23 reflect students’ concerns about not being able to receive the instruction necessary to understand class material and meet their goals. Topics 5, 6 & 7 reflect the students’ concerns that they would not be able to develop a relationship with the teacher to address any significant instructional concerns.

The descriptions of teachers ranged from professional to behavioral ineptitude. Professional flaws included a lack of focus, organization, and clarity. These descriptions of teachers may indicate that the teacher was overwhelmed which could be the cause of them acting out or refusing to help students. Also plausible is that teachers’ own immaturity, personal behavioral problems, or subject matter and pedagogical inadequacy could explain their inability to perceive students’ confusion or flaws in their own approach to instruction. Unsurprisingly, the expectations these teachers set for some of their students may have been out of touch or unreasonable.

These reviews reflected a self-awareness in students that could have been leveraged by the teachers who chose to listen. When teachers did not treat them with respect, students noticed and expressed how that got in the way of their learning through their reviews. Students perceived their own confusion but could not always recognize that their teacher was not meeting their goal of effectively teaching them. Given the sample, there is a chance that some teachers were doing a good job and that some of these students were among a small number who felt frustrated with their teachers’ attitude or approach. Even with this caveat, it is worth arguing that the teacher, particularly those who received complaints about their attitudes, were unable to identify these students and communicate with them effectively.

**Theme 2 - Bad Lessons that Waste Students’ Time**

The students who submitted these reviews were concerned that their classwork was not useful or constructive in terms of their success. Topics 4, 10 & 11 dealt with the perception by the students that little to nothing was being accomplished. Topic 2 highlighted that students’ main concern was the effect these classes had on evaluations (class tests, AP tests, SAT’s, etc.) and, in general, the irrelevance of evaluations given that classwork was so misaligned.

More than a description of teachers’ personal characteristics, these were descriptions of their work. Teachers described by these reviews seemed to fail in their planning, follow-through, or in their ability to evaluate and revise plans that did not make sense or stimulate their students. Most frequently, the teachers described by these reviews failed to help students connect classwork to the evaluations. Some teachers appear to be distracted too easily and their classwork became disconnected as a consequence. Others may simply lack the enthusiasm to make the class stimulating or consequential for students.

These reviews showed us that students were capable of recognizing poorly organized lessons. They may have also become frustrated when they felt the teacher got distracted or did not follow-through with their responsibilities or promises. Of special interest was that student comments suggested that they put great value in their own learning and how they spent their time in the classroom. These sentiments culminated in the reviews dominated by Topic 11 which described these classes as pointless and wastes of time. Given the reasoning above we named this theme “*Bad Lessons that Waste Students’ Time*.”

## Theme 3 - Difficult or Insensitive Teachers

The students who submitted these reviews were concerned with their grades and success. Topic 1, the most populated topic, described how students were worried that the workload and criteria in their courses were too severe. This was echoed by reviews dominated by Topics 13 & 21 in which students’ concerns revolved around an inability to get their questions answered and the help they needed in order to succeed.

The teachers described by these students may have been excessively demanding and insensitive to students’ needs. It is easy to give many of these teachers the benefit of the doubt as students with uneven workloads or poor time management skills may have been unable to deal with a reasonable but demanding workload. However, a healthy proportion of the reviews covered by these topics also describe teachers who were unwilling to answer questions or help those students who made it clear that they were struggling. Thus, these teachers may have lacked the sensitivity to gauge differences in student needs and, furthermore, they may have been insensitive to how that indifference exacerbated the stress put on students. This certainly appeared to be the case for teachers described by Topic 21 where reviews described teachers who did not even bother to learn their students’ names.

We found these topics to be most revealing of students. Topic 1 was the most dominant in the entire data set, revealing that students were primarily worried about success and how difficult it would be to achieve. Many students were willing to put in the work but became very frustrated when teachers showed a lack of sensitivity to the stress and demands that were imposed on them. Topic 17 aggregates reports which express hatred or dread, and includes stories about teacher insensitivity. The fact that it often appeared in comments which included Topics 1 & 13 supports the idea that students hold their own success as their highest priority. Teachers who are generally insensitive or thwart that powerful student motivation to succeed will be judged to be exceptionally bad. Given the reasoning above we named this theme “*Difficult or Insensitive Teacher*.” In comparison to Theme 1, where reviews discussed the obstacles to a productive learning environment, reviews dominated by topics in Theme 3 revolved around difficulty and often, directly tied the complaint to their own grades or success in the course.

## Theme 4 - Bad Leadership

The concerns represented in these reviews revolved around the students’ well-being. The main concern expressed was that the students were not being shown the respect they were due and that a hostile environment was prevalent in students’ experiences. Hostility was fostered by unfair, discriminatory, or otherwise biased behavior on behalf of the teacher (Topics 18 & 20). A kind of indifference was also perceived. These teachers and staff were seen by students and parents as uninterested in hearing any feedback about their work or behavior (Topics 8 & 16).

The teachers described in these reviews should probably not be either teachers or administrators. The harsh criticisms communicated in these reviews may have reflected the deep dissatisfaction with decisions educational staff make. Such decisions apparently and needlessly created a hostile environment. For those staff who were accurately described by this set of reviews, it seems clear that at some point their priorities became warped, and their own success and ego took precedence over their students’ needs.

These reviews reflect students’ social consciousness. They revealed that students could be very sensitive to how their teachers and principals’ actions altered the tightly woven social fabric of classrooms and schools. In particular, it reflected the expectations of students in the cultural and political systems that characterize classrooms and schools. They expected those who have been placed in charge to have their best interest in mind and when concerns arose they expected to be heard. Given the reasoning above we named this theme “*Bad Leadership*.”

## Theme 5 - Rigid Teachers from the Industrial Age of Teaching

The student concerns revealed in these reviews centered around the rigidity of some teachers in the classroom. The reviews took issue with the fact that there was little consideration of errors in grading, and that teachers did not engage in discussion or critical thinking. Topics 24 & 25 referred to rigidity in teachers’ instructional styles, whereas Topic 15 addressed other forms of inflexibility such as disapproval of specific political views or strict rules around dress code and the use of English in foreign language classes.

The teachers described by these reviews used methods that substantially limited the possibility of inspiring students. They appeared to be teachers who dogmatically adhered to the rules and promoted the required tasks, when perhaps they should have put more effort into engaging students dynamically. In some examples it was clear that these teachers were not familiar with the subject, seemed to have just learned the subject matter themselves, or were simply unprepared to teach a subject or curricula areas within that subject. Although these teachers possessed these salient faults, they also failed to help their students by teaching memorization techniques which can be framed as entertaining activities. Thus, it seems these teachers were lacking in time or enthusiasm. In addition, some teachers tried to dictate students’ attire and discourse, and this too harmed their chances at establishing a rapport with students. Regarding the latter, it is likely that teachers simply did not have the time or personality to address every form of student behavior in a manner that was seen as respectful by students. In sum, it is likely some of these teachers were overly committed to rules and regulations, and less committed to their students. Some descriptions were like Dickensian caricatures.

The students who submitted these reviews showed a desire to be trusted with their own exploration of ideas and behaviors. Their complaints were not outrageous. Rather, they were understandably dissatisfied with an environment that did not nurture their curiosity and instead prioritized the *appearance* of discipline and learning. It would be naive to dismiss these complaints as the typical intemperance of teenagers. The seriousness with which the criticisms were given suggests that it is appropriate to turn a critical eye on the teachers. Given the reasoning above, and one student’s insightful comment, we named this theme “*Rigid Teachers from the Industrial Age*.”

#### The Residual Topics (Topics 12, 19 & 22)

Unlike the underlying themes captured by other groups, these topics provided little in the way of an example or explanation. The most straight-forward was Topic 12 which simply described teachers as “the worst” and claimed that the teacher “cannot teach.” Topic 19 was a bit more substantial. Reviews dominated by Topic 19 described the teacher as a kind person that simply “cannot teach.” Whereas Topic 22 was a genuine residual topic, containing a variety of reviews that had idiosyncratic criticisms of specific teachers and also might have included positive language that either compared this bad teacher to good teachers, or were sarcastic in nature. Topic 22 also included some genuinely good reviews that were not perceivably sarcastic and may have been mistakenly submitted with a bad rating.

Despite their generality, the residual topics most overlapped with reviews dominated by Theme 1 (see Figure 1, T1 is closest to the residual topics). Since many reviews covering Theme 1 shared the language of “cannot explain,” “cannot learn,” and also “cannot teach,” and given that Topics 12 and 19 classified teachers as “the worst” or “good people but bad teachers,” topics 12 & 19 may be seen as additional qualifiers that suggest that students considered those concerns highlighted in Theme 1 as essential to teaching or their own learning.

#### A Special Note About Topic 14 - Student Forewarnings

One purpose of an open forum like RateMyTeacher.com where students and parents can describe their experiences with teachers is to inform future students (or others). This purpose was encompassed by Topic 14. Reviews dominated by this topic included direct warnings that fellow students who were considering classes taught by the teacher being reviewed should switch out. This topic overlapped most with Theme 1 and Theme 2 suggesting that students who took the time to criticize a teacher’s performance were most likely to post warnings to other students.

*Conclusion*

Our work elaborates on the research of Raufelder et al. (2016) and Chang-Kredl and Cloannino (2017), cited earlier. But as far as we know, the 25 topics that emerged from our analysis of ‘bad’ teachers’ behaviors provide the most complete compendium of the myriad ways that “bad” teachers can be described by students and, we assume, by anyone familiar with classroom teaching.

The six most frequent and salient of these topics, those with the highest likelihood of describing and encompassing the thousands of negative reviews of teachers that were proffered by students (and by a small number of parents) were: 1) The teacher is a harsh grader; assigns too much work with unrealistic timetables; 2) The teacher does not prepare students for a test or align classwork with tests; 3) The teacher provides confusing explanations or has a poor teaching style and holds unrealistic expectations; 4) The teacher talks too much; students don’t feel they are being taught; teacher talks about unrelated topics; 5) The teacher cannot “handle” students; cannot relate to them or develop good working rapport; and 6) The teacher is mean, intimidating, strict or never smiles; teacher is weird.

The descriptors used by the raters, and the topics in which we grouped them, provide valuable information to everyone with responsibility for the supervision, professional development, and evaluation of teachers. In particular, principals who note students talking about or describing a teacher using these descriptors should recognize that it may well be a signal that something about the classroom teacher, classroom practices, or classroom culture, is seriously amiss. Sadly, in their comments, students often wondered if anyone else noticed how awful these teachers or teacher practices were.

All 25 topics found were further analyzed using a statistical approach to visualize their distribution and see if a smaller set of themes could be found to represent the essential qualities, or basic substance, of the concept of “bad” teacher. Five such themes emerged from these students’ descriptions of their “bad” teachers. These we labelled: 1) Unable to create a productive learning environment; 2) Teaches bad lessons that waste students’ time; 3) Behavior is difficult or insensitive; 4) Displays bad leadership; and 5) Rigidity, as if coming from the industrial age.

Each theme was analyzed for what students appeared to want from their teachers, and also, for what seemed to be lacking in those teachers. We found that many students who voluntarily chose to comment about their teachers cared deeply about two particular things: *1) Achieving more in school, and 2) being treated fairly and with dignity*. These are not unreasonable requests, and would in all likelihood enhance student learning. Moreover, rating teachers as “bad” teachers because they stood in the way of student learning and, consequently, student achievement, is a finding that does not conform to the public’s image of contemporary K-12 students.

**Final Thoughts**

We found that almost half of these reviewers awarded the most positive rating that was possible. This should be reassuring to America’s educators. Although positive reviews were not the focus of our analyses, they serve as the parameters for bad teachers. We saw that the teachers characterized as “bad” teachers, were a numerically small number of teachers who are often the most salient in the concerns expressed by the public and the press. Of course, the goal is to improve their teaching or steer them toward another profession where they can be successful.

These data reveal that too many students feel dehumanized, and they judge many classroom practices to be ineffective. Nevertheless, their comments appeared to be more thoughtful, more sophisticated, and more reliable (other members of their class expressed similar opinions), than we had expected. This suggests that students can be critical and insightful evaluators of their teachers, *if they are asked*! They not only can describe the many “good” teachers and positive school cultures that exist, but are quite sensitive to school cultures where excellence in teaching is not expected, and where students are not the center of the work of the school. Because students’ demonstrated competence as observers and critics, we recommend that more districts and schools use students--those closest to daily classroom instruction-- to evaluate their teachers. Without being determinant of promotion and salary, students’ analyses appear quite capable of providing extraordinarily rich data about *their* lives in classrooms. We believe, as well, that with a little training, students could be even better prepared to provide both the positive and negative feedback that school administrators need to improve their schools. Thus, we urge further investigations into using student feedback in formative ways to improve instruction, and as summative evidence of school improvement.

Because many of the teacher behaviors or personal characteristics of teachers noted by their students can be remediated, the supervisor’s role should be to initiate such professional development swiftly. But it is also likely that some of the behaviors and attitudes displayed by teachers cannot be remediated easily, or in a cost-efficient way. If so, those teachers should be terminated. The descriptions of bad teachers we analyzed suggest that though their numbers may be small, there are, indeed, teachers working today that need to be removed from their classrooms. We also suggest that adults in the school pay attention to what students are saying. There are common descriptors which tell those who are listening, that “This is a bad teacher” who needs help.

Going back to Dorham (1987)who says that there are three distinct themes regarding the ratings of good teachers instruction, personality and classroom management), we see that these qualities were revisited by the students in this study. Again, administrators should ask teachers themselves and their students to operationalize them and be able to identify them in the classroom. Goodwin and Oyler (2008) say teachers have to have deep and understandable content knowledge, as did the students in our study. The research is available and should be used in the education and professional development of teachers and administrators.

**References**

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Statistical Association (2014, April 8). *ASA Statement on Using Value-Added Models for Educational Assessment*. Retrieved December 9, 2019 from: https://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf

Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. Philadelphia, PA: Routledge/Taylor and Francis Group.

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS® EVAAS®) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives*, 20 (12). retrieved December 9, 2019 from http://epaa.asu.edu/ojs/article/view/1096

Author (2018). Between Scylla and Charybdis: Reflections on and problems associated with the evaluation of teachers in an era of metrification. In María de Ibarrola (Ed.), *Key Issues in the Evaluation of Basic Education. Dialogues with the International Academy of Education*, México City, September 1-2, 2016.

Bill and Melinda Gates Foundation, (2012). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project*. Retrieved December 10, 2019 from https://docs.gatesfoundation.org/documents/preliminary-findings-research-paper.pdf

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. Sebastopol, CA., O'Reilly Media, Inc.,

Borko, H., Livingston, C. & Shavelson, R. J. (1990). Teachers' Thinking About Instruction. *Remedial and Special Education*, 11 (6), 40-49. Retrieved 12/10/2019 from<https://doi.org/10.1177/074193259001100609>

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology,* 3 (2), 77-101.

Butrymowicz, S. (2014, June 16). How many bad teachers are there? *The Hechinger Report.* Retrieved December 10, 2019 from <https://hechingerreport.org/many-bad-teachers/>

Chang-Kredl, S. & Cloannino, D. (2017). Constructing the image of the teacher on Reddit: Best and worst teachers. *Teaching and Teacher Education*, *64* (43-51). Retrieved December 10, 2019 from <http://dx.doi.org/10.1016/j.tate.2017.01.019>

Cruickshank, D. R., & Haefele, D. (2001). Good teachers, plural. *Educational Leadership,* 58 (5), 26-30.

Cuban, L. (2012). Students Evaluating Teachers. [Larry Cuban on School Reform and Classroom Practice](https://larrycuban.wordpress.com/). Retrieved December 10, 2019 from https://larrycuban.wordpress.com/2012/12/11/students-evaluating-teachers/

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Washington DC: Association for Supervision and Curriculum Practice.

Danielson, C. (2008). The Handbook for Enhancing Professional Practice: Using the Framework for Teaching in Your School. Alexandria, Virginia: Association for Supervision and Curriculum.

Dorham, J.K. (1987). [*Sixth Grade Students' Perceptions of Good Teachers*](https://search.lib.asu.edu/discovery/fulldisplay?docid=alma991037474109703841&context=L&vid=01ASU_INST:01ASU&lang=en&search_scope=MyInst_and_CI&adaptor=Local%20Search%20Engine&tab=Everything&query=any,contains,good%20teachers&offset=0)*.* S.l. : Distributed by ERIC Clearinghouse, Number: ED359164

Goodwin, L. (2016). Defining teacher quality. In Drew H. Gitomer & Courtney A. Bell (Eds.)

*Handbook of research on teacher education.* NY: Routledge, 399-403.

Goodwin, L. & Oyler, C. (2016). Teacher educators as gatekeepers. In In Drew H. Gitomer & Courtney A. Bell (Eds.) *Handbook of research on teacher education.* NY: Routledge, 468-489.

Grossman, P., Cohen, J., Ronfeldt, M, & Brown, L (2014). The Test Matters: The Relationship Between Classroom Observation Scores and Teacher Value-Added on Multiple Types of Assessment. *Educational Researcher,* 43(6), 293–303.

Gurl, T. J. J., Caraballo, L., Grey, L., Gunn, J. H., Gerwin, D. & Bembenutty, H. (2016). *Policy,professionalization, privatization, and performance assessment: Affordances and constraints for teacher education programs*. New York: Springer.

Haertel, E. H. (2013). Reliability and Validity of Inferences About Teachers Based on Student Test Scores, The 14th William H. Angoff Memorial Lecture, Princeton, NJ: Educational Testing Service. Retrieved December 10, 2019 from <https://www.ets.org/research/policy_research_reports/publications/publication/2013/jquq>

Hattie, J. (february, 2015). High Impact Leadership, *Educational Leadership*, Association for Supervision & Curriculum Development, 36-40.

Author. (2018) [Is the Teaching Profession Dominated by Bad Teachers? How Would We Know?](https://www.tcrecord.org/Content.asp?ContentId=22315) *Teachers College Record*, March 28, 2018. Retrieved March 15, 2019 from http://www.tcrecord.org/Content.asp?ContentId=22315.

Jackson, P. (1990). *Life in Classrooms*. New York: Teachers College Press.

Jacobs, H. H. (2012). Curriculum 21: Socrates Fails Teacher Evaluation. Retrieved December 10, 2019 from http://www.curriculum21.com/?s=Socrates

Kane, T. J., McCaffrey, D. F., Miller, T. & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* MET Project Research Paper. Retrieved May 4 2017 from<http://k12education.gatesfoundation.org/resource/have-we-identified-effective-teachers-validating-measures-of-effective-teaching-using-random-assignment/>

### Lawrence, J. W. (2018). Student evaluations of teaching are not valid. *Academe*, May-June 2018, American Association of University Professors. Retrieved December 10, 2019 from https://www.aaup.org/article/student-evaluations-teaching-are-not-valid

Lewis, D. D. (1992). Text representation for intelligent text retrieval: A classification-oriented view. In P. S. Jacobs (Ed). *Text-based intelligent systems: current research and practice in information extraction and retrieval (*pp. 179-197). Mahway, NJ: Lawrence Erlbaum.

Mimno, D., Talley, E., Leenders, M., Wallach, H. M., & McCullum, A. (2011). Optimizing semantic coherence in topic models (pp. 262-272). In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language processing*. Edinburgh, Scotland: Association for Computational Linguistics.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. Retrieved December 10, 2019 from *arXiv:1301.3781*.

Morgan, G. B., Hodge, K. J., Trepinksi, T. M., & Anderson, L. W. (2014). The stability of teacher performance and effectiveness: Implications for policies concerning teacher evaluation. Education Policy Analysis Archives, 22(95). <http://dx.doi.org/10.14507/epaa.v22n95.2014>

Morgan, G. B., Hodge, K. J., Trepinski, T. M., & Anderson, L. W. (2014). The Stability of Teacher Performance and Effectiveness: Implications for Policies Concerning Teacher Evaluation. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 22, 1-21. Retrieved December 10 2019 from https://www.redalyc.org/pdf/2750/275031898117.pdf

Newton, X.A., Darling-Hammond, L., Haertel, E., and Thomas, E. (2010). Value-Added Modeling of Teacher Effectiveness: An Exploration of Stability Across Models and Contexts. *Educational Policy Analysis Archives*, 18. Retrieved December 10, 2019 from [*http://dx.doi.org/10.14507/epaa.v18n23.2010*](http://dx.doi.org/10.14507/epaa.v18n23.2010).

Özgüngör, S., & Duru, E. (2015). Course and instructor characteristics distinguishing highest and lowest student ratings of instructors. *Eurasian Journal of Educational Research*, 61, 118-136.

Papadimitriou, C. H., Raghaven, P. & Tamaki, H. & Vempala, S. (2000). Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences*, *61*, (2), 217-235.

Peneul, W.R. & Shepard, L.A. (2016). Assessment and teaching in In In Drew H. Gitomer & Courtney A. Bell (Eds.) *Handbook of research on teacher education.* NY: Routledge, 787.

Pivopirova, M., Amrein-Beardsley, A., & Broatch, J. (2016). Value-Added Models (VAMs): Caveat Emptor. *Statistics and Public Policy*, 3 (1), 1-9. Retrieved December 9, 2019 from<https://doi.org/10.1080/2330443X.2016.1164641>

Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora.

Retrieved March 1, 2012 from <https://github.com/RaRe-Technologies/gensim>.

Řehůřek, R., & Sojka, P. (2011). Gensim—statistical semantics in python. NLP Centre, Faculty

of Informatics, Masaryk University.

Raufelder, D., Nitsche, L., Breitmeyer, S., Kessler, S., Hermann, E., & Regner, N. (2016).

Students’ perception of “good” and “bad” teachers—Results of a qualitative thematic

analysis with German adolescents. *International Journal of Educational Research*, 75,

31-44.

Ripley, A. (2012, October). Why kids should grade teachers. *The Atlantic*. Retrieved December 10, 2019 from <https://www.theatlantic.com/magazine/archive/2012/10/why-kids-should-grade-teachers/309088/>

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. Shanghai, China: *Proceedings of the eighth ACM international conference on Web search and data mining*, pp. 399-408. Retrieved December 10, 2019 from [Https://doi](https://doi).  [10.1145/2684822.2685324](https://doi.org/10.1145/2684822.2685324)

Rosner, F., Hinneburg, A., Röder, M., Nettling, M., & Both, A. (2014). *Evaluating topic coherence measures*. Retrieved 3 March, 2020 from [arXiv preprint arXiv:1403.6397](about:blank)

Scriven, M. (1994). Student Ratings Offer Useful Input to Teacher Evaluations. *Practical Assessment, Research, and Evaluation, 4*, Article 7. Retrieved December 10, 2019 from [https://scholarworks.umass.edu/pare/vol4/iss1/7](about:blank)

Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education, 8* (2), 151-184. Retrieved December 10, 20719 from [DOI: 10.1007/BF00972261](about:blank)

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the workshop on interactive language learning, visualization, and interfaces, 63-70*. Baltimore, MD: Association for Computational Linguistics.

Stevens, K., Kegelmeyer, P., Andrzejewski, P. & Buttler, D. (2012). Exploring topic coherence over many models and many topics. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning,* Jeju Island, Korea, *952-961*. Stroudsburg, PA.: Association for Computational Linguistics.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement." *Journal of Teacher Education*, *62* (4), 339-355.

Strunk, K., Weinstein, T., & Makkonen, R. (2014). Sorting out the signal: Do multiple measures of teachers’ effectiveness provide consistent information to teachers and principals? *Education Policy Analysis Archives, 22* (100). Retrieved December 10, 2019 from <http://dx.doi.org/10.14507/epaa.v22.1590>

Teh, Y. W., Jordan, M. J., Beal, M. J., & Blei, D. M. (2005). Sharing clusters among related groups: Hierarchical Dirichlet processes. In *NIPS'04 Proceedings of the 17th International Conference on Neural Information Processing Systems* (1385– 1392). Vancouver, BC and Cambridge, MA: MIT Press.

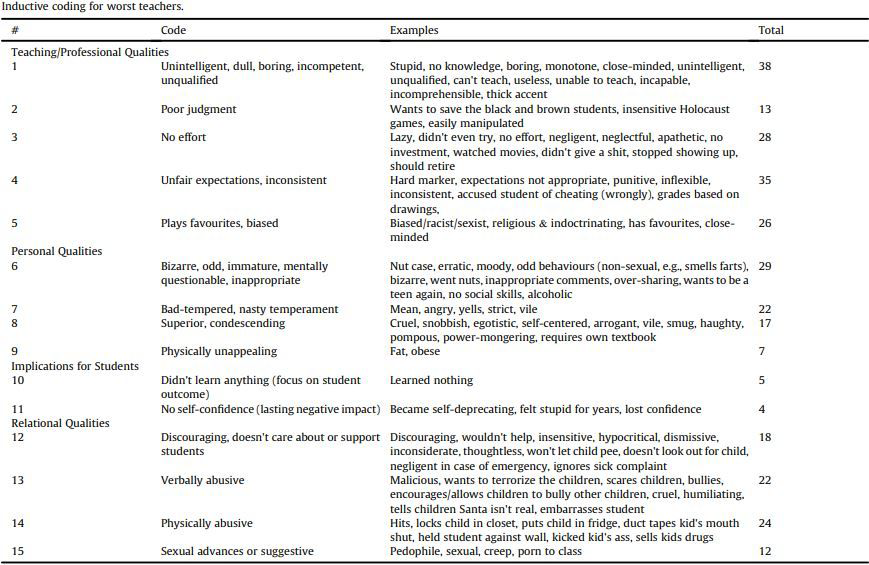


Table 1 - Coding of Reddit comments and descriptions about one’s “worst” teacher (From Chang-Kredl and Cloannino, 2017, p. 48)

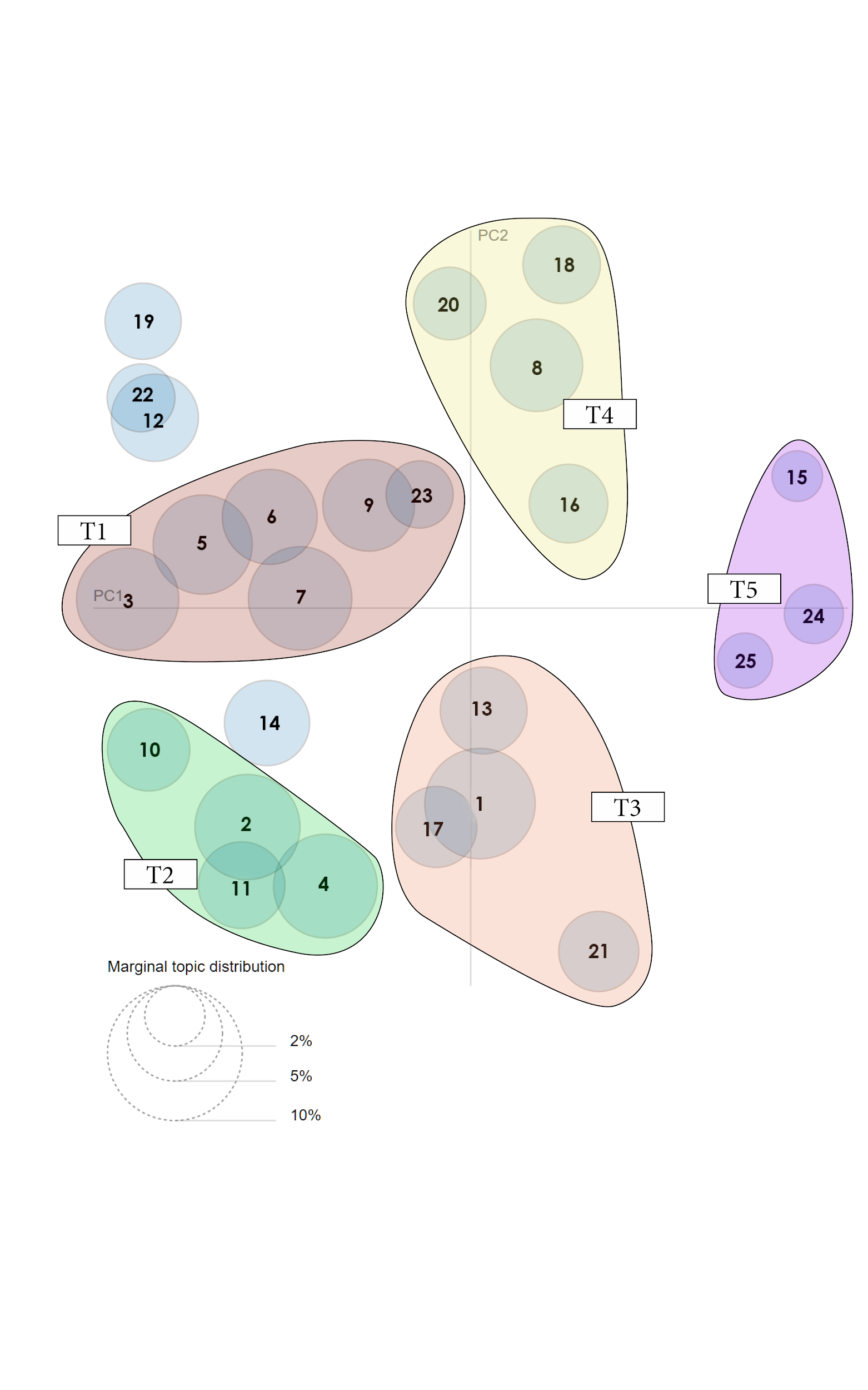
|  |  |
| --- | --- |
| Topic | **Topic 1** Teacher is a harsh grader, assigns too much work with unrealistic timetables. |
| Key  Words | give, grade, assign, homework, work, project, test, never, all, time, even, paper, hard, point, class, gave, essay, lose, teacher, due, check, bad, expect, turn, many, good, busy work, tell, make |
| Sample  Comments | *"Bad teacher. Picky grader and assigns ridiculous amounts of homework. But there are barely any tests and quizzes so if you do all your homework then you're good."* |
| *"She grades way too hard and gives us barely enough time to finish our assignments. She is not specific in what she wants from us."* |
| *"Not a good teacher at all. She gives too much homework and expects big projects to be due so early. Very strict at grading too."* |
| *“Very unclear about assignments and bad when it comes to projects grades way too harsh without even giving enough time or clarifying it”* |
| *“Her grading system makes no sense. She gives us a rubric and then grades the paper with a different one.”* |

TT

**Table 2 - Topic Descriptions and Frequency**

|  |  |
| --- | --- |
| Topic Title | Freq. |
| 1 - Teacher is a harsh grader, assigns too much work with unrealistic timetables | 7.9 |
| 2 - Teacher does not prepare students for test or align classwork with tests | 7.7 |
| 3 - Teacher provides confusing explanations or has a poor teaching style; unrealistic expectations | 7.4 |
| 4 - Teacher talks so much, students don’t feel they are being taught; talks about unrelated topics. | 6.8 |
| 5 - Teacher cannot “handle” students; cannot relate to them or develop good working rapport | 6.7 |
| 6 - Teacher is mean, intimidating, strict or never smiles and weird. | 6.3 |
| 7 - Teacher acts like a child or treats students like children; moody or short-tempered. | 5.7 |
| 8 - Teachers or staff have poor communication, personalities, or priorities. Includes parents’ reviews. | 5 |
| 9 - Teacher is unclear, unorganized or unfocused | 4.6 |
| 10 - Teacher is remarkably boring and/or classes are too easy | 4.3 |
| 11 - Teacher’s class is pointless; students feel they learn nothing and waste time | 4.2 |
| 12 - Teacher is described as “the worst” or some synonym of horrible | 4.2 |
| 13 - Teacher cannot or will not help or correct students which leaves them very frustrated | 3.8 |
| 14 - Students warning other students to switch out for any number of reasons | 3.2 |
| 15 - Teacher is defensive of a political position, enforces dress codes; Includes non-English reviews. | 3.2 |
| 16 - Teacher or staff are selfish and incompetent or do not put students first | 3.2 |
| 17 - Students express stories and feelings of genuine dread and hatred of a teacher | 3.1 |
| 18 - Teacher is disrespectful, unpleasant, rude or inappropriate; may show gender or religious bias | 2.9 |
| 19 - Teacher is a good person but a bad teacher. | 2.7 |
| 20 - Teacher picks favorites; many reviews about music teachers | 2.7 |
| 21 - Teacher pays little attention to students and/or their work. | 2.6 |
| 22 - Teachers compared to better teachers. Sarcastic comments. Good reviews mistakenly rated low. | 2 |
| 23 - Teacher does not provide help or advice to students | 1.8 |
| 24 - Teacher relies on powerpoints, notes and books; has an industrial teaching style. | 0.5 |
| 25 - Teacher relies on notes and external materials. Grade strictly with little room for critical thinking. | 0.45 |

**Table 3 - All Topic descriptions and the percent of reviews they dominated.**



### Figure 1 - Topics Collapsed into 2-Dimensional Space and Grouped into Themes.

|  |  |  |
| --- | --- | --- |
| Theme 1 - Unable to Create a Productive Learning Environment | | 32.5 |
| 3 | Teacher provides confusing explanations or has a poor teaching style; unrealistic expectations | 7.4 |
| 5 | Teacher cannot “handle” students; cannot relate to them or develop good working rapport | 6.7 |
| 6 | Teacher is mean, intimidating, strict, never smiles and “weird” | 6.3 |
| 7 | Teacher lacks maturity or treats students like children; moody or short-tempered. | 5.7 |
| 9 | Teacher is unclear, unorganized or unfocused | 4.6 |
| 23 | Teacher does not provide help or advice to students | 1.8 |
| Theme 2 - Teaches Bad Lessons that Waste Students’ Time | | 23 |
| 2 | Teacher does not prepare students for tests or align classwork with tests | 7.7 |
| 4 | Teacher talks too much or about unrelated topics; students don’t feel they are being taught. | 6.8 |
| 10 | Teacher is exceptionally boring and/or classes are too easy | 4.3 |
| 11 | Teacher’s class is pointless with students feeling they learn nothing. Waste of their time | 4.2 |
| Theme 3 - Behavior is Difficult or Insensitive | | 17.4 |
| 1 | Teacher is a harsh grader, assigns too much homework or classwork with unrealistic timetables | 7.9 |
| 13 | Teacher cannot or will not help or correct students | 3.8 |
| 17 | Students tell stories and relate feelings which describe dread and hatred of a teacher | 3.1 |
| 21 | Teacher pays little attention to students and/or their work. | 2.6 |
| Theme 4 -Displays Bad leadership | | 13.8 |
| 8 | Teachers or staff have poor communication, personalities or priorities. Includes parents’ reviews. | 5 |
| 16 | Teacher or staff are selfish, incompetent or do not put students first | 3.2 |
| 18 | Teacher is disrespectful, unpleasant, rude or inappropriate; may show gender or religious bias | 2.9 |
| 20 | Teacher picks favorites; in many reviews about music teachers | 2.7 |
| Theme 5 - Rigid as From the Industrial Age | | 4.2 |
| 15 | Teacher is defensive of a political position or opinion, enforces dress codes; reviews written in foreign languages. | 3.2 |
| 24 | Teacher relies on power-points, notes and books; has an “industrial” teaching style. | 0.5 |
| 25 | Teacher relies on notes and external materials. Grades rigidly with little room for critical thinking. | 0.45 |
| The Residual Topics | | 8.9 |
| 12 | Teacher is described as “the worst” or some synonym of horrible | 4.2 |
| 19 | Teacher is a good person but a bad teacher. | 2.7 |
| 22 | Reviews that fit nowhere else; these include some positive language. Teachers being compared to better teachers. Sarcastic comments. | 2 |

**Table 4 – Themes and the Topics that Comprise Them.**

# Appendix

# Experimental Results

The first table (Data Configuration Table) describes the different cleaning procedures that were implemented and their respective labels. For the current experiments, only configurations “A1” and “E1” were used, the others were not compatible with LDA but were used in preliminary testing of alternative modeling approaches (Fasttext). The main distinctions between “A1” and “E1” are the exclusion of “not,” “off” and “on,” the use of the stemmer and the size of the allowable n-grams (which made no difference since no 4+ n-grams were created). The input in an NLP model can have a dramatic effect on the outcome, after iterative testing with different sized samples, these were determined to be the most useful alternatives to test with the full text.

**Appendix Table 1**

The second table (Experimental Results for LDA Models) shows the coherence scores that resulted when different numbers of topics were pulled from each trial. The primary distinction across trials were **Data Configuration** and **Review Length**. Review length was more consequential than Data Configuration in producing higher coherence scores. Note that the number of passes for trials with review length 150 was larger 50. The outcome of the models improves with more trials, despite this advantage, the trials with minimum length 100 still outperformed the trials with length 150.

**Appendix Table 2**

**Complete List of Topics**

The table below shows key words and samples reviews for all 25 of the topics we found in our analysis. Each topic has a its topic number and title, the percentage of reviews it dominated, a list of key words and a selection of sample reviews from the reviews that had a probability of belonging to that topic of at least 80%.

**Appendix Table 3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Configuration | A1 | B1 | C1 | E1 |
| Duration | 6930.111 | 3096.355 | 6650.567 | 6146.227561 |
| N-Gram | 4 |  |  | 5 |
| Lemmatizer | TRUE | FALSE | TRUE | TRUE |
| Exclude |  |  |  | ['not', 'off', 'on'] |
| Appearance Threshold | 50 | 50 | 50 | 50 |
| Remove Contractions | TRUE | TRUE | TRUE | TRUE |
| Remove Stops | TRUE | FALSE | FALSE | TRUE |
| Repeated Character Removal | 2 |  |  | 2 |
| Spell Check | FALSE | FALSE | FALSE | FALSE |
| Stemmer | TRUE | FALSE | TRUE | FALSE |

Appendix Table 1 - Data Configuration Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trial | 1 | 2 | 3 | 4 |
| Number of Topics, Coherence Score | [(25, 0.502), (10, 0.475), (40, 0.462), (55, 0.449), (85, 0.440), (70, 0.434), (100, 0.432)] | [(25, 0.463), (10, 0.447), (40, 0.442), (55, 0.431), (70, 0.424), (85, 0.420), (100, 0.416)] | [(25, 0.488), (40, 0.461), (10, 0.457), (70, 0.444), (55, 0.443), (85, 0.440), (100, 0.423)] | [(10, 0.464), (40, 0.450), (55, 0.442), (25, 0.440), (100, 0.436), (70, 0.433), (85, 0.426)] |
| Corpus Length | 359387 | 359387 | 359387 | 359387 |
| Data Configuration | A1 | E1 | E1 | A1 |
| Duration | 5560.833 | 6882.202152 | 5612.498978 | 6736.142547 |
| Length of Cleaned Corpus | 211224 | 136193 | 211224 | 136193 |
| Exclude Words Above This % | 0.5 | 0.5 | 0.5 | 0.5 |
| Exclude Words That Appear Less Than | 30 | 20 | 30 | 20 |
| # of Topics to Test | [10, 25, 40, 55, 70, 85, 100] | [10, 25, 40, 55, 70, 85, 100] | [10, 25, 40, 55, 70, 85, 100] | [10, 25, 40, 55, 70, 85, 100] |
| Model Iterations | 32 | 50 | 32 | 50 |
| Review Length | 100 | 150 | 100 | 150 |
| Setting | LDA1 | LDA2 | LDA1 | LDA2 |

Appendix Table 2 - Experimental Results for LDA Models (0-35 Only)

|  |  |
| --- | --- |
| **Topic # - Description / Keywords / Sample Reviews** | **Freq** |
| **Topic 1 -** Teacher is a harsh grader, assigns too much work with unrealistic timetables | 7.9 |
| give, grade, assign, homework, work, project, test, never, all, time, even, paper, hard, point,  class, gave, essay, lose, teacher, due, check, bad, expect, turn, many, good, busy work, tell, make |
| * *"Bad teacher. Picky grader and assigns ridiculous amounts of homework. But there are barely any tests and quizzes so if you do all your homework then you're good."* * *"She grades way too hard and gives us barely enough time to finish our assignments. She is not specific in what she wants from us."* * *"Not a good teacher at all. She gives too much homework and expects big projects to be due so early. Very strict at grading too."* |
| **Topic 2 -** Teacher does not prepare students for test or align classwork with tests | 7.7 |
| test, teach, class, know, all, give, quiz, learn, taught, never, material, expect, study, anything, teacher, even, fail, prepare, exam, everything, physic, textbook, ap, review, note, stuff, chemistry, lab, well, inform |
| * *"Gets so off topic during class, tests us on a unit test which consists of 3 chapters, never finishes teaching us what she needs to during the class."* * *"she doesn't teach well; she doesn't present herself well. she just doesn't know how to teach; you may just study yourself. Too many tests are really not necessary since test is only a part of learning."* * *"just wish he could actually teach. none of us get what he is lecturing about, but when it comes time for a test we end up doing bad because he doesnt teach. none of us are even prepared for the ap test.”* |
| **Topic 3 -** Teacher provides confusing explanations or has a poor teaching style; unrealistic expectations | 7.4 |
| understand, explain, hard, teach, math, teacher, class, welcome, confuse, anything, good, know, make, really, cannot, student, try, time, material, lesson, expect, difficult, fast, bad, problem, everything, clearly, say, never |
| * *"I can't understand anything. I am usually good student but am sure to fail now with her as my teacher."* * *"hard to understand. most students do not understand the material even though the teacher repeats same material over and over."* * *"She is hardest teacher I have ever had. She knows math very well, so she is rushing over the material in the class and expects everyone to understand immediatel"* |
| **Topic 4 -** Teacher talks so much, students don’t feel they are being taught; talks about unrelated topics. | 6.8 |
| talk, class, time, all, teach, whole, spend, life, half, period, know, sit, story, watch, work, tell,  never, even, anything, waste, phone, listen, hear, spent, idea, stuff, joke, random, minute, entire |
| * *"Mrs.-- repeats herself all the time. My class made a bingo game with her most popular sayings/ topics. Also, if you want to waste a whole class period get her talking about Grancat."* * *“She didn't teach me anything. Kids in my class were teaching me. She would talk to us for the whole class”* * *“He knows alot of things.... just not how to teach them... he goes on and on and on sometimes and he doesnt stop talking and then you forget what he's talking about.... :|”* |
| **Topic 5 -** Teacher cannot “handle” students; cannot relate to them or develop good working rapport | 6.7 |
| teach, need, know, cannot, think, better, learn, student, speak, all, try, stop, actually, maybe, french,  say, retire, teacher, job, kid, people, well, anything, even, new, something, stuff, everything, realize, please |
| * *"She needs another job. She doesn't do well with students. Sorry Ms. XXXXX, but stick a fork in ya, your done!!!!!"* * *"I can't work with someone who won't pay attention to my needs as a student, also you just don't know when to take him seriously!"* * *"She can't handle being a teacher. she needs to learn how to deal with kids and not take things so personally."* |
| **Topic 6 -** Teacher is mean, intimidating, strict or never smiles and weird. | 6.3 |
| mean, hate, yell, really, annoy, all, think, teacher, ok, funny, like, say, alway, people, good, reason, bad, sometime, time, everyone, get, joke, cool, omg, ugh, kid, strict, laugh, even, well |
| * *"i like you as a teacher but sometimes (most of the time) you are very intimidating to me and im scared to really say anything"* * *"hes so mean. i really dislike him. he yells too much and is way to strict. never see him smile, EVER.”* * *"Um. A bit...Weird. Sometimes, rude. Annoying. Not all that good of a teacher..Sometimes alright, I guess. "* |
| **Topic 7 -** Teacher acts like a child or treats students like children; moody or short-tempered. | 5.7 |
| like, class, alway, fell, kid, fun, make, seem, student, act, teacher, bad, all, control, treat, people,  art, try, know, never, want, care, everyone, enjoy, good, look, old, something, work, dread |
| * *She's smart but I always feel like shes not exciting enough. she comes into class seeming depressed. she doesn't have an exciting class becuz of this* * *"One of the most moody and crankiest teachers out there. She is always sarcastic and acts like a child in class."* * *I don't understand why she is so critical. She won't let you say anything you want and treats everybody like they are five years old.* |
| **Topic 8 -** Teachers or staff have poor communication, personalities or priorities. Includes parents’ reviews. | 5 |
| student, teacher, child, parent, experience, educate, son, many, school, work, daughter, kid, time, issue, deal, concern, classroom, disappoint, well, problem, condescend, receive, home, class, severe, unprofessional, lack |
| * *“Mrs. XXXXXX is extremely unprofessional. She was very defensive in a parent meeting. My child has not made any progress with this teacher and I would not recommend her.”* * *"Does not take responsibility for her school. Does not present herself as a leader or a representative, yet demands respect which, due to her actions, is not given to her."* * *"Sets unrealistic expectations for children. Doesn't communicate her wishes or directions clearly to students or parents. Leaves a lot to be desired."* |
| **Topic 9 -** Teacher is unclear, unorganized or unfocused | 4.6 |
| teach, student, class, often, subject, knowledge, unclear, method, style, cannot, skill, extreme, lack, material, teacher, confuse, topic, interest, little, unhelp, complete, disorganize, well, lesson, lazy, lecture, instruct |
| * *"I'm going to be completely unprepared for the exam. His class may be interesting, but only because it's frequently off-topic."* * *"class is unfocused. lectures and assignments are not clear and unrelated to curriculum. provides little constructive feedback"* * *"This teacher is very interested in her topic. However, she is unorganized, and her exams are extremely poor in design. She is also often unprepared for class, and cannot answer questions clearly."* |
| **Topic 10 -** Teacher is remarkably boring and/or classes are too easy | 4.3 |
| really, class, bore, easy, pretty, hard, anything, learn, super, teacher, fun, god, bad, want,  think, pay attention, funny, sleep, teach, fall asleep, thing, never, extreme, lecture, care |
| * *"His class isn't that hard if you take notes, but he is EXTREMELY boring! It is hard not to fall asleep while listening to him."* * *"her class is sooooooo boring! she doesn't really teach that well and we cook really wierd stuff. I definately would never take any of her classes again!"* * *"hes a reaallly bad teacher but hes extremely fun to hang around with still i learned more from the text book then i did from him"* |
| **Topic 11 -** Teacher’s class is pointless; students feel they learn nothing and waste time | 4.2 |
| learn, class, noth, anything, all, history, absolute, honor, ap, teach, thing, waste\_time, taught, want, taken,  level, course, actually, joke, teacher, world, english, pointless, complete, take, except, took, regular, prepare |
| * *"I feel like I was placed in a basic class- and I had taken Honors English (Senior year). Talk about a mind-numbing, pointless experience. I learned nothing from him except how aggravating he can be. Worst English class EVER."* * *"If you want a blowoff class, this is the class for you. Mrs. XXXXXXX doesn't do anything and I haven't learned one thing in her class."* * *"Don't waste your time taking Intro to Business. I didn't learn anything. This class is more free time than anything. You're better off not having a class than taking this one"* |
| **Topic 12 -** Teacher is described as “the worst” or some synonym of horrible | 4.2 |
| teacher, worst, horrible, teach, far, probably, terrible, bad, know, never, all, even, anything, life, absolute, math, school, possible, fire, honestly, idea, world, been, hope, year, still, worst, meanest |
| * *"One of the worst teachers I've ever encountered. He has no clue how to teach and is very inflexible. He is a prime example of why teachers (at least some teachers) shouldn't be tenured. "* * *"She is the absolute WORST teacher I have ever had. She is horrible and she should get fired IMMEDIATELY! "* * *"The most incompetent teacher I have EVER had the misfortune to have. Her teaching liscense should be IMMEDIATELY revoked. Absolutely awful, and can't teach!!!!!"* |
| **Topic 13 -** Teacher cannot or will not help or correct students which leaves them very frustrated | 3.8 |
| question, answer, tell, ask\_question, wrong, say, right, ask, know, something, problem, alway, even, never, lok, answer\_question, give, correct, ignore, someone, thing, everything, whenever, think, try, figure, get\_mad, want |
| * *"She is never there. She is always busy. She said that she would get back to me and she never did!!!!!! She didn't fix the problem either."* * *"She often contradicts herself, mispeaks, makes errors--nonetheless, she always has a cocky attitude, she's never wrong and she knows everything. She says that."* * *"I don’t know if she even knows the material herself. Its at the point where i have to learn everything myself. If i ask a question she says email me. If I ask another question she says ask someone else. If I ask a question she says ‘idk should you?’"* |
| **Topic 14 -** Students warning other students to switch out for any number of reasons | 3.2 |
| will, class, good, spanish, teacher, fail, switch, want, luck, grade, pay, drop, try, unless, learn, side,  transfer, screw, recommend, say, next, language, never, life, take, anything, future, see |
| * *"She's a terrible teacher. If you get her you should transfer out unless you want a long and stressful year."* * *"For those people that actually want to learn, which yes, there are those of us out there. XXXX would not be a good choice for a teacher. "* * *"yes i had an A in her class but she’s not a good teacher at all....she did not prepare anyone for a higher level of spanish..not good "* |
| **Topic 15 -** Teacher isdefensive of a political position, enforces dress codes; Includes non-Enlgish reviews. | 3.2 |
| opinion, agree, mind, term, politic, speech, view, non, short, color, hey, day, biggest, hilarious, disagree, wear, memory, horrid, field, white, game, space, brain, pre, survive, train, remind, you ,with |
| * *"Ms. I is VERY opinionated and is ego-centric. She has the potential to be a great teacher, but her opinions get in the way. After having her, I am not sure what I learned was her opinion or facts."* * *"Bernie expects the Lakeside community to embrace his political views - there is no room for diversity of thought and the school is not inclusive if your views are not aligned with his socialistic views."* * *"she dress coded me. I hate her. she followed me into the bathroom just to dress code me. gfdvahsbjnhgvcfxdsjdgfhjksjfhgjdfksjfdghndjshfgd"* |
| **Topic 16 -** Teacher or staff are selfish and incompetent or do not put students first *[Peripheral]* | 3.2 |
| school, care, all, kid, around, coach, run, principle, rule, student, want, good, program, think, job, come, sport, team, see, business, sad, people, office, money, administer, leave, pay, even, athlete, away |
| * *"i have been in theater since I was seven years old. I quit after one year of her. she ruined theater for me, and i have lost all of the passion for it i once had."* * *"His machiavellian policies destroys the loyalty of all students. He spearheads a team of anti-semetic satanists bent on destroying all student's will to live. "* * *"all he did was sit in his office. and he wasn't there for the students, he was just there for the paycheck."* |
| **Topic 17 -** Students express stories and feelings of genuine dread and hatred of a teacher | 3.1 |
| got, class, told, friend, never, said, gave, hate, went, thank, thought, still, glad, even, good, back, gym, all, knew, freshman, came, walk, yell, year\_ago, senior, remember, counselor, gone, taught, reason |
| * *"she said that going 2 the bathroom is a privilege. -\_-. gave us silent lunch and made 7th grade miserable. i'm so happy i don't have 2 deal with that anymore."* * *"Blech. I had her last year. So-so, the only good thing is when she got mad at David. Ha! She threw a chair across the room."* * *"OMG! I am in high school now, and I still have nightmares about XXXXXXXXX. He made my eighth grade year a living nightmare!"* |
| **Topic 18 -** Teacher is disrespectful, unpleasant, rude or inappropriate; may show gender or religious bias | 2.9 |
| student, rude, respect, care, woman, make, think, people, toward, extreme, true, disrespect, constantly, life, sarcast, insult, front, punish, human, belief, intelligent, arrogant, attitude, comment, inappropriate, believe |
| * *"Rude and prejudiced. Makes jokes that are inappropriate (intolerant). Shouts down those who disagree with him. Is passionate, though."* * *"She is insensitive and will openly make fun of a student just for her own pleasure. She is one of the most unprofessional teachers I have ever had, and is also very childish."* * *"She is very rude to students when they try to ask a question. i don't believe that she cares for any students personal issues."* |
| **Topic 19 -** Teacher is a good person but a bad teachers. | 2.7 |
| nice, person, teacher, good, great, really, lady, teach, bad, sometime, kind, however, horrible, although, well, funny, over, seem, terrible, know, sweet, best, okay, come, love, strange, german, may, friendly, heart |
| * *"Although she could be a really nice and sweet person outside of class, she just wasn't a very good teacher."* * *"She just isn't a great teacher and that's me being honest. I am a native Spanish speaker and she really doesn't seem to know what she was doing...good efforts though"* * *"She may be a good club sponsor, but she has a job as a teacher. She has a bad attitude and is a horrible teacher."* |
| **Topic 20 -** Teacher picks favorites; many reviews about music teachers | 2.7 |
| favorite, play, pick\_favorite, unfair, student, music, pick, teacher, favor, good, people, band, other, grade, kid, hose, least, fair, one, all, definite, obvious, sing, popular, hate, everyone\_else, certain, girl, show, bias |
| * *"she loves me. I don't know why but it is not fair for everyone else who has to pay attention or get screamed at"* * *"She hates on some people with a freakish passion. You'll know that race def. plays a role in favoritism.Not recommended. At all."* * *"She favored some students and was really strict with others. She favors the good singers above the chorus. UNFAIR, even while I am one of those good singers "* |
| **Topic 21 -** Teacher pays little attention to students and/or their work. | 2.6 |
| class, last, first, week, time, took,name, minute, day, every, point, start, late, school, been,second, got, end, left, month, semester, take, even, sub, wait, still, long, notice, three, away |
| * *"She does not know what my name is and I have been in her class for 3 years."* * *"Very unreasonable. She lost one of my papers for a week found it again and said it was late taking off points for it. Despite its having been on time!"* * *"The last time I went to her for extra help, she kept me waiting in her room for an hour before I decided to leave. I go to other teachers for help b/c she doesn’t"* |
| **Topic 22 -** Teachers compared to better teachers. Sarcastic comments. Good reviews mistakenly rated low. | 2 |
| teacher, english, best, know, science, teach, good, real, compute, college, think, high\_school, worse, even, rate, bad, love, type, awesome, middle\_school, grammar, subject, world, pretend, social\_studies, terrible, surprise |
| * *"Such a bad teacher. I had her for Humanities. It was shocking how bad she was compared to the other teacher. She was totally checked out. "* * *"How did she get hired by XXXXXXX High School? Most of the teachers there are awesome but not this one! She didn't even teach us half of what we were supposed to be exposed to."* * *"Best teacher ever!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! Just Kidding"* |
| **Topic 23 -** Teacher does not provide help or advice to students | 1.8 |
| help, ask, student, all, never, avoid, teacher, extra, need, refuse, try, care, time, clear, rude, while,want, busy, struggle, offer, cost, work, class, even, come, tutor, instead, feel\_stupid, stay, problem |
| * *"she never had any clear answers to any of my questions... not very helpful either. she didn't really care about where my future was going. "* * *"She has no patience at all for her students. She makes students feel very stupid especially when they stay after for help. Afterschool she barely pays any attention to you and rushes kids out the door."* * *"He's lazy to the last degree. He doesn't make any effort to get to know his students. Don't even bother going to him for advice or assistance- you won't get either, at least not when you want them."* |
| **Topic 24 -** Teacher relies on powerpoints, notes and books; has an industrial teaching styles. | 0.5 |
| use, note, write, board, make, copy, paper, take, example, textbook, essay, lesson\_plan, wrote, internet, many, page, open, written, point, tech, online, number, info, refer, picture, use, text, quote, error, instead |
| * *"Does not allow the human mind to process math, but instead transforms people into calculator using, formula regurgitating minions. "* * *"Mainly, he throws notes up on the board, Talks about other things while you copy them, and hopes the information will stick. Bad, bad teacher. "* * *"A baffoon. Does not teach well, and has a very high opiniopn of himself. Overuses "technology," a.k.a. Microsoft Powerpoint, to the point of absurdity. "* |
| **Topic 25 -** Teacher relies on notes and external materials. Grade strictly with little room for critical thinking. | 0.45 |
| read, book, word, worksheet, hand, powerpoint, memory, say, read\_book, packet, cheat, every, study\_guide, page, use, all, vocab, slide, lecture, ill, power\_point, sentence, describe, moment, copy, text\_book, huge, mumble, list, fill |
| * *"Has a powerpoint of 100 or so slides for every chapter. WWII chapter is about 200! If you dont read, and dont copy the powerpoint you will fail."* * *"All she does is put slides and explain them. All her worksheets have the other teachers name on it. None of it was hers.”* * *"Busy work and a strict teacher. He just reads a slideshow word for words copied from the book word for word, then hands out a worksheet for you to do the rest of class. Every. Single. Day. "* |

## 

## Appendix Table 3 – All 25 topics with key words and sample reviews