ELICITATION OF EMOTIONAL RESPONSE FROM MUSIC LYRICS

A Thesis

by

NATALIE ANNE DAVIS

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Billy Ray McKim
Committee Members,	Clare A. Gill
	Andrew OP McCubbins
Head of Department,	Scott Shafer

August 2019

Major Subject: Agricultural Leadership, Education, and Communications

ABSTRACT

The emotional effect of music lyrics on humans has not been widely researched. There were 53 participants that ranged from ages 19-34 in this study. Of the 53 participants, 43 identified as female, 10 identified as male, and music education varied amongst participants. Three methods were used to conduct this study including a sentiment analysis, a perception analysis, and a survey. The sentiment analysis analyzed words in 497 songs from the Billboard year-end music charts for the country, pop, and urban genre formats. The perception analysis used the Dialsmith Perception Analyzer software to measure moment-to-moment participant feedback for nine selected songs from the year-end music charts. The survey was used as a participant self-report for opinions on each of the nine songs.

Bivariate correlations were used to find relationships among sentiment scores and moment-to-moment scores, internal and external factors and moment-to-moment scores, and sentiment scores and song performance on the Billboard charts. There were significant relationships for all correlations except year-end chart performance. Sentiments and human response were not an indicators of year-end chart rank performance for the nine songs that the participants listened to.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Billy R. McKim and Professor Andrew OP McCubbins of the Department of Agricultural Leadership, Education, and Communications, and Professor Clare A. Gill of the Department of Animal Science

The research for the sentiment analyses were supervised by Professor Clare A. Gill of the Department of Animal Science. The research related to the Dialsmith Perception Analyzer and the survey responses were supervised by Professor Billy R. McKim of the Department of Agricultural Leadership, Education, and Communications.

All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

Funding for this research project was only contributed by the graduate student researcher. No external or internal sources were contributed to this research.

NOMENCLATURE

AFINN	Finn Årup Nielsen
AM	Amplitude Modulation
BDS	Broadcast Data Systems
CHR	Contemporary Hits Radio
CRAN	Comprehensive R Archive Network
CRM	Continuous Response Measurement
DMRDL	Digital Media Research and Development Lab
EDM	Electronic Dance Music
FCC	Federal Communications Commission
FM	Frequency Modulation
IQ	Intelligence Quotient
MB	Megabyte
MSC	Memorial Student Center
MTM	Moment-To-Moment
NWC	Negative Word Count
NRC	National Research Council
PA	Perception Analyzer
POS	Positive Word Count
PRO	Performance Rights Organizations

Q&A	Questions & Answers
R&B	Rhythm & Blues
RAB	Radio Advertising Bureau
RAM	Random Access Memory
RIAA	Recording Industry Association of America
RMT	Radio Media Tour
RQ	Research Question
RTR	Real-Time Rating
SW	Similar Words
TV	Television
US	United States
USB	Universal Serial Bus
WD	Word Count
XM	No Modulation

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Americans are listening to more music, now, than ever before. The average hours of music consumed by Americans each week has increased every year, with 23.5 hours in 2015, 26.6 hours in 2016, 32.1 hours in 2017, and 33.5 hours in 2018 (McIntyre, 2017; Rys, 2018). The increase in music consumption can be attributed to technological advances throughout the years. Technology has enabled a wide range of the population to listen to music in numerous places and social contexts, and at any time of day (Juslin, Liljestrom, Vastfjall, Barradas, & Silva, 2008). Music can be consumed on devices including smartphones, laptops, and tablets, which have made music more accessible and enabled listeners to engage with music anywhere, and at any time (Nielsen, 2017).

In 2018, the US recorded music portion of the music industry generated $9.8 billion of revenue— growth at a double-digit pace for the third year in succession (Recording Industry Association of America [RIAA], 2019). There are more than 50 million paid subscribers to streaming digital platforms, including Spotify, Apple Music, and Amazon, and satellite (SiriusXM) radio. Even with streaming numbers increasing, terrestrial radio is still the main source for music consumption. In fact, US terrestrial radio broadcasters operate more than 11,000 commercially licensed radio stations that reportedly reach 92% of Americans each week (Radio Advertising Bureau [RAB], 2019).

Recorded music content distributed through streaming, satellite, and terrestrial platforms are monitored at many levels and by numerous entities, including writers, composers, publishers, performers, producers, record labels, media distributors, performance rights organizations, and royalty administrators. Because of federal legislation (e.g., Music Modernization Act of 2018 and Local Radio Freedom Act of 2019 [H.R. 20, 2019; S. 5, 2019]) and regulatory agencies (e.g., FCC), recorded music content is often measured by unit counts (e.g., album, single, spin) or audience metrics (e.g., reach, frequency, time spent listening). However, less attention has been directed to the content in each song and consumers' indirect response to the content.

There are notable connections among human thoughts, emotions, and music. The human mind is continuously engaged with internal dialogue and emotional thoughts throughout the course of the day (Eerola, 2012). Humans experience between 60,000 to 80,000 thoughts and emotions per day (Chopra, 2016). Types of media including television, radio, newspapers, billboards, the Internet, and social media communicate endless stories, messages, and offers. With hundreds of television channels, radio stations, and products to choose from, there are also thousands of words to process (Dialsmith, 2013). There are as many as 5,000 impressions made per individual a day (Johnson, 2006). The competition and easy access to technology has made it harder for communicators to capture the attention of their audience and leave a lasting impression. If these communicators fail to convince people to keep tuning in, the opportunity to make a meaningful impression is gone (Dialsmith, 2013). This is a reason why researchers want to understand what messages connect, impact, and provoke thoughts and emotions in their consumers. There are several ways in which emotions can be measured (see figure 1). However, what specifically causes thoughts and emotions has long been a debated topic among scholars.

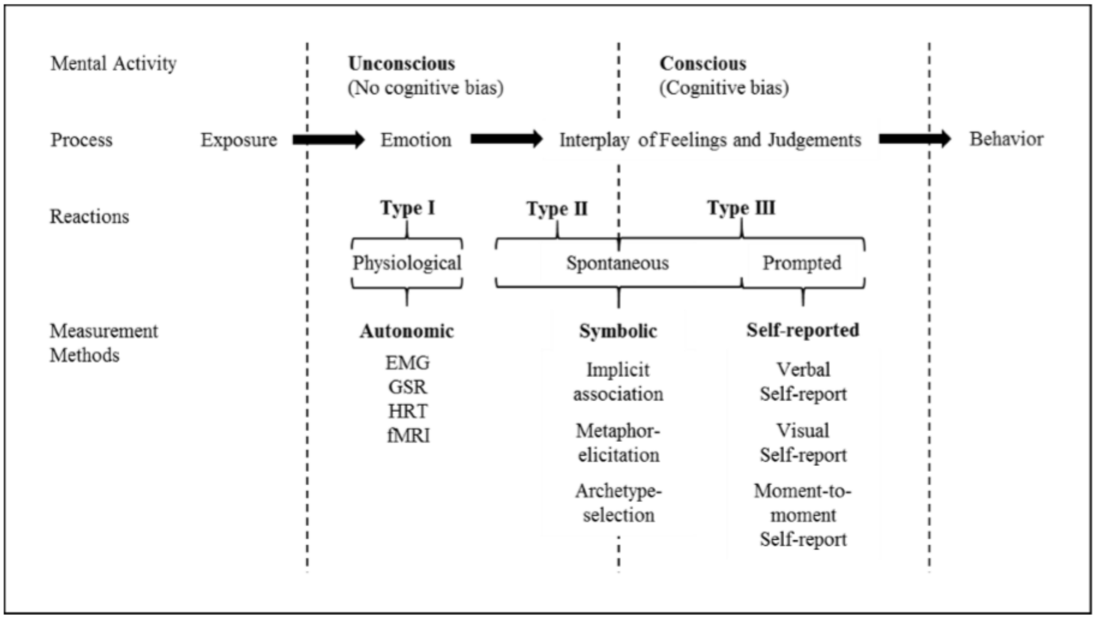

Figure 1. These are just a few ways in which emotional reaction can be measured. Emotions can be processed by a multitude of means, but each process results in change in behavior (Micu & Plummer, 2010; Figure obtained with permission*[1]).

Music has been reported to alter an individual's emotional state by changing certain elements. Fast tempo, high pitch, high frequency energy content, and fast tone attacks reportedly elicit anger. Fast tempo, medium sound level, fast tone attacks, and bright timbre reportedly elicit happiness. Slow tempo, dark timbre, low pitch, and slow

[1] *Reprinted with permission from "Measurable emotions: How television ads really work" by Micu & Plumber, 2010 Journal of Advertising Research, 50(2), 137-153, Copyright 2010 by The ARF.

attacks reportedly elicit sadness (Eerola, 2012). Although, researchers seem to agree music can elicit emotions in listeners (Juslin, Liljestrom, Vastfjall, Barradas, & Silva, 2008), few researchers have investigated how the responses occur.

Researchers (e.g., Strauss & Allen, 2008) have studied how words are related to emotion, but they have mainly focused on degree of pleasantness. Other factors including emotional intensity, word frequency, and word length in music have not been widely studied or as in depth (Strauss & Allen, 2008). Additionally, music lyrics are often not distinguished from other musical elements including tone, melody, and genre (Limbaugh, 2019), which has resulted in few researchers reporting the effect of music lyrics on emotional experience (Hu, Downie, & Ehmann, 2009).

2. DESIGN AND METHODOLOGY

2.1. Summary of Methods

The framework, build, and data collection for this study spanned a 17-month period between February 2018 and June 2019. Data collection began by examining, collecting, and researching popular music, meeting with industry professionals, and attending radio and music conferences and events (Country Radio Seminar [2018, 2019], NAB/RAB Radio Show [2018, 2019], NAB Show [2018], MusicBiz [2018], Worldwide Radio Summit [2018], and Country Music Awards [2018]). Data collection then advanced to the narrowed research of the three most consumed music genre formats for my target age group, resulting in three methods used in this study.

The three methods used in this study, included a sentiment analysis, a perception analysis, and a survey. The sentiment analysis was performed using the statistical platform, R, to measure the polarity of impactful words in music lyrics. The words were measured in clusters of five words which were assigned a score or category based on three tidytext lexicons.

I used Dialsmith's Perception Analyzer software and hardware to measure human response to music by collecting and analyzing real-time, continuous audience feedback. Each participant used a wireless, hand-held dial that reported a second-by-second score from 0 to 100. To the left, 0 indicated an extreme dislike score, and to the right, 100 indicated an extreme like score. While listening to the predetermined song selection, the participant moved their dial from left to right to indicate their opinion of what they were hearing. Reponses were recorded continuously.

Each participant completed two surveys. The first was a survey composed of 16 questions to aid in better understanding the participant. Questions about music preference, music background, age, and media consumption were asked. Age reported in the first survey was used to determine the eligibility of the person to participate in the study (i.e., between the ages of 18-34).

The second survey was administered in conjunction with the perception analysis of music and consisted of four introductory questions, followed by four distinct questions that were repeated after each of the nine songs. Four questions that were repeated after each song related to song familiarity, song polarity, emotions elicited from the song, and the degree to which the emotions were felt. The participants who ranked high in degree were then asked a follow up question asking what caused that emotional reaction. This section of the survey was used to add follow up information to help the researcher better understand why the participant rated the song the way they did with the Dialsmith dials.

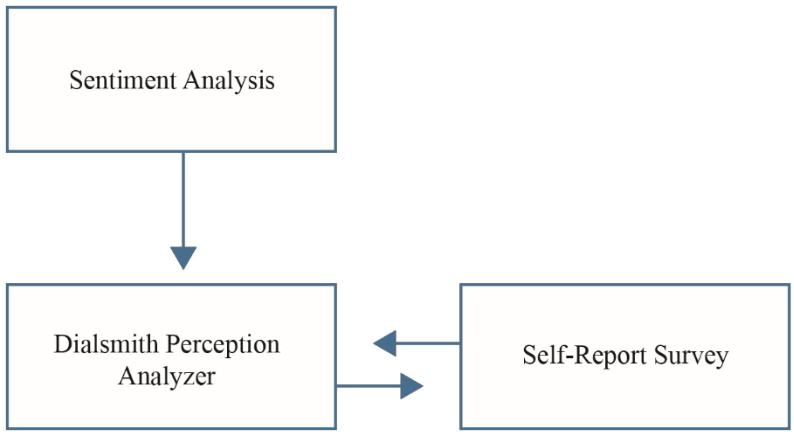

Figure 2. This is the order in which the methods are described. The sentiment analysis will be the first step, then the Dialsmith Perception Analyzer, and the self-report survey. The survey and the dials will be used together.

When measuring variables accurately, ideally the measured values should have the same meaning over time and across situations. There are often discrepancies (measurement error) between numbers used to represent what is being measured, and the actual value of what is being measured (Field, 2016).

There are two properties that are used to help keep measurement error to a minimum and suggest that the measure is appropriate. These properties include validity and reliability. Validity indicates whether the instrument or questionnaire measures what it is intended to measure. Reliability indicates whether the instrument or questionnaire can be interpreted consistently across different situations. In order to be valid, the measurement must first be reliable. Validity and reliability are addressed for each method.

2.2. Research Aims, Questions, and Objectives

The purpose of this study was to explore the lyrical content of music and human responses to music lyrics. Three aims guided this study to develop an in-depth understanding of how sentiment and human response to music lyrics lead to the success and emotional response of music. Five research questions were used to direct this study.

Aim 1: Describe the relationship between sentiment analysis and human response to music lyrics.

RQ.1: What sentiments are present in contemporary lyrical music?

RO1.1: Describe the positive and negative aspects of sentiments in lyrics.

RO1.2: Describe the emotional elements in lyrics.

RQ.2: How do humans respond to lyrics in music?

RO2.1: Describe how humans respond to positive and negative aspects of sentiments in lyrics.

RO2.2: Describe how humans respond to emotional elements in lyrics.

RQ.3: What are the relationships among sentiments in lyrics and humans' response to lyrics?

RO3.1: Describe the relationship between positive aspects of sentiments in lyrics and humans' responses to sentiments in lyrics.

RO3.2: Describe the relationship between negative aspects of sentiments in lyrics and humans' responses to sentiments in lyrics.

Aim 2: Describe the internal and external factors that play a role in response to music lyrics.

RQ.4: Are demographic, psychographic, environmental, and/or behavioral factors predictors of human response to lyrical music?

RO4.1: Describe the relationship among demographic factors and humans' responses to sentiments in lyrics.

RO4.2: Describe the relationship among psychographic factors and humans' responses to sentiments in lyrics.

RO4.3: Describe the relationship among environmental factors and humans' responses to sentiments in lyrics.

RO4.4: Describe the relationship among behavioral factors and humans' responses to sentiments in lyrics.

Aim 3: Describe how sentiment, human response, and internal and external factors predict the performance of music on the Billboard charts.

RQ.5: Are sentiments in lyrics predictors of year end performance on music charts?

RO5.1: Describe the relationship among sentiments in lyrics and year end performance on music charts.

H_{O1}: There will be no significant differences in sentiments among genres.

$\mu_1 = \mu_2 = \mu_3$

Country (μ1) = Pop (μ2) = Urban (μ3)

$H_{A1.1}$: There will be significant differences in sentiments among genres

$\mu_1 \neq \mu_2 \neq \mu_3$

Country (μ1) ≠ Pop (μ2) ≠ Urban (μ3)

$H_{A1.2}$: Positive sentiments in pop (μ_2) will be greater than in country (μ_1)

$\mu_2 > \mu_1$

Pop (μ2) > Country (μ1)

$H_{A1.3}$: Positive sentiments in pop (μ_1) will be greater than in urban (μ_3)

$\mu_2 > \mu_3$

Pop (μ2) > Urban (μ3)

$H_{A1.4}$: Positive sentiments in urban (μ_3) will be greater than in country (μ_1)

$\mu_3 > \mu_1$

Urban (μ3) > Country (μ1)

$H_{A1.5}$: Negative sentiments in country (μ_1) will be greater than in pop (μ_2)

$\mu_1 > \mu_2$

Country (μ1) > Pop (μ2)

$H_{A1.6}$: Negative sentiments in urban (μ_3) will be greater than in pop (μ_1)

$\mu_2 > \mu_3$

Urban (μ3) > Pop (μ2)

$H_{A1.7}$: Negative sentiments in country (μ_1) will be greater than in urban (μ_3)

$\mu_1 > \mu_3$

Country (μ1) > Urban (μ3)

H_{O2}: There will be no significant differences in sentiments between years.

$\mu_1 = \mu_2$

2017 (μ1) = 2018 (μ2)

$H_{A2.1}$: There will be significant differences in sentiments between years.

$\mu_1 \neq \mu_2$

2017 (μ1) ≠ 2018 (μ2)

H_{O3}: There will be no significant differences in sentiments between sentiment analysis and perception analyzer.

$\mu_1 = \mu_2$

Sentiment analysis (μ1) = perception analyzer (μ2)

H_{A3}: There will be significant differences in sentiments between sentiment analysis and perception analyzer.

$\mu_1 \neq \mu_2$

Sentiment analysis (μ1) ≠ perception analyzer (μ2)

3. SENTIMENT ANALYSIS

The next section describes the pieces of a sentiment analysis, how it is used, how it was used in the study, and the sources that the sentiment analysis pulled from. A sentiment is a long-term opinion of a certain topic, person, or entity (Soleymani, Garcia, Jou, Schuller, Change, & Pantic, 2017). A sentiment analysis focuses on the measure of an opinion's polarity, being positive or negative, through automatic recognition. Sentiment analysis is currently being used commercially to summarize reviews and customer opinions. Being able to understand people's opinions toward products, topics, and/or entities is vital for companies to understand how their brand or product is perceived (Soleymani et al., 2017).

Sentiment analysis has become essential in enhancing sales and improving a company's marketing strategies, identifying shifts, and analyzing trends in political strategy planning, and even forecasting stock market momentum. Algorithms within the analysis employ Natural Language Processing to use resources including sentiment or emotion-based lexicons to model textual documents (Giatsoglou, Vozalis, Diamantaras, Vakali, Sarigiannidis, & Chatzisavvas, 2011).

Sentiment analysis can be conducted at three different levels. The first is document level, this analyzes a whole document. The second is sentence level, this looks at individual sentiments within a sentence. The third is feature level, which analyzes attributes or features of products within a sentence or a document (Li, 2017). This research used sentence level analysis, specifically looking at words in clusters of five.

A sentiment analysis was used to investigate emotions associated with words in song lyrics. I measured sentiment associated with words by assigning a numeric value between -5 and +5. Words considered to be neutral were not assigned a value and were not included in the analysis. Whereas, words considered to be other than neutral were assigned values based on intensity (1 to ±5) and direction (negative or positive) of the sentiment. Three tidy data sentiment lexicons were used that are included in the tidytext [*sic*] package (built under R version 3.5.2). Packages are extensions for R source code that can be downloaded from the CRAN (Comprehensive R Archive Network) repository. Once packages are downloaded, they are then installed using RStudio, which is a free and open-source integrated development environment for R, a programming language for statistical computing and graphics. Each sentiment lexicon and analysis are described in a subsequent section.

3.1. Reliability and validity

Reliability

The reliability of a measure means that it can be repeated while staying consistent over and over again (Li, 2017). Sentiment analyses for the three lexicons were tested more than one dozen times, with the results changing once. The change in this result was because of the addition of eight more stop words to the analysis. After the change, the analysis was run four more times with the same results.

Sentiment analyses are good at detecting mood signals such as positive or negative sentiment, but are not very reliable when it comes to tracking and inferring a relationship between moods and targets. (Ding & Pan, 2016). This research study found what sentiments were present in lyrical music in 2017 and 2018, and it did not look at tracking a relationship.

Validity

Counter to measuring consistency the way reliability does, validity refers to the extent to which an instrument measures what it is supposed to (Li, 2017). The three tidytext sentiment lexicons used in this research (Bing, AFINN, NRC), were constructed by either crowdsourcing or by one of the authors, and were validated using combinations of more crowdsourcing, restaurant or movie reviews, or Twitter data (Silge & Robsinson, 2017). It may be difficult to apply these sentiment lexicons to older pieces of work (more than 100 years), because language has changed since then, and will result is less accurate results. However, this research focused on music lyrics from 2017 and 2018. Therefore, the sentiment results were likely to be accurate.

There are a few caveats that hinder a sentiment analysis. One, not all English words are in the lexicons because many English words are pretty neutral (Silge & Robinson, 2017). Two, qualifiers before words are not accounted for, for example “not true” or “no good”. Sarcasm and negation are not sustained. Third, large text documents have an effect on the analysis. Paragraphs may range from being positive in one, to the next being negative, and average out to zero. In this case, sentence-sized or paragraph-sized text works best.

3.2. Media characteristics

Billboard and Nielsen

Song lyrics were the unit of analysis for this subsection. Therefore, the characteristics of the songs selected are described, rather than the subjects or participants. Songs included in this subsection appeared on the Billboard charts between January 2017 and December 2018. The Billboard charts rank and define the most popular music consumed, by genre. Billboard charts are the most referenced music ranking system in the world (Billboard, 2019a).

Each week, Billboard ranks the most popular albums, songs, and artists by key fan interactions. Interactions include album sales, audio streaming, digital sales and downloads, radio airplay, and streaming social interactions through Facebook, Twitter, Vevo, YouTube, Spotify, and other music streaming services (Billboard, 2019b).

Each Billboard chart includes songs ranked by a) radio airplay and audience impressions measured by Nielsen Music; b) sales data provided by Nielsen Music; and c) streaming activity, reported by online music sources (Billboard, 2019a). Billboard charts include current album, song, and/or artist ranks, previous rank, number of weeks charted, artist's name, and peak rank (Bhattacharjee, Gopal, Lertwachara, & Marsden, 2006).

Billboard and their partners, including Nielsen Broadcast Data Systems (BDS), Nielsen SoundScan, and Next Big Sound, track radio airplay, online streaming, and music sales data year-round. Nielsen Music tracks point-of-sale transactions in the same way merchants track inventory. Therefore, when merchandise sales happen at a concert venue during an artist's performance, the purchase, in essence, serves as a ballot being cast for the artist on the Billboard charts (Billboard, 2019c).

Nielsen is considered the authority in tracking music sales in stores and digitally (Nielsen, 2019). Nielsen Music compiles data for Billboard's sales charts from music retailers that make up more than 90% of the United States' retail market (Billboard, 2019c). With 39,000 retail stores (as of 2019), Nielsen helps record labels, music publishers, recording artists, artist management, and performance rights organizations (PROs) quantify and understand what and where consumers are buying music. In addition to music consumption data, Nielsen reports radio airplay and online streaming data, which are then used by Billboard to produce their charts. The most popular Billboard charts include, *Year-end 2018*, *Holiday*, *Hot 100 60th Anniversary*, *Hot100*, *Billboard 200*, *Artist 100*, *Greatest of All Time*, *Pop*, *R&B/Hip-Hop*, *Latin*, *Dance/Electronic*, *Country*, *Rock*, *Web*, *Breaking and Entering*, *Christian/Gospel*, and *International*.

Genius

The Genius website was the source for song lyrics used in this study. Genius Media Group is a media company based out of Brooklyn, New York that collects its information from more than 2 million contributors, including editors, the community, their in-house creative team, and recording artists. More than 100 million people access Genius each month (Genius, 2019a). Genius currently has integrations with YouTube, Facebook, Twitter, and Spotify, making it a resourceful site for streaming and social platforms. Other music lyric retrieval sites that were looked at included Metro Lyrics, Sing365, A-Z Lyrics, and Lyrics Mode. These other sites did not include the same amount of background information about the song or artist and did not have the same kind of social interaction that Genius offered.

Genius was first introduced in 2009 as a site solely based on annotating rap lyrics. Since then, the Genius site has been expanded to include all music genres and provides more insight to songs than the lyrics alone. With the "Intel" (Intel is a Genius specific term) from their contributors, Genius offers information from more than 25 million verified songs, albums, artists, and annotations, which makes it the most accurate lyrics and knowledge database on the internet (Genius, 2019a). The Genius in-house creative team works with artists and the community to create videos, events, features, news, and a variety of music knowledge. These works are then published across the website.

Once a user (scholar), creates an account in Genius, he or she is then able to follow other scholars and artists. Each scholar has a number attributed to his or her username, which is referred to as an "IQ". Scholars earn IQ points by adding knowledge to the site. Knowledge can be earned in a number of ways, including annotated lyrics and cover artwork, artist, song, or album bios, adding song facts, transcribing lyrics, profile edit suggestions, and answering Q&A questions (Genius, 2019b). This information is then verified by Genius editors before it is accepted and published. IQ points are given when the Genius editors accept the annotation and points are also lost if an editor rejects the annotations. Each piece of knowledge added by a scholar is worth a certain amount of points. With the collection of IQ points, milestones are accomplished, and a level of credibility and access is added to a scholar's account. Milestones start at 100 points and increase up to 3,000 points. It is possible to accumulate IQ points beyond 3,000. The more points a scholar has, the more credible they are. Everyone on the site can view each other's points.

All songs and artists have these contributors on each page. Users are able to see leaderboards on each page, as well as recent activity. Within this activity, users can see how many people are currently viewing that page, how many total views the page has, and how many contributors have added to the page. With the data from these activity reports, Genius, much like Billboard, has their own set of charts. These charts vary by song, artist, album, lyrics, and genre and are classified by today, this month, and all time. However, unlike Billboard, Genius ranks these charts based on page views, not by airplay, streams, or sales. Therefore, an artist's top charted song on Billboard, may not necessarily be his or her top charted song on Genius.

3.3. Sampling procedures

Song lyrics analyzed in this study were identified from the 2017 and 2018 Billboard music charts. In 2017, 52 weekly charts were published. Whereas, 53 weekly charts were published in 2018. The number of positions available on each chart varies by genre (e.g., the weekly *Country Airplay* chart includes 60 positions; whereas, the *Adult Top 40* chart includes 40 positions). Billboard archives decades of music charts, week-by-week and year-by-year, all of which are publicly accessible through their website.

In 2018, Nielsen Music (Kelly) reported that 92% of American Millennials, between the ages of 18 and 34, consumed music through AM/FM radio during 2017.

Further, AM/FM radio has been reported to reach more than 71,000 Millennials each month (Kelly, 2018). In Nielsen Music's report (Kelly, 2018), three of the top consumed radio formats in 2017 were country (15.6%), pop CHR (contemporary hits radio, 12.9%), and urban contemporary (6.7%). Therefore, to narrow the scope of this study, the three highest ranked genres for American Millennials (ages 18 to 34) were included in the analysis.

Billboard's airplay charts (AM/FM radio) use Nielsen Music's information to compile their charts using radio stations in more than 140 markets across the U.S. Airplay charts are comprised of a song's number of plays in a given format for that specific week. The Nielsen Music data and the audience charts are then cross-referenced with listener information created by the Arbitron ratings system. This determines the exact number of audience impressions made for each song play. Therefore, "a song that plays at 4 a.m. does not count as much as one played at 4 p.m., and a station with a large audience will influence the chart more than either a station in a smaller market or one with a specialized format that attracts less audience" (Billboard, 2019c).

Country

There are seven country charts in Billboard: hot country songs, country airplay, country digital song sales, country streaming songs, top country albums, bluegrass albums, and Americana/folk albums. Because this study focused on the top three consumed formats for Millennials, hot country songs (Radio Airplay + Sales Data + Streaming Data) will be the chart included in this analysis. Nielsen Music measured this chart by audience impressions, sales data, and streaming activity data from online music sources. The hot country songs chart has 50 positions available per week for songs to chart. In 2017, there were 52 charting weeks published, making a total of 2,600 charting song positions available. In 2018, there were 53 charting weeks published, resulting in a total of 2,650 positions available on the charts (Billboard, 2017a; 2018a).

Pop CHR

Pop CHR is commonly known and referred to as “Pop” on the Billboard charts. There are three pop music charts on Billboard, including adult contemporary, adult top 40, and mainstream top 40. Mainstream top 40 is referred to as “Pop Songs” and “Top 40/CHR” on Billboard, and it is ranked by mainstream top 40 radio airplay detections as measured by Nielsen Music. Only 40 positions are available on each week’s chart. There were 52 published charting weeks in 2017, making a total of 2,080 available song charting positions. There were 53 published charting weeks in 2018, making a total of 2,120 available song charting positions (Billboard 2017b; 2018b).

Urban Contemporary

Billboard refers to urban contemporary as R&B/Hip-Hop. There are 13 charts for R&B/Hip-Hop on Billboard, including hot R&B/Hip-Hop songs, R&B/Hip-Hop airplay, R&B/Hip-Hop digital song sales, R&B/Hip-Hop streaming sales, hot R&B songs, R&B streaming songs, hot rap songs, rap streaming songs, top R&B/Hip-Hop albums, R&B albums, rap albums, adult R&B songs, and rhythmic songs. The hot R&B/Hip-Hop songs (Radio Airplay + Sales Data + Streaming Data) chart was the chart used. Nielsen Music measured this chart by audience impressions, sales data, and streaming activity data from online music source. This chart has 50 song positions available each week to chart. In 2017, there were 52 published charting weeks, making a total of 2,600 available song charting positions. In 2018, there were 53 published charting weeks, making 2,650 available song charting positions.

From these three formats, there were a total of 7,280 total song charting positions in 2017, and a total of 7,420 song charting positions available in 2018. An overall total of 14,700 song charting positions were available during the years 2017 and 2018 in the top three genres consumed by Millennials (Billboard 2017c; 2018c).

Summary statistics for each lexicon were reported to address RQ1, including frequency, percent, total, mean, median, mode, standard deviation, minimum, and maximum. Once the sentiment analysis was complete, I selected nine songs (three from each genre) for the moment-by-moment testing (RQ2) and the survey (RQ3). From each genre, there is one song with a high positive score and a low negative score (to represent a positive song), one with a high negative score and a low positive score (to represent a negative song), and one with a low positive and low negative score (to represent a neutral song. Songs that were not categorized for the six basic emotions (happiness, surprise, sadness, fear, anger, and disgust [Poels & Dewitte, 2006]) were eliminated.

There were five criteria when making the song selection that were not considered in previous research. The first criterion was that each genre needed to have a female artist. The second criterion was that songs selected should not be present in more than one genre. The third criterion was that one song in each genre should have charted in both 2017 and 2018. The fourth criterion was that songs should range in chart rank. Those that charted in both 2017 and 2018 should have a varying rank number. The fifth criterion was that songs should have more than a 50 analyzed word count, and less than a 200 analyzed word count. Words were analyzed in clusters of five. Songs with less than 50 analyzed words were too repetitive, whereas songs with more than 200 analyzed words were overwhelming to listen to. The key to keeping listeners engaged with a song is the repetition of hooks and choruses (Spencer, 2019), but too much repetition can make the song boring.

Criterion 1

The first criterion was that there needed to be a female artist in each genre. Gender inequality amongst several genre formats in the music industry has been a topic of discussion in recent years. Rolling Stone conducted a study on the most popular music on the Billboard Hot 100 chart over a seven-year period (2012-2018). They found that only seventeen percent of the music was made by women. This includes all facets of music such as performers, songwriters, and producers (Wang, 2019).

Table 1 is a summary of female artist that charted on the Billboard Year-End charts across the three studied genre formats in 2017 and 2018. There are three female categories in Table 1, solo artist, solo artist repeated, and band/group/duet. The first category, solo artist, is a single performer. Examples of a solo artist include Carrie Underwood, Taylor Swift, and Cardi B [*sic*].

The second category, solo artist repeated, is a solo artist who is on the same chart more than once for different songs. In some cases, an artist was listed on a chart more than one time in one year (e.g., Maren Morris performed *I Could Use a Love Song* in

2017, and Maren Morris also performed *80s Mercedes* in 2017). Female presence will only be counted once per artist in Table 1 to describe the representation of female artists.

The third category, band/group/duet, are multiple artists performing together. A band in Table 1 has two or more artists that have all female members, or both male and female members, e.g. Maddie & Tae, and Sugarland. A group in this table is made up of multiple artists who do not belong to a single band. This can be present in songs featuring more than one person, for example *Eastside* by benny blanco [*sic*] featuring Halsey and Khalid, or *Bang Bang* by Jessie J featuring Nicki Minaj and Ariana Grande. A duet in this table is two artists or bands, either both female, or male and female that are present in one song. An example of this would be *Meant to Be* by Bebe Rexha featuring Florida Georgia Line, or *Wild Thoughts* by DJ Khaled featuring Rihanna.

Table 1
Female Representation

	Country			Pop			Urban		
Category - Year	*n* [a]	*f*	% [b]	*n* [a]	*f*	% [b]	*n* [a]	*f*	% [b]
Solo Artist									
2017	100	10	10	50	11	22	100	4	4
2018	100	10	10	50	13	26	100	11	11
Solo Artist Repeated [c]									
	100	3	3	50	0	0	100	1	1
	100	3	3	50	5	10	100	5	5
Band/ Group/ Duet									
	100	3	3	50	6	12	100	2	2
	100	4	4	50	9	18	100	7	7

Notes. [a] = total songs included in the respective sample for the noted genre and year; [b] = percent of total songs included in the respective sample for the noted genre and year; c = artists who recorded more than one song that appeared on the same chart

In 2017 and 2018, solo female artists represented 10% ($f = 10$) of the Billboard Year-End Hot Country Songs chart ($n = 100$). Out of those 10, 3 of the female artists who charted had multiple songs, which means 7% ($f = 7$) of the 2017 and 7% ($f = 7$) of the 2018 country charts were solo female artists. Outside of the solo female artists, the band/group/duet category had a 3% ($f = 3$) female presence in 2017. In 2018, there was a 4% ($f = 4$) female presence.

The Pop Songs chart had a few more females present than the Hot Country Songs chart. Of the 50 charted songs, female solo artists recorded 22% ($f = 11$) in 2017, and 26% ($f = 13$) in 2018. Of the 11 female solo artists in 2017, none of them appeared more than once, but 5 of 13 had more than one song on the chart in 2018. In the band/group/duet category, 12% (6/50) had a female artist presence. In 2018, there was an 18% (9/50) female presence.

There was an increase in female presence from the 2017 to the 2018 Hot R&B/Hip-Hop Songs chart. In 2017, there was a 4% (4/100) female solo artist presence, whereas in 2018 there was an increase to 11% (11/100) female solo artist presence. From the 4 female solo artists in 2017, 25% (1/4) appeared more than once. In 2018, 45% (5/11) of the female solo artists were repeated. Apart from the female solo artists, 2% (2/100) females were present in 2017 in a band, group, or duet. In 2018 there was an increase to 7% (7/100) of female presence in a band, group, or duet.

Female artists make up a very low percentage of the total positions available on the Billboard Year-End charts. Because of this, it was important to include at least one female in each of the three studied genres to provide a wider variety of artists among each genre.

Criterion 2

The second criterion was that the songs selected should not be present in more than one genre. Each music genre needs to be independent of itself. If a song crosses over between genres, then it would be difficult to test differences amongst genres. The artist may be present in more than one genre, however each of the nine songs selected belonged to only one genre in the three charts. Among the 500 songs from the three music charts in 2017 and 2018, 28 songs were present in two different genres.

Criterion 3

The third criterion was that one song in each genre should have charted in both 2017 and 2018. Each of the three genres had a number of songs that charted in both 2017 and 2018. There are a few reasons for this repetition. First, the three charts selected are off of sales data + streaming data + radio airplay. If a song continued to have high numbers in each or any of these categories that would prolong its position on the charts.

A second reason for this repetition is a song's release date. A song that was released later in the year, for example November or December, has a better chance of staying on the charts the following year over a song that was released early in the year like January or February. This is because the song most likely still has momentum even after a couple months of being released.

A third reason for this repetition is touring. Whether it be a radio media tour (RMT) or a concert tour, the artist continues to promote their music or song. A radio media tour is a series of pre-arranged interviews at targeted radio stations that reach listeners across the nation (Media Tracks Communications, 2019). A concert tour is a string of live performances by an artist or band to gain exposure, increase revenue, build a fan base, and promote releases (Frost School of Music, 2016).

A song that charted in both 2017 and 2018 has a better chance of a higher familiarity with the listeners. The longer the song has been around and has shown success, the more of a chance the listener has heard it. Familiarity was used in a correlation with song polarity and emotions experienced by participants during their music test.

In the Hot Country Songs chart, 49 (n = 200) songs charted in both 2017 and 2017. In the Pop Songs chart, 33 (n = 100) songs charted in both 2017 and 2018. In the Hot R&B/Hip-Hop Songs chart, 56 (n = 200) songs charted in both 2017 and 2018.

Criterion 4

The fourth criterion was that songs should range in chart rank. Those that charted in both 2017 and 2018 should have a different rank number. Songs were selected that spanned in range in chart rank amongst their genre. This was a key factor in level of popularity and familiarity. Songs that charted in both 2017 and 2018 also had different chart ranks between those two years.

Criterion 5

The fifth criterion was that songs should have more than a 50 analyzed word count, and less than a 200 analyzed word count. Word count was a large component of sentiment scores. Songs that had a larger word count, had a greater number of positive and negative categorized words in the Bing lexicon. They also had more individually valued words for the AFINN lexicon, and a greater range in emotion categories for the NRC lexicon.

Songs that had a low word count tended to be more neutral than positive or negative. They did, however, vary amongst chart rank. Word were analyzed in clusters and five and excluded profanity, numbers, explicit language, non-English words, slang, and words with one or two letters.

3.4. Sample size, power, and precision

To condense this study even further, instead of looking at the charting song positions available week-by-week in these three formats, the year-end charts were compiled. There were 100 song position spots available in both 2017 and 2018 for the country and R&B/Hip-Hop genre formats. The Pop genre had 50 song position spots available in both 2017 and 2018 (Table 2).

Table 2
Billboard Charting Positions

Year	Country	Pop	R&B/Hip-Hop
2017	100	50	100
2018	100	50	100

The year-end charts included a total of 500 songs. These year-end charts are the most popular songs within each genre ranked by radio airplay, sales data, and streaming data detections reported by Nielsen Music.

Web Scraping

To start the sentiment analysis, I collected the lyrics of each song included on Billboard's 2017 and 2018 year-end charts for the country, pop CHR, and urban contemporary genres. To do this, I used the tidyverse (Tidyverse, n.d.) and rvest (Wickham, 2016) packages in R, as well as a web scraping tutorial (Dsilva, 2018). The data were then changed to match the tidy text format. The tidy text format is a table with one token per row (Silge & Robinson, 2017). A token is defined as being a meaningful unit of text, such as a word, that is used in analysis. Tokenization takes that text and splits it into tokens. Other ways that text is stored in analyses is as strings or in a document-term matrix. However, in text mining the token stored in each row is usually a word, n-gram, sentence, or paragraph (Silge & Robinson, 2017).

RStudio's rvest package was used to locate and scrape information from specific web pages (Wickman, 2016). The tidyverse package is a collection of R packages that all share the same structure for design, grammar, and data (Tidyverse, n.d.). Web scraping is a process of retrieving unstructured data from websites, then formatting it into structured data that can be used for further analysis (Dsilva, 2018). The tidy text format makes data easier to analyze and much more effective. By using a specific structure: 1) each variable is a column, 2) each observation is a row, and 3) each type of observation unit is a table (Silge & Robinson, 2017).

R script was written to reference the Billboard and Genius websites with an inner and outer loop coded that allowed searches by year and by genre. Using the rvest package (Wickham, 2016), the R script searched specific charts (i.e., 2017 year-end Hot

Country chart), and then generated a table with the rank order of artist and song. Next, code in the R script searched each artist page on Genius, found the charted song it needed, and collected the lyrics for that song. Then, the R script generated a new table with the rank order of artist, song, and the lyrics associated with that song. A find and replace within the script was used to delete any unwanted text from the lyrics (e.g., [Verse 1] and [Chorus]). After the lyrics were collected, the tidyverse package was then used to transform the data from the table into a usable tidy text format, which was necessary for the sentiment analysis.

Figures 3, 4, 5, and 6 are the steps for the sentiment analysis loop. Figure 3 is an image from the 2017 year-end pop chart with the top five songs listed. The Billboard charts made up the outer loop in the R code. Figure 4 is an image from Genius of *Shape Of You* by Ed Sheeran song lyrics. The song lyrics from Genius made up the inner loop in the R code. Figure 5 is an image of the R code that I generated for the inner and outer loops that retrieve the Billboard charts and subsequent Genius lyrics for the pop 2017 year-end Billboard chart. The code was run in order to create the data frame (tibble) that will include song rank, song title, artist name, and song lyrics. Figure 6 is the data frame (tibble) that R generated from the inner and outer loop.

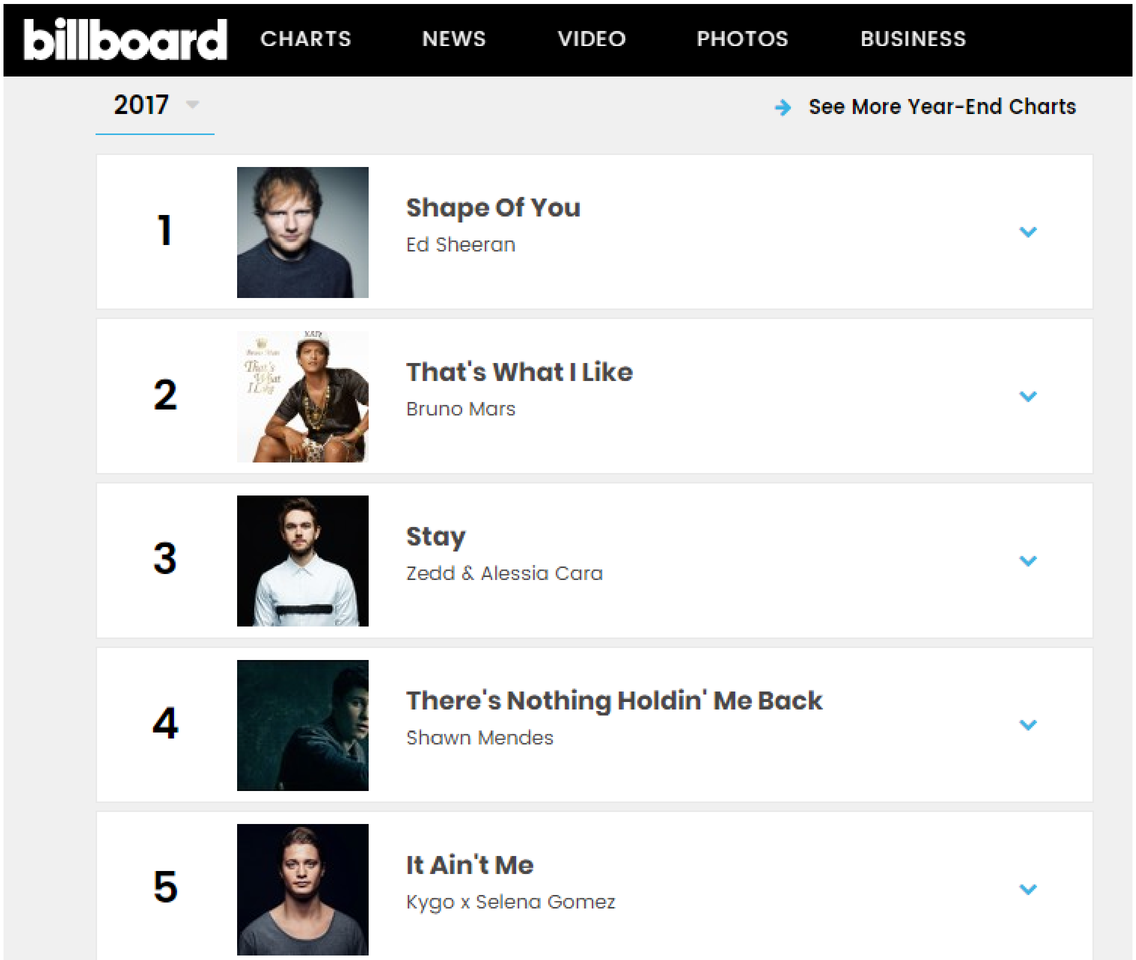

Figure 3. These are the top five charted songs for the 2017 year-end Billboard pop chart. The rank number, song title, artist name, and artist photo are available in order from most successful, to least (Billboard, 2017a).

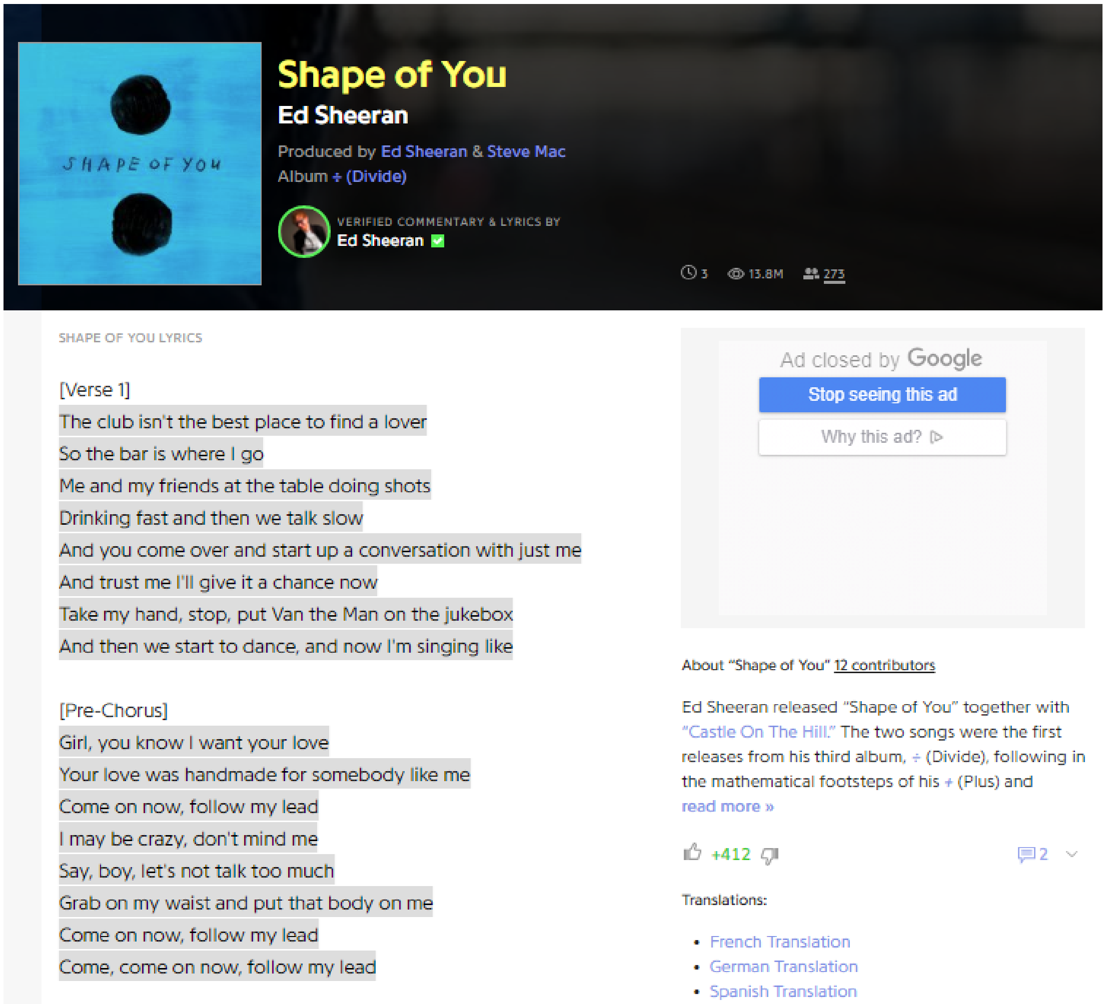

Figure 4. The lyrics for *Shape of You* are verified on Genius. Insights into the song and artist are provided on the song page as well. The lyrics continue to the end with each stanza (verse, chorus, etc.) indicated in sequence of the song (Genius, 2019c).

```r
for (i in 1:50) {

  pop_lyrics_scraped <- read_html(pop_genius_urls[i]) %>%
    html_nodes(".lyrics p") %>%
    html_text()

  pop_lyrics_scraped <- gsub("\\[[&,:,A-Z,a-z, ,0-9,\\-]*\\]","",pop_lyrics_scraped)
  pop_lyrics_scraped <- gsub("\n"," ",pop_lyrics_scraped)

  pop_song_name <- read_html(pop_genius_urls[i]) %>%
    html_nodes(".header_with_cover_art-primary_info-title") %>%
    html_text()

  pop_artist_lyrics <- rbind(pop_artist_lyrics, tibble(Pop_Rank = top_pop_artists$Pop_Rank[i],
                                                       Pop_Artist = top_pop_artists$Pop_Artist[i],
                                                       Pop_Song = pop_song_name,
                                                       Pop_Lyrics = pop_lyrics_scraped ))


  # Sys.sleep(15)
}

pop_artist_lyrics
```

Figure 5. This is the R code that was generated to create the inner and outer loop for the Billboard 2017 year-end pop chart. The loop continuously scrapes information from the selected Billboard chart, then is directed to locate each song on Genius, and scrapes the song's lyrics. R then creates a data frame (tibble) with all the information it scraped from both Billboard and Genius.

```
> pop_artist_lyrics
# A tibble: 50 x 4
   Pop_Rank Pop_Artist            Pop_Song               Pop_Lyrics
      <dbl> <chr>                 <chr>                  <chr>
 1        1 Ed Sheeran            Shape of You           The club isn't the best place to find a lover So the bar is where I go Me and my friends at t~
 2        2 Bruno Mars            That's What I Like     " Hey, hey, heyI got a condo in Manhattan Baby girl, what's  hatnin'? You and your ass invite~
 3        3 Zedd & Alessia Cara   Stay                   " Waiting for the time to pass you by Hope the winds of change will change your mind I could ~
 4        4 Shawn Mendes          There's Nothing Holdi~ " I wanna follow where she goes I think about her and she knows it I wanna let her take contr~
 5        5 Kygo x Selena Gomez   It Ain't Me            I had a dream We were sipping whiskey neat Highest floor, The Bowery Nowhere's high enough So~
 6        6 Charlie Puth          Attention              " Woah-oah, hm-hmm You've been runnin' 'round, runnin' 'round, runnin' 'round Throwin' that d~
 7        7 Luis Fonsi & Daddy ~  Despacito              "[Letra de \"Despacito\" ft. Luis Fonsi] Ay, ¡Fonsi! ¡D.Y.! Ohhh, oh, no, oh, no, oh ¡Hey, ye~
 8        8 Niall Horan           Slow Hands             " We should take this back to my place That's what she said right to my face 'Cause I want yo~
 9        9 Zayn / Taylor Swift   I Don't Wanna Live Fo~ " Been sitting eyes wide open behind these four walls, hoping you'd call It's just a cruel ex~
10       10 Alessia Cara          Scars to Your Beautif~ "She just wants to be beautiful She goes unnoticed, she knows no limits She craves attention,~
# ... with 40 more rows
```

Figure 6. These are the data frame generated from the inner and outer loop for the Billboard 2017 year end pop chart. There are four columns that make up the data from, these columns include the song rank on the Billboard chart, then song artist, then song title. The song lyrics from Genius make up the last column.

Sentiment Analysis

Three tidy data sentiment lexicons are available and commonly used for sentiment analyses in R; AFINN (Liu, 2012), Bing (Nielsen, 2011), and NRC (Mohammad & Turney, 2013; Silge & Robinson, 2017). All three lexicons are based on single words. The AFINN lexicon assigns each word a value ranging from -5 to +5, with negative scores reflecting a negative sentiment and positive scores reflecting a positive sentiment. The Bing lexicon assigns each word one of two categories; either positive or negative. The NRC lexicon assigns each word one of eight emotional categories; joy, anger, sadness, disgust, fear, anticipation, positive, or negative (Silge & Robinson, 2017).

3.5. Design and measure

Research question one addresses the descriptions of what sentiments exist in contemporary music lyrics. In this cross-section research design, 500 songs in the Billboard Year-End charts from 2017 and 2018 were evaluated. As a case study research design, the 500 songs, and the lyrics associated with them, spanned across three formats, Country, Pop, and Urban. Three lexicons were used in a sentiment analysis that assigned either a score or category to each of the 500 songs. The lexicons that made up the sentiment analysis were AFINN, Bing, and NRC. The three lexicons ignored 124 stop words. These stop words included numbers, words with two letters or less, slang terms, profanity, gender specific words (boy, girl, woman, man), numbers, and curse words.

The cases were the Billboard year-end 2017 and 2017 Hot Country, Pop, and Hot Urban/Hip-Hop charted songs. The cross section were the years 2017 and 2018. The measures were the sentiment lexicons and the word scores and categories that each one provided. Music genre was given a variable name and a number for each genre (*MX0003* = 1, 2, or 3). Time was also given a variable name and was categorized by year (*T0001* = 2017 or 2018). Analysis was drawn from each lexicon, genre, and year. Reported was mean (*M*), standard deviation (*SD)*, minimum word count (min), maximum word count (max), and total word count (*MX0008),* including stop words, by genre (MX0003) and year (T0001).

AFINN

The AFINN lexicon found the most significant word in clusters of five, then assigned that word a predetermined integer value between -5 and 5. Negative scores were assigned to words scoring a -5 to -1. Positive scores were assigned to words scoring between 1 and 5. Words that scored neutral, or 0, were not included in the analysis. When analyzing the song as a whole, the sum of all the AFINN scores each word in the song were added together to assign the song as being either positive or negative. Variable codes are listed below.

MX0023 (AFINN): Method

MX0010 (AFINN – Included Word Count): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001])

MX0011 (AFINN - Index): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001])

MX0012 (AFINN – Lexicon Score): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001])

Bing

The Bing lexicon also found the most significant word in clusters of five. It then assigned the selected words in one of two categories, “positive” or “negative”. Words that did not categorize were given a value of “#N/A”, then later left blank. Variable codes are listed below.

MX0024 (Bing): Method

MX0013 (Bing – Included Word Count): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001])

MX0015 (Bing – Negative word count): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001])

MX0016 (Bing – Negative word ratio): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001])

MX0017 (Bing – Positive word count): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001])

MX0018 (Bing – Positive word ratio): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001])

NRC

The NRC lexicon looked at all words in the music lyrics. It then found predetermined categories to place the words into. There are ten emotion categories that NRC uses, of which are, "anger", "anticipation", "disgust", "fear", "joy", "sadness", "surprise", "trust", “positive” or “negative”. Some words (such as “feeling”) were categorized as more than one emotion. Variable codes are listed below.

MX0025 (NRC): Method

MX0019 (NRC – Sentiment): 01 = "anger"; 02 = "anticipation"; 03 = "disgust"; 04 = "fear"; 05 = "joy"; 06 = "sadness"; 07 = "surprise"; 08 = "trust"; 09 = “positive”; 10 = “negative”

MX0020 (NRC – Unique sentiment word number): *M*, *SD*, min., max. (total and by genre [MX0003] and year [T0001] and sentiment [MX0019])

MX0021 (NRC – Unique sentiment word text): qualifying word

MX0022 (NRC – Unique sentiment word count): Number of occurrences (total and by genre [MX0003] and year [T0001] and sentiment [MX0019])

4. DIALSMITH PERCEPTION ANALYZER

The next section describes the Dialsmith Perception Analyzer, what it is used for, how it was used in the study, what conclusions can be drawn from using this software. To address RQ2, I used continuous response measurement (CRM)—which is also known as the real-time rating (RTR) method—to measure participants' responses to music. This method continuously measures a moment-by-moment individual experience by evaluating the participant's opinion on a sample of media content (Seiffert-Brockman, & Jarolimek, 2017). RTR measurement does not require the respondent to evaluate an object at a single moment in time, instead it gives constant feedback from the respondent during an entirety of an episode (Keil & Ottler, 2017). As an individual formulates and changes his or her opinion of a piece of media, the RTR tracks and records each second of his or her response. With these recorded responses, it is then possible to see what caused shifts in opinion, what were the high and low ratings from the participants, and what aspects of the media made an impression.

CRM systems date back to the 1930s and are most commonly used today for media and advertising research, and are a major asset in political debates (Saks, Compton, Hopkins, & El Damanhoury, 2016). In social science, RTR bridges the gap between communication science and psychology (Seiffert-Brockman, & Jarolimek, 2017). Because the RTR analysis provides information about opinion change it, gives an insight on the "…moment-to-moment flow of mediated communication" (Seiffert-Brockman, & Jarolimek, 2017, p. 1). CRM research has been able to capture unfiltered, visceral reactions in respondents to all forms of media. The rich, quantitative data it provides informs and complements qualitative discussions, and helps both researchers and analysists see what specific elements, messages, arguments, and even mannerisms impact their audience (Dialsmith, 2013). This method is used extensively in audience and consumer research, message testing, and public opinion polling. With the aid of questionnaires, qualitative, and quantitative content analysis, these measurements draw a more in-depth level of insight in the cognitive processing of how media content is evaluated and the level of impression it creates.

Dialsmith

Dialsmith's Perception Analyzer (PA) software can be used to gather continuous, in-the-moment feedback in a variety of research questions. The PA system is purportedly an engaging and interactive research device that is currently used by 45 different countries (Dialsmith, 2019). Dial testing with the Perception Analyzer has been used for live moment-to-moment dial testing for real-time results in focus groups, in-depth interviews, media feedback, and other live research sessions. The portable PA system can be used to test groups ranging from one to more than one hundred participants (Dialsmith, 2019) and can be virtually set up in any environment.

The PA system was used while conducting an auditorium test to evaluate respondents' continuous change in opinion while listening to music. The PA system includes hand held, wireless dials and uses an interactive audience response system for collecting real-time, moment-to-moment feedback from each participant. The hand-held dials can be turned left or right to indicate a range of responses (e.g., like/dislike, agree/disagree, etc.; Dialsmith, 2013). The dials are designed to be easy to use and intuitive. There is also an online platform available for this, which includes a mouse-controlled on-screen slider to measure responses to media in surveys. This study used the Model V dial indicated in figure 7.

Figure 7. The Dialsmith dials have improved since they were introduced in 1985. For this study, I used the Model V dials. The dials are simple to use with very few buttons. The screen on the dials indicates the number the participant has turned their dial to for ease. The three buttons on the bottom were not used (Dialsmith, 2016*[2]).

[2] *Reprinted with permission from "Moment-to-moment research playbook" by Dialsmith, 2016. Copyright 2016 by Dialsmith.

The PA dials are set to record the responses once per second. For this section of the study, I used three discrete questions, nine moment-to-moment questions, and eleven instruction pages. Media being used in the moment-to-moment questions varied in length between three and four minutes.

4.1. Reliability and validity

Reliability

Attitudes and opinions are always changing. CRM measures a momentary evaluation, cognitive process, or mental state indicated by the participant. Most reliability tests assume that the current mental state of the participant is long-lasting, but CRM measures are not always stable (Lang, 1994). Overarching opinions about a particular communication event may be stable, but reactions measured each second vary because audience response measures the flow of cognitive response.

Ecological Validity

Ecological validity is a problem that exists commonly in laboratory measures of the communication process (Lang, 1994). In most research designs, the evaluation of messages made by participants are in a setting outside of the environment where they are typically received. During CRM, or moment-to-moment studies, messages are viewed by participants through an inputted device. Because of this, demands on attention produced by the task may alter participant response to the message (Lang, 2014).

Remote data collection systems, such as the Dialsmith PA software, can be moved from one location to another so that participants can respond in their natural environment such as work, home, school, and auditoriums. This does help alleviate the issue of ecological validity, but it does not eliminate it entirely.

Content Validity

Content validity is the degree to which a measure captures the full range of a concept (Lang, 1994). Attitudes, hedonic responses, or attention are unlikely to be captured by a single response scale, so if participants can only respond with a range of nominal selections, a very limited number of distinctions are able to be measured (Lang, 1994).

4.2. Media characteristics

The PA dials (hand held, wireless devices) are part of an interactive audience response system for collecting real-time, moment-to-moment feedback from respondents. The PA software provides instant feedback through the dials that is immediately available for analysis, enhanced discussion, and group interaction (Dialsmith, 2018). Participants used PA dials to respond to questions displayed on a screen using the PA software. The dials transmit a radio frequency to the receiver (console) that was connected to a Windows computer, which then recorded the answers.

Getting Started

Before starting the data collection process, equipment and software meeting or exceeding the minimum system requirements (see Dialsmith, 2018) were secured: The PA software was installed on a Windows-based laptop computer (Windows 10 with an Intel i7 processor, 32GB RAM, 1TB SSD, and a 2TB HDD). There also had to be one USB 3.0 port available for the external display device. An addition USB 3.0 or 2.0 port was also required for the PA to connect to the console. A third USB 2.0 or 3.0 port was recommended, but not required, for an external flash drive connection to transfer project files. To prevent a stall or delay in media display and/or data logging during collection internet connections, hibernation, and screen shut-downs were disabled or suspended during data collection to avoid any unanticipated disruptions from pop-ups, virus checks, disk scans, and/or disk back-ups.

After installing and launching the software package, there were three programs available to choose from in the main dialogue box. These programs are Edit, Collect, and Report. Each of these programs take the user through the process of executing a research project.

Edit

Edit is the first step in creating a new research project and was used to create the questionnaire. There are three different options for question types to choose from and an instruction option. The three different types of question options include discrete, scale, and moment-to-moment. All questions and answer choices need to be coded with variable names before starting this step.

Discrete questions use nominal or ordinal level data in which the participant can select a single answer from a set of possible responses. For example, a participant could be asked what gender they identify with. This is a demographics question, so the assigned variable code for the question will be D0002. The answer choices, along with their variable code are as follows: 1 = Male; 2 = Female; 3 = Non-binary / third gender;

4 = Prefer to self-describe, 5 = Prefer not to say. Figure 8 depicts the sections in which this information is to be placed.

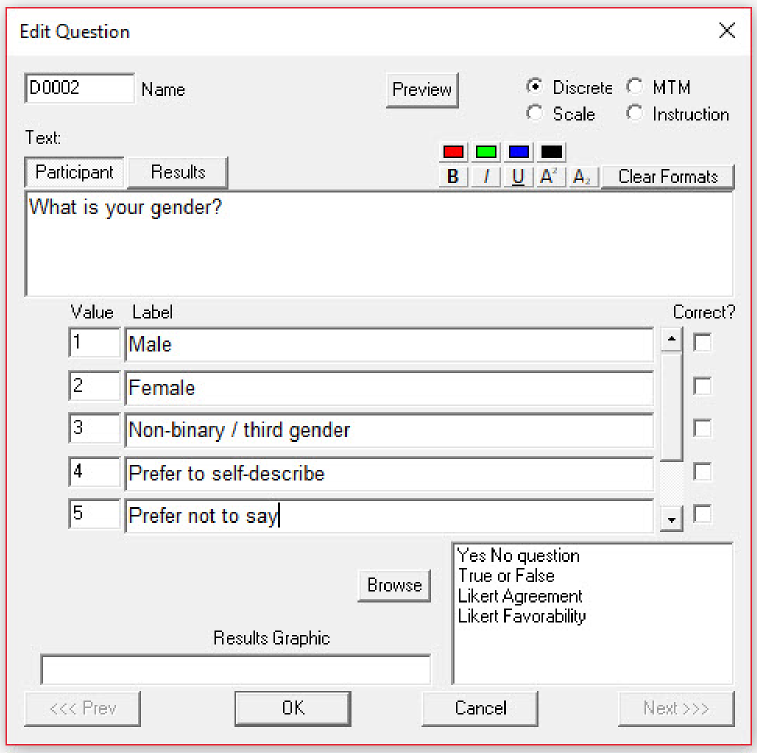

Figure 8. In edit mode, the Dialsmith PA software can be used to create questions with discrete (nominal) response options. This is a screenshot of the edit mode interface being used to create the gender question used in the auditorium dial tests.

Scale questions allow participants to rate their level of preference or opinion on a numeric scale. This type of question uses ordinal or interval level data. It is important to designate labels, or anchors, on opposing ends of the scale. For example, if the participants are asked how they would describe a song (UC0081), the answer choices would use a scale, ranging from 0 to 5, with 0 labeled as extremely negative, and 5 labeled as extremely positive. Figure 9 depicts the sections in which this information is to be placed.

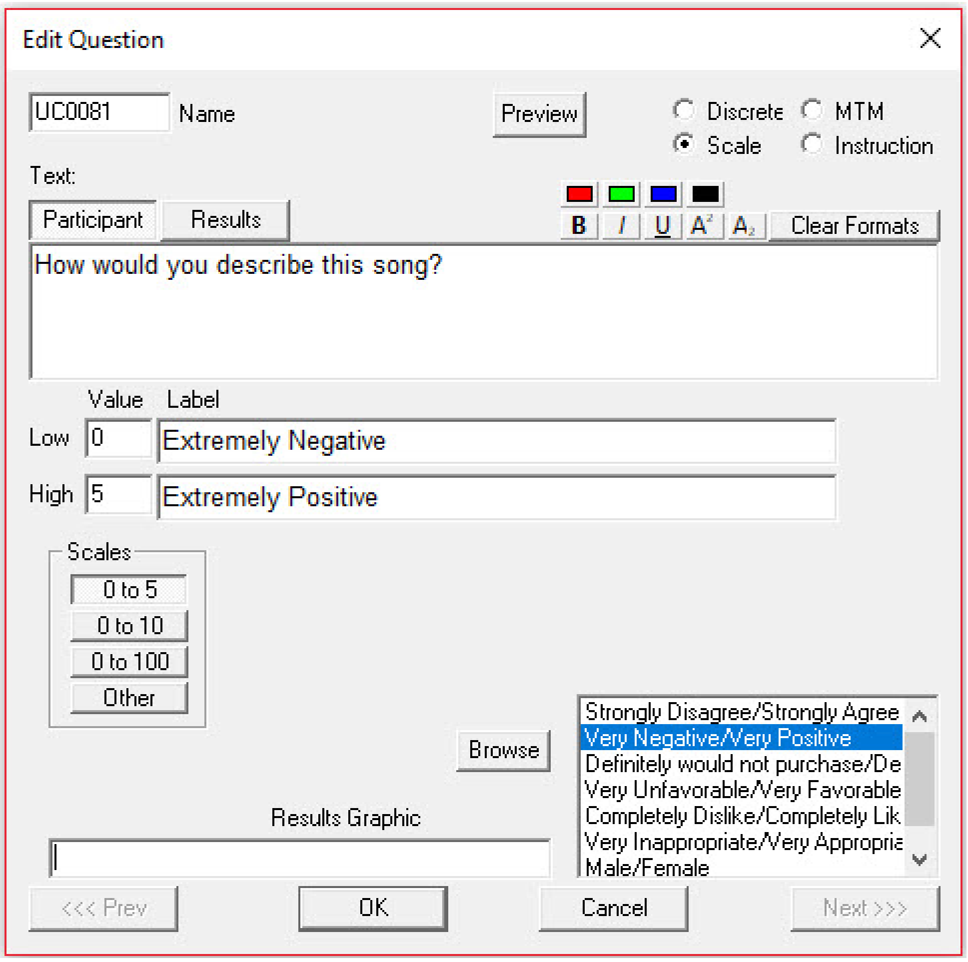

Figure 9. In edit mode, the Dialsmith PA software can be used to create questions with continuous (ordinal or interval) response options. This is a screenshot of the edit mode interface being used to create the polarity question used in the auditorium dial tests.

Moment-to-Moment questions are similar to scale questions. They are used to continuously evaluate time-based materials like music, television commercials, speeches, and lectures. In conjunction with viewing the material, participants are able to continuously rate, or evaluate it using an intensity scale. The dials are set to record the responses once per second. It is also important to label the ends of each scale, just like with scale questions. For example, if the participants were asked to continuously rate their opinion of a 30 second clip of a song (UC0095) on a scale from 0 to 100, the label for 0 would be very negative, and the label for 100 would be very positive. Figure 10 depicts the sections in which this information is to be placed.

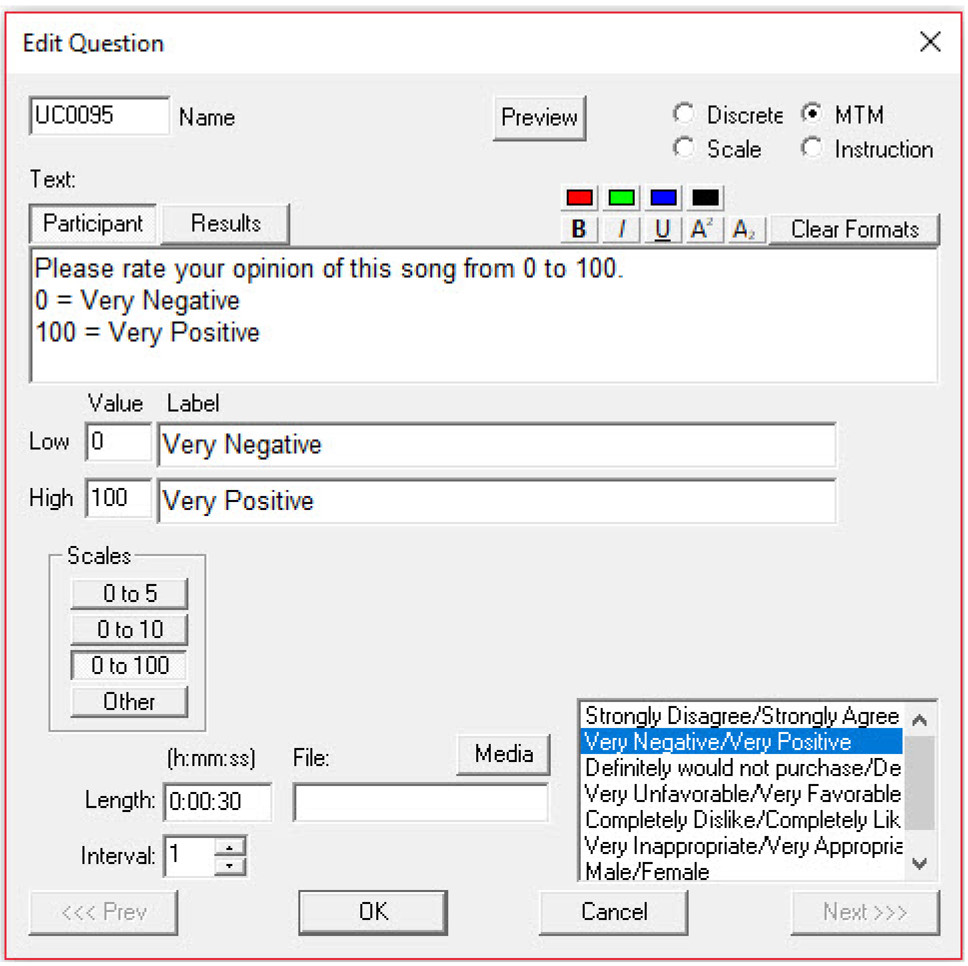

Figure 10. In edit mode, the Dialsmith PA software can be used to create questions with continuous (ordinal or interval) response options. This is a screenshot of the edit mode

interface being used to create the moment-to-moment question used in the auditorium dial tests.

The instruction option was not a question. It was a piece of the questionnaire that allowed for the researcher to display information they wanted the participant to know. The instruction also needed to have a variable code. For example, at the beginning of the data collection when participants were coming into the room and settling in, a welcome screen with beginning instruction was displayed. This instruction was “Welcome! Please take a seat and put your cell phone in airplane mode. We will get started shortly” (X0054). Throughout the data collection, new instructions may need to be communicated with the participants, whether that be for them to set their dials to 50, to take a 5-minute break, or to thank them for coming. Figure 11 depicts the sections in which this information was placed.

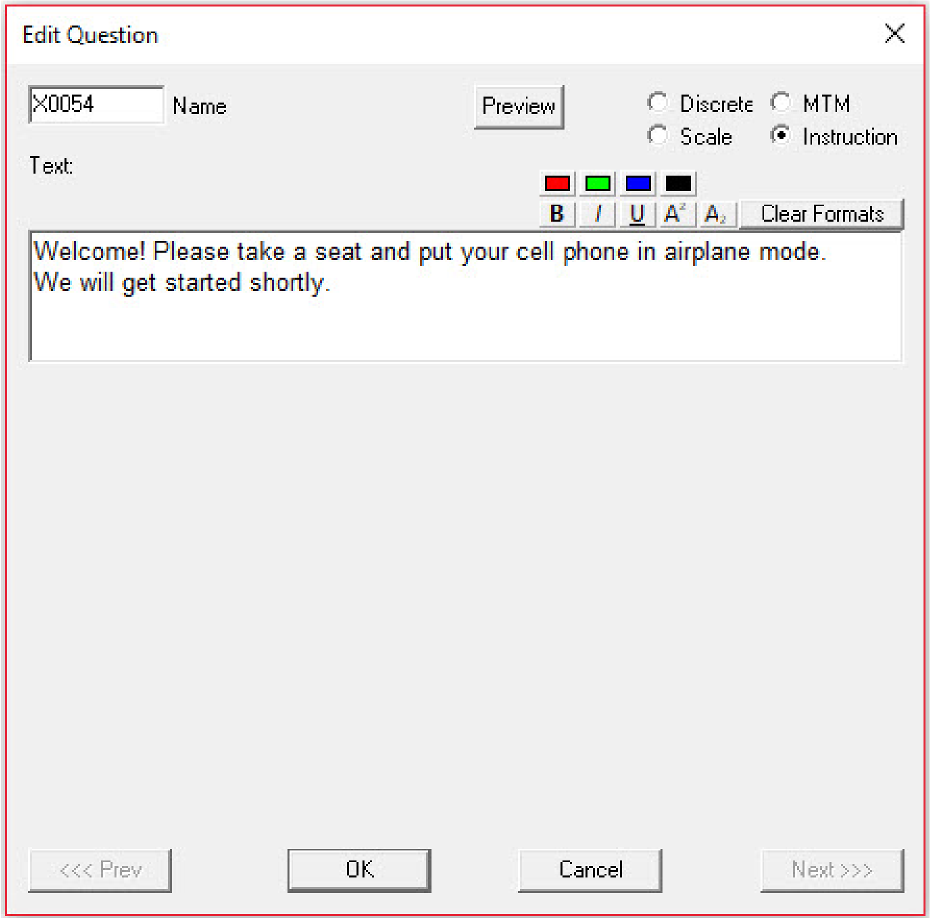

Figure 11. In edit mode, the Dialsmith PA software can be used to create instructions. This is a screenshot of the edit mode interface being used to create the instruction example used in the auditorium dial tests.

Collect

The questionnaire was tested once the displays were setup and configured for the PA software. Setting up the displays for the participant and the results was a long process. There were 10 different devices, nine different cords, and three power cords (not including laptop charger, screen power, or extension cords). Figure 12 illustrates the set-up process for each of the devices and cords, as well as their input and outputs.

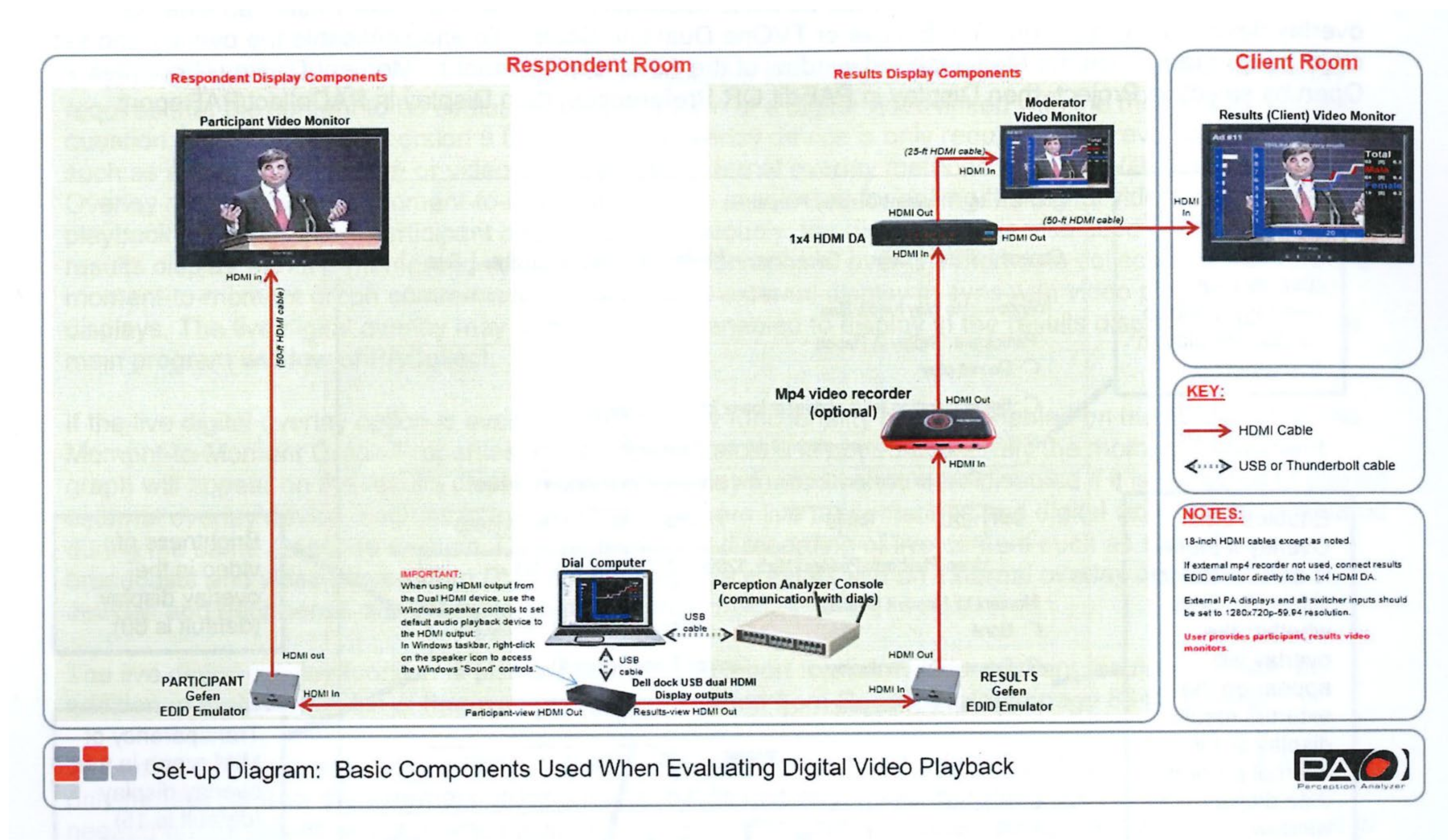

Figure 12. This set-up is optional for those who want to use a live digital overlay on their MTM questions (Dialsmith, 2018*[3]).

[3] * Reprinted with permission from "Comprehensive training guide for the perception analyzer" by Dialsmith, 2018. Copyright 2018 by Dialsmith.

After setting up the video playback devices, the external displays were configured through the PA system. There were three displays that were set up including primary, participant, and results. The primary display was the PA system that was running on the laptop. The participant display was the screen that the participants saw, so questions were viewed from this screen. The results display showed the results for each question, as well as the MTM overlay. Figure 13 shows what the external display screen looked like in set-up.

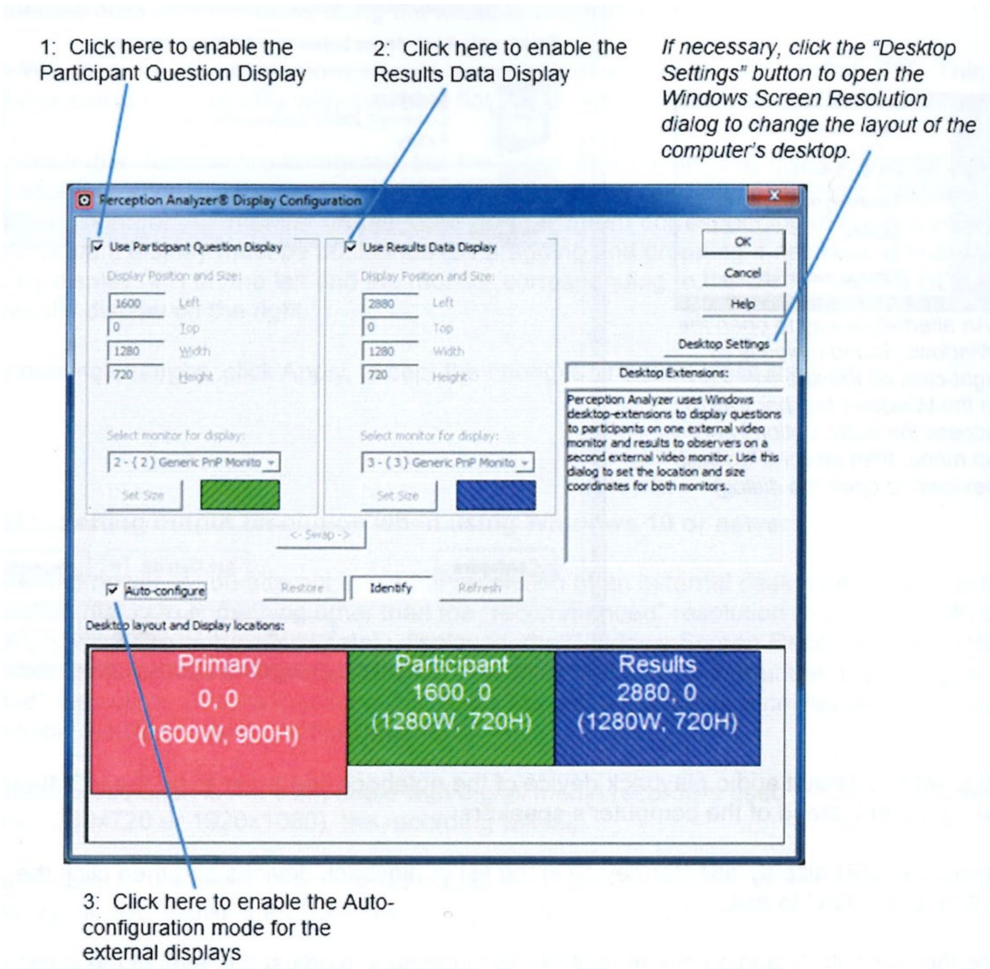

Figure 13. The most common configuration uses the auto configure function. The PA system will automatically size and position the participant and results displays to fill two full screens (Dialsmith, 2018*[4]).

[4] * Reprinted with permission from "Comprehensive training guide for the perception analyzer" by Dialsmith, 2018. Copyright 2018 by Dialsmith

When prompted, participants were shown each question. Using their dials, the participants selected the answer choice that best represented them and the answers were recorded. Data were saved automatically after each question was answered. I observed the participants' responses in real-time on the administrator's screen while the participants reacted to the music by turning their dials. The ratings graph appeared as an overlay on top of the material that is being tested. Each of the previous questions that were asked were used as subsets to see which groups were reacting to the materials. The subset view could be changed by the researcher at any time during the moment-to-moment process, and all responses were recorded.

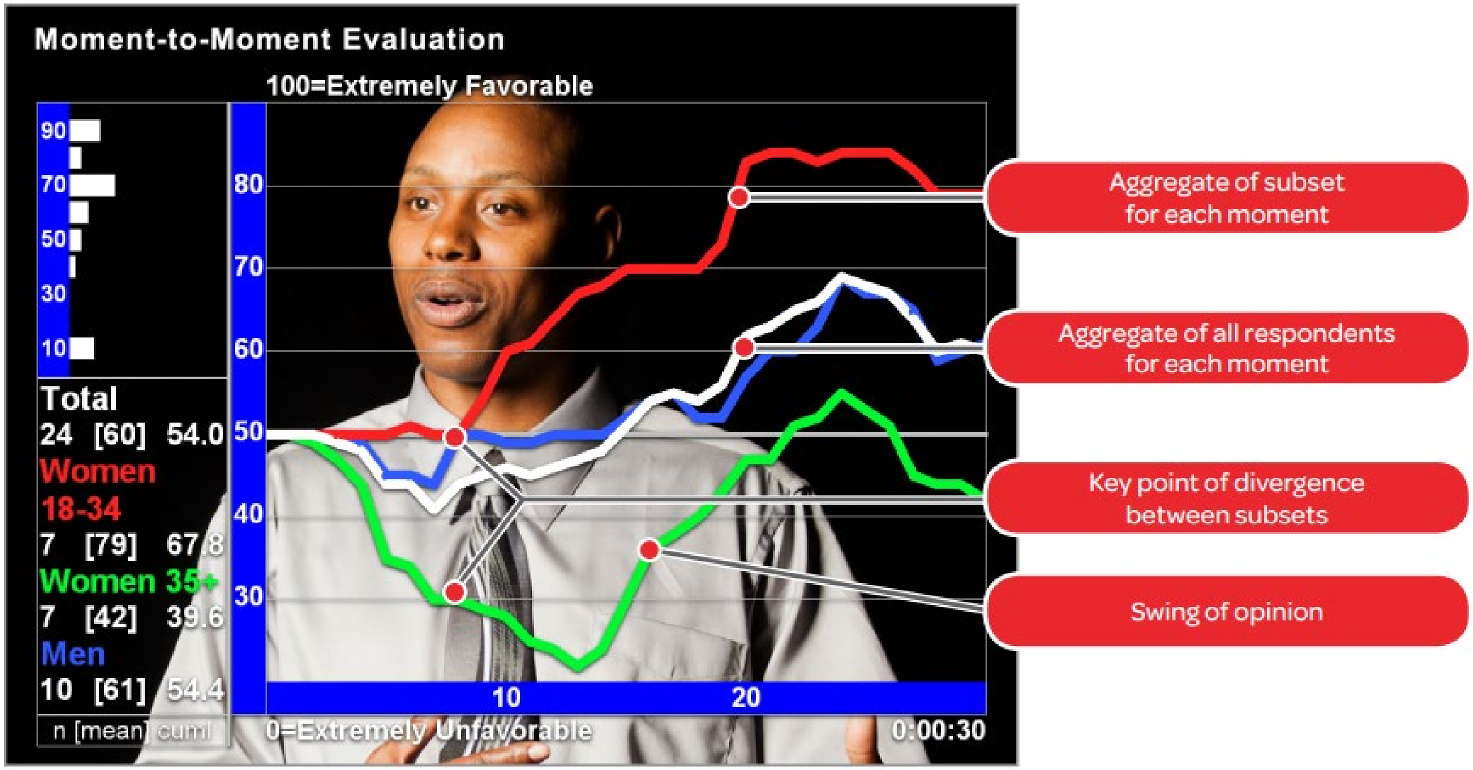

Figure 14. This is an image of the results view with an overlay of moment-by-moment responses. The subsets indicated in this image include women 18-34, women 35+, and men (Dialsmith, 2013*[5]).

[5] *Reprinted with permission from "Essentials of moment-to-moment research" by Dialsmith, 2013. Copyright 2013 by Dialsmith.

When in a testing session, participants had a different view of the question than the researcher. Figures X and Y provide the two separate views that the researcher and the participant experience. The researcher was able to see immediate results of the question, whereas the participant could only see the question and answer choices. These two separate views were available for each of the four different types of questions. Figure 15 is an example of a discrete question view, and figure 16 is an example of a MTM question view.

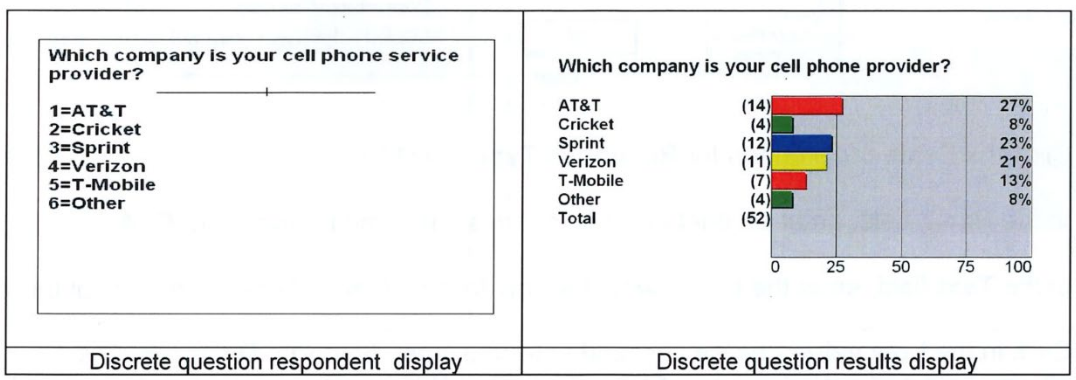

Figure 15. The discrete question respondent display is what the participants see on their screen. The discrete question results display is what the researchers see on their screen. The results are masked from participants (Dialsmith, 2018*[6]

---

[6] * Reprinted with permission from "Comprehensive training guide for the perception analyzer" by Dialsmith, 2018. Copyright [2018] by Dialsmith.

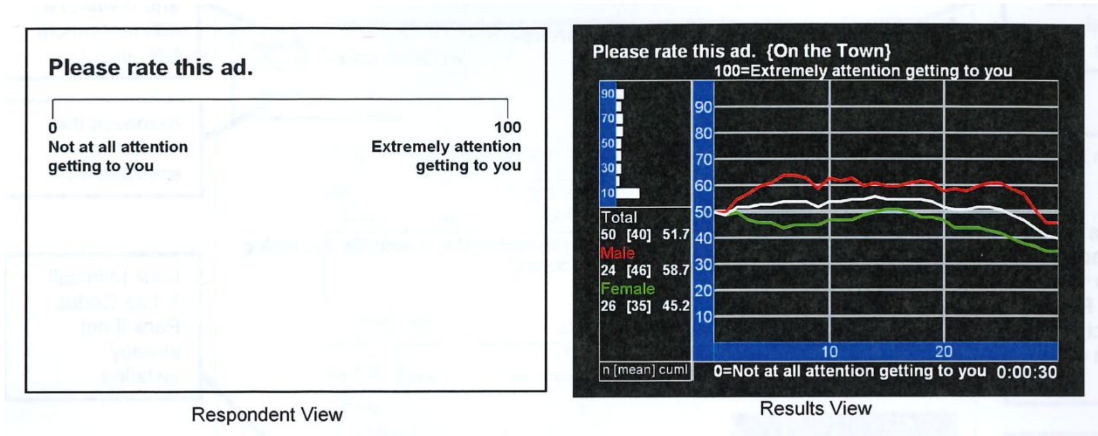

Figure 16. Participants see the respondent view on their screen, whereas researchers see the results view on their screen. The results will continue to extend along the x-axis as the media progresses on (Dialsmith, 2018*[7]).

Report

There are four formats in which data can be displayed when in the results program. These formats are graphical, quick frequencies, crosstab, and participant. The graphical display provides a visual display of the results that were collected. Both the participant and the results external displays are active and facilitate the moment-to-moment data. The quick frequencies display gives a simple analysis which includes mean, median, standard deviation, standard error, confidence intervals and other statistics. The crosstab display allows for the subgroups' (subset) responses to be compared for each question using numeric tables. Selecting more than one question from the subset list creates a crosstab by the current selected question. Participant display

[7] * Reprinted with permission from "Comprehensive training guide for the perception analyzer" by Dialsmith, 2018. Copyright 2018 by Dialsmith.

mode shows the questions exactly how the participant saw them. The reports can then be exported to open in Microsoft Excel or PowerPoint.

This study used three discrete questions, nine moment-to-moment questions, and eleven instruction pages. The media used in the moment-to-moment questions varied in length between three and four minutes each, with the shortest being 3:09 (minutes: seconds), and the longest being 3:54. Participants were those selected from the survey method section.

4.3. Sampling procedures

The media (songs) selected for this section were based off of their overall Bing sentiment analysis scores. The Bing sentiment analysis was used because it analyzed more words than the AFINN and NRC lexicons. Three genres (Country, Pop CHR, Urban Contemporary [Kelly, 2018]) were analyzed from the year-end 2017 and 2018 Billboard charts, with a positive, negative, and neutral song representing each genre. This made a total of nine songs that were systematically selected. There was a total of 500 songs reviewed, and because only nine were selected, so there was a 1.8% sample rate of the total population (Billboard, 2017a; Billboard, 2018a; Billboard, 2017b; Billboard, 2018b; Billboard, 2017c; Billboard, 2018c).

Using the PA system dials, each participant continuously responded to each selected song, from start to end, on a scale from 0-100 (i.e., dial scores) and all participants listened to the same songs. Participants' responses to each song were interval-level data with one observation per second, per song.

4.4. Sample size, power, and precision

All MTM measures obtained from Dialsmith were renamed to begin with the letter I, followed by six digits assigned to a specific media asset (song), followed by an underscore (_), and the sequence of measurement in seconds (xxx). For example, Lauv's *I Like Me Better* (Leff & Matosic, 2017) was assigned an asset code of A0022. Beginning at the first second (00:01), the PA system software captures responses to media. Therefore, MTM responses in this study were assigned to each second of media beginning with assigned a variable name of I1001_001. The subsequent measure of the same song (Lauv, *I Like Me Better*, Leff & Matosic, 2017) at two seconds would be I1001_002, followed by I1001_003.

Each song varied in length. Every song had a two second timer inserted before the song was played, and two seconds of extra time added to the end of it. The times of each song are listed in table 3.

Table 3
Positive, Neutral, and Negative songs

Order	Time	Category	Song	Artist	Citation
1	03:22	Pop Neutral	*I Like Me Better*	Lauv	Leff & Matosic, 2017
2	03:26	Country Positive	*Blue Tacoma*	Russel Dickerson	Brown, Dickerson, & Nohe, 2016
3	03:22	Urban Negative	*Young Dumb & Broke*	Khalid	Little, Riley, & Robinson, 2017
4	03:22	Country Negative	*Criminal*	Lindsay Ell	Ell, Stevens, & Wilhelm, 2017
5	03:54	Pop Positive	*Shape of You*	Ed Sheeran	Briggs, Burruss, Cottle, Mac, McDaid, & Sheeran, 2017
6	03:35	Urban Positive	*LOVE.*	Kendrick Lamar	Lamar, Kurstin, Pacaldo, Spears, Tiffitc, & Walton, 2017
7	03:09	Country Neutral	*What If's*	Kane Brown feat. Lauren Alaina	Brown, McGinn, & Schmidt, 2017
8	03:56	Urban Neutral	*Boo'd Up*	Ella Mai	Dopson, Howell, James, & McFarlane, 2017
9	03:23	Pop Negative	*Sorry Not Sorry*	Demi Lovato	Brown, Douglas, Felder, Lovato, & Simmons, 2017

4.5. Design and measures

The three discrete questions for this section were based off of a screening survey that participants filled out prior to testing. These questions were used as subsets in their analysis that identified trends amongst groups. The first question was about gender identification and used nominal level data. Same as the example listed above, the question was, “What is your gender?” The answer choices were as follows: *1 = Male; 2 = Female; 3 = Non-binary / third gender; 4 = Prefer to self-describe*, and *5 = Prefer not to say.*

The second discrete question was about music preference. It also used nominal level data. The question was, “What is your **most** listened to music genre?” The list of answer choice options are as follows: *R&B / Hip-Hop, Rock, Pop, Country, Latin, Dance/Electronic*, and *Christian / Gospel*. The participant selected one out of the following genres.

The third discrete question followed in line with the second one. This time the question was, “What is your **least** listened to music genre?” This too used nominal level data. The answer choice options were also the same as question two and followed the same order: *R&B / Hip-Hop, Rock, Pop, Country, Latin, Dance / Electronic*, and *Christian / Gospel*. Participants were able to select one genre as their answer.

The moment-to-moment questions were on a scale from 0-100. The participants listened to a song, and then continuously rated their opinions of it as they listened. There were nine songs, resulting in nine moment-to-moment questions. The songs, in order, are those listed in Table **3**. These songs were all predetermined from the sentiment analysis section. The moment-to-moment data offered insight in the human response to media, which was later compared to the computational analysis.

To address RQ3, data from RQ1 and RQ2 were used. Time markers aligning with elements identified in the AFINN, Bing, and NRC analyses used in RQ1 were added to each participant's dial score data (RQ2). The addition of time markers created time-sequenced data that allowed for periodic comparisons of dial scores in the same temporal (time) space as elements identified in the sentiment analysis.

Mean and standard deviation scores associated with the nine emotional categories (joy, anger, sadness, disgust, fear, anticipation, trust, positive, or negative) in the NRC lexicon (Silge & Robinson, 2017) were reported. Additionally, mean and standard deviation scores associated with the AFINN and Bing lexicons were reported for each song and genre. For each of the lexicons, bivariate correlations were calculated and reported to describe the relationship between lexicon values and dial scores.

Research question two was a case study with 53 participants and nine songs. These nine songs were the same duration and in the same order for every participant. The variable code for each of the nine songs were:

A0014: Kane Brown, ft. Lauren Alaina, *What ifs*

A0015: Russell Dickerson, *Blue Tacoma*

A0016: Lindsay Ell, *Criminal*

A0017: Ella Mai, *Boo'd Up*

A0018: Demi Lovato, *Sorry Not Sorry*

A0019: Kendrick Lamar, *LOVE.* [*sic*]

A0020: Khalid, *Young, Dumb, & Broke*

A0021: Ed Sheeran, *Shape of You*

A0022: Lauv, *I Like Me Better*

The measure for this case study was the human response, or dial score, to each of the songs. A total of 53 participants listened to the same nine pre-selected songs based on the song's positive, negative, and neutral values. While listening to each song, participants continuously rated their level of like/dislike of what they were hearing with the Dialsmith dials. After each song, the participant answered a series of questions pertaining to what they had just heard. Although the participant scored the songs based on the music, I focused on their response to the lyrics themselves. The *M*, *SD*, *min.*, and *max.* values were reported for each second of the song, and the whole song. Music genre (MX0003) and year (T0001) were variables used in this.

Research question three was both a correlational study and case study. The selected cases were all 53 participants, music genre, years, AFINN results, and Bing results. The selected measure was to describe the relationship among music lyrics and human response.

After collecting the sentiment analysis scores, and the human response scores, it was possible to calculate a relationship between the scores and the music lyrics. AFINN sentiments were plotted by word, time, and value in each song. Markers were placed on a wave form of each of the nine songs. These markers indicated the moment that each of the specific words from the sentiment analysis took place. Human responses to the polarity of each were also correlated with the Bing sentiment analysis results. A relationship among song polarity through the sentiment analysis and the human response was described.

The *M* scores for dial response for each second of each was plotted on a line graph. Each of the AFINN word scores were plotted along the same line graph at the time they occurred, to see a visual relationship between the sentiments themselves, and the human response.

5. SELF-REPORT SURVEY

The next section describes the self-reported responses that each participant gave in two different surveys. The first survey was collected before individuals arrived for the music test, and a multiple section survey was collected once each person arrived at the music-testing site.

1. A screening survey was administered before the participant enrolled in the study and before he or she was scheduled to participate in a music testing session. Each person's responses to the screening survey was used to determine his or her eligibility to participate in the study (e.g., at least 18 years of age) and assessed his or her music listening habits and opinions.
2. The second survey was administered once participants arrived at the testing site. The survey was divided into multiple sections. First, participants answered some contact information questions to match their response from their screening survey. They then listened to a piece of music, and after answered a series of questions. This was the music testing survey.

I used the Qualtrics survey platform. Qualtrics is an online survey development software company that has become a leader in market research and enterprise feedback management (Goodwin, Chiarelli, & Irani, 2011). Survey studies can be very beneficial for collecting data because they provide an estimation of the different types of responses that may occur. They also take into account the many factors and opinions that play an important role in the responses (Juslin, Liljestrom, Vastfjall, Barradas, & Silva, 2008).

Survey responses help to identify background variables that take part in explaining the individual differences between participants, especially in a large sample.

The social sciences have often heavily relied on self-report measurement methods to capture reactions from respondents (Micu & Plummer, 2010; Poels & Dewitte, 2006; Scherer, 2005). In fact, the most prominent research methods are questionnaires self-report designs to measure emotions experienced by an individual (Stieler & Kriegl, 2018). These self-reports often rely on post-stimulus measurements and have often been neglected (Stieler & Kriegl, 2018).

The accuracy and reliability of self-reports are not always valid in research. However, self-report data is necessary for unobservable behaviors. Only the individual can provide their own interpretations of events or feelings (Garcia & Gustavson, 1997). This study was designed to investigate how individuals react to music. A self-report of the polarity of the song (extremely negative to extremely positive), the emotions an individual experienced while listening to the song (happiness, sadness, anger, surprise, fear, disgust, other [Poels & Dewitte, 2006]), and the reason behind why they had the selected feelings, were fit appropriately. Factors thought to influence the listener's emotional response to music include music preference, musical training, personality, and the listener's motive for listening to music (Juslin et al., 2008).

5.1. Reliability and validity

Reliability

Self-reports can often times be an unreliable measure because some people may find it difficult to interpret how they feel. People may be unaware of their feelings, or they alter their responses because of social desirability (Poels & Dewitte, 2006). Self-reports are likely to have a relatively large measurement error, because other factors that are not measured tend to influence how people respond to questionnaires (Field, 2016).

The test-retest method is a way of testing the stability of a measure (Bryman, 2015). A test-retest involves giving the test, or survey in this case, on one occasion, and then giving it again to the same sample a second time. To test reliability, the same people can be tested twice in a test-retest. A reliable instrument should be able to produce similar scores in both tests (Field, 2016). Participants were asked four questions twice to ensure stability in their response. They were asked for their first name, last name, and email on three separate occasions (SignUpGenius, screening survey, music test survey). They were asked for their year of birth on two occasions (screening survey, music test survey). They were asked for their gender, most preferred music genre, and least preferred music genre on two occasions (screening survey, PA).

The self-report questions were used to identity elements within music that cause a polar and emotion reaction in participants. They were included in a qualitative analysis. Questions related to music education, music psychographics, and music preference were used for descriptive, comparison, and regression data.

Validity

The survey questions were opinion-based, so there is not much to go off of for if the respondent's answers were truthful or not. However, because they were asked about whether they thought the songs they listened to were positive or negative, their responses to the survey can be compared to their dial responses to determine if how they rated the song overall reflects their survey response. Self-report measures sometimes look at content validity. Content validity determines whether each individual item represents and covers the full range of the construct being measured (Field, 2016).

1. Pilot testing

A development survey and two pilot tests were completed to test the quality, reliability, and validity of the questions asked. It is important to test the survey process in order to assess its successes, and to see what could be improved (Dillman, Smyth, & Christian, 2014). Pilot tests are used to test procedures, monitor response rate, manage respondent inquiries, track incentives, examine problems within the questionnaire, and can give the researcher a sense of how the respondents experienced the survey (Dillman, Smyth, & Christian, 2014). After assessing the survey with the pilot test, changes can be made to better streamline the survey and fix any errors. Experiments with human subjects were approved by the Institutional Review Board (IRB).

Development survey

A development survey was sent out in web urls via Facebook, text message, email, Twitter, LinkedIn, and GroupMe. Questions were also distributed via iPad at the Texas State Fair, and at the Memorial Student Center (MSC) at Texas A&M University during the 2018 fall semester. The development survey was used to test several questions to see which ones had the greatest number of quality responses that could then be used for the pilot study. Some key points I had to keep in mind while developing the survey were the clarity and length of the questions, the usability of the survey across multiple devices, the variability of the answer choices to each question, the question order, and the design of the questionnaire. Each question was derived from published journal refereed articles, and some had already been tested in previous studies. The responses to these questions were not reported, just the questions and answer choices were evaluated.

The main takeaways from the development survey were that some of the questions worked as is, some needed rewording, and some needed to be completely eliminated. When asked about music genre listening habits, there were a total of 32 sub-genre choices. This was overwhelming for the participant, and those surveyed on an iPad did not know that they needed to scroll down. These music genre answer choices were changed to seven categories that contained each of the 32 original sub-genres. Another issue that I ran into was the overuse of text response questions. A common occurrence in responses was participants skipping the third or more open-ended question. One more problem that needed to be fixed was clarifying the answer choices. When asked what activities the participant engages in while listening to music, some participants did not consider driving to be in the same category as traveling. Questions and answer choices need to be as clear and as straight forward as possible. The questions in the development survey are included in Table 4.

Table 4

Development survey questions

Order	Question	Answer Choice	Type of data
1.	What is your name?	Text entry	String
2.	What is your email address?	Text entry	String
3.	What year were you born? (YYYY)	Text entry	String
4.	Please review the music genres, below, and then select the types of music you listen to at least once per month. (Scroll down to see all options, and select all that may apply)	Alternative Rock, Bluegrass, Blues, Classic Country , Classic Rock , Classical , Country , Disco , Dubstep , Easy Listening , EDM , Folk , Funk , Gospel , 90s Grunge , Heavy Metal , Hip Hop , Jazz , Opera , Orchestra , Pop , Punk Rock , R&B , Red Dirt Country , Reggae , Rock , Techno , Texas Country , Top 40 , Latin/Spanish , Christian .	Nominal
5.	Do you listen to playlists to enhance your experience? *For example, these are some highly-followed playlists on Spotify: "Workout", "Good Vibes", "Happy Hits!", "Down in the Dumps", "Low-Key", "Tender", and "Creativity Boost".*	Yes, No	Nominal
6.	Do you think that music can change a person's emotional state?	Yes, No, I don't know	Nominal
7.	Can you think of some songs that can cause you to feel… (matrix)	1. happy (yes, no) 2. sad (yes, no) 3. surprised (yes, no) 4. fearful (yes, no) 5. angry (yes, no) 6. disgusted (yes, no)	Nominal

Table 4 (continued)

Order	Question	Answer Choice	Type of data
8.	1. What is a song that makes you feel happy? 2. What is a song that makes you feel sad? 3. What is a song that makes you feel surprised? 4. What is a song that makes you feel fearful? 5. What is a song that makes you feel angry? 6. What is a song that makes you feel disgusted?	1. Song title (text entry) 2. Artist (text entry)	String
9.	Do you think music lyrics play a part in creating an emotional experience in music?	Yes, Maybe, No	Nominal
10.	Do you think past experiences can cause an emotional reaction while listening to music?	Yes, Maybe, No	Nominal
11.	What time(s) of day do you listen to music? (*select all that may apply*)	1 = Early Morning 4:00 a.m. to 8:59 a.m.; 2 = Morning (9:00 a.m. to 12:59 p.m.), 3 = Afternoon (1:00 p.m. to 5:59 p.m.); 4 = Evening (6:00 p.m. to 10:59 p.m.); 5 = Over Night (11:00 p.m. to 3:59 a.m.); 6 = Other (text entry)	Nominal, string
12.	What are you doing while listening to music? (select all that may apply)	01 = Working/Studying; 02 = Hanging out with friends; 03 = Relaxing; 04 = Watching TV or a movie; 05 = Housework; 06 = Traveling; 07 = Eating; 08 = Cooking; 09 = Working out or playing sports; 10 = Shopping; 11 = Partying; 12 = Playing video games; 13 = At a concert; 14 = Other (Text entry)	Nominal, string
13.	Do you play a musical instrument?	Yes, No	Nominal

Table (4 continued)

Order	Question	Answer Choice	Type of data
14.	Do you have any formal music education?	None at all, A little, A moderate amount, A lot, A great deal	Nominal
15.	Please select the forms of music platforms you use. (*select all that may apply*)	Radio AM or FM, Sirius - XM Satellite Radio, YouTube, Spotify, Pandora, iTunes, iHeart Radio app, Tidal, Apple Music, Google Play, SoundCloud, Vimeo, TuneIn, Rhapsody	Nominal
16.	How often do you listen to each of the of the following music platforms? (Matrix)	Radio AM or FM, Sirius - XM Satellite Radio, YouTube, Spotify, Pandora, iTunes, iHeart Radio app, Tidal, Apple Music, Google Play, SoundCloud, Vimeo, TuneIn, Rhapsody 1 = Several times a day; 2 = Every day, but not several times a day; 3 = Every week, but not every day; 4 = A couple times a month, but not every week; 5 = Once in a while	Ordinal
17.	What is your gender?	Male, Female, Non-binary / third gender, Prefer to self-describe (text entry), Prefer not to say	Nominal, String

After one week of the development survey being sent out via web urls and app-based platforms, approximately 138 responses were collected. After two days of data collection at the Texas State Fair, and five days of data collection at the MSC, approximately 252 responses were collected. That made a total of 390 respondents for the development survey.

Pilot test 1

The first pilot test was constructed before the sentiment analysis was run on the year-end Billboard charted songs. With the use of the Dialsmith PA software, this survey was tested with 12 undergraduate Agricultural Communications & Journalism students at Texas A&M University. The student's ranged in music education and were between the ages of 20-22 years old. Three songs were tested among the three genres being studied (Country, Pop CHR, Urban Contemporary), with a survey and moment-to-moment perception analysis. After testing, the participants engaged in a discussion where they were asked what worked and what didn't work during the test. They were then asked for any additional feedback that they thought would improve the study.

During the pilot test, participants were asked their first and last name and dial number. They were then instructed to read the lyrics that belonged to the first song. Three follow-up questions were asked about the lyrics. After answering the lyrics questions, the participants were directed to listen to the song they had just read the lyrics for, rating it from 0 (very negative) to 100 (very positive) with the Dialsmith dials, moving from left to right every time their opinion changed. After listening to the song, the same three follow up questions were asked again, but this time about the song they listened to. These steps repeated three times for each of the songs.

There were many issues that presented themselves in this first pilot study that needed to be addressed. The first of which being the copied variable coding in Qualtrics. Instead of creating new variable codes for each set of questions, they were just copied and pasted three times. This was a problem in Qualtrics, because instead of collecting data for each time the question was asked, it just overwrote the previous question that used the same code.

The second problem was the influence that reading the song lyrics had on the listening of the song itself. When the participants read the lyrics, it changed the way in which they heard the song. Instead of rating the song based on whether they liked it or not, they based their rating off of whether they liked the lyrics or not. This changes the way in which data were collected, giving the song an unfair rating because the participant already knew the context and message of the song before they listened to it, which created bias. It was important not to separate the lyrics from the music to produce a natural response from the participants.

The third problem was that the participants were not asked their level of familiarity with the song. This is important because if a person is very familiar with a song, they probably already know if they like it or not, and therefore already have an opinion (or bias) formed about the song. This could greatly affect their rating of the song in the moment-to-moment analysis.

The last problem with the survey was the level on instruction. The participants were confused at the beginning of the dial testing, because they did not know what specifically they were rating the song based off of. They were also not familiar with the

Dialsmith software because none of them had used it before. Prewritten instructions needed to be clearly defined and announced to the participants before the test even started.

The three songs tested in order were:

1. *On My Way To You*, by Cody Johnson (James & Lane, 2018).
2. *1-800-273-8255*, by Logic (Caracciolo, Hall, Ivatury, Robinson, Taggart, & Wiggins, 2017).
3. *I'm A Mess*, by Bebe Rexha (Brooks, Jussifer, Peiken, Rexha, & Tranter, 2018).

The questions from pilot test 1 survey are included in table 5.

Table 5
Pilot test 1 questions

Order	Question	Answer Choice	Type of data
1.	What is your name?	Text entry	String
2.	What is your dial number?	Text entry	String
Read lyrics			
3.	What emotions did you feel while reading these lyrics? (*Select all that apply*)	Happiness, sadness, fear, anger, surprise, disgust	Nominal
4.	How would you rate these lyrics?	Extremely positive, Somewhat positive, Neither positive nor negative, Somewhat negative, Extremely negative	Nominal
5	What do you think the message of these lyrics was?	Text entry	String
Listen to song			
6.	What emotions did you feel while listening to this song? (*Select all that apply*)	Happiness, sadness, fear, anger, surprise, disgust	Nominal
7.	How would you rate this song?	Extremely positive, Somewhat positive, Neither positive nor negative, Somewhat negative, Extremely negative	Nominal
8.	What do you think the message of this song was?	Text entry	String

Questions 2-8 were then repeated two more times for each of the next two songs. Looking back on the questions and original pilot test, the issues are very evident. Pilot test 2 was designed to fix all the presented problems.

Pilot test 2

The second pilot test was developed after the Bing sentiment analysis was run on the 2017 and 2018 year-end Billboard charted songs for the three selected genres. After analyzing the lyrics, songs were chosen based off of their positive, negative, and neutral score, as well as their clustered word count. Of the 500 songs analyzed, nine were chosen; one positive, one negative, and one neutral song from each of the three genres. Making the necessary changes from the development survey and pilot test one, pilot test two was constructed and tested.

Nine Agricultural Communications & Journalism students at Texas A&M University participated in this pilot test. Participants ranged from ages 20-22 and had some formal music education background. All students completed a pre-test screening survey before entering the classroom.

Before the music test was administered, instructions on what was to be expected, what the test consisted of, and how to use the Dialsmith dials were verbally given. Included was a range of numbers from which the dial responses reflected (figure 17). Participants were asked to answer a few introductory questions on the dials to help familiarize themselves with how the Dialsmith dials worked. The answers to the questions were also used as subsets later on when reviewing the results in the Dialsmith

PA. These questions included their gender identification, their most preferred music genre, and their least preferred music genre.

Instructions communicated to the participants are included in figure 17.

Beginning Dial Instructions:

- In front of you, you have one of these [hold up dial]. It's called a dial and allows us to better understand what you're thinking.
- We'll use the dial in two different ways today.
- The first is really simple. First, you will answer a few questions using the dial.
- Just turn the knob on the dial, left or right, until the number appears on the dial that best represents your choice for the question on the screen.
- Once you have it, leave it there until we move on to the next question.
- Don't worry about the buttons on the bottom, they have all been disabled for today's event. You don't even have to push the 'Enter' button.
- If your dial does something unexpected, let me know right away and we'll get it resolved.
- If you have a question about a question on the screen, I want to know that too.
- Any questions before we give it a try?
- Great, here's the first question…

Questions appear on screen…

Continuous Dialing Instructions

- Now that you have the feel for the dials, we're going to use them in a different way.
- With the previous questions, I wanted you to give me a single answer for that question.
- But in a moment, we'll be listening to some music.
- And while you are listening, I want to know *EVERY SECOND on* how you feel about what you are hearing, and if LIKE or DISLIKE what you are hearing
- It's NOT whether or not you like or dislike the artist or genre, but if you like or dislike the song itself.
  Are you with me so far? Great!

*Figure 17***.** Instructions provided to students during pilot test two.

- To do this, you'll be moving the knob on your dial left and right *THROUGH OUT* the song *EVERY TIME* your opinion changes.
- Still with me? Okay.
- We'll start every song at '50'. Go ahead and put your dials at '50' now.
- 50' is our neutral point on the scale. Everything above '50' indicates that you LIKE with what the song, and everything below 50 indicates that you DISLIKE it. The more you turn the dial towards "0" or "100" reflect how strongly you dislike or like the song at the moment.
- Is this all making sense? I ask because this is very important for us to better understand what types of communications work or don't work with an audience.
- When the song starts, that's your cue to start evaluating it.
- Again, we'll start each song at '50' and I want you to continuously move the knob as your opinion changes.
- And I ask that you use the full range of the dial.
    - Any movement in the 40-60 range is fairly neutral. You only slightly dislike or like what the speaker is saying.
    - Movements in the 60-80 range mean that you like the song at that moment.
    - Where movements 80-100 mean that you strongly like it.
    - On the flipside, movements in the range of 20-40 mean that you dislike the song at that moment.
    - And 0-20 means that you strongly dislike it.
    - So PLEASE don't just keep the dial in the central range or I won't learn a thing. Use the full range.
- Everyone understand that? Great!
- Also, larger dial movements with tells us more than little movements.
- It doesn't matter to me if you're at a '62' or a '64', I'm more interested in the range that you are dialing. Keep your focus on the song and respond by sweeping through the dial ranges.
-

*Figure 18***.** Instructions provided to students during pilot test two (continued).

After the participants completed the questions in the PA, they were directed to their iPads. On their iPad, they then answered some questions that were previously covered in the screening survey. These questions were asked again just in case an issue came up where we needed a second source for their information.

The process of having the participants read and answer questions about the song lyrics was eliminated. Instead, they listened to the song, were asked about their familiarity with it, and then were asked some similar questions that were in the first pilot test. The questions for the second test are in table 6.

Table 6
Pilot test 2 questions

Order	Question	Answer Choice	Type of data
1.	What is your dial number?	Text entry	String
2.	What is your name?	Text entry	String
3	What year were you born? (YYYY)	Text entry	String
		Listen to song	
4.	Which of the best describes your level of familiarity with this song?	I have never heard it before, I may have heard it before, I have heard it before, I have heard it many times	Nominal
5.	How would you rate this song?	Extremely positive, Somewhat positive, Neither positive nor negative, Somewhat negative, Extremely negative	Nominal
6.	What emotions did you feel while listening to this song? (*Select all that apply*)	Happiness, sadness, fear, anger, surprise, disgust, other (text entry)	Nominal, string
7.	To what degree did you feel... (carry forward)	Happiness, sadness, fear, anger, surprise, disgust, other 0 = None, 1 =1, 2 =2, 3 = 3, 4 = 4, 5 = Extremely Intense	Ordinal
8.	What in the song caused you to feel … (carry forward emotion selection)?	Text entry	String

Questions 4-8 were then repeated after each of the nine songs. A five-minute break was given between songs five and six to give the participants a reprieve. Five minutes was not enough time for the participants, so the break time was changed to 10 minutes.

This second pilot test was successful. The participants were told in their next class why their responses were important, and why these specific questions were asked. They were also given a report of their answers, and what the moment-to-moment analysis concluded. They then participated in a class discussion and were satisfied with the test.

5.2. Media/ participant characteristics

The media characteristics will be the same as those identified in the Dialsmith method section. Nine songs were tested based on their Bing score from the sentiment analysis. Those nine songs came from three music genres: Country, Pop CHR, and Urban Contemporary. Each genre had three songs associated with it, a positive score, a negative score, and a neutral score.

The participants are the same as those in the Dialsmith method section. Participants were Millennials between the ages of 18-34. They varied between that age range, in sex, racial group, education level, socioeconomic level, and gender identity. They all spoke English and were able to read and write.

5.3. Sampling Procedures

A total of 53 participants was included in the study. An original number of 55 were tested, but one participant was eliminated because he had taken the test before, and the second was eliminated because he did not fill out the music test survey. Participants ranged between the ages of 19-30 and identified as either male or female. There were 10 males and 43 females. Although this was the target population, the ages 31-34 were not reached, and there were more female than male participants.

Participants were put into a classroom, spread apart among multiple tables. They all faced a projector that had instructions and questions displayed on it. Each participant had an iPad in front of them for their music test survey, a dial for the PA, and a piece of paper that had the ranges from 0 – 100 listed (0-*20 = strongly dislike, 21-40 = dislike, 41-60 = neither like nor dislike, 61-80 = like, 81- 100 = strongly like*). Participants filled out the beginning of the music test survey, listened to the first song, then answered questions pertaining to the song. This was repeated nine times for each of the nine songs.

5.4. Design and measure

2. Screening Survey

Data from RQ2 (dial scores) and data collected from two surveys were used to address RQ4. Before individuals arrived for the in-person music test, they completed a screening questionnaire using a web-based survey. Each person's responses to the screening survey were used to determine his or her eligibility to participate in the study (e.g., at least 18 years of age). Additionally, participants were asked about their media consumption habits and media preferences.

Participants signed up for dates and times via an online sign up form, SignUpGenius. Required fields for signing up included first name, last name, email and an optional phone number. After signing up for a time, the screening survey was given out via email to each participant within 48 hours of the study.

Contact information

This survey included basic contact and demographic questions, as well as some media preference questions. The contact information questions included first name, last name, and email address. This information helped to identify individuals and provided the means to follow up with them if needed.

Demographics

The next questions about demographics included year of birth, and gender (male, female, non-binary / third gender, prefer to self-describe, and prefer not to say). Gender was a subset used to identify trends. Year of birth was very important in this study, because I was evaluating Millennials, ages 18-34 (Kelly, 2018). Rather than just asking the participants how old they are, year-of-birth gives an inarguable number.

Music psychographics

There were four music psychographics questions that all used nominal level data.

The first question in this group was, "Do you listen to playlists to enhance your experience? *For example, these are some highly-followed playlists on Spotify: 'Workout', 'Good Vibes', 'Happy Hits!', 'Down in the Dumps', 'Low-Key', 'Tender', and 'Creativity Boost'* (Spotify, 2019)." The two answer choices were either "Yes" or "No", making it a dichotomous selection. Playlists are most popular among 16-34-year-olds and account for approximately 35% of their total listening time (Delmonte, 2018). Music has the ability to gratify the emotional needs of its listener. According to the

Zillman theory of mood management, people strategically, even if not consciously, select specific types of music that allow them to achieve or prolong hedonic states (Lippman & Greenwood, 2012; Zillmann, 2000). There is also evidence that supports the use of music to adjust current mood (Knobloch, 2003).

The second question in this group was, "Do you think that music can change a person's emotional state (Juslin, et al., 2008)?" The three answer choices were "Yes", "No", or "I don't know". The use of "I don't know" allows the participant a neutral selection. Whether music has the ability to evoke emotion in its listeners is controversial. There is, however, evidence that when self-reported, listeners do have a feeling of emotional change (Juslin & Sloboda, 2001). This question helps gauge whether or not the participant has experienced this before or believes this to be true.

The third question in this group was, "Do you think music lyrics play a part in creating an emotional experience in music (Hu, 2009)?" The three answer choices were "Yes", "Maybe", and "No". If unsure, the participant had an option of selecting "maybe" for neutral. Lyrics can have an abundance of textual features and can outperform audio features when the semantic meaning of the lyrics tie into the emotion experienced by the participant (Hu, 2009).

The last question in this group was, "Do you think past experiences can cause an emotional reaction while listening to music (Salimpoor, Benovoy, Longo, Cooperstock, & Zatorre, 2009)?" The three answer choices were "Yes", "Maybe", and "No". If unsure, the participant had an option of selecting "maybe" for neutral. Emotional reaction while listening to music can be very individualized depending on personal preference and previous experiences (Salimpoor, et al., 2009). For most people it can be easy to recall specific songs that can be associated with a previous experience, moment in time, or past relationship (Lippman & Greenwood, 2012). The emotional significance provoked while listening to music does not rely on just the elements of the music itself. It is a function of social and psychological contexts that are heard, then associated with personally (Lippman & Greenwood, 2012; Frith, 1987; van Dijck, 2006). The elicited emotional reaction can be associated with past or current moments of significance.

Music preference

The top three consumed music genre formats for Millennials ages 18-34 were studied. Music genre preference is a key factor to understand and select individuals for the study. Participants were asked to rank their genre preference on a scale from 1 to 7 using the “drag and drop” function in Qualtrics. This function allows the user to rank items by dragging and dropping them into place. The Billboard top music genre charts were selected for the answer choice criteria. These are: Pop, Rock, Latin, R&B/Hip-Hop, Dance/Electronic, Country, and Christian/Gospel (Billboard, 2019d). The answer choices were set in a randomized order, so that they would not be automatically ranked or alphabetical. Those who prefer pop music may have a more positive reaction to pop music, and those who dislike country music may have a high negative reaction to country music. Genre preference helps to understand individual response to each piece of music.

Media consumption

The media consumption group was broken up into seven questions. Four of these questions were multiple select, two were single answer, and one was a matrix.

The first question in this group was a multiple answer selection using nominal level data. The question was, "What music genres do you listen to? (*select all that may apply*)." This was different than the genre rank question, because this one gave the participant instruction to only select genres that they actually listen to, instead of responding to all of them. The answer choices will contain the same seven music genres as the previous question. The answer choices were in a randomized order.

The second and third questions in this group were also a multiple answer selection using nominal level data. These two questions pertained to time of day spent listening to music, and activities being done while listening to music (Juslin, et al., 2008). Current activity, external factors in the environment around them, and time of day highly influence the music that an individual chooses to listen to (Cunningham, et al., 2008).

The first of the two questions was, "What time(s) of day do you listen to music? (*select all that may apply*)." The answer choices for this question included *Early Morning (4:00 a.m. to 8:59 a.m.)*, *Morning (9:00 a.m. to 12:59 p.m.)*, *Afternoon (1:00 p.m. to 5:59 p.m.)*, *Evening (6:00 p.m. to 10:59 p.m.)*, *Over Night (11:00 p.m. to 3:59 a.m.)*, and *Other*. Each time of day is stamped with a number to indicate a specific moment throughout the day.

The second of the two questions was, "What are you doing while listening to music? (*select all that may apply*)." The answer choices for this question included

Working/Studying, *Hanging out with friends*, *Relaxing*, *Watching TV or a movie*, *Housework*, *Driving/Traveling*, *Eating*, *Cooking*, *Working out or playing sports*, *Shopping*, *Partying*, *Playing video games*, *At a concert*, and *Other*. With the increasing use of portable music players (mobile phone, mp3 players, radio), and the variety of music platforms available (Spotify, Pandora, Apple Music), has contributed to the ability to listen to music practically anywhere, and at any time (Juslin, et al., 2008).

Questions four and five in this group were single answer and referred to music background. The first question used nominal level data. The question was, "Do you play a musical instrument?" The answer choices for this were "Yes" and "No", making it dichotomous.

The second question was, "Do you have any formal music education?" The answer choices used a five point Likert scale, with anchors on each of the five points. This used ordinal level data. The answer choices included *1 = None at all*; *2 = A little*; *3 = A moderate amount*; *4 = A lot*; *5 = A great deal.*

Participants who had a background in music, studied music, or play a musical instrument may hear music in a different way than those who do not have a music background. They could be able to pick apart the different musical elements, may have a personal tie or connection to an artist or song, or may catch on quick to the emotions and messages being put forth by the song.

Question six and seven in this group asked about music platform use and frequency. Question six was a multiple answer selection that used nominal level data. Question seven carried forward using the selected responses from question six in using ordinal level data. Qualtrics' carry forward function is typically used when asking multiple response follow up questions. In the initial question, participants were asked to select which options they use or prefer. The selected answer choices were then included in a future question in the survey.

For example, a participant may be asked to indicate which social media platforms he or she uses from a list of likely options. Then, if he or she is asked about how frequently he or she uses the selected social platform or platforms in a subsequent question, only the selected options will be displayed. Reducing the quantity of options a person must consider has been reported to reduce respondent fatigue (Dillman, Smyth, & Christian, 2014) and will likely reduce the overall time needed to complete a questionnaire.

Question six was, "Please select the forms of music platforms you use. (*select all that may apply*)." The music platforms in the answer choices were all popular streaming based service. The answer choices were *Radio AM or FM*, *Sirius - XM Satellite Radio*, *YouTube*, *Spotify*, *Pandora*, *iTunes*, *iHeart Radio app*, *Tidal*, *Apple Music*, *Google Play*, *SoundCloud*, *Vimeo*, *TuneIn*, and *Rhapsody*.

This set of answer choices were developed and tested by researches in the Digital Media Research and Development Lab (DMRDL) at Texas A&M University. DMRDL researchers reviewed communication-industry metrics, including monthly and quarterly reports by Nielsen Audio, to create and refine these media consumption questions and answers (Hill, 2016). Multiple pilot tests were conducted, and revisions were made to accurately describe and evaluate these options. In question seven, participants were asked, “How often do you listen to each of the of the following music platforms?” Only the selected answer choices from question six were displayed. The answer choices were in a five-point Likert scale, with each point containing an anchor. The answer choice were 1 (*Several times a day*), *2 = Every day, but not several times a day*; *3 = Every week*, *but not every day*; *4 = A couple times a month, but not every week*; and *5 = Once in a while.*

Time spent listening to music and music platform use can influence familiarity with the music being presented. Those who spend more time listening, or have access to multiple music platforms, may recognize the song selection more than those who do not.

3. Music Testing Survey

The second survey was administered using iPads, once participants arrived at the testing site. The questionnaire was divided into two sections. The first section included questions used to connect each participant’s responses to the screening survey and his or her responses provided through the PA system. A second section was repeated after each song was played. Each time, respondents were asked about his or her familiarity with the song and his or her opinion of the song.

The second survey (music testing survey) was a self-report following each of the nine songs that were listened to. Self-reports can sometimes be unreliable because some people may have a hard time describing how they feel when asked to report. Either they are unaware, or they are unwilling to report their emotions because of social desirability (Poels & Dewitte, 2006). To help eliminate these concerns, each participant had their own seat spaced out from one another and had their own iPad to answer questions. The risk of other people in the room seeing an individual's answers was negligible.

Aggregate scores were reported to describe the group demographic, psychographic, environmental, and/or behavioral data collected in the surveys. Additionally, bivariate correlations were calculated and reported to describe the relationship between participants' demographic, psychographic, environmental, and/or behavioral data, and their dial scores.

Introduction

The survey started with four introductory questions. They were string level data, where the participant wrote in their text response. The first question asked for their dial number. Each of the Dialsmith PA dials have an individual number (1-40) assigned to them. The participant was able to type their dial number in.

The next three questions were used to match the screening survey results with the music testing survey results. These questions included first and last name, email address, and year of birth. This was a backup in case there were any issues that arise from either survey.

Moment-to-moment subsets

After completing the introductory questions, the participants answered three questions through the PA software. The three questions all had a single answer selection and used nominal level data. The answers were used as subsets for the actual moment-to-moment results to see trends. The participant was asked their gender, their most preferred music genre, and their least preferred music genre. All of the answer choices were the same as in their screening survey. There was then an instruction for the participant to move their dial from left to right (0-100) to get a feel for the dial. They were able to see the difference in spacing between each number and practiced moving the dial. After practicing, the participant was instructed to set their dial to 50 (neutral).

Song listening

After completing the PA subset questions, the participant was directed to their iPad. The next steps were repeated nine times for each song selection

A stop sign was displayed on the participant's iPad. Under the stop sign were instructions for proceeding. The instructions said, "Please listen to the song you are about to be presented. Use your dials to indicate your continuous level of interest ranging from very negative (0 - Dislike) to very positive (100 - Like)." The participant then listened to the first song, and continuously rated it from 0-100 on their dial. Every time their opinion changed, they moved the dial to the left or right.

The order of the songs was selected based on the negative and positive score from the Bing lexicon in the previous sentiment analysis. There was a 10-minute break between songs five and six to give the participants a reprieve.

Table 7
Tested song list

Order	Category	Song	Artist	Citation
1.	Pop Neutral	*I Like Me Better*	Lauv	Leff & Matosic, 2017
2.	Country Positive	*Blue Tacoma*	Russel Dickerson	Brown, Dickerson, & Nohe, 2016
3.	Urban Negative	*Young Dumb & Broke*	Khalid	Little, Riley, & Robinson, 2017
4.	Country Negative	*Criminal*	Lindsay Ell	Ell, Stevens, & Wilhelm, 2017
5.	Pop Positive	*Shape of You*	Ed Sheeran	Briggs, Burruss, Cottle, Mac, McDaid, & Sheeran, 2017
6.	Urban Positive	*LOVE.*	Kendrick Lamar	Lamar, Kurstin, Pacaldo, Spears, Tiffitc, & Walton, 2017
7.	Country Neutral	*What If's*	Kane Brown feat. Lauren Alaina	Brown, McGinn, & Schmidt, 2017
8.	Urban Neutral	*Boo'd Up*	Ella Mai	Dopson, Howell, James, & McFarlane, 2017
9.	Pop Negative	*Sorry Not Sorry*	Demi Lovato	Brown, Douglas, Felder, Lovato, & Simmons, 2017

Song evaluation

After listening to each song, the participants were asked four questions pertaining to what they had just heard.

The first question was about the participant's familiarity with the song. This had a single select answer and used ordinal level data. The question was, "Which of the best describes your level of familiarity with this song?" Familiarity with music can lead to preconceived opinions (bias) already established by the listener. Although music that is familiar can be pleasing, it can also be considered boring by the listener due to repetition (Salimpoor, et al., 2008). The answer choices or this question were *01 = I have never heard it before*, *02 = I may have heard it before*, *03 = I have heard it before*, *04 = I have heard it many times*.

The second question in this group allowed the participant to identify whether they thought the song to be positive or negative using ordinal level data. This had a single select answer and was measured on a five-point Likert scale, with anchors identifying each scale point. The question was, "How would you rate this song?" The answer choices will be *01 = Extremely positive, 02 = Somewhat positive, 03 = Neither positive not negative, 04 = Somewhat negative*, and *05 = Extremely negative*.

The third question in this group asked the participant to pinpoint the exact emotions experienced while listening to the specific song. The participants were able to select multiple answer choices ranging across the six basic emotions (Poels & Dewitte, 2006). Music can be classified as an emotional trigger (Panksepp, 1998), so in order to understand which specific emotions are being triggered, this question being asked was, “What emotions did you feel while listening to this song? *(Select all that apply)*.” The six basic emotions to select from were *happiness, sadness, fear, anger, surprise*, and *disgust.* The participant also had the opportunity to select “other” and write in an emotion that was not offered. Linguistic analyses, much like a sentiment analysis, support the use of a fairly short list of primary emotions. When asking people to report four or five basic emotions they experience, there is a consistent agreement on a short list of options (Panksepp, 1998).

The fourth question carried forward the selected choices from question three. Only the choices selected by the participant in the previous question were options. The participant ranked the degree to which they felt the specified emotions indicated. The rank went from 0-5 with 0 equaling “none” and 5 equaling “Extremely intense”. Points 1, 2, 3, and 4 remained numbers. These were ordinal level data.

Emotion context

This next group of questions were fully dependent on the participant’s response to question four, the degree to which they felt specific selected emotions. If the participant selected four or five as their rank for an emotion, they proceeded to this section. This section was all string level data, and the participant typed their response into a text box. The question was, “What in the song caused you to feel?” Whatever emotion that the participant ranked as a four or five, was used in the blank area. Each emotion had its own separate question in this section.

A text response from the participant helped the researcher better understand the context as to why a strong level of intensity was felt about the song. The degree of intensity could be because of personal and interpersonal factors such as past experiences, positive or negative preconceived opinion about the song, current mood or emotional state personal preferences, familiarity with the song, and expectations all contribute to emotional responses to musical stimuli (Lippmann & Greenwood, 2012 ;Salimpoor, et al., 2008).

Other factors can contribute to degree of intensity of elements of the song itself. These include beats-per-minute, lyrics, tone, pitch, tempo, frequency, energy, etc. (Eerola, 2012; Gabrielson & Lindstrom, 2010; Juslin & Laukka, 2003). The self-report helped break down the reasoning behind why the participant felt strongly towards a specific emotion. The music testing questions will repeat after each of the nine songs.

Research question four was used as a case study. The selected cases were the 53 participants and music genre. Variables used in this were genre (MX0003) and years (T0001). The selected measures were used to explore human response to music lyrics.

Based on demographics, psychographics, environmental, and behavioral factors, human response to music lyrics were explored with potential to be predicted. Participants responded to multiple questions asking about these different factors in their screening and music test surveys. With the responses collected from the surveys and their dials, future predictions can be made about human response to music lyrics.

Research question five was also used as a case study. The selected cases were all 53 participants, music genres (MX0003), years (T0001), Billboard chart (MX0003), and chart rank (MX0004). The selected measure was to explore chart performance. Each of the nine songs tested had varying levels of Billboard Year-End chart performance. Using the sentiment analysis for the lyrical content, human response from the dials, and self-report responses from the surveys, Billboard chart performance was explored with the potential to be predicted.

6. FINDINGS

There are different types of data that can be collected, and different ways to collect data. One way to test hypotheses are by observing what naturally happens, and by manipulating the environment and then observing what kinds of effects it has on the variable, known as correlational research. Correlational research provides a very natural view of the question being researched because nothing is being intentionally interfered with, so the variables should not be biased by the researcher being there (Field, 2016).

Two ways to conduct correlational research are by a cross-sectional study or by a longitudinal study. A cross-sectional study is done by taking a snapshot of many variables at a single point in time. Unfortunately, cross-sectional research does not show the contiguity between the different variables. A longitudinal study is done by measuring variable repeatedly at different points in time. Longitudinal research addresses contiguity between variables to some extent, but there is still an issue that the other variables that haven't been measured (confounding variables) might influence both variables that have been measured (Field, 2016).

Variables can be discreet or continuous and can be measured on different scales. Variables can be on a nominal, ordinal, interval, or ratio scale. A nominal scale assigns variables into categories, for example gender, occupation, or eye color. An ordinal scale is the same as a nominal scale, but has a logical order, for example months in a year or grade level. An interval scale has an equal difference in intervals, for example miles between point A and point B or height in inches. A ratio scale is the same as an interval scale, but "… there is a meaningful zero representing a complete absence of the construct being measured" (Field, 2016, p. 67). An example of this provided by Field (2016) would be beats per minute, because a drummer can play at 100 beats per minute, and a drummer can also play at 0 beats per minute.

In this chapter, I used bivariate correlations for each of the three methods to answer each research objective. A bivariate correlation is a correlation between two variables. The platforms used to analyze the variables were Microsoft Excel, IBM SPSS, Rstudio, and Tableau. Some variables had different scales of measurement so an effect size was used in order to compare variables across different research objectives. An effect size is "…an objective and usually standardized measure of an observed effect" (Field, 2016, p. 371). Because the measure is standardized, this give the effect size the ability to be compared across multiple studies that have measure different variables.

To calculate effect size, the Pearson correlation coefficient (*r*) was used. This is a measure of "…the strength of a relationship between two continuous variables, or between one continuous variable and a categorical variable containing two categories" (Field, 2016, p. 377). The measure can vary between -1 and 1. A perfect negative relationship has an $r = -1$, which means that as the first variable increases, the second variable decreases. A positive relationship has an $r = 1$, which means that as one variable increases, the second variable increases as well. A positive value that is less than one (e.g., $r = 0.5$) represents a situation where the first variable increases and the second variable increases, but not by a proportionate amount. If a correlation coefficient is ($r = 0$), that means that there was no relationship at all between the variables (Field, 2016).

Hypotheses were tested for each bivariate correlation by using a non-directional statistical test. When testing a null hypothesis, scores can be higher or lower in certain conditions, which indicates that a relationship between variables is in a specific direction. Hypotheses can either be directional or non-directional that can be determined by a one-tailed or a two-tailed statistical test (Field, 2016). A one-tailed test (directional) only looks at one end of a distribution, whether it be the upper or lower tail. If the results of a one-tailed test are the opposite direction of what was expected, the null hypothesis cannot be rejected. A two-tailed test (non-directional) looks at both ends of a distribution and take into account both positive and negative test statistics. When doing a two-tailed test and the p is 0.06, then the results are considered not significant because 0.06 is larger than the critical value of 0.05 (Field, 2016). Only two-tailed tests were used in the correlations within this study.

Error is inherent in all research. When testing the significance of a null hypothesis, there are two types of error that are present, Type I and Type II. Test statistics are used to tell us about the true state of the world, and whether or not there is an effect in our population. The two possibilities that cannot be known to be true are that there is an effect in the population, or that there is no effect in the population. Although these are not known, the test statistics and their associated probability are used to describe which of the two possibilities is more likely (Field, 2016).

Type I error is the belief that there is a genuine effect in the population, when there actually is not (Field, 2016), or concluding that the treatment has an effect when it in fact does not (Campbell & Stanley, 1963). If the conventional criterion is used, then the probability of this error is 0.05 (5%) when there is no effect in the population (α-level). If there is no effect in the population, and the α-level is set at 0.05, if the data is replicated 100 times, "…you could expect that on five occasions you would obtain a test statistic large enough to make you think that there was a genuine effect in the population even though there is not" (Field, 2016, p. 348).

Type II error is the opposite of Type I error. It is the belief that there is no effect in the population when there actually is (Field, 2016), or failing to reject the null hypothesis of no effect (Campbell & Stanley, 1963). Ideally, the probability of this error should be small. It has been suggested that the maximum acceptable probability of a Type II error is 0.2 (20%, β-level). If 100 samples of data from a population in which an effect exists were taken, "…you would fail to detect that effect in 20 of those samples; you would miss 1 in 5 genuine effects" (Field, 2016, p. 348).

Type I and Type II errors are based on different assumptions, so the relationship between the two is not straightforward. In order to make a Type I error, there has to be no effect in the populations, but to make a Type II error, there has to be an effect that was missed. As the probability to make a Type I error decreases, the probability to make a Type II error increases.

If the sample size is too low, because of the two types of errors there may not be enough power to detect a significant effect. In a sample of observations for any given observation, the confidence interval is "…a range of values around that statistic that are believed to contain, in a certain proportion of samples (e.g., 95%), the true value of that statistic (i.e., the population parameter)" (Field, 2016, 703).

Calculating a confidence interval for *r* can be done by adding and subtracting 2 times the standard error of the parameter. If the confidence interval is 95% (0.95), then that means in 95% of the samples the confidence interval will, "…contain the true value of the parameter" (Field, 2016, p. 466). What that also means is that for the other proportion of the sample, the true value will not be determined by the confidence interval.

6.1. Recruitment

Information about the study was sent out through mediums including Facebook, text message, email, GroupMe, and word of mouth, encouraging people to participate in the music test. People were directed to a SignUpGenius link with the available dates and times for the test. The data collection took place on seven different days in nine different sessions, with morning (e.g. 10:30 a.m.) and afternoon (e.g. 6:00 p.m.) time slots. Each session varied in number of participants, with the least being one, and the most being fourteen. The maximum number of participants allowed to test at one time was 15 due to number of iPads available.

6.2. Statistics and data analysis

Aim 1: Describe the relationship between sentiment analysis and human response to music lyrics.

6.2.1. **RQ.1: What sentiments are present in contemporary lyrical music?**

RO1.1: Describe the positive and negative aspects of sentiments in lyrics.

In a sentiment analysis, the Bing lexicon can be used to categorize words as either positive or negative. The Bing lexicon was used to analyze the positive and negative sentiments for the three music genre formats (country, pop, and urban) for the years 2017 and 2018. Table 8 gives a summary of these results. Described within the table are sample size (n), mean (m), standard deviation (SD), minimum (min), and maximum (max). The total word count column indicates the total number of words analyzed by the Bing lexicon for the specified year and genre. The positive column describes what was categorized within the total word count. The negative column describes what was categorized as negative within the total word count. The table also gives an indication that genres that had the greatest number of words analyzed had a higher number of positive and negative words.

Table 8

Bing Lexicon

				Total Word Count				Positive					Negative				
Year	Genre	*n_song*	*n-WC*	*M*	*SD*	*min*	*max*	*M*	*SD*	*min*	*max*	*f*	*M*	*SD*	*min*	*max*	*f*
2017	Country	99	8305	84.0	23.2	34	134	6.6	4.6	0	18	586	8.8	7.0	0	48	858
	Pop	49	5732	112.0	46.4	28	239	11.4	8.5	0	36	519	13.5	11.1	0	59	620
	Urban	100	15465	155.1	64.7	53	440	10.0	8.6	0	46	933	15.6	11.3	0	65	1523
2018	Country	99	8404	85.3	25.2	32	195	7.7	6.0	0	29	684	9.8	6.3	0	26	916
	Pop	50	5585	111.7	53.3	48	363	9.0	7.1	0	30	432	12.4	9.7	1	39	618
	Urban	100	17940	179.4	79.1	49	412	9.7	9.0	0	50	915	16.4	10.4	2	48	1638
Total		497	61431	727.5	291.9	244	1,783	58.0	43.8	0	209	5241	76.5	55.8	3	285	6173

Note. Excluded Easton Corbin, *A Girl Like You* in country 2017 & 2018; Excluded Luis Fonsi & Daddy Yankee, *Despacito* in pop 2017

There were 100 songs analyzed for the urban genre for both 2017 and 2018. There were 99 songs analyzed for the country genre for both 2017 and 2018. One song that charted in both years was eliminated because the song *A Girl Like You* by Easton Corbin did not have a page on the Genius website. There were 49 songs (*n*) analyzed for the pop genre in 2017, and 50 songs analyzed for the pop genre in 2018. The song *Despacito* (Luis Fonsi & Daddy Yankee, 2017) was excluded from the 2017 analysis because it was primarily written in Spanish. According to (Silge & Robinson, 2017), the AFINN, Bing, and NRC lexicons contain many English words with scores already assigned.

The mean number of negative words was greater in each genre during 2018 than in 2017. Similarly, the standard deviations for negative words were also greater in each genre during 2018 than in 2017. The minimum number of negative words (*min*. = 0) was consistent in each genre during 2017 and for country in 2018. However, the minimum number of negative words increased in pop (*min*. = 1) and urban (*min*. = 2) in 2018. The maximum number of negative words was greater in each genre during 2017 than in 2018. The maximum number of negative words was greater in pop (*max*. = 39) during 2018, but less in country (*max*. = 26) and urban (*max*. = 48) during 2018. Based on the results of the Bing sentiment analysis, songs were more negative in 2018 than in 2017 across all genres.

Table 9 is a summary of the Bing sentiment analysis in the selected songs that were tested. Each song was selected based on its number of positive words vs. negative words from the Bing sentiment analysis for each genre. Songs that were considered positive had a combination of one of the highest positive word counts of out its genre and year, with one of the lowest negative word counts for the same genre and year (e.g. pop 2017, *Shape of You*, positive word count = 36, negative word count = 9). Whereas words that were considered negative had a combination of one of the highest negative word counts out of its genre and year, with one lowest positive word counts for the same genre and year (e.g. country 2018, *Criminal*, positive word count = 1, negative word count = 22). Songs that were considered neutral had the same number of positive and negative word counts (plus or minus one), that were closest to zero (e.g. pop 2018, *I Like Me Better*, positive word count = 2, negative word count = 2).

Positive and negative word counts were derived from the total word count for each song analyzed by the Bing sentiment analysis. Songs with a larger total word count were more likely to have more positive and negative words counted toward their song. As mentioned in the sentiment analysis section in the methods, there were five criteria that songs had to meet in order to be considered.

Table X includes selected songs in order of genre popularity (Kelly, 2018) and polarity. Country was the most consumed genre for Millennials; therefore, country songs are in order first by positive, neutral and then negative overall polarity. Pop was the second most consumed genre, so pop songs are in order second by positive, neutral, and then negative overall polarity. Urban was the third most consumed genre, making urban in order third by positive, neutral, and then negative overall polarity.

Table 9

Bing Song Selection

Title	Artist(s)	Years	Genre	Rank	Polarity	WC	PWC	PR	NWC	NR
Blue Tacoma	Russell Dickerson	2018	Country	30	(+)	115	24	0.209	1	0.009
What Ifs [a]	Kane Brown ft. Lauren Alaina	2017, 2018	Country	4	(0)	65	8	0.123	7	0.108
Criminal	Lindsay Ell	2018	Country	77	(-)	92	1	0.011	22	0.239
Shape of You	Ed Sheeran	2017	Pop CHR	1	(+)	178	36	0.202	9	0.051
I Like Me Better	Lauv	2018	Pop CHR	22	(0)	52	2	0.039	2	0.039
Sorry Not Sorry [a]	Demi Lovato	2017, 2018	Pop CHR	29	(-)	134	6	0.045	30	0.224
LOVE.	Kendrick Lamar	2017, 2018	Urban	49	(+)	163	46	0.282	4	0.025
Boo'd Up	Ella Mai	2018	Urban	10	(0)	116	1	0.009	2	0.017
Young Dumb & Broke [a]	Khalid	2017, 2018	Urban	41	(-)	97	11	0.113	48	0.495

Note. Rank = Chart Rank Number; WC = Word Count; PWC = Positive Word Count; PR = Positive Ratio; NWC = Negative Word Count; NR = Negative Ratio; Polarity: + = positive; 0 = neutral; - = negative; [a] Charted in 2017 and 2018 with the 2017 chart rank indicated

Figures 18, 19, and 20 are plot trajectories of the net sentiment scores (positive – negative) for each of the nine songs listed in Table 9, above (Silge & Robinson, 2017). These plots place the positive and negative sentiment scores in separate columns with positive sentiments extending upwards on the Y-axis from zero (0), and negative sentiments extending downwards on the Y-axis from zero (0). The plot trajectory expands across the X-axis in sections of time, or cluster order in this case (Silge & Robinson, 2017).

The blue bars extending upwards from zero on the Y-axis indicate positive words. Blue bars that go up to one (1) on the Y-axis represent one positive word, bars that go up to two (2) on the Y-axis represent two positive words, bars that go up to three (3) on the Y-axis represent three positive words, and so forth. On the opposite end, the red bards extending downwards from zero on the Y-axis indicate negative words. Red bars that go down to negative one (- 1) on the Y-axis represent one negative word, bars that go down to negative two (- 2) on the Y-axis represent two negative words, bars that go down to negative three (- 3) on the Y-axis represent three negative words, and so forth. The gaps included in the plot trajectories were spaces where neutral words were present.

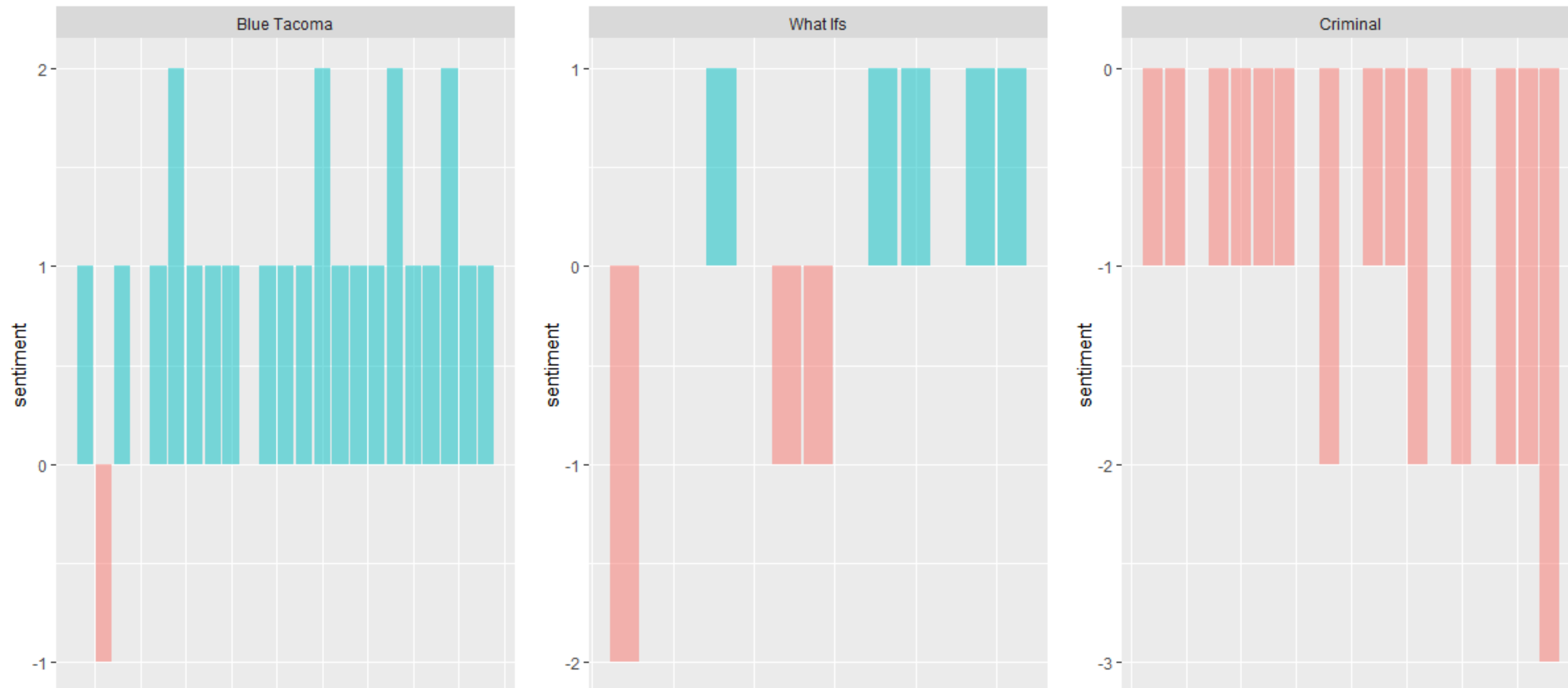

Figure 19. The plot trajectories for each of the three country songs are in order of polarity. *Blue Tacoma* represents the song with a positive polarity, *What Ifs* represents the song with the neutral polarity, and *Criminal* represents the song with the negative polarity.

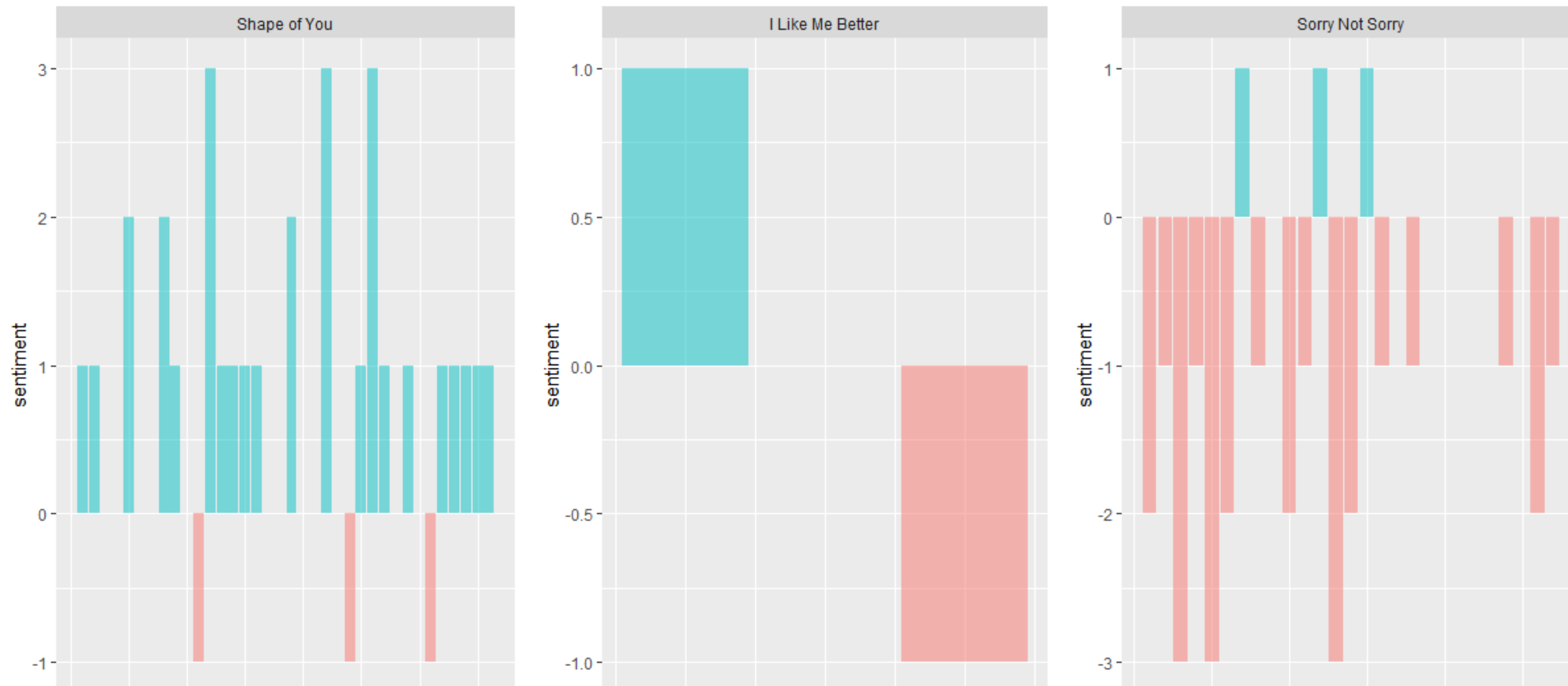

Figure 20. The plot trajectories for each of the three pop songs are in order of polarity. *Shape of You* represents the song with a positive polarity, *I Like Me Better* represents the song with the neutral polarity, and *Sorry Not Sorry* represents the song with the negative polarity.

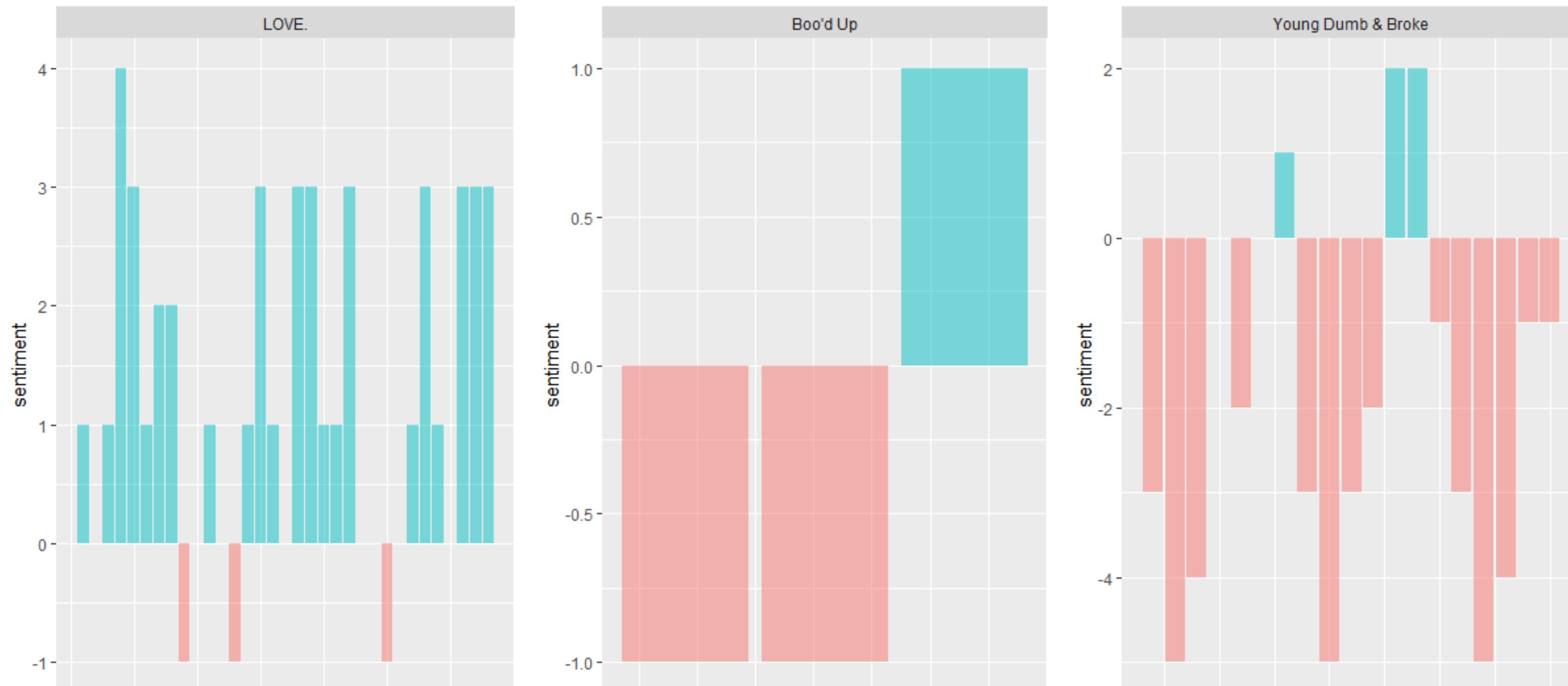

Figure 21. The plot trajectories for each of the three urban songs are in order of polarity. *LOVE.*[*sic*] represents the song with a positive polarity, *Boo'd Up* represents the song with the neutral polarity, and *Young Dumb & Broke* [*sic*] represents the song with the negative polarity.\

RO1.2: Describe the emotional elements in lyrics.

Words included in the sentiment analysis using the NRC lexicon were assigned to eight emotion categories, as well as a positive or negative polarity. The eight emotions that NRC assigns in a sentiment analysis are anger, anticipation, disgust, fear, joy, trust, sadness, and surprise, as well as positive and negative. Some words were categorized as multiple emotions; e.g., feeling was included in all eight emotion categories.

Table 10 is a summary of the NRC lexicon sentiment analysis scores for category of emotion by year for the country genre. Songs were more negative in this genre for both 2017 and 2018. Trust was the highest emotion category for both 2017 and 2018. From the six basic emotions (happiness [joy], sadness, fear, anger, disgust, surprise), fear was the highest emotion for 2017, while happiness [joy] was highest in 2018. The lowest emotion category (for the six basic emotions) in 2017 was surprise. The lowest emotion category for 2018 was disgust.

Table 10

NRC Lexicon - Country

	2017		*2018*		*Total*	
Sentiment	*f*	*%*	*f*	*%*	*f*	*%*
Anger	87	8	79	8	166	8
Anticipation	86	8	99	9	185	9
Disgust	59	6	51	5	110	5
Fear	101	10	94	9	195	9
Joy	93	9	98	9	191	9
Trust	107	10	104	10	211	10
Sadness	96	9	91	9	187	9
Surprise	51	5	59	6	110	5
Positive	181	17	186	18	367	17
Negative	192	18	188	18	380	18
Year Total	1,053	100	1,049	100	2,102	100

Note. Excluded Easton Corbin, *A Girl Like You* from 2017 and 2018; There were 198 songs analyzed.

Table 11 is a summary of the NRC lexicon sentiment analysis scores for category of emotion by year for the pop genre. Songs were categorized almost as positive as they were negative (2017 positive $f = 158$, % = 18 vs. 2017 negative $f = 157$ % = 18) in the pop genre for 2017, and more negative (2018 positive $f = 149$, % = 17 vs. 2018 negative $f = 164$, % = 19) in 2018. Trust was the highest emotion category for 2017, while both fear and surprise were highest in 2018. From the six basic emotions (happiness [joy], sadness, fear, anger, disgust, surprise), fear was the highest emotion for 2017, whereas fear and surprise were the highest in 2018. The lowest emotion category (for the six basic emotions) in 2017 was disgust. Disgust and sadness were two emotion categories that were lowest in 2018.

Table 11

NRC Lexicon - Pop

	2017		*2018*		*Total*	
Sentiment	*f*	%	*f*	%	*f*	%
Anger	72	8	80	9	152	9
Anticipation	77	9	68	8	145	8
Disgust	39	5	42	5	81	5
Fear	83	10	83	10	166	10
Joy	77	9	72	8	149	9
Trust	84	10	78	9	162	9
Sadness	68	8	42	5	110	6
Surprise	46	5	83	10	129	7
Positive	158	18	149	17	307	18
Negative	157	18	164	19	321	19
Year Total	861	100	861	100	1,722	100

Note. Excluded Luis Fonsi & Daddy Yankee, *Despacito* from 2017; There were 99 songs analyzed.

Table 12 is a summary of the NRC lexicon sentiment analysis scores for category of emotion by year for the urban genre. Songs were more negative in this genre for both 2017 and 2018. Fear was the highest emotion category for 2017, while trust was highest in 2018. From the six basic emotions (happiness [joy], sadness, fear, anger, disgust, surprise), fear was the highest emotion for 2017, while happiness [joy] was the highest in 2018. The lowest emotion category (for the six basic emotions) in 2017 was surprise and disgust was lowest in 2018.

Table 12

NRC Lexicon - Urban

	2017		*2018*		*Total*	
Sentiment	*f*		*f*		*f*	
Anger	147	8	79	8	226	8
Anticipation	144	8	99	9	243	9
Disgust	103	6	51	5	154	6
Fear	180	10	94	9	274	10
Joy	130	8	98	9	228	8
Trust	170	10	104	10	274	10
Sadness	147	8	91	9	238	9
Surprise	80	5	59	6	139	5
Positive	298	17	186	18	484	17
Negative	333	19	188	18	521	19
Year Total	1,732	100	1,049	100	2,781	100

Note. There were 200 songs analyzed.

6.2.2. **RQ.2: How do humans respond to lyrics in music?**

RO2.1: Describe how humans respond to positive and negative aspects of sentiments in lyrics.

Participants rated each of the nine songs by their polarity after listening. Table 13 is a summary of the participant ratings on a five-point Likert scale. The highest rated positive songs by participants were *Criminal* and *Young Dumb & Broke*, which does not align with the Bing sentiment analysis. The highest negative rated song by participants was Sorry Not Sorry, which does align with the Bing sentiment analysis scores.

Table 13

Song Polarity

	Extremely positive		Somewhat positive		Neither positive nor negative		Somewhat negative		Extremely negative	
Song	*f*	%	*f*	%	*f*	%	*f*	%	*f*	%
I Like Me Better	1	3	10	13	4	5	21	13	17	14
Blue Tacoma	2	7	8	10	4	5	25	16	14	11
Young Dumb & Broke	3	10	10	13	10	12	15	9	15	12
Criminal	5	17	14	18	17	20	13	8	4	3
Shape of You	4	14	4	5	3	4	20	13	22	18
LOVE.	2	7	12	15	14	16	9	6	16	13
What Ifs	4	14	8	10	14	16	18	11	9	7
Boo'd Up	7	24	10	13	15	18	15	9	6	5
Sorry Not Sorry	1	3	4	5	4	5	24	15	20	16
Total	29	100	80	100	85	100	160	100	123	100

RO2.2: Describe how humans respond to emotional elements in lyrics.

After listening to each of the nine songs, the participants were asked to select which emotions they experienced while listening to the song. They were able to select as many choices as they wanted. Table 14 is a summary of the emotions selected by participants. The song with the greatest happiness rating was *Sorry Not Sorry*, and the song with the greatest sadness rating was *Young Dumb & Broke*. The song with the greatest fear rating was *Young Dumb & Broke*, and the song with the greatest anger rating was *Sorry Not Sorry*. The song with the greatest surprise rating was *Criminal*, and the song with the greatest disgust rating was *Criminal*.

Table 14

Emotions Elicited

Song	Happiness		Sadness		Fear		Anger		Surprise		Disgust		Other	
	f	%	*f*	%	*f*	%	*f*	%	*f*	%	*f*	%	*f*	%
I Like Me Better	41	77.4	7	13.2	0	0	0	0	2	3.8	2	3.8	18	34
Blue Tacoma	41	77.4	3	5.7	0	0	0	0	3	5.7	8	15.1	12	22.6
Young Dumb & Broke	26	49.1	12	22.6	4	7.5	0	0	6	11.3	8	15.1	18	34
Criminal	19	35.8	8	15.1	0	0	4	7.5	11	20.8	12	22.6	18	34
Shape of You	43	81.1	0	0	0	0	4	7.5	4	7.5	6	11.3	8	15.1
LOVE.	26	49.1	8	15.1	1	1.9	3	5.7	4	7.5	10	18.9	19	35.8
What Ifs	30	56.6	4	7.5	2	3.8	1	1.9	8	15.1	11	20.8	13	24.5
Boo'd Up	25	47.2	4	7.5	0	0	1	1.9	9	17	11	20.8	15	28.3
Sorry Not Sorry	47	88.7	0	0	0	0	6	11.3	5	9.4	4	7.5	12	22.6

Notes. *f* = frequency; % = percent.

In the “Other” category, participants were able to provide an emotion that was not listed from the six basic emotions. There were 134 combined responses for “other” among all nine songs. Similar words (SW) were combined into one category, e.g., “did not feel anything” and “nothing” were combined into the neutral category.

The word “annoyed” (SW: annoying, annoy) was reported 32 times. The word “neutral” (SW: nothing, did not feel anything) were reported 14 times. The word “bored” was reported 13 times. The words “calm”, “empowered”, and “nostalgia” were reported 5 times each. The words “indifference”, “irritated”, and “love” were reported four times each. The words “confused”, “dancy” [*sic*], “disappointment”, “carefree”, and “relaxed” were reported three times each. The words “careless”, “confidence”, “dislike”, and “excited” were reported two times each. All other words indicted in figure 21s below were reported one time each.

Figure 22. Word cloud of words reported for the "other" category. The most frequently used word was "annoyed". Larger words were reported more than smaller words.

6.2.3. **RQ.3: What are the relationships among sentiments in lyrics and humans' response to lyrics?**

> **RO3.1**: Describe the relationship between positive aspects of sentiments in lyrics and humans' responses to sentiments in lyrics.
>
> **RO3.2:** Describe the relationship between negative aspects of sentiments in lyrics and humans' responses to sentiments in lyrics.

Both objectives for RQ.3 were analyzed and described together, and interchangeably, so they were put together in this section. The words identified from the AFINN sentiment analysis were plotted and markers were added for each time one of the specified words occurred in a selected song. Words were timestamped at the exact time they appeared in the selected song. AFINN words were given a value between +5 and -5, with neutral words (0) eliminated. Words were then recoded so that they could be plotted along with the participants' dial scores. Because 50 on the dials was neutral, 50 needed to be the neutral (0) score for the AFINN recodes. The data were recoded to 5 → 100, 4 → 90, 3 → 80, 2 → 70, 1 → 60, 0 → 50, -1 → 40, -2 → 30, -3 → 20, -4 → 10, and -5 → 0.

The mean dial scores for each second of each song were then plotted on a line graph using the data visualization software, Tableau. Each of the AFINN word scores were also plotted on the same graph at each second that they occurred. The graphs for each song are located in figures 22 - 30. The mean dial score response is indicated in the horizontal orange lines. The AFINN words are indicated as blue circles. These graphs were plotted to give a visual for participant response to sentiments. Something to take into account, is that there may be a 1 to 2 second delay in participant response. This is due to reaction time between hearing the music and moving the dial. Figures 22 – 33 are given below.

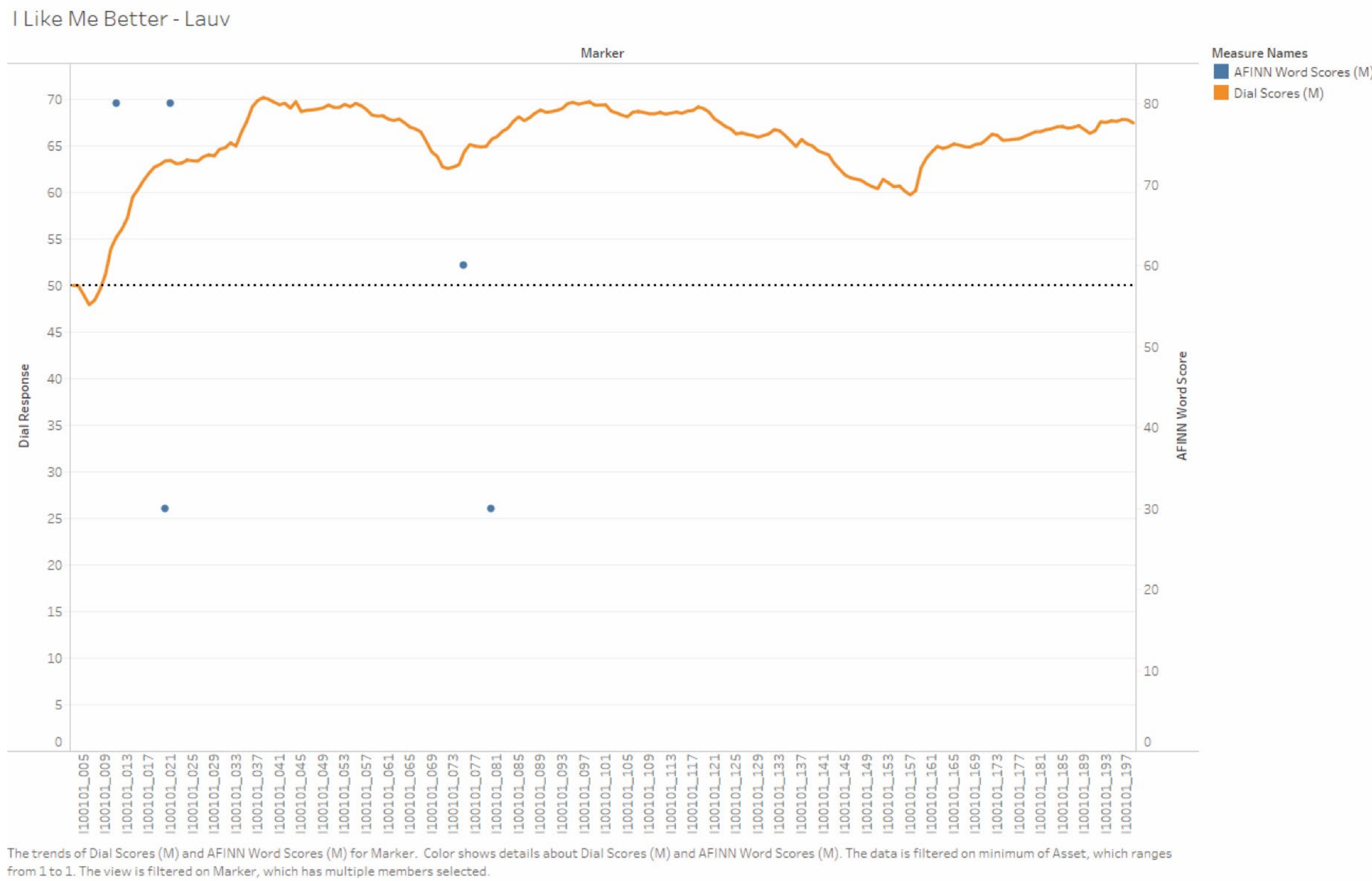

Figure 23. The plotted song in this figure is *I Like Me Better* by Lauv.

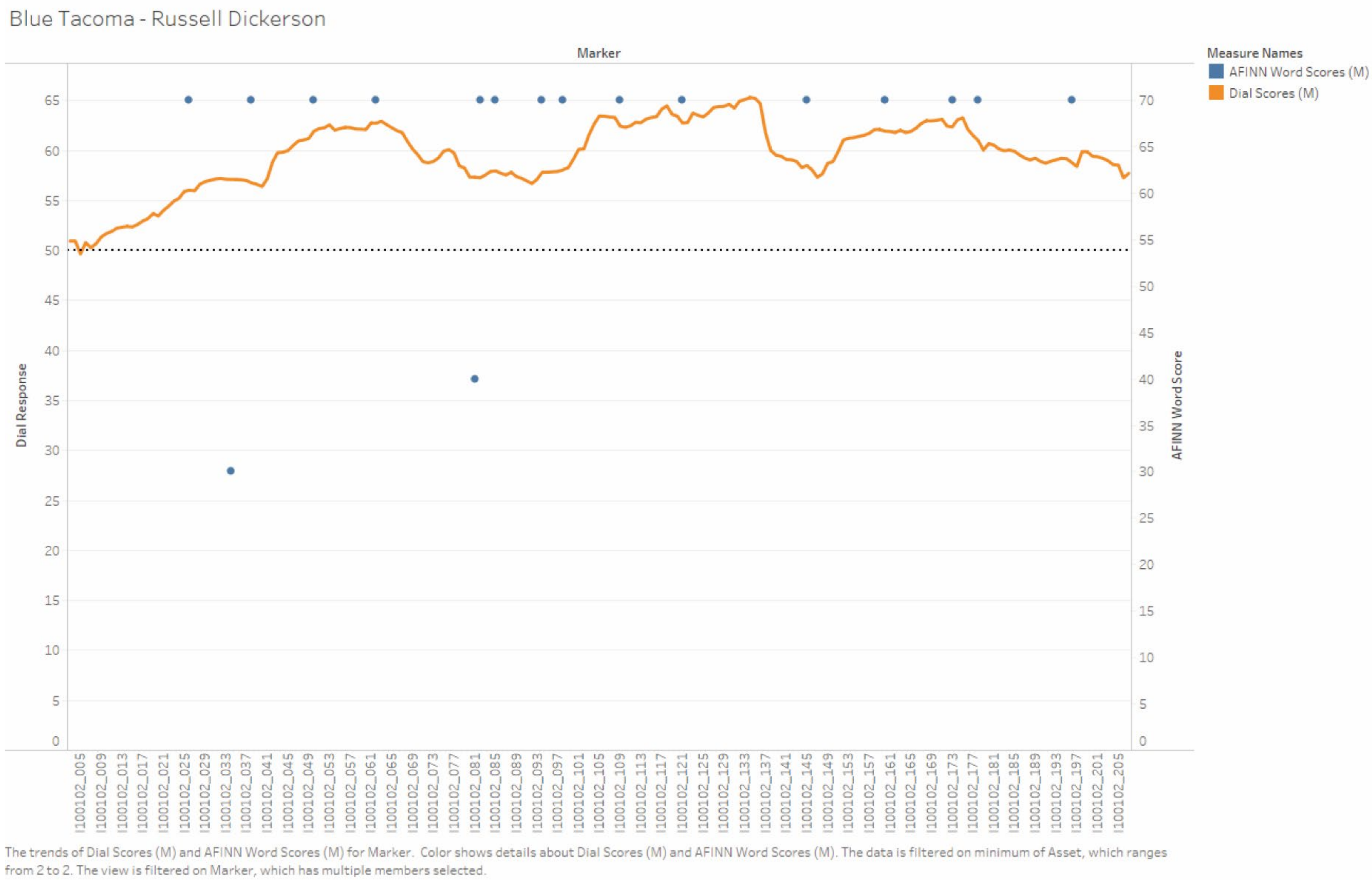

Figure 24. The plotted song in this figure is *Blue Tacoma* by Russell Dickerson.

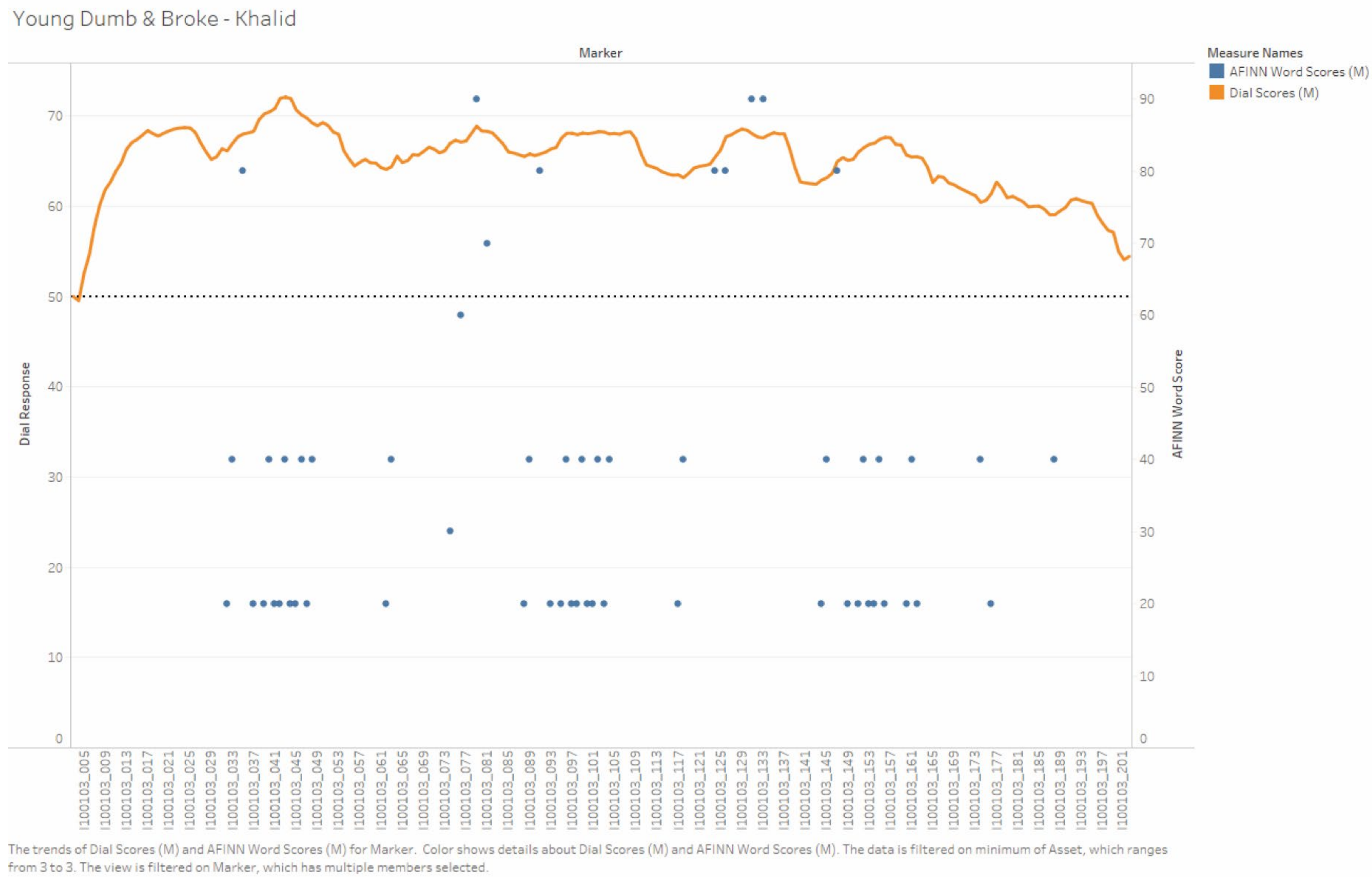

Figure 25. The plotted song in this figure is *Young Dumb & Broke* by Khalid.

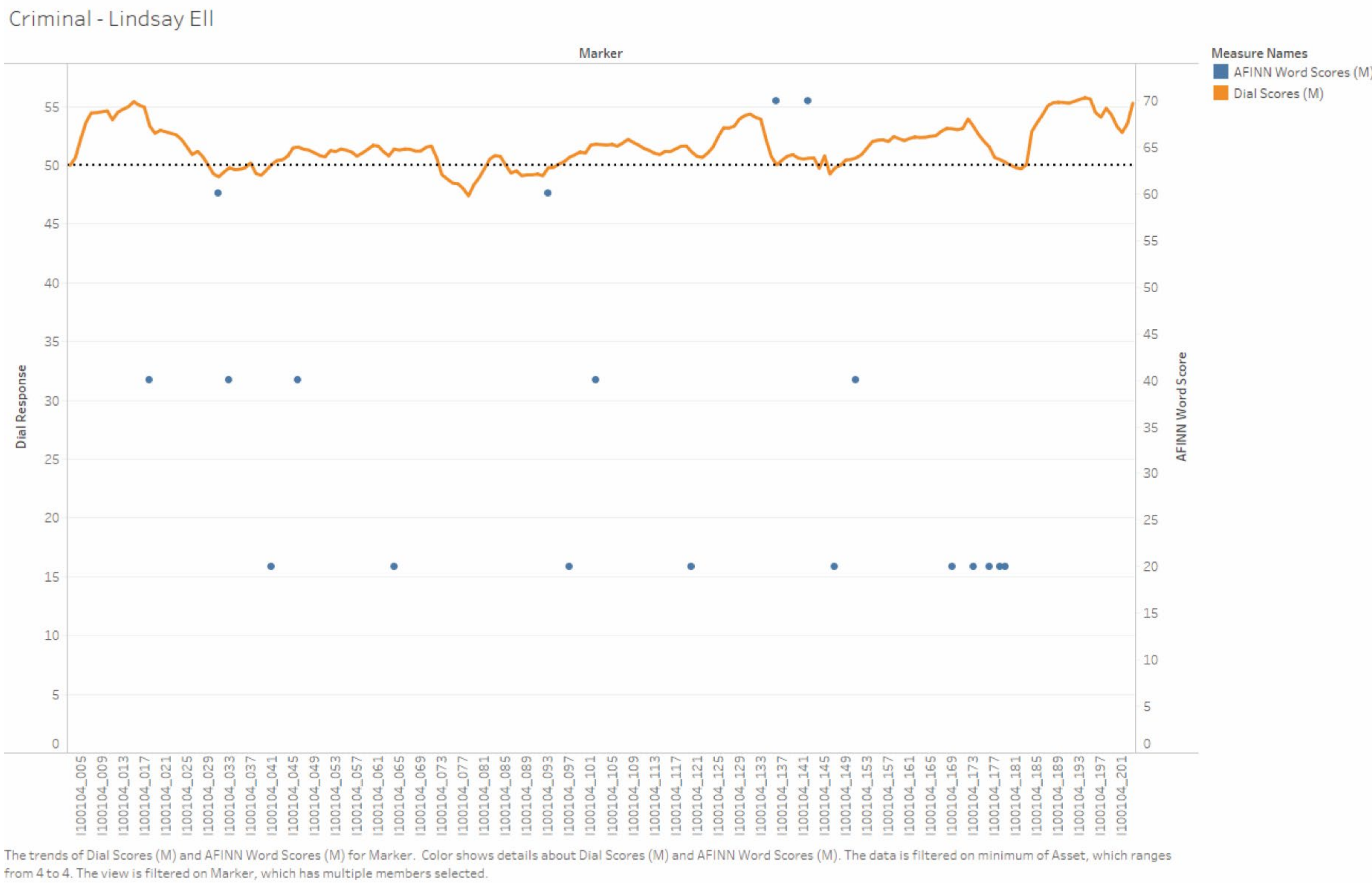

Figure 26. The plotted song in this figure is *Criminal* by Lindsay Ell.

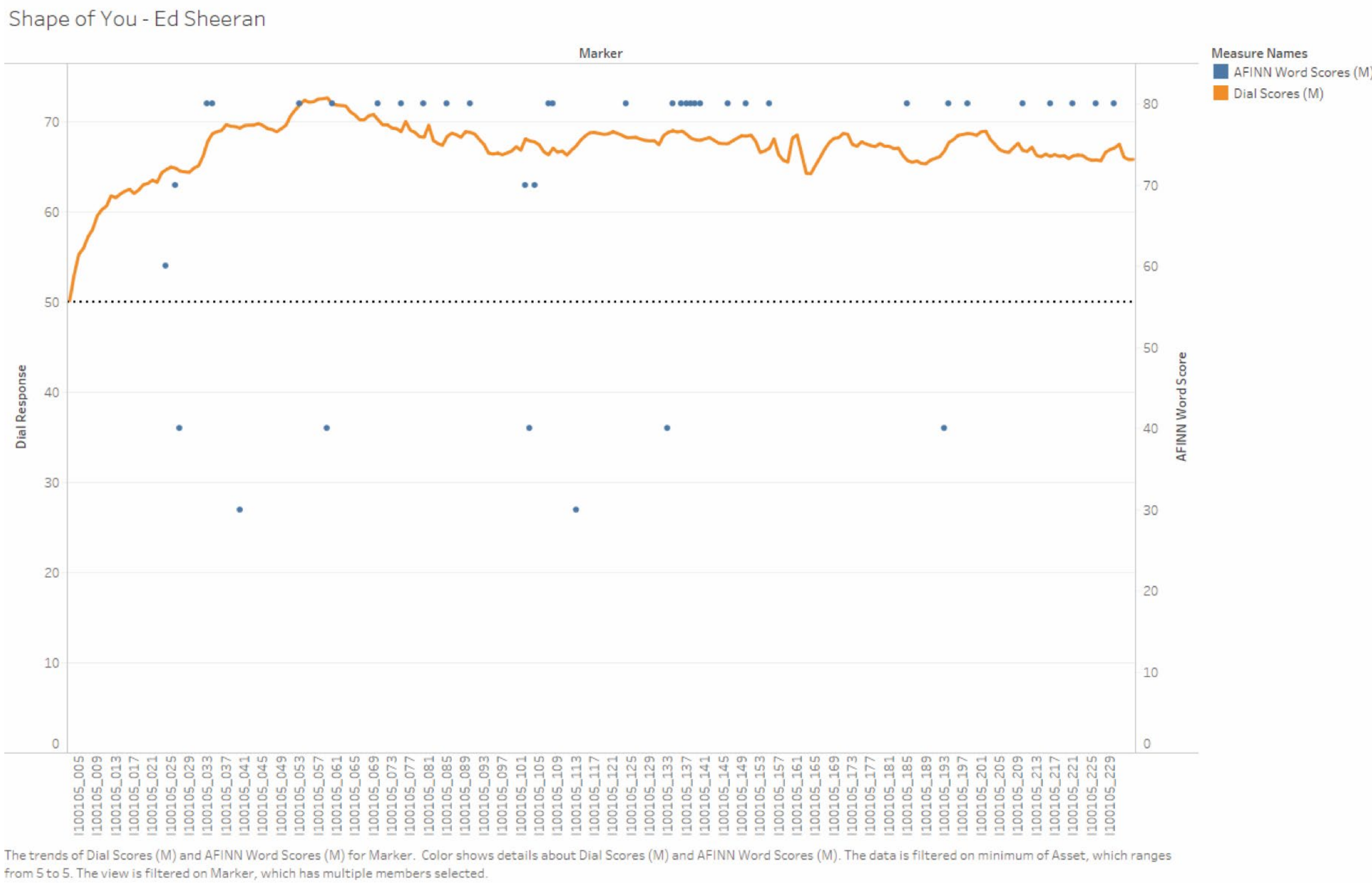

Figure 27. The plotted song in this figure is *Shape of You* by Ed Sheeran.

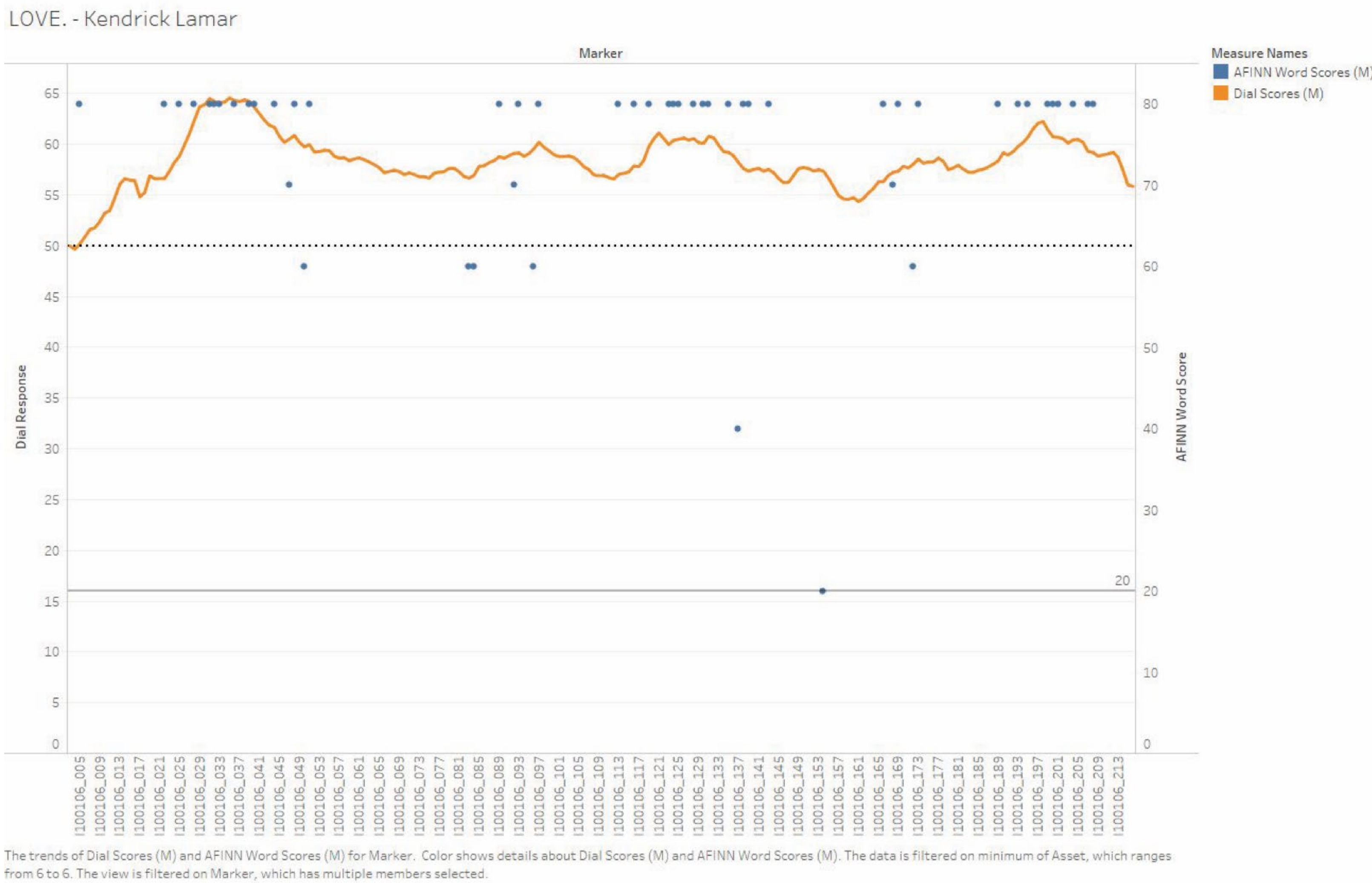

Figure 28. The plotted song in this figure is *LOVE.* by Kendrick Lamar.

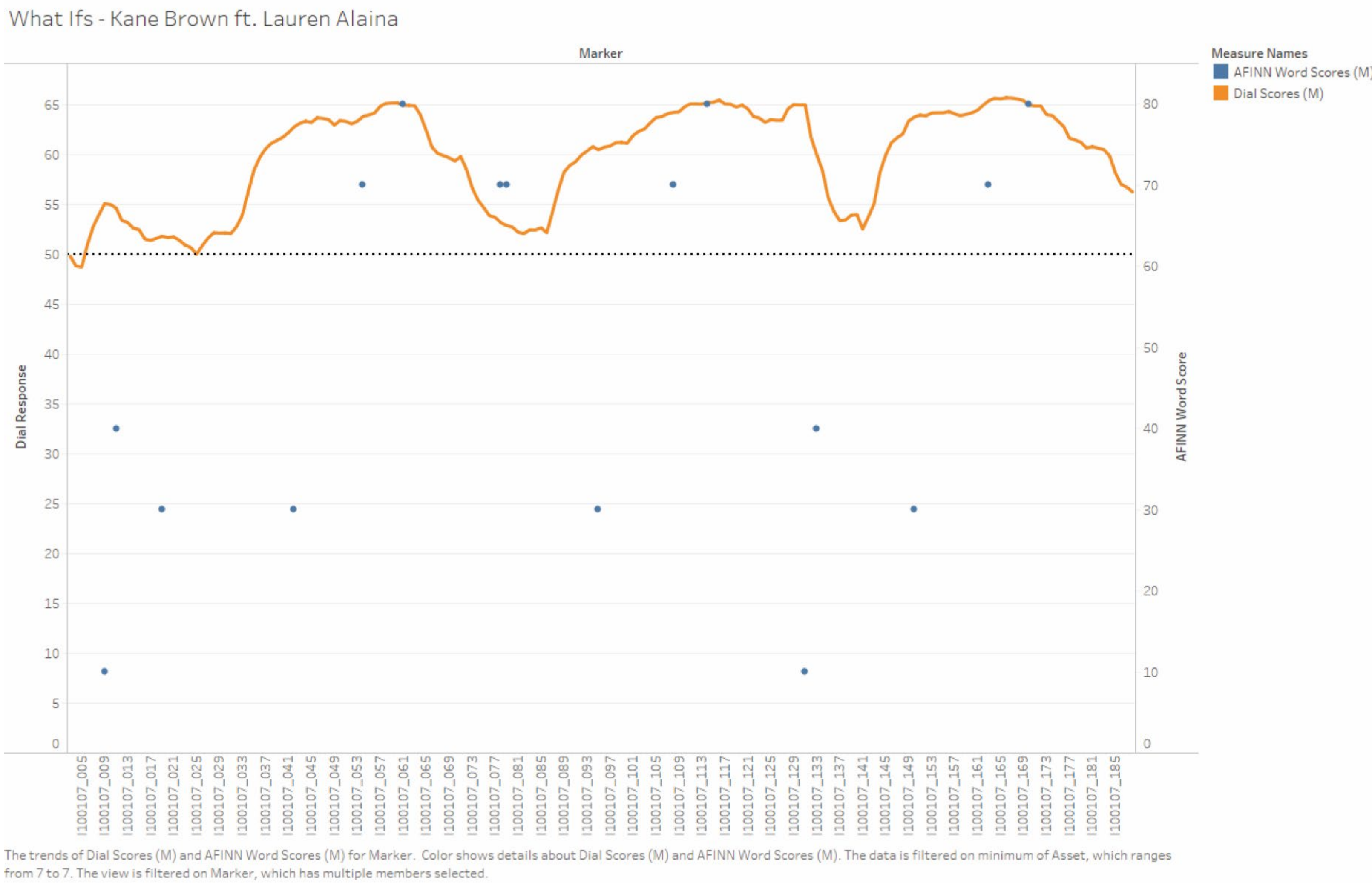

Figure 29. The plotted song in this figure is *What Ifs* by Kane Brown ft. Lauren Alaina.

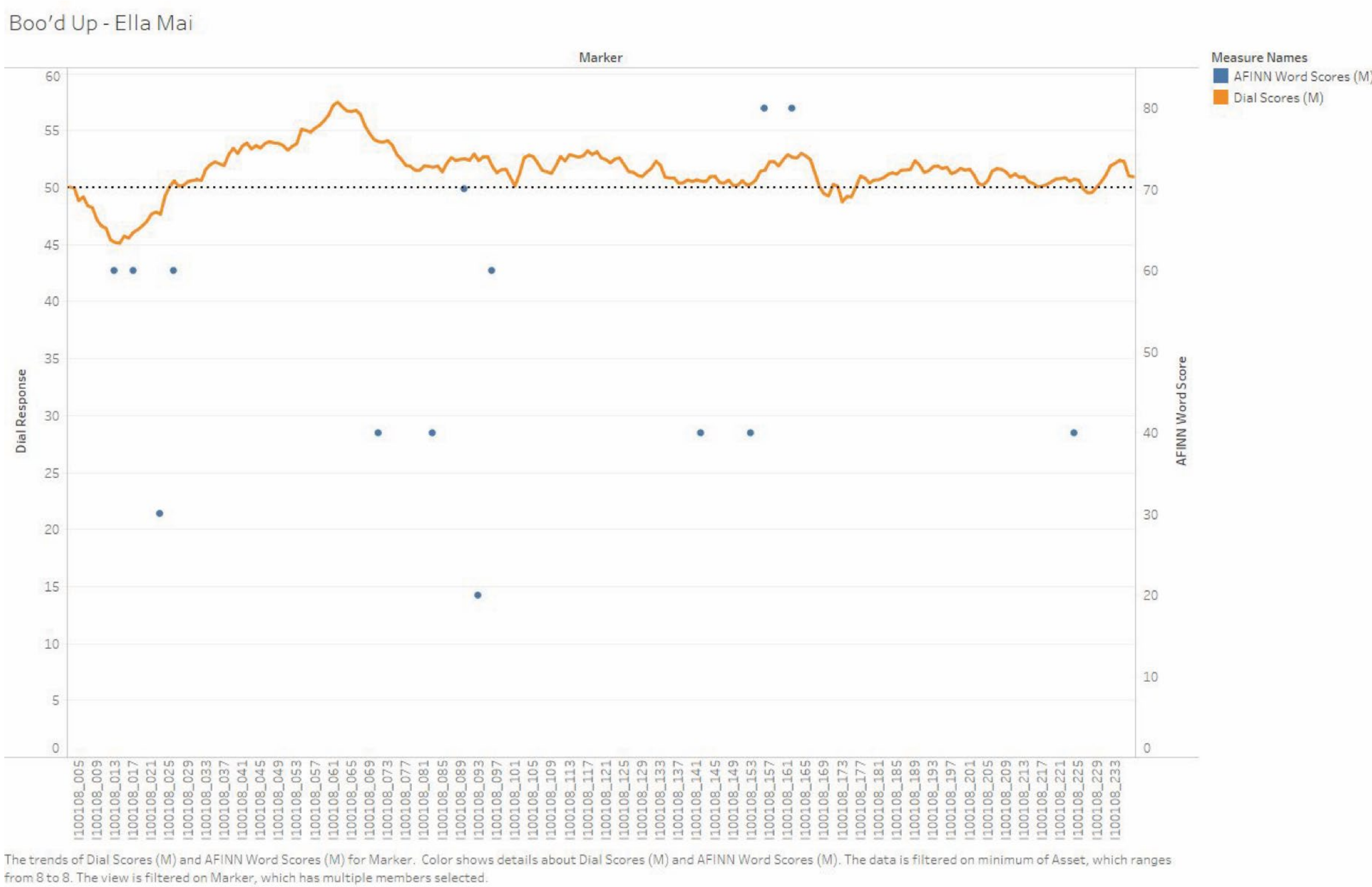

Figure 30. The plotted song in this figure is *Boo'd Up* by Ella Mai.

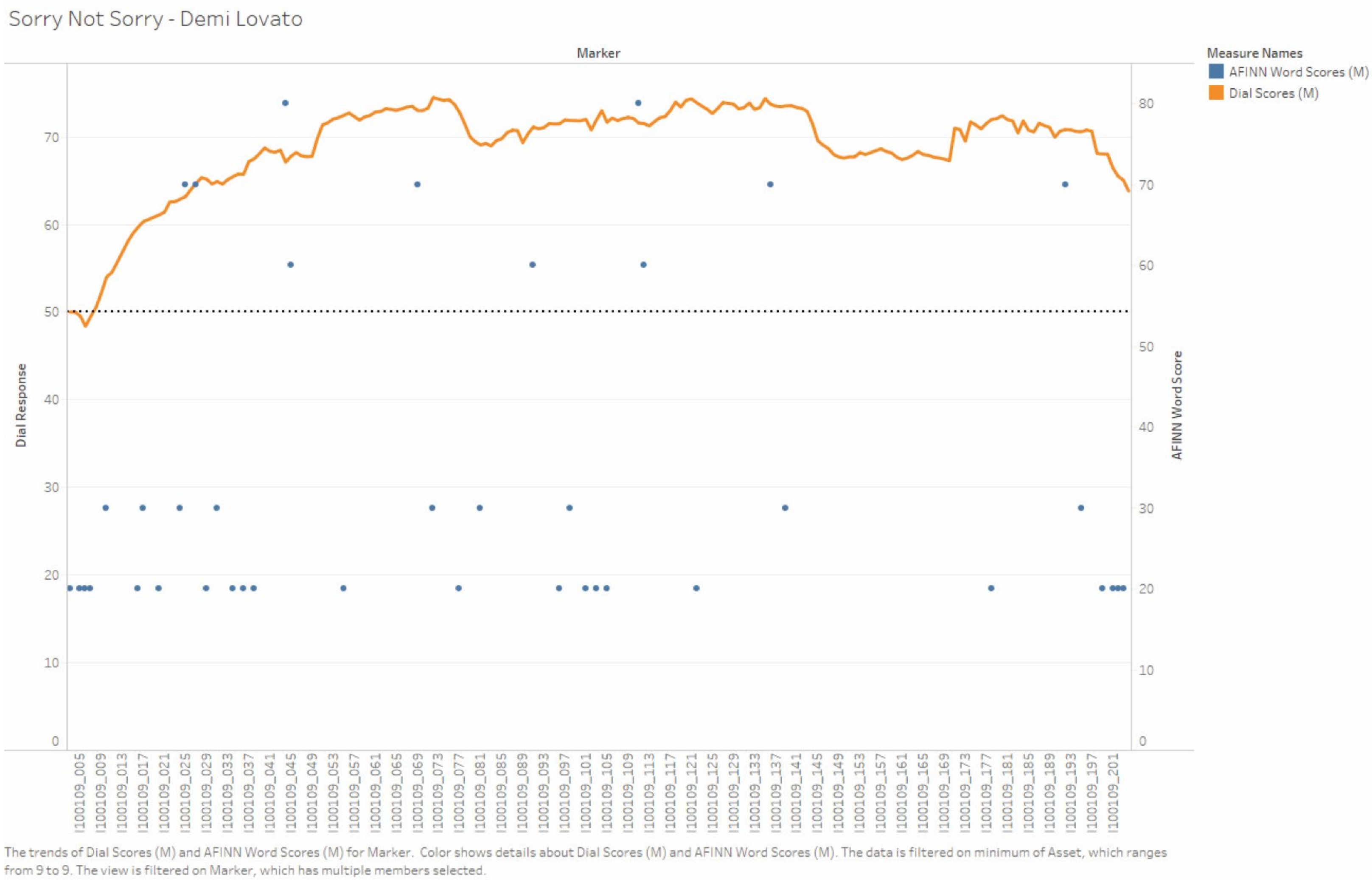

Figure 31. The plotted song in this figure is *Sorry Not Sorry* by Demi Lovato.

A bivariate correlation was run on the AFINN scores with the second-by-second mean dial scores. Three seconds were accounted for each AFINN score to cover if there was a delay in response time. Correlation was not significant in eight of the nine songs. However, correlation was significant in *LOVE.* [*sic*] by Kendrick Lamar (Urban, 2017 & 2018) because the correlation coefficient is significantly different from zero.. A negative correlation was reported in five of the nine songs and they were not significant. Results are located in table 15.

Table 15
AFINN and mean dial scores in bivariate correlation

Song	*n*	*r*	*p*
I Like Me Better	5	-.619	.265
Blue Tacoma	17	.341	.181
Young Dumb & Broke	56	-.013	.922
Criminal	19	-.305	.204
Shape of You	40	-.005	.975
LOVE.	51	.280*	.047
What Ifs	16	.325	.22
Boo'd Up	14	-.03	.92
Sorry Not Sorry	41	.268	.09

Note. n = sample of AFINN words per song; r = Pearson correlation; p = significance (two – tailed); * = Correlation is significant at the 0.05 level (two-tailed).

Participants were asked how to describe the polarity of each song they listened to on a five-point Likert scale ranging from extremely negative to extremely positive. Answers were recoded so that there was only negative, neutral, and positive (extremely negative and somewhat negative → negative; neither positive nor negative → neutral;

somewhat positive and extremely positive → positive). Each participant chose one answer choice per song. Eight of the nine songs were rated more positive than negative or neutral. Four of the nine songs had a rating of being overall positive (above 60%). Two of the nine songs had a majority of being positive (above 50%, less than 60%), however, the majority was not a consensus. Two of the nine songs resulted in a plurality, with more positive ratings than negative or neutral, however less than half of the sample rated the songs as positive. One of the nine songs also resulted in a plurality, with more negative ratings than positive or neutral, however less than half of the sample rated the song as positive. Results are located in Table 16.

Table 16

Polarity of Bing sentiment analysis and participant music survey response

Song	Genre	Bing	*n*	Positive *f*	Positive %	Neutral *f*	Neutral %	Negative *f*	Negative %
I Like Me Better	Pop CHR	(0)	53	38	72	4	8	11	21
Blue Tacoma	Country	(+)	53	39	74	4	8	10	19
Young Dumb & Broke	Urban	(-)	53	30	57	10	19	13	25
Criminal	Country	(-)	53	17	32	17	32	19	36
Shape of You	Pop CHR	(+)	53	42	79	3	6	8	15
LOVE.	Urban	(+)	53	25	47	14	26	14	26
What Ifs	Country	(0)	53	27	51	14	26	12	23
Boo'd Up	Urban	(0)	53	21	40	15	28	17	32
Sorry Not Sorry	Pop CHR	(-)	53	44	83	4	8	5	9

To measure the correlation between the participants' polarity survey responses to each song, a bivariate correlation was run on the survey response (UC0081) and the dial scores. The dial scores were recoded to stay consistent with the survey responses. The M dial scores were recoded to 0 – 20 → 1, 21 – 40 → 2, 41 – 60 → 3, 61 – 81 → 4, 81 – 100 → 5. All survey responses and dial scores had a significant correlation at the .01 level (two – tailed). The mean (*M*), standard deviations (*SD*), and Pearson correlation (*r*) are located in table 17.

Table 17

Survey polarity response with dial score mean (n = 53)

			Survey Responses		*Dial Responses*	
Song	Genre	*r*	*M*	*SD*	*M*	*SD*
I Like Me Better	Pop CHR	0.82**	3.81	1.144	58.796	22.015
Blue Tacoma	Country	.822**	3.77	1.12	55.557	19.52
Young Dumb & Broke	Urban	.836**	3.55	1.249	60.27	21.116
Criminal	Country	.799**	2.94	1.099	47.936	18.244
Shape of You	Pop CHR	.873**	3.98	1.217	60.657	23.035
LOVE.	Urban	.898**	3.47	1.25	52.772	23.31
What Ifs	Country	.793**	3.38	1.164	54.395	21.515
Boo'd Up	Urban	.697**	3.06	1.216	47.044	20.815
Sorry Not Sorry	Pop CHR	.737**	4.09	0.966	62.12	23.043

Note. Dial responses were recoded to 1 – 5; **. Correlation is significant at the 0.01 level (2-tailed).

Aim 2: Describe the internal and external factors that play a role in response to music lyrics.

6.2.4. **RQ.4: Are demographic, psychographic, environmental, and/or behavioral factors predictors of human response to lyrical music?**

RO4.1: Describe the relationship among demographic factors and humans' responses to sentiments in lyrics.

There were three demographics variables used in this study including, gender, age, and music education. The gender sample was inadequate because there was not enough of a range in participants (10 male, 43 female). Gender was used as a descriptive, for groups, but was not used in the analysis.

Participants met the target demographic and ranged between 19 and 30 years old. Ages were recoded into two categories; 19 – 22 →1, 23 – 30 → 2. A bivariate correlation was then used to determine the *r* and *p* of age group and dial score. There was one significant inverse correlation at the .01 level (two –tailed) between *Young Dumb & Broke* by Khalid and age. There was also one significant inverse correlation (.05 level [two –tailed]) between *LOVE*. By Kendrick Lamar and age (table 18).

Table 18
Age correlated with dial score (n = 53)

Song	*r*	*p*
I Like Me Better	-.124	.377
Blue Tacoma	.041	.769
Young Dumb & Broke	-.460**	.001
Criminal	.188	.176
Shape of You	-.001	.997
LOVE.	-.330*	.016
What Ifs	.118	.402
Boo'd Up	.031	.828
Sorry Not Sorry	-.007	.959

Notes. **. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.0 level (2-tailed); *r* = Pearson correlation; *p* = significance (two – tailed).

Table 19 is a summary of music education levels from all 53 participants. Participants were asked their level of music education on a five-point Likert scale. A total of 43% of participants rated themselves as having a little, 26% rated themselves as having none at all, 21% rated themselves as having a moderate amount, 6% rated themselves as having a lot, and 4% rated themselves as having a great deal. Females rated themselves as having a higher level of music education than males.

Table 19
Summary of self-reported levels of music education (n = 53)

Music Education	Total		Male		Female	
	f	%	*f*	%	*f*	%
None at all	14	26	3	30	20	47
A little	23	43	3	30	11	26
A moderate amount	11	21	3	30	8	19
A lot	3	6	1	10	2	5
A great deal	2	4	0	0	2	5
Total	53	100	10	100	43	100

A bivariate correlation was used to see if there were any similarities between music education level and dial score. Music education answer choices were collapsed to smaller groups to have greater cell size (1 → 1; 2 → 2, 3 → 3; 4, 5 → 4). Table 20 indicates that there were three inverse relationships at the .05 level (two – tailed) between dial scores and music education for three of the nine songs including, *Blue Tacoma*, *What Ifs*, and *Sorry Not Sorry*. As dial scores (song rating) went up, music education went down.

Table 20

Music education correlated with dial score (n = 53)

Song	*r*	*p*
I Like Me Better	-.212	.128
Blue Tacoma	-.344*	.012
Young Dumb & Broke	-.041	.769
Criminal	-.020	.887
Shape of You	-.175	.209
*-.316*LOVE.*	.032	.821
What Ifs	-.316*	.021
Boo'd Up	.137	.329
Sorry Not Sorry	-.315*	.022

Notes. *. Correlation is significant at the 0.05 level (2-tailed); *r* = Pearson correlation; *p* = significance (two – tailed).

RO4.2: Describe the relationship among psychographic factors and humans' responses to sentiments in lyrics.

There were four variables used to describe psychographics among participants including if they play a musical instrument, if they think music lyrics play a role in emotional response, if they think past experiences play cause an emotional reaction while listening to music, and their familiarity with each song.

Table 21 is a summary of the participants' the results as either yes or no to if the participant played a music instrument. A total of 38% of participants selected yes, while 62% selected no. More participants in this did not play an instrument than those who did.

Table 21
Plays a musical instrument

	Total		Male		Female	
Musical Instrument	*f*	%	*f*	%	*f*	%
Yes	20	38	6	60	14	33
No	33	62	4	40	29	67
Total	53	100	10	100	43	100

Participants were asked if they think that past experience causes an emotion reaction while listening to music. Table 22 describes that 98% (n = 52) selected yes, while 2% (n = 1) selected no.

Table 22
Past experiences cause an emotional reaction while listening to music and gender

	Total		Male		Female	
Past Experiences	*f*	%	*f*	%	*f*	%
Yes	52	98	9	90	43	100
Maybe	0	0	0	0	0	0
No	1	2	1	10	0	0
Total	53	100	10	100	43	100

Participants were asked if they think that music lyrics contribute to emotional experience. Table 23 is a summary of their responses with either yes, maybe, or no. Yes was selected by 87% of the participants, maybe was selected by 13%, and no was not selected.

Table 23
Music lyrics contribute to emotional experience

	Total		Male		Female	
Music Lyrics	*f*	%	*f*	%	*f*	%
Yes	46	87	9	90	37	86
Maybe	7	13	1	10	6	14
No	0	0	0	0	0	0
Total	53	100	10	100	43	100

A bivariate correlation was used to find similarities among the three variables described in tables 21, 22, and 23 and the dial scores for all nine songs. There were two significant relationships at the .05 level (two – tailed) with musical instrument and *I Like Me* Better and *Shape of You*. Dial scores increased for people who did not play a music instrument. There were no significant relationships between past experiences and dial score. There was one significant inverse relationship at the .05 level (two – tailed) between music lyrics and *Criminal*.

Table 24

Past experiences cause an emotional reaction while listening to music and dial score

	Instrument		Past		Lyrics	
Songs	*r*	*p*	*r*	*p*	*r*	*p*
I Like Me Better	.344*	.012	.094	.505	.066	.966
Blue Tacoma	.261	.059	.174	.212	.060	.670
Young Dumb & Broke	.225	.106	.171	.221	.009	.950
Criminal	-.144	.418	.074	.600	-.327*	.017
Shape of You	.276*	.046	.074	.601	-.096	.495
**LOVE.*	.235	.090	.166	.236	.078	.578
What Ifs	.092	.514	.154	.270	.167	.233
Boo'd Up	.093	.509	.040	.775	.019	.893
Sorry Not Sorry	.187	.181	.072	.609	.225	.065

Notes. *. Correlation is significant at the 0.0 level (2-tailed); *r* = Pearson correlation; *p* = significance (two – tailed).

After listening to each of the nine songs, participants were asked to rate their level of familiarity of the song. Table 25 is a summary of the familiarity levels for each of the nine songs. The three pop songs (I like me better, shape of you, and Sorry Not Sorry) were the most familiar songs with a percentage of "I have heard it many times" 75% and up. Criminal by Lindsay Ell was the least recognized song.

Table 25
Song Familiarity f and %

	1		2		3		4	
Song	*f*	%	*f*	%	*f*	%	*f*	%
I Like Me Better	4	8	1	2	8	15	40	75
Blue Tacoma	15	28	8	15	10	19	20	38
Young Dumb & Broke	4	8	4	8	8	15	37	70
Criminal	25	47	11	21	8	15	9	17
Shape of You	1	2	0	0	2	4	50	94
LOVE.	13	25	6	11	6	11	28	53
What Ifs	14	30	0	0	9	20	23	50
Boo'd Up	21	40	9	17	11	21	12	23
Sorry Not Sorry	2	4	2	4	9	17	40	75

Notes. 1 = I have never heard it before; 2 = I may have heard it before; 3 = I have heard it before; 4 = I have heard it many times; *f* = frequency.

Bivariate correlation (two – tailed) was used to see if there was a significant correlation between song familiarities and dial scores. Results from correlation analysis were that six out of the nine sounds (66.66%) had a significant relationship at the .001 level, one out of the nine songs (.111%) had a significant relationship at the .05 level, and two out of the nine songs (.222%) did not have a significant relationship. As familiarity with the song increased, so did the dial scores.

Table 26

Song Familiarity and dial score

Song	*M*	*SD*	*r*	*p*	*n*
I Like Me Better	3.58	.865	.603**	<.001	53
Blue Tacoma	2.66	1.255	.208	.136	53
Young Dumb & Broke	3.47	.932	.487**	<.001	53
Criminal	2.02	1.152	.048	.731	53
Shape of You	3.91	.450	.447**	.001	53
LOVE.	2.92	1.284	.726**	<.001	53
What Ifs	2.77	1.266	.521**	<.001	53
Boo'd Up	2.26	1.211	.331*	.016	53
Sorry Not Sorry	3.64	.736	.516**	<.001	53

Notes. 1 = *. Correlation is significant at the 0.05 level (2-tailed); **. Correlation is significant at the 0.01 level (2-tailed); *r* = Pearson correlation; *p* = Significance (two – tailed); *n* = sample

RO4.3: Describe the relationship among environmental factors and humans' responses to sentiments in lyrics.

Participants were asked what activities they engage in while listening to music. Table 27 summarizes 13 categories, and "other", the participants selected from. Participants were able to select as many activities as they wanted. Driving/Traveling was the most selected, working/studying and housework were the second most selected, and relaxing was the third most selected.

Table 27
Activities while listening to music

Activity	*f*	%
Working/Studying	47	89
Hanging out with friends	41	77
Relaxing	44	83
Watching TV or a movie	4	8
Housework	47	89
Driving/Traveling	52	98
Eating	26	49
Cooking	38	72
Working out or playing sports	43	81
Shopping	11	21
Partying	42	79
Playing video games	14	26
At a concert	43	81
Other	4	8

Participants were able to select multiple options for activities they are doing while listening to music question. A bivariate correlation was used to find similarities between activities and dial scores. Three songs had significant relationships with activities at the .05 level (two-tailed), and one song has a significant relationship with an activity at the .01 level (two-tailed). Two songs had an inverse significance with activities .05 level (two-tailed), and four songs had inverse significant relationships at the .05 (two- tailed) with activities. Results can be seen in table 28.

Table 28
Activities while listening to music with dial scores

Activity		Songs								
		1	2	3	4	5	6	7	8	9
1	*r*	-.202	-.124	-.297*	.205	-.257	-.200	-.104	.115	-.065
	p	.147	.375	.031	.140	.063	.150	.459	.414	.646
2	*r*	-.089	-.038	.111	-.095	-.076	.171	-.149	-.137	-.074
	p	.526	.790	.427	.497	.590	.220	.286	.326	.597
3	*r*	-.181	-.207	-.005	-.087	-.134	-.144	-.234	-.051	-.154
	p	.195	.137	.973	.535	.338	.303	.091	.717	.271
4	*r*	-.059	-.159	.086	-.081	.081	-.055	.008	-.011	-.166
	p	.673	.255	.539	.563	.565	.697	.954	.938	.234
5	*r*	-.060	-.295*	.060	-.175	-.193	.083	-.258	.107	-.082
	p	.668	.032	.668	.210	.166	.555	.062	.444	.561
6	*r*	.065	-.141	.163	-.166	-.216	.306*	.013	.135	.386**
	p	.645	.315	.244	.234	.121	.026	.928	.336	.004
7	*r*	-.163	-.196	.069	-.090	-.199	-.001	-.241	.036	-.054
	p	.244	.160	.625	.523	.152	.994	.082	.800	.703
8	*r*	-.264	-.434**	-.072	.036	-.200	-.097	-.380**	.105	-.082
	p	.056	.001	.606	.800	.150	.489	.005	.455	.561

Table 28 (continued)

		1	2	3	4	5	6	7	8	9
9	*r*	.120	.021	.115	-.062	-.107	.059	.114	-.081	.069
	p	.393	.882	.410	.657	.447	.675	.416	.564	.626
10	*r*	.019	-.268	.164	-.108	.018	.005	-.259	.189	-.115
	p	.894	.053	.240	.440	.900	.971	.061	.176	.411
11	*r*	.205	.016	-.038	.026	.147	.123	.235	.144	.296*
	p	.141	.911	.785	.853	.294	.381	.090	.302	.031
12	*r*	-.062	-.374**	.298*	.050	-.078	.138	-.147	-.070	-.057
	p	.660	.006	.030	.725	.581	.324	.293	.616	.684
13	*r*	<.001	-.180	-.034	-.151	-.355**	.058	-.151	.010	.019
	p	.999	.198	.808	.282	.009	.679	.279	.946	.893
14	*r*	-.297*	-.229	-.222	-.043	-.207	-.128	-.198	-.155	-.209
	p	.031	.099	.110	.760	.137	.362	.156	.267	.134

Note. *. Correlation is significant at the 0.05 level (2-tailed); **. Correlation is significant at the 0.01 level (2-tailed); *r* = Pearson correlation; *p* = Significance (two – tailed); Songs: 1 = *I Like Me Better*; 2 = *Blue Tacoma*; 3 = *Young Dumb & Broke*; 4 = *Criminal*; 5 = *Shape of You*; 6 = *LOVE.*; 7 = *What Ifs*, 8 = *Boo'd Up*; 9 = *Sorry Not Sorry*; Activities: 1 = Working/Studying; 2 = Hanging out with friends; 3 = Relaxing; 4 = Watching TV or a movie; 5 = Housework; 6 = Driving/Traveling; 7 = Eating; 8 = Cooking; 9 = Working out or playing sports; 1 = Shopping; 11 = Partying; 12 = Playing video games; 13 = At a concert; 14 = Other.

11

Participants were asked what times of they listen to music and were able to select multiple options. Below, table 29 is a summary of the times of day selected. The data are ordinal with the times of day in sequence of occurrence. Evening and afternoon were ranked as first; morning was ranked as second; early morning and overnight were ranked as third.

Table 29
Times of day listening to music

Time of Day	*Total*		*Male*		*Female*	
	f	%	*f*	%	*f*	%
Early Morning (4:00 a.m. to 8:59 a.m.)	29	55	3	6	26	49
Morning (9:00 a.m. to 12:59 p.m.)	44	83	7	13	37	70
Afternoon (1:00 p.m. to 5:59 p.m.)	48	91	9	17	39	74
Evening (6:00 p.m. to 10:59 p.m.)	48	91	8	15	40	75
Over Night (11:00 p.m. to 3:59 a.m.)	29	55	4	8	25	47

A bivariate correlation was used in table 30 to find the similarities between times of day spent listening to music and dial scores. There were two significant relations on the .05 level (two – tail) between afternoon and *Sorry Not Sorry*, and also evening and *Young Dumb & Broke.* There was one significant relationship on the .05 level (two – tail) between evening and *LOVE*. Two inverse relationships were identified between evening and *Criminal*, and between over-night and *Criminal*.

Table 30
Times of day listening to music and dial scores

Time of day		Songs								
		1	2	3	4	5	6	7	8	9
1	*r*	0.071	-0.176	0.033	-0.082	0.110	0.173	0.011	0.143	0.099
	p	0.616	0.207	0.814	0.561	0.434	0.217	0.936	0.308	0.480
2	*r*	-0.073	-0.022	0.155	-0.196	-0.068	0.048	-0.026	0.041	0.221
	p	0.603	0.875	0.267	0.159	0.629	0.735	0.852	0.773	0.112
3	*r*	-0.168	0.084	0.054	-0.025	-0.224	0.076	0.103	0.047	.284*
	p	0.229	0.550	0.700	0.856	0.107	0.587	0.463	0.736	0.040
4	*r*	0.140	-0.186	.327*	-.311*	-0.160	.444**	-0.038	0.118	0.189
	p	0.318	0.183	0.017	0.024	0.251	0.001	0.786	0.401	0.176
5	*r*	-0.113	-0.167	0.227	-.275*	-0.077	0.017	-0.016	0.071	-0.014
	p	0.419	0.231	0.102	0.047	0.584	0.906	0.911	0.613	0.921

Note. *. Correlation is significant at the 0.05 level (2-tailed); **. Correlation is significant at the 0.01 level (2-tailed); *r* = Pearson correlation; *p* = Significance (two – tailed); Songs: 1 = *I Like Me Better*; 2 = *Blue Tacoma*; 3 = *Young Dumb & Broke*; 4 = *Criminal*; 5 = *Shape of You*; 6 = *LOVE.*; 7 = *What Ifs*, 8 = *Boo'd Up*; 9 = *Sorry Not Sorry*; Time of day: 1 = Early Morning 4:00 a.m. to 8:59 a.m.; 2 = Morning (9:00 a.m. to 12:59 p.m.), 3 = Afternoon (1:00 p.m. to 5:59 p.m.); 4 = Evening (6:00 p.m. to 10:59 p.m.); 5 = Over Night (11:00 p.m. to 3:59 a.m.).

RO4.4: Describe the relationship among behavioral factors and humans' responses to sentiments in lyrics.

Participants were asked if they listen to playlists to enhance their music experience to help describe listening behavior. Table 31 summarizes their responses as either yes or no. Yes was selected by 74% of the participants; whereas, no was select by 26% of participants.

Table 31
Playlists enhance music experience and gender

	Total		Male		Female	
Playlists	*f*	%	*f*	%	*f*	%
Yes	39	74	5	50	34	79
No	14	26	5	50	9	21
Total	53	100	10	100	43	100

To determine if there were similarities between playlist use and dial scores, a bivariate correlation was used. Reported in table 32 are *r, p, M, SD*. One of nine songs had an inverse relationship at the .05 level (two – tailed). People who listened to playlists to enhance their musical experience scored *Sorry Not Sorry* lowered on their dials.

Table 32
Playlists enhance music experience and dial score

Song	*r*	*p*	*M*	*SD*
I Like Me Better	-.184	.187	58.796	22.057
Blue Tacoma	<.001	.999	55.557	19.520
Young Dumb & Broke	.082	.558	60.270	21.116
Criminal	.073	.605	47.936	18.244
Shape of You	-.147	.294	60.657	23.035
LOVE.	-.075	.593	52.772	23.310
What Ifs	-.105	.455	54.395	21.515
Boo'd Up	-.176	.208	47.044	20.815
Sorry Not Sorry	-.284*	.039	62.120	23.043

Notes.*. Correlation is significant at the 0.05 level (2-tailed); *r* = Pearson correlation; *p* = Significance (two – tailed); *M* = mean; *SD* = standard deviation.

Aim 3: Describe how sentiment, human response, and internal and external factors predict the performance of music on the Billboard charts.

6.2.5. RQ.5: Are sentiments in lyrics and/or human responses to lyrics predictors of year end performance on music charts?

RO5.1: Describe the relationship among sentiments in lyrics and year end performance on music charts.

There was not a significant relationship between Billboard chart rankings and Bing negative word ratio, $r = .04$, and p (two-tailed) = .35. There was also not a significant relationship between Billboard chart rankings and Bing positive word ratio, $r = .05$, and p (two-tailed) = .24. There was not a significant relationship between Billboard chart rankings and AFINN word score, $r = .06$, and p (two-tailed) = .18.

Table 33

Bing lexicon and Billboard chart rank with bivariate correlation

Measure	1	2	3
1. Billboard	--	- .04	- .05
2. Negative Word Ratio	- .04	--	.03
3. Positive Word Ratio	- .05	.03	--

7. CONCLUSION

This study explores and describes what sentiments are present in contemporary lyrical music, their polarity, and their emotional category. This study also explores and describes human response to music, and how the described sentiments contribute, or differ from human response. People have a reaction to music. There are many elements in music that can be attributed to those reactions. This study was intended to provide information about how people react to sentiments in music lyrics, provide insight into the similarities and differences between computer-generated responses and human response, and provide possible predictions based on the results. Descriptive statistics were used in each of the research questions to describe what was found in the collected sample. Using these descriptions, inferential statistics were made about the population based on the sample data.

There were, however, some limitations that occurred in this study. Starting with the first method, the sentiment analysis had issues of its own. The three Billboard year-end charts selected, Hot Country (country), Pop (pop CHR), and Hot R&B/Hip-Hop (urban), did not have an equal number of charting positions available. Country and urban both had 100 charting positions available, whereas pop only had 50. This did not make for an equal representation of each genre for the sentiment analysis. The reduced number of songs in the pop chart made for a decreased word count, a decrease in artist and song variety, and a lower number of sentiments recognized within the genre.

This leads to the second limitation; word count. Each song analyzed had a different number of recognized words per sentiment analysis. The same words that were recognized in the AFINN lexicon, were not necessarily the same words recognized in the Bing or NRC lexicons. The word count was different for every song and for every genre collectively. Urban tended to have the most words, so there were more sentiments recognized in that genre than in country and pop. The sentiment scores changed for each song because of the number of recognized words in the word count. A factor for number of words analyzed was the use of stop words.

Stop words are words that each lexicon excludes from its analysis. There were 124 stop words that were eliminated from all lexicons. These stop words included profanity, curse words, numbers, slang, one or two letters words, abbreviated words, and gender specific words. Some words included in that list may have multiple meanings or interpretations. For example, the word "ice" may be considered negative in a sentiment analysis, whereas, could be considered positive in the song context because it is slang for jewelry and diamonds. Stop words reduced the word counts for songs, which then decreased the number of words analyzed. If the study was repeated, multiple testes should be used for each of the correlations, and then numbers should be normalized to be on the same sample size scale.

The next method that included limitations was the Dialsmith Perception Analyzer. This software is expensive, and not attainable to the everyday user. It also requires an immersive three full days of training to learn how to operate the software. Although it is fairly easy to use on the participant end with just the turn of the dial, instructions still need to be explained and practiced in order to get accurate results. Usually by the third song, participants were acclimated to the dials. For the researcher, it takes a lot of practice, time, and concentration to run a test smoothly. For me, it took about 20 rounds of testing to get a smooth, accurate, and reliable stream of tests.

For the data collection and results, a major limitation was the number of participants. There were originally 55 total participants that took the music test. Two people were eliminated because one had taken the test before in a previous pilot study, and the other failed to answer the survey questions that went along with the dial test. Two participants that were signed up to take the test did not show up for their testing times because of external circumstances. Although 53 respondents is a respectable number, the analysis required a much larger number (100 or more) to calculate predictions and trends. Another factor that contributed to predictions and trends, were the number of songs tested. Only nine songs in this study were tested across two years and three genres. With a higher sample size of songs, more analysis could have been done to improve results. The number of participants needed can be determined by power analysis. A power analysis can determine how many participants should be included in a study "… to have a good chance of rejecting the null hypothesis" (Cooper, 2011, p. 32). When conducting a power analysis, a priori (a) is decided for the expected effect size, (b) what p level will be used to reject the null hypothesis, and (c) the likelihood with which the null hypothesis will be able to be rejected to see if there is truly a relationship in the population (Cooper, 2011).

Participants did fall into the target audience range, but did not reach both ends of the range. The youngest participant was 19 years-old, and the oldest participant was 30 years-old. There were not participants that were 18, and no participants that were between 31 and 34 years-old. There also was not an equal variability among genders reported by participants. Of the 53 participants, 10 of them were male and 43 of them were female. This made the reporting between genders difficult due to the one-sided representation.

7.1. Research Question One

The purpose of RQ1 was to see what sentiments were present in music lyrics from the country, pop, and urban genres in 2017 and 2018. Sentiments included emotions and the value associated with these emotions, and polarity frequency throughout these genres. The "participants" for this were all 500 songs that charted on the Billboard Year-End charts for the three genres. Three songs were eliminated from this study, making a total of 497 songs analyzed.

Along with the rise of technology, people are more, now than ever, expressing their opinions and thoughts via web using review site, blogs, social networks, and forums (Ding & Pan, 2017). A sentiment analysis can capture these emotional responses to gain insight and make informed decisions. Many open source and commercial sentiment analysis tools have been developed to help determine marketing strategies, consumer product preferences, and determine reaction to political debates (Ding & Pan, 2017). Massive amounts of opinion-rich data are widely available causing more and more businesses, organizations, and individuals to utilize the power of a sentiment analysis and applying it to their data.

To address RQ.1, a sentiment analysis was performed to describe the different sentiments that were present in contemporary lyrical music. Three tidytext lexicons were used in the sentiment analysis, Bing, AFINN, and NRC. Each lexicon identified different lexicons within the lyrics. The Bing lexicon assigned a polarity score of either positive or negative to significant words. The AFINN lexicon assigned a value ranging from -5 to

5 to each significant word. The NRC lexicon categorized significant words into eight different emotion categories, as well as categorizing them as positive or negative.

The three sentiment analyses varied amongst years and genres, and each lexicon analyzed different amounts of words. Pop had one half of the number of songs charted in both 2017 and 2018, which did make a difference in word count analysis for each lexicon.

Bing Lexicon

The mean scores from the Bing sentiment analysis were different in positive and negative score in each genre across 2017 and 2018. Songs were more negative in all three genres in 2018 than in 2017. However, there was a higher frequency of positive words in the country genre for 2018 than in 2017. The most positive genre in 2017 was urban, and the most positive genre in 2018 was pop. The most negative genre for 2017 and 2018 was urban. There most analyzed words for 2017 and 2017 was in the urban genre. The genre with the lowest number of analyzed words for 2017 and 2018 was pop.

Hypothesis was supported that there would be a significant difference between sentiments between genres.

$H_{1.1}$: The null hypothesis ($H_{O1.1}$) was rejected in favor of the alternate hypothesis (($H_{A1.1}$) because significant differences in sentiment score means existed between years. Sentiment score means (AFINN and Bing) were greater in [Year] than in [Year]. Additionally, significant differences in NRC emotion categories existed between years.

$H_{1.2}$: The null hypothesis ($H_{O1.2}$) was retained because no significant differences existed between positive sentiments in the pop and country genres. Positive sentiments were not greater in pop songs than in country songs during 2017 and 2018.

$H_{1.3}$: Hypothesis was not supported that positive sentiments in pop were greater than in urban for both 2017 and 2018.

$H_{1.4}$: Hypothesis was supported that positive sentiments in urban were greater than in country for both 2017 and 2018.

$H_{1.5}$: Hypothesis was supported that negative sentiments in country were greater than in pop for both 2017 and 2018.

$H_{1.6}$: Hypothesis was supported that negative sentiments in urban were greater than in pop for both 2017 and 2018.

$H_{1.7}$: Hypothesis was not supported that negative sentiments in country were greater than in urban for both 2017 and 2018.

AFINN Lexicon

The AFINN sentiment analysis varied between year and genre for both country and pop, but not for urban. A sum of values for the each genre per year was calculated. In the country genre, 2017 had a negative sum. Whereas, songs in the country genre during 2018 had a positive sum. In the pop genre, 2017 had a positive sum, whereas 2018 had a negative sum. In the urban genre, both 2017 and 2018 had very close negative sums.

$H_{A1.1}$: Hypothesis was supported that there was a significant difference in sentiments between years for country and pop. Hypothesis was not supported that there was a significant difference in sentiments between years for urban.

$H_{A1.2}$: Hypothesis was supported that positive sentiments in pop were greater than in country for 2017. Hypothesis was not supported that positive sentiments in pop were greater than in country for 2018.

$H_{A1.3}$: Hypothesis was supported that positive sentiments in pop were greater than in urban for both 2017 and 2018.

$H_{A1.4}$: Hypothesis was not supported that positive sentiments in urban were greater than in country for both 2017 and 2018.

$H_{A1.5}$: Hypothesis was supported that negative sentiments in country were greater than in pop for 2017. Hypothesis was not supported that negative sentiments in country were greater than in pop for 2018.

$H_{A1.6}$: Hypothesis was supported that negative sentiments in urban were greater than in pop for both 2017 and 2018.

$H_{A1.7}$: Hypothesis was not supported that negative sentiments in country were greater than in urban for both 2017 and 2018.

NRC Lexicon

The eight emotion categories in the NRC lexicon are anger, anticipation, disgust, fear, joy, trust, sadness, surprise, and two polarities positive and negative. Because this research was based off of the six basic emotions (happiness [joy], sadness, fear, disgust, surprise, and anger) this discussion will focus on those emotion categories only.

Happiness and surprise fall under positive sentiments, whereas sadness, fear, disgust, and anger fall under negative sentiments. All three genres had more negative sentiments present in both 2017 and 2018. There was not a significant difference in negative sentiment occurrences between 2017 and 2018 for country and pop, however, there was a significant difference in negative sentiment occurrences between 2017 and 2018 for urban. There was not a significant difference in positive sentiment occurrences between 2017 and 2018 for country and pop, however, there was a significant difference in positive sentiment occurrences between 2017 and 2018 for urban.

$H_{A1.1}$: Hypothesis was not supported that there was a significant difference in sentiments between years for country and pop. Hypothesis was supported that there was a significant difference in sentiments between years for urban.

$H_{A1.2}$: Hypothesis was not supported that positive sentiments in pop were greater than in country for both 2017 and 2018.

$H_{A1.3}$: Hypothesis was not supported that positive sentiments in pop were greater than in urban for both 2017 and 2018.

$H_{A1.4}$: Hypothesis was supported that positive sentiments in urban were greater than in country for 2017. Hypothesis was null that positive sentiments in urban were greater than in country for 2018. Positive sentiments for urban and country were the same for 2018.

$H_{A1.5}$: Hypothesis was supported that negative sentiments in country were greater than in pop for both 2017 and 2018.

$H_{A1.6}$: Hypothesis was supported that negative sentiments in urban were greater than in pop for both 2017 and 2018.

$H_{A1.7}$: Hypothesis was not supported that negative sentiments in country were greater than in urban for 2017. Hypothesis was null that negative sentiments in country were greater than in urban for 2018. Country and urban had the same number of negative sentiments for 2018.

7.2. Research Question Two

The purpose of RQ2 was to describe how humans respond to lyrics in music. Participants were asked to listen to a group of nine songs ranging in polarity and chart rank. Participants then rated their response to the music moment-by-moment, using a dial scale. After listening to each song, the participants then answered questions about their opinion of the song. Songs were selected based on their Bing sentiment analysis score by either being strongly positive, strongly negative, or fairly neutral.

From the sentiment analysis and previous research, words have been described to carry a degree of emotional value. Emotions can be expressed through various words. These emotions can differ in the contexts in which they are used. Formed from words, music lyrics are consumed through both written form and auditory means. They communicate messages both cognitively and emotionally (Limbaugh, 2019). A limitation with the sentiment analysis is that context is not accounted for when analyzing, assigning value, or categorizing the words.

Participants engaged in a series of dial tests where, while listening to each song, they continuously rated each song using the Dialsmith dials. Dial testing is often used in the media industry to help broadcasters identify specific moments, topics, and contexts in which the audience responded to most positively and negatively (Dialsmith, 2013).

After listening to, and dial testing each song, participants were asked their opinion on the polarity of the song, and what emotions they reported feeling while listening to the song. The six basic emotions were answer choices in the questionnaire, along with “other”. The following hypotheses will not take into account “other”, because it is not part of the six basic emotions and is based on the participants’ responses to the nine songs they listened to.

Positive to Negative

Participants were asked to rate the nine selected songs on a five-point Likert scale with 1 = Extremely positive, 2 = Somewhat positive, 3 = Neither positive not negative, 4 = Somewhat negative, and 5 = Extremely negative. The positive to negative choices were recoded to mirror the Bing lexicon outcomes used in the sentiment analysis. To maximize cell sizes, responses were further collapsed: 01 and 02 were considered “positive”, 04 and 05 were considered “negative”, and 03 was considered neutral. Based on the participants’ reported choices, pop was reported, on average, to be the most negative genre, followed by country, and then urban. Based on the participants’ reported choices, urban was the believed to be the most positive genre, followed by country, then by pop.

The NRC sentiment analysis tested eight different emotions cater goes, as well as positive and negative sentiment. Because this study focused on the six basic emotions, the outlying emotions from NRC were eliminated. Participants were able to select as many emotions they wanted in response to what they felt during each song. After selecting responses, participants were then asked to what degree they felt it [emotion]. This question was eliminated from analysis because not all participants received it, and the sample size was not large enough to test.

$H_{A1.2}$: Hypothesis was supported that positive sentiments in pop were greater than in country in the six songs selected.

$H_{A1.3}$: Hypothesis was supported that positive sentiments in pop were greater than in urban in the six songs selected.

$H_{A1.4}$: Hypothesis was not supported that positive sentiments in urban were greater than in country in the six songs selected.

$H_{A1.5}$: Hypothesis was supported that negative sentiments in country were greater than in pop in the six songs selected.

$H_{A1.6}$: Hypothesis was supported that negative sentiments in urban were greater than in pop in the six songs selected.

$H_{A1.7}$: Hypothesis was not supported that negative sentiments in country were greater than in urban in the six songs selected.

7.3. Research Question Three

The purpose of RQ.3 was to describe the relationships between sentiment and human response to lyrical music. The sentiment analysis used for this research question were the AFINN and Bing lexicons.

AFINN scores were recoded and plotted along a graph at each second that a word occurred. The mean dial scores for each song were also plotted along the same graph for each second. The first two seconds of each song were cut off because a countdown was put in place to prepare participants for the start of each song. Because MTM responses are sometimes delayed, when running the correlation on these two variables, the second from the dial scores that were associated with the occurrence of the AFINN recognized word, was combined with two seconds following the word occurrence. So, if a word like "bad" occurred at one minute and forty-five seconds (1:45) into a song, the dial score means were recoded to collect responses from one minute and fort five seconds through one minute and forty eight seconds (1:45 – 1:48).

Of the nine songs, only one had a significant correlation between the AFINN scores and dial scores. Songs that had a higher AFINN word count, tended to have closer to significant results than those that did not. This could be due to the fact that more seconds of the song that were analyzed gave a better representation of the song scores as a whole.

The Bing sentiment analysis scores were the unit of analysis that were used to pick the song selection. One song from each music genre was positive, one was negative, and one was neutral, based off of the Bing scores. After listening to each of the nine songs, participants were asked to rate the polarity of the song on a five-point Likert scale from extremely negative to extremely positive. Eight of the nine songs had a survey response of a positive polarity. Although these eight songs were rated more positive than negative or neutral, that doesn't mean the majority of the respondents chose positive, it just means that positive was more present than negative or neutral.

The mean dial scores for every song were positive (0-50 = negative; 51 – 100 = positive). Mean and standard deviation for the polarity survey response from each song were calculated then compared to the dial scores. Every song was significant at the .001 level (two-tailed). As the dial scores went up, the polarity of the survey responses went up (positive and positive). Alternatively, as the dial scores went down, so did the polarity of the survey response (negative and negative).

7.4. Research Question Four

The purpose of RQ.4 was to explore the demographic, psychographic, environmental, and behavioral factors of human response to lyrical music. Survey responses were analyzed with dial mean scores to determine what factors were and were not significant.

The three demographic variables were gender, age, and music education. There was not enough of a gender representation, so gender was not part of the analysis. Age was recoded to fit into two groups (1 = 19 – 22; 2 = 23 – 30). Music education was placed on a four-point scale with answers ranging from none at all, to a lot.

Age was correlated with mean dial score. There were two significant inverse correlations that occurred, including urban negative (*Young Dumb & Broke*) and urban positive (*LOVE.*). As the dial scores went down, age went up. Because both songs were from the urban genre, it can be said that the older group (2) did not like urban music.

Music education levels were correlated with mean dial scores. Three songs had an inverse correlation including, country positive (*Blue Tacoma*), country neutral (*What Ifs*), and pop negative (*Sorry Not Sorry*). As the dial scores went down, the music education level went up. Other than participants with a higher music education level not liking these three songs, there was no other determinants for the significance.

There were four variables that made up the psychographic factors, including if the participant played a musical instrument, if the participant thought music lyrics played a role in emotional response, if the participant thought that past experiences causes an emotional reaction while listening to music, and the participant's familiarity with the song.

Frequency and percent were reported for musical instrument, past experiences, and music lyrics. These questions were used to describe the sample. There were 38% (f = 20) of the participants that reported playing a music instrument, and 62% (f = 33) that reported not playing a musical instrument. A total of 98% (f = 52) of participants said yes to past experiences causing an emotional reaction while listening to music. Only one participant (f = 1, % = 2) that said maybe. None of the participants said no. When asked if music lyrics contribute to emotional experience, 87% (f = 46) reported yes, 13% (f = 7) said maybe, and no one said no.

A bivariate correlation was used to find similarities between these three variables and the mean dial scores. For people who played a musical instrument, the songs *I Like Me Better* and *Shape of You* had a significant relationship. People who selected that they did play music instrument rated those two songs higher than those who did not. There was also a negative inverse relationship between the mean dial score of Criminal, and those who said yes to music lyrics playing a role in emotional reaction. As the dials scores went down, those who selected yes went down.

Song familiarity affected this study. Participants answered if they had heard the song before on a four-point scale ranging from I have never heard it before to I have heard it many times. If the song was not familiar to people, they had a new experience with it during data collection. If it was something they have heard, or have heard many times, they most likely already had preconceived notions about it. Also, it may feel very repetitive to them or may have had an impactful experience with the song prior to this study. Among the nine songs, the highest familiarities were with *Shape of You* (f = 50; % = 94), *I Like Me Better* (f = 40; % = 75), and *Sorry Not Sorry* (f = 40, % = 75). These three songs were all part of the pop genre. It can be assumed from this study, that pop music was either listened to more, had a higher reach than the other genres, and was exposed to the audience more than the other genres. Genre preference also plays a role in that assumption. Pop was the second most preferred genre from the participants. The two least recognized songs by the participants were *Criminal* (f = 25; % = 47) and *Boo'd Up* (f = 21; % = 40). Subsequently, these songs made up two thirds (2/3) of the female artists in this study. This goes back to inadequate gender representation within the music industry. The genres of these songs were country and urban. The two genres had the lowest number of female artists for the three genres looked at.

A bivariate correlation was used to find similarities between song familiarity and the means of the dial scores. Seven of the nine songs had a significant correlation, with six of them being at the .05 level (two-tailed), and one of them being at the .01 level (two-tailed). It was reported that as the dial scores increased, the level of familiarity of the song increased. Those who had heard the song before rated the song positively. The two songs that did have a correlation were *Criminal* and *What Ifs*, both of which are from the country genre.

Two variables were used to describe the environmental factors that influence human responses to sentiments in lyrics. The first one was what activities the participant engages in while listening to music, and the second one was what times of day the participant listens to music. Participants could select as many options they wanted for both questions.

The most selected activities included working/studying, housework, and driving/travelling. A total of 89% ($f = 47$) of participants said that they listened to music while working/studying, 98% ($f = 52$) said they are driving/travelling, and 89% ($f = 47$) said they were doing housework.

There were 10 significant relationships among the activities and the means of the dial scores. Working/studying had an inverse relationship with *Young Dumb & Broke*. Housework had an inverse relationship with *Blue Tacoma*. Driving/travelling had a positive relationship with *LOVE.* and *Sorry Not Sorry*. Cooking had an inverse relationship with *Blue Tacoma* and *What Ifs*. Partying had a positive relationship with *Sorry Not Sorry*. Playing video game had an inverse relationship with *Blue Tacoma* but had a positive relationship with *Young Dumb & Broke*. At a concert had an inverse relationship with *Shape of You*. Other had an inverse relationship with *I Like Me Better.*

Although there were a lot of significant relationships, these correlations don't really tell us anything specific about the music genres.

Participants reported what times of day the listen to music, with hours in the day broken into five categories. Most participants said that they listen to music in the afternoons and evenings ($f = 48$; % = 91). The fewest participants said that they listen in the early mornings and overnight ($f = 29$; % = 55). A correlation between the dial scores and the times of day was used. Urban songs *Young Dumb & Broke*, and *LOVE*. were significantly listening to in the evenings. The participants who selected evenings reacted positively to this song in their dial scores. There was an inverse relationship between *Criminal* and in the evenings and overnight. Participants who selected evenings and overnights reacted negatively to this song in their dial scores. *Sorry Not Sorry* had a significant relationship with afternoons. Participants who selected afternoons reacted positively to this song in their dial scores.

One variable used to measure behavior amongst participants and their response to sentiments in lyrics was if they listened to playlists to enhance their music experience. A total of 74% ($f = 39$) said yes, whereas 26% ($f = 14$) said no. There was only one significant inverse relationship between playlists and dial mean scores, which was for *Sorry Not Sorry*.People who selected that yes, they do listen to playlists to enhance their emotional experience rated the song lower on their dials.

7.5. Research Question Five

The purpose of RQ.5 was to describe the relationship among sentiments in lyrics and year end performance on music charts. The positive and negative word ratio scores from the Bing sentiment analysis, were correlated with the year-end Billboard charts, in all three music genres (country, pop, urban0.

A bivariate correlation was used to see if the Bing sentiment analysis had a relationship with the Billboard chart rank for year-end chart performance. There were no significant relationships between these two variables. Therefore, a Bing sentiment analysis cannot predict chart performance. Researchers have not looked into relationships between chart performance and sentiments, so this correlation at least gives a baseline for future research.

7.6. Recommendations for future research

All three methods in this study have not been used all together in previous research. Methods have either been used independently, or only two of the three have been used together when researching music. Because of this, this study, or pieces of this study, can be built off or create a baseline for future research. Some recommendations for future research include collecting both a sentiment analysis, and dial scores, considering both a sentiment analysis and dial scores, refining survey questions for use in this context, using different sentiment analysis lexicons, and context of lyrics in music.

It is important to collect and consider both the sentiment analysis and dial scores when making comparisons or finding relationships among them. Preference, consumption, and polarity to songs cannot just be purely decided based on only a sentiment analysis. Even with the advancement of technology, there is still a need for human response. A sentiment analysis at least provides information about what to specifically look for in a song, which way (positive or negative) individual words sway, and a general computational analysis. However, it cannot predict actual, physical, human response. Dials will still need to be used for this.

Survey questions should be developed before and after the sentiment analysis has been done. The questions used in this study were made before the sentiment analysis was complete. Making questions after completing the sentiment analysis allows for specific inquiry about what was found in the analysis including word recognition, emotions or polarity associated with specific words, and which pieces in the song elicited certain emotions.

This study used only three sentiment lexicons from the tidytext package. There are, however, multiple other lexicons to choose from. The lexicons used in this study were very limiting when it came to stop words and lyric context. Sometimes, profanity and slang terms help to give songs a little more of a stronger message, which was not able to be conveyed with my lexicon choice. These sentiment lexicons only looked at individual words, and did not interpret sarcasm or multiple meanings, making the context of the words unknown. Consider using other lexicons as a different approach.

REFERENCES

Bhattacharjee, S., Gopal, R., Lertwachara, K., & Marsden, J. R. (October 01, 2006). Whatever happened to payola? An empirical analysis of online music sharing. *Decision Support Systems*, *42*(1), 104-120.

Billboard. (2017a). Year-end charts: Pop songs. Retrieved from https://www.billboard.com/charts/year-end/2017/pop-songs

Billboard. (2017b). Year-end charts: Hot country songs. Retrieved from https://www.billboard.com/charts/year-end/2017/hot-country-songs

Billboard. (2017c). Year-end charts: Hot R&B/hip-hop songs. Retrieved from https://www.billboard.com/charts/year-end/2017/hot-r-and-and-b-hip-hop-songs

Billboard. (2018a). Year-end charts: Pop songs. Retrieved from https://www.billboard.com/charts/year-end/2018/pop-songs

Billboard. (2018b). Year-end charts: Hot country songs. Retrieved from https://www.billboard.com/charts/year-end/2018/hot-country-songs

Billboard. (2018c). Year-end charts: Hot R&B/hip-hop songs. Retrieved from https://www.billboard.com/charts/year-end/2018/hot-r-and-and-b-hip-hop-songs

Billboard. (2019a). Billboard charts: about. Retrieved from http://billboardlicensing.licensestream.com/LicenseStream/Portal/Billboard/About/Seals.aspx

Billboard. (2019b). Frequently asked questions: about. Retrieved from http://billboardlicensing.licensestream.com/LicenseStream/Portal/Billboard/Faq.aspx

Billboard. (2019c). Billboard charts legend. Retrieved from https://www.billboard.com/biz/billboard-charts-legend

Billboard. (2019d). Charts. Retrieved from https://www.billboard.com/charts

Briggs, K., Burruss, K. L., Cottle, T. D., Mac, S., McDaid, J., & Sheeran, E. (2017). Shape of you [Recorded by Ed Sheeran]. On ÷ [WAV file]. New York City, NY: Atlantic Records. (2017).

Brooks, M., Jussifer, Peiken, S., Rexha, B., & Tranter, J. (2018). I'm a mess [Recorded by Bebe Rexha]. On *Expectations* [WAV file]. Los Angeles, CA: Warner Bros. Records (2018).

Brown, C. R., Dickerson, R. E., & Nohe, P. (2016). Blue Tacoma [Recorded by Russell Dickerson]. On Yours [WAV file]. Franklin, TN: Dent Records. (2016).

Brown, K., McGinn, M., & Schmidt, J. M. (2017). What If's [Recorded by Kane Brown feat. Lauren Alaina]. On *Kane Brown* [WAV file]. Nashville, TN: RCA Records Nashville. (2017).

Brown, T., Douglas, S., Felder, W., Lovato, D., & Simmons, W. Z. (2017). Sorry Not Sorry [Recorded by Demi Lovato]. On *Tell Me You Love Me* [WAV file]. Burbank, CA: Hollywood Records. (2017).

Bryman, A. (2015). *Social Research Methods - 5th Edition*. Oxford: OXFORD University Press.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand-McNally.

Caracciolo, A. Hall, R. B., Ivatury, A., Robinson, K. D., Taggart, A., & Wiggins, D. (2017). 1-800-273-8255 [Recorded by Logic feat. Alessia Cara & Khalid]. On *Everybody* [WAV file]. New York City, NY: Def Jam Recordings (2017).

Chopra, D. (2016). Why meditate? *The Chopra Center*. Retrieved from http://chopra.com/sites/default/files/24%20BYB-S3-WhyMeditate.pdf

Cooper, H. M. (2011). *Reporting research in psychology: How to meet journal article reporting standards*. Washington, D.C: American Psychological Association.

Cunningham, S., Caulder, S., & Grout, V. (2008). Saturday night or fever? Context aware music playlists. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.489.7774&rep=rep1&type=pdf

Delmonte, R. (2018). Audio monitor U.S.: The overall music landscape 2018. *MusizBiz*. Retrieved from https://musicbiz.org/wp-content/uploads/2018/09/AM_US_2018_V5.pdf

Dialsmith. (2013). *Essentials of moment-to-moment research* [E-reader version]. Retrieved from http://www.dialsmith.com/wp-content/uploads/2015/07/Dialsmith-Essentials-of-MtM-eBook-070813.pdf

Dialsmith. (2016). Moment-to-moment research playbook [E-reader version]. Retrieved from http://www.dialsmith.com/request-ebook-essentials-of-moment-to-moment-research/

Dialsmith (2018). Comprehensive training guide for the perception analyzer. *Training Guide - Version: 9.0.*

Dialsmith (2019). Perception Analyzer. Retrieved from http://www.dialsmith.com/dial-testingfocus-groups-products-and-services/perception-analyzer-dial-research

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). Internet, phone, mail, and mixed-mode surveys: The tailored design method (4th ed.). Hoboken, NJ, US: John Wiley & Sons Inc.

Ding, T., & Pan, S. (2016, April). How reliable is sentiment analysis? A multi-domain empirical investigation. In *International Conference on Web Information Systems and Technologies* (pp. 37-57). Springer, Cham.

Dopson, L. L., Howell, E. M., James, J. M., & McFarlane, D. I. (2017). Boo'd up [Recoded by Ella Mai]. On *Ready* [WAV file]. Los Angeles, CA: 10 Summers Records. (2017).

Dsilva, D. (2018, May 4). Learn to create your own datasets—web scraping in R. *Towards Data Science*. Retrieved from https://towardsdatascience.com/learn-to-create-your-own-datasets-web-scraping-in-r-f934a31748a5

Eerola, T. (2012). Modeling listeners' emotional response to music. *Topics in Cognitive Science*, *4*(4), 607-624. doi: 10.1111/j.1756-8765

Ell, L., Stevens, C. & Wilhelm, F. (2017). Criminal [Recorded by Lindsay Ell]. On *The Project* [WAV file]. Nashville, TN: Stoney Creek Records. (2017).

Field, A. (2016). *An adventure in statistics: The reality enigma*. Los Angeles: Sage.

Frith, S. (1987). Towards an aesthetic of popular music. In R. Leppert & S. McClary (Eds.), *Music and society: The politics of composition, performance and reception*, (pp. 133-149). Cambridge, MA: Cambridge University Press.

Frost School of Music. (2016). What is involved in putting together a big music tour? *University of Miami*. Retrieved from https://frostonline.miami.edu/articles/what-is-involved-in-putting-together-a-big-music-tour.aspx

Garcia,J. & Gustavson, A. (1997). The science of self-report. *Association For Psychological Science*. Retrieved from https://www.psychologicalscience.org/observer/the-science-of-self-report

Genius. (2019a). About Genius. Retrieved from https://genius.com/Genius-about-genius-annotated

Genius, (2019b). How Genius works. Retrieved from https://genius.com/Genius-how-genius-works-annotated

Genius. (2019c). Shape of you lyrics. Retrieved from https://genius.com/Ed-sheeran-shape-of-you-lyrics

Giatsoglou, M., Vozalis, G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications*, *69*, 214-224.

Goodwin, J. N., Chiarelli, C., & Irani, T. (2011). Is perception reality? Improving agricultural messages by discovering how consumers perceive messages. Journal of Applied Communications, *95*(3), 3.

Hill, J. (2016). Reaching mid-millennial produce shoppers at Whole Foods Market, Inc. In Texas (Unpublished master thesis). Texas A&M University, College Station, Texas.

Hu, X., Downie, J. S., & Ehmann, A. F. (2009). Lyric text mining in music mood classification. International Society for Music Information Retrieval Conference, 1-6. Retrieved from https://pdfs.semanticscholar.org/e658/ec86e033aae370ba680118a04431071cafe1.pdf

James, B. & Lane, T. (2018). On my way to you [Recorded by Cody Johnson]. On *Ain't Nothin' to It* [WAV file]. Nashville, TN: Warner Music Nashville. (2018).

Johnson, C. (September 17, 2006). Cutting through advertising clutter. *CBS News*. Retrieved from https://www.cbsnews.com/news/cutting-through-advertising-clutter/

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *5*(129), 770–814

Juslin, P. N., & Sloboda, J. A. (2001). Music and emotion: Theory and research. Oxford, Angleterre: Oxford University Press.

Juslin, P. N., Liljestrom, S., Vastfjall, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion*, *8*(5), 778-683. doi: 10.1037/a0013505

Kelly, B. (2018, April 5). Audio today 2018: how America listens. The Nielsen Company. Retrieved from https://www.nielsen.com/content/dam/corporate/us/en/reports-downloads/2018-reports/audio-today-report-apr-2018.pdf

Knobloch, S.(2003). Mood adjustment via mass communication. *Journal of Communication*, *53*(2), 233-250

Lamar, K., Kurstin, G., Pacaldo, Z., Spears, M., Tiffitc, A., & Walton, T. (2017). LOVE. [Recorded by Kendrick Lamar]. On *DAMN.* [WAV file]. Los Angeles, CA: Top Dawg Entertainment. (2017).

Lang, A. (1994). Measuring Psychological Responses to Media Messages. Hoboken: Taylor and Francis.

Leff, A. & Matosic, M. (2017). I like me better [Recorded by Lauv]. On *I Met You When I Was 18* [WAV file]. Los Angeles, CA: no label. (2017).

Li, G. (2017). *Application of sentiment analysis: Assessing the reliability and validity of the global airlines rating program*, B.S. thesis, University of Twente.

Limbaugh, E. M. (2019). *Does every life have a soundtrack?: A study on how music lyrics shape self-narrative* (Order No. 13426260). Available from ProQuest Dissertations & Theses Global. (2171103452). Retrieved from http://ezproxy.library.tamu.edu/login?url=https://search.proquest.com/docview/2171103452?accountid=7082

Lippmann, J. R., & Greenwood, D. N. (2012). A song to remember: emerging adults recall memorable music. *Journal of Adolescent Research, 27(*6), 751-774. doi: 10.1177/0743558412447853

Little, J., Riley, T., & Robinson, K. D., (2017). Young dumb & broke [Recorded by Khalid]. On *American Teen* [WAV file]. Beverly Hills, CA: RCA Records. (2017).

Local Radio Freedom Act of 2019, H.R. 20, 116a Cong. (2019).

McIntyre, H. (2017, Nov 9.) Americans are spending more time listening to music than ever before. *Forbes*. Retrieved from https://www.forbes.com/sites/hughmcintyre/2017/11/09/americans-are-spending-more-time-listening-to-music-than-ever-before/#2d3475712f7f

Media Tracks Communications. (2019). Services: Radio media tours (RMTs). Retrieved from https://mediatracks.com/services/radio-media-tours-rmts/

Micu, A. C., & Plummer, J. T. (2010). Measurable emotions: How television ads really work. *Journal of Advertising Research, 50*(2), 137-153

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence, 29*(3), 436-465.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Orrin G. Hatch–Bob Goodlatte Music Modernization Act of 2018, Pub. L. No. 115–264, §1(a), 132 Stat. 3676 (2019).

Panksepp, J. (1998). Series in affective science. Affective neuroscience: The foundations of human and animal emotions. New York, NY, US: Oxford University Press.

Poels, K. & Dewitte, S. (2006). How to capture the heart? Reviewing 20 years of emotion measurement in advertising. *Journal of Advertising Research*, *46*(1), 18-37. doi: 10.2501/S0021849906060041

Radio Advertising Bureau. (2018). Why Radio. *RAB*. Retrieved from http://www.rab.com/whyradio.cfm

Recording Industry Association of America. (2018). RIAA 2018 year-end music industry revenue report. *RIAA.* Retrieved from https://www.riaa.com/wp-content/uploads/2019/02/RIAA-2018-Year-End-Music-Industry-Revenue-Report.pdf

Rys, D. (2018). Nielsen releases in-depth statistics on live music behavior: 52 percent of Americans attend shows. *Billboard*. Retrieved from https://www.billboard.com/articles/business/8485063/nielsen-releases-in-depth-statistics-live-music-behavior-360-report

Saks, J., Compton, J. L., Hopkins, A., & El Damanhoury, K. (2016). Dialed in: Continuous response measures in televised political debates and their effect on viewers. *Journal of Broadcasting & Electronic Media, 60*(2), 231-247.

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R.J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS ONE*, *4*(10). doi: 10.1371/journal.pone.0007487

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695-729.

Seiffert-Brockman, J. & Jarolimek, S. (2017). Real-time ratings. In J. Matthes, C. S. Davis and R. F. Potter (Eds.),*The International Encyclopedia of Communication Research Methods* (pp. 1-8). Hoboken, NJ: John Wiley & Sons, Inc. doi: 10.1002/9781118901731.iecrm0206

Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. Beijing: O'Reilly.

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Change, S., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*, 3-14.

Spencer, S. (2019). Song structure: A quickstart guide for beginning songwriters. *Musician on a Mission*. Retrieved from https://www.musicianonamission.com/song-structure/

Stieler, M. & Kriegl, B. (July 2018). How do consumers experience the emotional rollercoaster? A smartphone app to measure emotions continuously. *Werbeforschung & Praxis, 64* (2), 43-53. Retrieved from https://www.researchgate.net/publication/326265288_How_do_consumers_experience_the_emotional_rollercoaster_A_smartphone_app_to_measure_emotions_continuously

Strauss, G.P. & Allen, D.N. (2008). Emotional intensity and categorisation ratings for emotional and nonemotional words. *Cognition and Emotion*, *22*(1), 114-133. doi: 10.1080/02699930701319154

The Nielsen Company. (2017). Time with tunes: how technology is driving music consumption. *Nielsen.* Retrieved from https://www.nielsen.com/us/en/insights/news/2017/time-with-tunes-how-technology-is-driving-music-consumption.html

The Nielsen Company. (2019). Music sales measurement. Retrieved from https://www.nielsen.com/us/en/solutions/measurement/music-sales-measurement.html

Tidyverse. (n.d.). Retrieved from https://www.tidyverse.org/

van Dijck, J. (2006). Record and hold: Popular music between personal and collective memory. *Critical Studies in Media Communication*, 23, 357-374. doi:10.1080/07393180601046121

Wang, A. (February 5, 2019). Just 17 percent of 2018's top 100 songs were made by women. *Rolling Ston*e. Retrieved from https://www.rollingstone.com/music/music-news/women-in-music-usc-study-789599/

Wickham, H. (2016, Jun 17). Rvest. Retrieved from https://github.com/hadley/rvest

Zillmann, D . (2000). Mood management in the context of selective exposure theory. *Communication Yearbook*, *23*, 103-124.