Anatomical Sciences Education

Using natural language processing to explore health profession student reflections about the significance of anatomy to themselves and their donors' lives

Journal:	*Anatomical Sciences Education*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Report
Date Submitted by the Author:	n/a
Complete List of Authors:	Rechowicz, Krzysztof; Old Dominion University, Virginia Modeling, Analysis, and Simulation Center Elzie, Carrie; Eastern Virginia Medical School, Pathology and Anatomy; Eastern Virginia Medical School, Pathology and Anatomy
Keywords:	Anatomy education, machine learning, natural language processing, reflections, sentiment analysis

Research Report – Special Edition AI in Anatomy Education

Using natural language processing to explore health profession student reflections about the significance of anatomy to themselves and their donors' lives.

Krzysztof J. Rechowicz[1] and Carrie A. Elzie[2]

[1] Virginia Modeling, Analysis, and Simulation Center, Old Dominion University, Suffolk, VA

[2] Department of Pathology and Anatomy, Eastern Virginia Medical School, Norfolk, Virginia

Running Title: Natural language processing anatomy reflections

Correspondence to: Dr. Carrie A. Elzie, Department of Pathology and Anatomy Eastern Virginia Medical School. 700 W. Olney Rd, Lewis Hall 3148, Norfolk VA. 23501. USA. E-mail: elzieca@evms.edu

Keywords: (3-8 words) anatomy, reflections, sentiment analysis, natural language processing, machine learning,

ABSTRACT

Reflective writing, a critical and valuable skill for health professionals, is often utilized with gross anatomy courses. This study compared health professional students' reflections on how various regions of the body contributed to their life versus the proposed importance of anatomy to their anatomical donors. The most commonly written about body regions were the hands, heart, and brain. Natural language processing, a nascent field of research that uses the application of computational techniques for the analysis and synthesis of text, was used to analyze the reflections. Binary sentiment analysis revealed the reflections had an overwhelming positive sentiment with major contributing words "love" and "loved". Predominant words such as "pain" contributed to the negative sentiments and reflected various ailments experienced by students and revealed through dissections of the donors. Lexicon-based emotion classification utilizing the National Research Council Emotion Lexicon (NRC EmoLex) was used to classify writings into eight emotion categories: anger, fear, sadness, disgust, surprise, anticipation, trust, and joy. Analysis revealed the top three emotions to be trust, joy and anticipation. Each body region envoked a unqique combination of emotions. Similarities between student self reflections and donor reflections were evident suggesting a shared view of humanization and person-centeredness toward the anatomical donor. Given the pervasiveness of reflections in anatomy, adopting a natural language processing approach to analyzing reflections could provide a rich source of new information related to students' previously undiscovered experiences and competencies.

INTRODUCTION

The importance of reflection and reflective practice is regarded by many as an essential characteristic for professional development (Schön, 1983). Reflection allows for integrating concepts or a combination of skills, knowledge, attitudes, and values with the learners' cognitive framework (Kolb, 1994; Westberg & Jason, 1994; Smith & Irby, 1997). This integration usually requires an understanding or exploration of one's personal beliefs and experiences to be interpreted and expanded to form new knowledge in an active, experiential manner. Meaning can then be constructed within a community of professional discourse, encouraging learners to achieve and maintain critical control over the more intuitive aspects of their experience (Mann et al., 2009). It is considered an essential aspect of lifelong self-learning and core to professional competency (Wald & Reis, 2010). Because of this, reflective activities are becoming integral to curricula at all levels of health professional and medical education with positive impacts (Mann et al., 2009). Reflection by undergraduate medical students has been shown to increase self-reported measures of self-awareness, professional thinking skills and the skills required for intimate examinations (Sandars, 2009). Additionally, reflection has been shown to have a positive effect on cross-cultural understanding (Lie et al., 2010), diagnostic accuracy (Mamede et al., 2008), empathy (DasGupta & Charon, 2004), feedback integration (Sargeant et al., 2009), and well-being (Rabow & McPhee, 2001).

As one of the first courses in medical and health professional curricula, gross anatomy provides a rich setting for the practice of reflection to promote professionalism, integrate clinical concepts, and process emotions around death and dying (Lachman & Pawlina, 2004). Examples of reflective activities in anatomy have included anatomical donor reports (Wisner et al., 2004), attitudinal questionnaires (Crow et al., 2012), creative projects (Shapiro et al., 2009), journals

(Lazarus et al., 2017), personal or fictional narratives (Coulehan et al., 1995), and problem-oriented dissection (Chan, 2015). Students' reflective writings related to anatomical donors have revealed complex and difficult emotions, as well as an appreciation for the dissection experience (Abrams et al., 2021; Wu et al., 2021).

While it is widely acknowledged that reflective practices are beneficial, assessing reflection can be challenging, often requiring tremendous resources and time (Hatton & Smith, 1995; Wong et al., 1995; Kember et al., 1999; Williams et al., 2000). Some researchers have employed a quantitative system for evaluation such as questionnaires (Wittich et al., 2013), scales (Aukes et al., 2007) and rubrics (Wald et al., 2012). While these are efficient to grade, they provide a limited, structured, and often shallow reflective analysis. Assessment of narratives has also been done by stating judgments about the learners' abilities or engagement with the exercise (Learman et al., 2008). Numerous models and frameworks with varying complexity have been developed to analyze reflections (Préfontaine et al., 2021). However, many of these mechanisms require interrater and internal consistency reliability to develop scores, often lack validity, and are demanding in terms of resources (Pee et al., 2002; McMullan, 2006; Préfontaine et al., 2021). Even when thematic analysis is utilized, there can be a tension between the reliability of coding schemes and their ability to discriminate between learners (Wong et al., 1995).

Natural language processing (NLP) is a promising and nascent way to potentially overcome some of the challenges with analysis of student reflections, especially for large sets of open-ended text and limited time. NLP enables computers to understand natural language as humans do. Whether the language is spoken or written, NLP uses artificial intelligence to take real-world input, process it, and make sense of it in a way a computer can understand (Kumar, 2011). While this does not replace the richness of information that can be extracted using qualitative analysis, the use of NLP to analyze text data can be a valuable means of discovering underlying sentiments and themes. The most widely used technique in NLP is sentiment analysis (SA). SA is an automated process in which natural language processing, text analysis, computational linguistics, and biometrics are used to systematically identify, extract, quantify, and study affective states and subjective information (Liu, 2012). Also known as opinion mining, it is the automated task of determining what feelings a participant expresses in text, typically framed as the binary distinction of positive and negative sentiments (Zhang & Liu, 2017). SA can be applied to a document, sentence, or aspect (levels) and produce a score, rating or polarity (positive, negative or neutral). While SA aims to detect positive, neutral, or negative feelings from text, emotion detection (ED) aims to detect and recognize types of feelings through the expression of texts, such as anger, disgust, fear, happiness, sadness, and surprise (Medhat et al., 2014). ED is often performed using two main techniques: machine learning, which extracts features using a criterion or a combination of criteria, and a lexicon-based approach involving dictionaries with mapped words to emotions (Balakrishnan et al., 2019). Thus, leveraging SA and emotional recognition, underlying attitudes associated with open-ended response data can be determined, or feelings of groups of people can be assessed without directly asking for purposes of formative or summative assessments. Applying these principles to students' reflections can produce a rich data source. For example, Lin et al. (2016) used Linguistic Inquiry and Word Count to investigate overall language patterns and gender differences in reflective writings of medical students regarding pediatric patients and the psychosocial challenges faced by the patients and their family members. They found that students predominantly used cognitive words in their reflections and that female students used more words related to positive emotions and sadness than male students. Another study used a process-oriented text mining approach to better understand meanings of learner experiences within a community placement by connecting key concepts in extended student reflective essays. They found that impressions of the practicing area environment were strongest in students, and these impressions were influenced by the hospital workplace, treatment provision, and training (Lebowitz et al., 2020).

Given that students' sentiments in the anatomy course can have a lasting impact on the future of their medical practice and professionalism (Shalev & Nathan, 1985; Gustavson, 1988; Arráez-Aybar et al., 2008), it is essential to capture and analyze them early in training. Therefore, we used NLP to analyze students' reflections on the importance of their anatomy to their life and the proposed importance of anatomy to their anatomical donors. In particular, writings were analyzed for sentiment and emotions were extracted using a lexicon-based approach.

METHODS AND MATERIALS

The methodology employed was based on a commonly used NLP pipeline (**Error! Reference source not found.**) which included formatting a collection of documents, pre-processing, and analysis. The process outputs visualizations, datasets and models.

Reflections

Health professional students (physician assistant (n=82), pathologist assistant (n=25) and surgical assistant (n=25)) enrolled in a semester-long gross anatomy course with dissection were instructed to provide reflections about themselves and their donors. At the beginning of the course, students were asked to reflect on what exactly their anatomy has meant to them thus far in their life. They were asked to choose five out of 17 provided body regions and write a short paragraph about how/why it is important to them. The body regions included the arms, back, brain, ears, eyes, face, feet, gastrointestinal system, gluteals, hands, heart, knees, lungs, mouth, nose, pelvis, and skin. There was also an option of "other" if they wanted to write about something not listed. Students were given the same writing prompt at the end of the semester and completion of a whole-body dissection but asked to speculate how anatomy had contributed to their anatomical donors' lives using the same body regions. Students were also asked to provide their donor table number, as groups of 10-12 students shared an anatomical donor. There were a total of 12 donors dissected, 4 males and 8 females. Course participation points were awarded for the completion of reflections indicated by an attestation within the learning management system. However, reflections were gathered in a separate link using Google Forms to provide anonymity. A total of 1365 anonymous reflections (677 about a donor, 688 about self) were collected. This study was deemed exempt by the institutional review board.

Data Pre-Processing

Reflections were downloaded into R (Core Team R version 4.1.2 with supplementing packages), a free, open source, programming environment for statistical computing and graphics. Data were reviewed and corrected for typographical and grammatical errors. Data were formatted to ensure it was correctly encoded in UCS Transformation Format 8, a Unicode standard best suitable for representing text for mining. The pre-processing transformed the running text into a structured format that allowed computations. In addition to the narratives, data were categorized as about self or the donor and the body part referred to in the reflection.

Additionally, all reflections were broken into sentences. Tokenization, lemmatization and stop words removal were all performed. Tokenization divides a document into characters grouped by a functional semantic unit. A token could be a word or a set of words, called n-grams. Lemmatization is a subset of stemming that reduces words to their roots (e.g., "am" becomes "be," and "better" becomes "good"). Subsequently, stop words were removed that did not carry much discriminative content using three lexicons (SMART, onix, snowball) containing 1149 commonly recognized stop words (Aggarwal 2018) To illustrate, we expect that word "and" would appear in all reflections with a similar frequency and, therefore, would not allow us to infer about reflection's characteristics or content. Lastly, a custom dictionary of words had to be generated to remove the body regions' names (hands, arms, etc.) since they frequently appeared (appendix).

Data analysis

All analyses were conducted using R version 4.1.2 with supplementing packages (Feuerriegel and Proellochs 2021; Rinker 2021a; Ripley 2001). Word counts were initially calculated for each reflection without taking into account stop words. Then, the average reflection length and standard deviation were calculated for all reflections. Subsequently, median word counts of donor-related and self-reflections were compared. A Wilcoxon test was performed to establish if the difference between the medians was statistically significant at a p-value < .05. Median word counts were also compared for all body regions. The relative frequency of what body region the students chose to reflect on was calculated. Additionally, a comparison was made between of regions chosen in self and donor reflections. The top ten most common words in all reflections were identified. Like the previous step, ten words with a high occurrence rate were broken down by the reflection type, i.e., self vs. donor. Stop words and words included in the custom dictionary were not counted. Word frequencies across reflections about self and donors were compared. Pearson's correlations between words occurrences in distinct sets of reflections were calculated.

Next, sentiment analysis was performed utilizing an approach that considers valence shifters (VSs) when the polarity is calculated (Rinker 2021b). VSs, such as negators (e.g., not, no, never), amplifiers (e.g., really, very, precisely), de-amplifiers (e.g., hardly, maybe, probably), and adversative conjunctions (e.g., but, still, yet) can affect words' sentiment (Polanyi & Zaenen 2006). This method was implemented in R's sentimenr package (Rinker 2021a). Each reflection was broken down into sentences and subsequently into words in this method. Words, excluding stop words, were tagged according to their lexical polarity. The lexicon used was the Jockers and Rinker (Rinker 2019) polarity lookup table. The dictionary used was an augmented combination of the lexicon developed by Jokers (Jokers 2017) and Riker's augmentation of the lexicon developed by Hu and Liu (Hu and Liu 2004). The dictionary contained 11,710 words identified as polarized positively or negatively. Around each polarized word, a cluster including four preceding and two following words was created. The algorithm (Rinker 2021b) searched that neighborhood for VSs, and, if found, the sentiment value was recalculated. Punctuation (comma, colon, semicolon) was used to delimit the effect of the VSs on the word. Cluster's value resulted from applying weights based on the VS type and their number in the cluster. For example, an amplifier increased the base polarity score by a number greater than 1, whereas a de-amplifier by a number greater than 0 and smaller than 1. Negation was determined by raising -1 to the power of the number of negators plus 2. This resulted from a belief that an even number of negatives equal a positive, whereas an odd number gives a negative. The adversative conjunction up-weighted or down-weighted the context cluster depending on their location, before or after the polarized word. Similarly to negators, it was essential to consider the number of preceding and subsequent conjunctions. To obtain an *unbounded polarity score* for each sentence in a reflection, the weighted context clusters were summed and divided by the square root of the word count. In the end, a reflection's sentiment was calculated as a mean of sentences' polarity scores. The top ten words contributing to positive and negative sentiment were identified for all reflections and separately for those about self and a donor. Again, the lexicon developed by Rinker was utilized in this step (Rinker 2019).

Lastly, lexicon-based emotion classification was performed utilizing the National Research Council (NRC) Emotion Lexicon (EmoLex) (Mohammad and Turney 2013). The NRC EmoLex lists 14,182 English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by crowdsourcing. The results of the tokenization process were used for this analysis, i.e., a set of tokens corresponding to all words used in reflections without the stop words. Then, an inner join operation was performed on the list of words and the EmoLex dictionary (Mohammed & Turney, 2013) to assign an emotion to each word in the reflections, unless the particular word was not part of the lexicon. Those words were omitted. The resulting list of words with assigned emotions was used to characterize what feelings were present when the student reflected on a particular body region about themselves or their donor.

RESULTS

Word count

The average word count for all reflections was 34.6 [SD=18.8] words (excluding stop words). The shortest reflection was four words, and the longest was 169 words long. On average, reflections about the donor averaged 29.4 [SD=14.9] words, whereas self-reflections averaged 39.6 [SD=20.7] words (**Error! Reference source not found.**). Since the word count distributions for both types of reflections were not normally distributed according to the Shapiro-Wilk test ($p < 0.05$), a Wilcoxon test was performed to identify a significant difference between the word count of reflections. The p-value of the test ($p < 0.05$) implied that the median word count for the donor-related reflections was significantly different from the median word count for the self-reflections. The difference between the word count is also visible when the data was broken down by the type of reflection and the body regions (**Error! Reference source not found.**). Reflections about donors' noses were the least elaborated. In contrast, reflections about students' knees were the most elaborated based on median word count.

Body regions

Every option for a body region was reflected on by students (Figure 4). Overall, students choose to write most about the hands, heart and brain. When reflecting on self, the students chose hands most of the time (12%), eyes (11%) and brain (10%). When reflecting on the donor**Error! Reference source not found.**, the students chose to write most frequently about the heart (13%), hands (12%), and brain (11%). In the case of donor and self-reflections, gluteals accounted for 1% and 1.3% reflections, respectively. However, the least selected body region was the donor's nose (0.9%). The gastrointestinal system and pelvis were chosen more often when the reflections were about the donor than the self. The opposite held valid, for the back. Other areas (0.02%, n=14) chosen by students to write about included their appendix, breast, hair, kidneys, larynx, legs and nails (data not shown). Only two students choose "other" for their donors (larynx and kidneys).

Word Frequency Comparison

Word frequencies were calculated across reflections after stop words were removed. The most frequently written words in all reflections were "life," followed by "body" and "allowed" (**Error! Reference source not found.**). The order of those words changed when the reflections were separated into donor-related and self-reflections. Words "body" and "life" were the most common when the students wrote about themselves (**Error! Reference source not found.**). The word "allowed" was not present in the list of the ten most common words in those reflections. When the donor-related reflections were considered, "life" and "allowed" were the most occurring words, followed by the word "loved" (Figure 7). **Error! Reference source not found.** illustrates the comparison between frequencies of words appearing in reflections about self and donors.

Words closer to the diagonal line appear in both types of reflections with similar frequency. Terms further from the line are found more in one set of texts than another. Words that appear in only one type of reflections were omitted. The data shows a more extensive set of words that often appear in reflections about self than about donors. Pearson's correlation coefficient was calculated between word frequency for reflections about self and donor. The coefficient of 0.76 indicated a strong positive correlation between the two datasets.

Sentiment Analysis

Using a binary sentiment analysis, most reflections were positive for self and donor types (Figure 9). For three levels of polarity (negative, neutral, positive), where "neutral" polarity was assigned to sentiment values between -0.2 and 0.2, most reflections were neutral followed by positive (Figure 10). Words "pain" and "issues" were the most often occurring words contributing to negative sentiment for the donor-related reflections, whereas words "love" and "loved" to positive sentiment (Figure 11). In general, there were more occurrences of words contributing to the positive than negative sentiment, which is explained by the overall positive sentiment of the reflections. When the words contributing to positive and negative sentiment were broken down by the reflection type, i.e., about self or a donor, "pain" was still the most often occurring word with negative polarity (Figure 12). At the same time, "cancer" moved down to the eighth position for reflections about self. On the positive side, changes to the order of words were smaller. "Love" and "ability" were the top two words with positive polarity. When the words contributing to the positive and negative sentiment of reflections about donors were considered (Figure 13), "pain" and "cancer" were the top two words. At the same time, "hard" was the third most often occurring word with negative polarity. "Loved" and "care," followed by "helped," were high-occurring positive words. Differences in the top ten terms, negatively and positively influencing sentiment, likely reflect differences in the demographics of medical students and donors and their medical history.

Emotions

To extract emotions from reflections, a lexicon-based approach was utilized. An emotion lexicon is a specific type of linguistic resource that maps the emotive or affective vocabulary to a fixed set of emotion labels. Each entry in the dictionary associates a word with zero or more emotion labels. The EmoLex lexicon (Mohammed & Turney, 2013) applied here was based on Plutchik's (1980) classification of emotions into eight categories: anger, fear, sadness, disgust, surprise, anticipation, trust, and joy. When emotions were analyzed, the levels of emotional sentiment per emotion were similar when reflections about the donor and self were compared. The most considerable difference was for anger 8.2% (self) vs. 5.9% (donor) (Figure 14). Regardless of the reflection type, the top three emotions were trust, joy and anticipation. Disgust, anger and surprise were the bottom three emotions. **Error! Reference source not found.Error! Reference source not found.**5 shows the distribution of emotions for all reflections broken down by body region. In general, trust, joy, and anticipation are the dominating emotions for almost all body regions. Trust was present for all body regions, with lungs, nose and pelvis being the lowest. Surprise was relatively low for all body regions except the mouth. Sadness varied depending on the body part, with the nose having none. Joy was high for the arms, mouth, and ears and consistent for all body regions except reflections about the back, knees, and lungs. Fear was higher for several body regions, such as the lungs, knees, pelvis, and back. It was low for the arms, ears, hands and nose. Disgust was relatively low for all body regions except the nose, which was likely associated with words related to smells. The high proportions of disgust decreased the ratios of other emotions. Anticipation was relatively high for almost all body regions, with the arms being the highest and the nose the lowest. Anger was the highest for the nose, which is already skewed towards negative emotions. When the reflections are broken down by the type (self vs. donor) (Figure 16), the emotion distributions follow similar patterns. In student self reflections, the nose had a higher proportion of “sadness” and “fear” when compared to the same donor’s body region.

DISCUSSION

Within medicine, the explicit awareness of one’s own experiences deepens the capacity to respond empathically to patients and is the crux of relationship-centered care. (Charon, 2001; Beach & Inui, 2006). As such, exploring one’s narrative through reflection and self-awareness allows better listening to and accepting of another’s narrative (Charon, 2001). Since the cadaver-student relationship is thought to mirror the patient-physician relationship (Gustavson, 1988), our approach to gross anatomy has been intentionally humanistic (Hildebrandt, 2016). As such, the goals of this exercise were to garner self-awareness and to shift the thinking from “specimen-minded” to “person-minded”, as this has been shown to be associated with empathy, emotional engagement, and a better performance within the course (Goss et al., 2019). While the use of reflections in anatomy to this end is not new, this study used reflections in a novel way to explore the contribution of various anatomical body regions to one’s life. Given a large amount of text submitted, we employed NLP, a nascent field within medical education, to extract sentiments and emotions.

The mobilization of personal experience in academic writing is essential for professional identity formation (Das Gupta a& Charon, 2004). As such, this study asked students to reflect on how their anatomy has contributed to their own lives and to speculate how anatomy mattered to their anatomical donor’s life. This illuminated specific body areas that resonated most with students, such as the heart, hands and brain, which were consistent between self and donor reflections. This is not surprising concerning the donor narratives, as excerpts from other published anatomy reflections have commonly mentioned these body regions (Coulehan et al., 1995; Goss et al., 2019; Shiozawa et al., 2020; Wu et al., 2021). Interestingly, many students wrote about the love/hate relationship with their skin and its importance to beauty and identity, which was much less common in the donor reflections. Other differences in regions chosen to write about may have been influenced by the fact that students were provided with their donor's occupation at the introduction of the course. The use and hindrance of their donor's anatomy on donors' careers were commonly mentioned (data not shown).

Survey and thematic analyses of sentiments related to anatomy courses have revealed both positive and negative responses. While a positive experience for the majority of students (Penney, 1985; Dinsmore et al., 2001; McGarvey et al., 2001; O'Carroll et al., 2002; Arráez-Aybar et al., 2008), gross anatomy can also be distressing for some (Finkelstein & Mathers, 1990; Horne et al., 1990; Evans & Fitzgibbon, 1992; Druce & Johnson, 1994; Dickinson et al., 1997; Quince et al., 2011). However, this previous data was collected with respect mostly to the overall course experience or dissection process. Our data are the first (to our knowledge) to report various sentiments conjured by distinct body regions and their impact on one's life (self and donor). Most body regions elicited a neutral response related more to their function and utility. However, when using a dual scale, most self and donor reflections sentiments were positive, with subtle differences in word contributions. While both expressed words of love, self-reflections focused on the "ability" of one's anatomy as a positive contributor to their life. This was reversed when looking at the contribution of negative words which were more disease-focused (words such as "cancer" and "issues") in donor reflections. This was not unexpected given the advanced age of our donors and disease accumulated throughout a lifetime compared to students' relative youth and healthiness. Student self-reflections with negative sentiments were mostly related to previous personal "pain" and "issues." These personal illness stories are often central to students' motivations for becoming healthcare professionals. They may critically inform the nature of students' professional caregiving (DasGupta & Charon, 2004). Additionally, a detailed analysis of the types of language used in reflections can indicate the reflectors' preferred mode of processing their experience and constructing their narratives (Lin et al., 2016). Thus, further analysis of what factors may influence the language style used by students in their anatomy reflections is appealing. It could inform students' psychological well-being and ability to empathize with patients by enabling them to access and accept their feelings.

Humanistic anatomy curricula, such as those practiced in this course, can engage students by validating the full range of emotional experiences of dissection and acknowledging the protective role of emotional detachment in medical practice (Hildebrandt, 2016). Further, understanding of the emotions of students provides insights into their cognitive processes (Keltner et al., 2014; LeBlanc et al., 2015), learning (Pekrun et al., 2002), perceived self-efficacy (Bandura, 1997), and professional identify formation (Helmich et al., 2014; Dornan et al., 2015). As such, numerous studies have used a qualitative approach to describing students' complex emotional and moral experiences in gross anatomy (Finkelstein & Mathers, 1990; Hafferty, 1991; Coulehan et al., 1995; Sinclair, 1997; Lempp, 2005). While the reflections reported here were not directly related to the anatomical dissection, they still prompted a similar range of emotional responses experienced in the class as a whole. This study utilized a lexicon to analyze word associations in student reflections with emotions about specific body regions. While very little has been reported in this area, particular patterns between emotions and corresponding bodily organ systems have been studied with respect to East Asian Medicine and surprisingly coordinate quite well with the data presented here (Lee et al., 2017). Overall the most frequently extracted emotions were trust, joy and anticipation. This aligns with the positive nature of the sentiment analysis reported here and appreciation for one's body which is a common theme in anatomy reflections (Abrams et al., 2021). The general similarity of the words and emotions used in self writings compared to donor writings suggests a shared view of humanization and person-centeredness toward the anatomical donor. Because this was not a direct longitudinal study (donor vs. donor over time), we cannot say viewing the donor as a person over time did not decline over the course length as previously reported (Arráez-Aybar et al., 2008). However, the data still support a high level of personhood toward the donors after sixteen weeks of full dissection. Further, insights gained here provide guidance to students' emotional responses to various regions of the body, which may allow anatomists to respond in an appropriate manner to create a positive impact on the learning process as each region is discussed. Assessment of emotional responses over time may also provide a holistic picture of the students' emotional trajectory throughout the course and beyond.

As the use of reflections in medical and health professional education rises, the challenge will remain how to analyze large sets of text and provide valuable feedback to learners. The true advantage of using NLP lies in an evaluator's ability to use a quantitative approach in analyzing text. It is especially valuable where the breadth of a dataset or constraints associated with time and available data scientists make it very difficult for practitioners to perform a comprehensive analysis. Since the method is purely quantitative, comparing sentiments across and between groups and longitudinally can be accomplished easily. The quantitative extraction of emotions may be helpful considering evidence exists that self-reports about feelings may be misleading, as students may not be aware of the depth of their own emotions (Grochowski et al., 2014). This type

386 of objective measure of emotions from writings may provide educators with better insight for
387 providing student feedback about self-awareness.

388 Building larger datasets related to student reflection could make deep learning (DL) more
389 accessible to educators. Over the last decade, DL has become a powerful method for creating SA
390 models (Agarwal et al., 2020; Zhang et al., 2018). DL builds on multi-layer artificial neural
391 networks (ANNs) trained to perform the sentiment classification task. In short, DL mimics how
392 the human brain works. Given enough examples, an ANN can identify distinct features in the text
393 and use them to differentiate between various sentiments carried by, in our case, reflections. Once
394 the model is trained, it can infer the sentiment of reflections that were not presented during the
395 training process with a certain level of accuracy. Agrawal et al. (Pathak et al., 2020) referenced
396 many studies applying DL to build sentiment models using various data sets, such as Amazon
397 reviews, movie reviews or social media posts. When applied to education, DL can allow for more
398 personalized learning approaches and enable educators to tailor learning pathways to individual
399 students. DL environments can also intelligently analyze data across all personalized training
400 instances to recommend improvements and highlight inefficiencies that would not be possible
401 otherwise (Muniasamy & Alasiry, 2020).

402 ***Limitations***

403 This study is not without limitations. To encourage personal and honest reflections, no
404 demographic data on the students was collected; however, this may have limited our results. While
405 previous work suggests that women and men have similar emotional responses to dissection
406 (Snelling et al., 2003), differences have been reported with respect to the need for detached concern
407 (Dickinson et al., 1997) and person-centeredness (Hafferty, 1991; Goss et al., 2019). Further,
408 previous experience with dissection, which was not recorded, may also have influenced students'

409 emotions especially toward their donor. Future studies will gather more information about
410 students to futher classify the data.

411 The use of NLP for analysis also has some inherent restrictions. Although the selected
412 approach overcomes some of the challenges associated with this method (Rinker 2021b), this study
413 had limitations inherent to the SA approach. The collected reflections were relatively short, 35
414 words on average, excluding stop words. Although SA is often applied to social media messages,
415 such as tweets that are 280 characters long (Agarwal and others 2011), a few polarized words can
416 decide the sentiment of the whole reflection, especially if the remaining sentences have neutral
417 polarity. Requiring a minimum reflection length may help minimize this possibility.
418 Additionally, the neighborhood where VSs were searched for was four words before and two after
419 a polarized word. Sensitivity analysis could generate insight into the most optimal text cluster size
420 for students' reflections.

421 Another constraint is the use of general-purpose lexicons which were created by
422 crowdsourcing polarity assignment. As with any NLP, that selection significantly impacted the
423 results. It has to be mentioned that there are also several biomedical domain-related lexicons
424 (Satapathy et al., 2017). Medical WordNet (MWN) (Smith & Fellbaum, 2004) and WordNet of
425 Medical Events (WME) (Mondal et al., 2015; Mondal et al., 2016) are two lexicons based on
426 WordNet, which is a domain-free extensive dictionary developed at Princeton University
427 (Fellbaum, 2010). MWN consists of medically relevant terms used by and intelligible to non-
428 expert subjects and supplemented by natural-language sentences that are designed to provide
429 medically validated contexts. MWN primarily focuses on relationships between sets of synonyms,
430 does not provide polarity, and is not publically available yet. WME, in its 2.0 version, concentrates
431 on conveying information, such as event definition and polarity, to medical experts and laypeople.

Its main objective is to support expert systems in extracting medically relevant insights leading to identifying diseases and their relationships to various populations (Mondal et al., 2018). However, neither MWN nor WME have been constructed to reflect emotions while becoming a medical professional and being exposed to humanistic side of the profession. The existence of a such lexicon is a prerequisite to furthering the use of stories and reflections in medical education.

Analaogically, it also applies to the lexicon-based emotion classification. In the EmoLex, used for this analysis, words were assigned one out of eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust). Recent studies showed that a different model of emotions, such as anxiety, apprehension, confusion, fear, uncertainty, upset, calmness, curiosity, enthusiasm, excitement, fascination and interest (Chiou and others 2021)., which could also be applicable in a gross anatomy course setting. Since the students had not reported sentiment or emotions accompanying the reflection process, it is also challenging to verify and validate the results of this study. A mixed research design, combining qualitative and quantitative methods, might provide further insights into the current results.

CONCLUSIONS

The emotionally charged experience of anatomical dissection provides a rich opportunity for students to thoughtfully reflect on their own experiences and those of their silent teachers. The automated analysis of such writings is a highly valuable tool for student feedback, educators, and researchers, especially considering the importance of reflective writing to foster one's professional development. While still a maturing tactic, assessing reflections through NLP is a promising method for uncovering themes, the connectedness of student responses, and determining what areas warrant future investigations. In this study, utilization of NLP on student reflections successfully presented information in an easy-to-understand manner about the sentiment and emotions experienced while writing about anatomical contributions to self and donors' lives. Given the pervasiveness of reflections in anatomy, adopting a NLP approach to analysis could provide a rich source of new information related to students' previously undiscovered experiences and competencies. Not without impact will be the concurrent development of machine learning approaches to SA and domain-specific lexicons, which will increase the accuracy of the results.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the anatomical donors, donor families and donor programs for their pivotal contributions to anatomical education. The authors have no conflicts of interest or competing interests.

NOTES ON CONTRIBUTORS

KRZYSZTOF RECHOWICZ, Ph.D. is a research assistant professor in the Virginia Modeling, Analysis, and Simulation Center of Old Dominion University in Suffolk, Virginia. His research interest is in applying machine learning, virtual reality, and human-computer interfaces to create inclusive and accessible environments for people of all abilities.

CARRIE ELZIE, Ph.D., M.A. is an associate professor at Eastern Virginia Medical School in Norfolk, Virginia. Her research interests include humanistic approaches such as narrative medicine and virtual embodiment to anatomical education, as well as, teaching innovations such as gamification, art and adaptive learning.

REFERENCES

Abrams MP, Eckert T, Topping D, Daly KD. 2021. Reflective Writing on the Cadaveric Dissection Experience: An Effective Tool to Assess the Impact of Dissection on Learning of Anatomy, Humanism, Empathy, Well-Being, and Professional Identity Formation in Medical Students. Anat Sci Educ 5:658-65.

Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau RJ. 2011. Sentiment analysis of twitter data. In: Proceedings of the workshop on Language in Social Mmedia, p 30-38.

Agarwal B, Nayak R, Mittal N, Patnaik S. 2020. Deep learning-based approaches for sentiment analysis. Singapore: Springer.

Aggarwal CC. 2018. Machine learning for text. Cham: Springer.

Arráez-Aybar LA, Castaño-Collado G, Casado-Morales MI. 2008. Dissection as a modulator of emotional attitudes and reactions of future health professionals. Med Educ 42:563–571.

Aukes LC, Geertsma J, Cohen-Schotanus J, Zwierstra RP, Slaets JP. 2007. The development of a scale to measure personal reflection in medical practice and education. Med Teach 29(2-3):177-82.

Balakrishnan V, Martin MC, Kaur W, Javed A. 2019. A comparative analysis of detection mechanisms for emotion detection. In: Journal of Physics: Conference Series 1339(1), p. 012016. IOP Publishing.

Bandura, A, 1997. Self-efficacy: The exercise of control. W H Freeman & Co.

Beach MC, Inui T. 2006. Relationship-centered care. J Gen Intern Med 1:3-8.

Chan LK. 2015. Promoting reflection in the laboratory. In: Teaching Anatomy. Springer, Cham. p 203-211.

Charon R. 2001. Narrative medicine: a model for empathy, reflection, profession and trust. JAMA 286:1897–902.

Chiou R-J, Tsai P-F, Han D-Y. 2021. Impacts of a gross anatomy laboratory course on medical students' emotional reactions in Taiwan: the role of high-level emotions. BMC Med Educ 1:1-13.

Coulehan JL, Williams PC, Landis D, Naser C. 1995. The first patient: reflections and stories about the anatomy cadaver. Teach Learn Med 1:61-6.

Crow SM, O'Donoghue D, Vannatta JB, Thompson BM. 2012. Meeting the family: promoting humanism in gross anatomy. Teach Learn Med. 1:49-54.

DasGupta S, Charon R. 2004. Personal illness narratives: using reflective writing to teach empathy. Acad Med 4:351-6.

Dickinson GE, Lancaster CJ, Winfield IC, Reece EF, Colthorpe CA. 1997. Detached concern and death anxiety of first-year medical students: Before and after the gross anatomy course. Clin Anat 10:201–207.

Dinsmore CE, Daugherty S, Zeitz HJ. 2001. Student responses to the gross anatomy laboratory in a medical curriculum. Clin Anat 14:231–236.

Dornan T, Pearson E, Carson P, Helmich E, Bundy C. 2015. Emotions and identity in the figured world of becoming a doctor. Med Educ 2:174-85.

Druce M, Johnson MH. 1994. Human dissection and attitudes of preclinical students to death and bereavement. Clin Anat 7:42–49.

Evans EJ, Fitzgibbon GH. 1992. The dissecting room: Reactions of first year medical students. Clin Anat 5:311–320.

Fellbaum C. 2010. WordNet. Theory and applications of ontology: computer applications. Springer. p. 231-243.

Finkelstein P, Mathers L. 1990. Post-traumatic stress among medical students in the anatomy dissection laboratory. Clin Anat 3:219–226

Feuerriegel S, Proellochs N. Sentiment Analysis [Internet]. Available from: https://cran.r-project.org/web/packages/SentimentAnalysis/SentimentAnalysis.pdf

Goss AL, Viswanathan VB, DeLisser HM. 2019. Not just a specimen: a qualitative study of emotion, morality, and professionalism in one medical school gross anatomy laboratory. Anat Sci Educ 4:349-59.

Grochowski CO, Cartmill M, Reiter J, Spaulding J, Haviland J, Valea F, Thibodeau PL, McCorison S, Halperin EC. 2014. Anxiety in first year medical students taking gross anatomy. Clin Anat 27:835–838.

Gustavson, N. 1988. The effect of human dissection on first-year students and implications for the doctor-patient relationship. Acad Med 1:62–4.

Hafferty FW. 1991. Into the Valley: Death and the Socialization of Medical Students. 1st Ed. New Haven, CT: Yale University Press. 234 p.

Hatton, N., and D. Smith. 1995. Reflection in teacher education: Towards definition and implementation. Teach Teach Educ1:33-49.

Helmich E, Bolhuis S, Laan R, Dornan T, Koopmans R. 2014. Medical students' emotional development in early clinical experience: a model. Adv Health Sci Educ 3:347-59.

Hildebrandt S. 2016. Thoughts on practical core elements of an ethical anatomical education. Clin Anat 27:839–843.

Horne DJL, Tiller JW, Eizenberg N, Tashevska M, Biddle N. 1990. Reactions of first-year medical students to their initial encounter with a cadaver in the dissecting room. Acad Med 65:645–646.

Hu M, Liu B. 2004. Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. p. 168-177.

Jokers ML. Syuzhet: Extract sentiment and plot arcs from Text [Internet]. Available from: https://rdrr.io/cran/syuzhet/

Keltner D, Oatley K, Jenkins JM. 2014. Understanding emotions. Hoboken, NJ: Wiley.

Kember, D, Jones A, Loke A, Mckay J, Sinclair K, Tse H, Webb C, Wong F, Wong M, Yeung E. 1999. Determining the level of reflective thinking from students ' written journals using a coding scheme based on the work of Mezirow. Int J Lifelong Educ 1:18–30.

Kolb DA. 1984. Experiential Learning. Englewood Cliffs, NJ: Prentice Hall.

Kumar E. 2011. Natural language processing. IK International Pvt Ltd; Oct 6.

Lazarus L, Sookrajh R, Satyapal KS. 2017. Reflective journals: unmasking student perceptions of anatomical education. Folia morphologica 3:506-18.

Learman LA, Autry AM, O'Sullivan P. 2008. Reliability and validity of reflection exercises for obstetrics and gynecology residents. Am J Obstet Gynecol 4:461-e1.

LeBlanc VR, McConnell MM, Monteiro SD. 2015. Predictable chaos: a review of the effects of emotions on attention, memory and decision making. Adv Health Sci Educ 1:265-82.

Lee YS, Ryu Y, Jung WM, Kim J, Lee T, Chae Y. 2017. Understanding mind-body interaction from the perspective of east Asian medicine. Evid-based Complement Altern Article ID 7618419.

Lempp HK. 2005. Perceptions of dissection by students in one medical school: Beyond learning about anatomy. A qualitative study. Med Educ 39:318–325.

Lie D, Shapiro J, Cohn F, Najm W. 2010. Reflective practice enriches clerkship students' cross-cultural experiences. J Gen Intern Med 25(Suppl 2):S119–S125.

Lin CW, Lin MJ, Wen CC, Chu SY. 2016. A word-count approach to analyze linguistic patterns in the reflective writings of medical students. Med Educ Online 21(1):29522.

Liu B. 2012. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies 5(1):1-67.

Mamede S, Schmidt HG, Penaforte JC. 2008. Effects of reflective practice on the accuracy of medical diagnoses. Med Educ 42:48–75.

Mann K, Gordon J, MacLeod A. 2009. Reflection and reflective practice in health professions education: a systematic review. Adv in Health Sci Educ 4:595-621.

McGarvey MA, Farrell T, Conroy RM, Kandiah S, Monkhouse WS. 2001. Dissection: A positive experience. Clin Anat 14:227–230.

McMullan M. 2006. Students' perceptions on the use of portfolios in pre-registration nursing education: a questionnaire survey. Int J Nurs Stud. 3:333–343

Medhat W, Hassan A, Korashy H. 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams Eng J 4:1093-113.

Mohammad SM, Turney PD. 2013. Crowdsourcing a word–emotion association lexicon. Computational intelligence 29(3):436-465.

Mondal A, Cambria E, Das D, Hussain A, Bandyopadhyay S. 2018. Relation extraction of medical concepts using categorization and sentiment analysis. Cognitive Computation 10(4):670-685.

Mondal A, Chaturvedi I, Das D, Bajpai R, Bandyopadhyay S. Lexical resource for medical events: A polarity based approach. 2015 IEEE International Conference on Data Mining Workshop (ICDMW); 2015: IEEE. p. 1302-1309.

Mondal A, Das D, Cambria E, Bandyopadhyay S. WME: Sense, polarity and affinity based concept resource for medical events. Proceedings of the 8th Global WordNet Conference (GWC); 2016. p. 243-248.

Muniasamy A, Alasiry A. 2020. Deep learning: The impact on future elearning. Int J Emerg Technol 1: 188-199.

O'Carroll RE, Whiten S, Jackson D, Sinclair DW. 2002. Assessing the emotional impact of cadaver dissection on medical students. Med Educ 36:550–554.

Pathak AR, Agarwal B, Pandey M, Rautaray S. 2020. Application of Deep Learning Approaches for Sentiment Analysis. In: Agarwal B, Nayak R, Mittal N, Patnaik S, editors. Deep Learning-Based Approaches for Sentiment Analysis. Singapore: Springer Singapore. p. 1-31.

Pee B, Woodman T, Fry H, Davenport ES. 2002. Appraising and assessing reflection in students' writing on a structured worksheet. Med Educ. 6:575–585.

Pekrun R, Goetz T, Titz W, Perry RP. 2002. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educ Psychol 2:91-105.

Penney JC. 1985. Reactions of medical students to dissection. J Med Educ 60:58–60.

Plutchik, R. 1980. A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), Emotion: Theory, research and experience, Theories of emotion (Vol. 1, pp. 3–33). New York: Academic Press.

Préfontaine C, Gaboury I, Corriveau H, Beauchamp J, Lemire C, April MJ. 2021. Assessment tools for reflection in healthcare learners: A scoping review. Med Teach 12:1-7.

Quince TA, Barclay SI, Spear M, Parker RA, Wood DF. 2011. Student attitudes toward cadaveric dissection at a UK medical school. Anat Sci Educ 4:200–207.

Rabow MW, McPhee SJ. 2001. Doctoring to heal: Fostering well-being among physicians using personal reflection. West J Med 174(1):66–69.

Rinker T. lexicon [Internet]. Available from: https://github.com/trinker/lexicon

Rinker T. sentimentr [Internet]. Available from: https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf

Rinker T. sentimentr Github Repository [Internet]. Available from: https://github.com/trinker/sentimentr

Ripley BD. 2001. The R project in statistical computing. MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network 1(1):23-25.

Satapathy R, Cambria E, Hussain A. 2017. Sentiment Analysis in the Bio-Medical Domain. Springer Interntional Publishing AG, Gwerbestrasse 11:6630.

Sandars J. 2009. The use of reflection in medical education: AMEE Guide No. 44. Med Teach 8:685-95.

Sargeant JM, Mann KV, Vander Vleuten CP, Metsemakers JF. 2009. Reflection: A link between receiving and using assessment feedback. Adv Health Sci Educ Theory Pract 14(3):399–410.

Schön DA. 1983. The reflective practitioner: How professionals think in action. New York: Basic Books.

Shalev, A., Nathan, H., 1985. ‘Medical students’ stress reactions to dissections. Isr J Psychiatry Relat Sci 1–2:121–133.

Shapiro J, Nguyen VP, Mourra S, Boker JR, Ross M, Thai TM, Leonard RJ. 2009. Relationship of creative projects in anatomy to medical student professionalism, test performance and stress: an exploratory study. BMC Med Educ 9(1):1-0.

Shiozawa T, Glauben M, Banzhaf M, Griewatz J, Hirt B, Zipfel S, Lammerding-Koeppel M, Herrmann-Werner A. 2020. An insight into professional identity formation: qualitative analyses of two reflection interventions during the dissection course. Anat Sci Educ 3:320-32.

Sinclair S. 1997. Making Doctors: An Institutional Apprenticeship. 1st Ed. New York, NY: Berg Publishers. 320 p.

Smith B, Fellbaum C. 2004. Medical WordNet: a new methodology for the construction and validation of information resources for consumer health.

Smith CS, Irby DM. 1997. The roles of experience and reflection in ambulatory care education. Acad Med 72:32–5.

Wald HS, Reis SP, Monroe AD, Borkan JM. 2010. ‘The Loss of My Elderly Patient’: Interactive reflective writing to support medical students’ rites of passage. Med Teach 32(4):e178–e184.

Wald HS, Borkan JM, Taylor JS, Anthony D, Reis SP. 2012. Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. Acad Med 87:41-50.

Westberg J, Jason H. 1994. Fostering learners’ reflection and self-assessment. Fam Med 26:278–82.

Williams, R., G. Sundelin, E. Foster-Seargeant, G.R. Norman. 2000. Assessing the reliability of grading reflective journal writing. J Phys Ther Educ 14:23–26.

Wisner BP, Cravens GT, Holley KE, Carmichael SW, Pawlina W. 2004. Autopsy report: A valuable tool for developing observational skills. Clin Anat 17:197.

Wittich CM, Pawlina W, Drake RL, Szostek JH, Reed DA, Lachman N, McBride JM, Mandrekar JN, Beckman TJ. 2013. Validation of a method for measuring medical students' critical reflections on professionalism in gross anatomy. Anat Sci Educ 4:232-8.

Wong, F.K.Y., D. Kember, L.Y.F. Chung, and L. Yan. 1995. Assessing the level of reflection from reflective journals. J Adv Nurs 1: 48-57.

Wu A, Yu AC, Chang CS, Goel R, Sagoo MG, Sakurai T, Viranta-Kovanen S, Chien CL, Traxler H, Waschke J, Kitahara S. 2021. Initiating Students' Reflections on Life's Passing in the Anatomy Course–an International Observation at 14 Universities. Ann Anat 237:151741.

Zhang L, Wang S, Liu B. 2018. Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8(4):e1253.

Zhang, L., & Liu, B. (2017). Sentiment analysis and opinion mining. In C. Sammut & G. I. Webb (Eds.), Encyclopedia of machine learning and data mining (pp. 1152–1161). Springer.

FIGURE LEGENDS

Figure 1. Reflections processing pipeline. Student reflections were formatted and pre-preprocessed by stemming and lemmatization, tokenization, and stop word removal. Lexicon-based analysis was performed to visualize the results.

Figure 2. Comparison of the word count per reflection type (donor vs. self) excluding stop words. Reflections about the donor (n=677) averaged 29.4 [SD=14.9] words, whereas self-reflections (n=688) averaged 39.6 [SD=20.7] words.

Figure 3. Comparison of word count per body region per reflection type (donor (n=677) vs. self (n=688)) excluding stop words.The width of the boxplots corresponds to the number of reflections about the particular body region. Reflections about donors' noses were the least elaborated. In contrast, reflections about students' knees were the most elaborated based on median word count.

Figure 4. Comparison of relative frequencies of body regions students (n=132) chose to reflect on in reflections about donors (n=677) and self (n=688). Frequencies for each body region were calculated by dividing the number of reflections the students chose to reflect on by the number of reflections of the particular type. While self-reflecting, students chose to reflect mostly on hands (n=88), in contrast to reflections about the heart (n=87) when reflecting about the donor.

Figure 5. Top 10 most common words in all reflections (n=1,365) excluding stop words and words in the custom dictionary (n=21).

Figure 6. Top 10 most common words in self-reflections (n=688) excluding stop words and words in the custom dictionary (n=21).

Figure 7. Top 10 most common words in donor-related reflections (n=677) excluding stop words and words in the custom dictionary (n=21).

Figure 8. Comparison of word frequency in reflections about self (n=688) and donors (n=677) in the log-log scale. Only words appearing in both types of reflections were plotted (n=1,952).

Figure 9. Comparison of two-level sentiment between donor-related (n=677) and self-reflections (n=688). The relative frequency of positive and negative sentiment reflections was calculated as a ratio between the number of reflections with particular sentiment and the number of reflections of a particular type (self vs. donor). For both reflection types, the majority of reflections had positive polarity.

Figure 10. Comparison of three-level sentiment between donor-related (n=677) and self-reflections (n=688). The relative frequency of positive, neutral and negative sentiment reflections was calculated as a ratio between the number of reflections with particular sentiment and the number of reflections of a particular type (self vs. donor). For both reflection types, the majority of reflections had neutral polarity.

Figure 11. Top 10 words contributing to negative and positive sentiment of all reflections. Out of the total number of polarized words (n=1,786), “pain” occurred the most (n=121) amongst the negative words (n=837), whereas “love” occurred the most (n=162) amongst the positive words (n=949).

Figure 12. Top 10 words contributing to negative and positive sentiment of self-reflections. Out of the total number of polarized words (n=1,404), “pain” occurred the most (n=74) amongst the negative words (n=662), whereas “love” occurred the most (n=110) amongst the positive words (n=742).

Figure 13. Top 10 words contributing to negative and positive sentiment of donor-related reflections. Out of the total number of polarized words (n=1,034), “pain” occurred the most (n=47) amongst the negative words (n=662), whereas “loved” occurred the most (n=127) amongst the positive words (n=742).

Figure 14. Proportions of emotions per reflection type. Proportions of emotions were a ratio of the sum of occurrences of words associated with a particular emotion to the sum of occurrences of words associated with any out of eight emotions. Out of all occurrences in reflections about a donor of words associated with an emotion (n=5,307), words related to “trust” (n=1,350), “joy” (n=1,027) and “anticipation” (n=941) were the most common. Out of all occurrences in reflections about self of words associated with an emotion (n=7,140), words related to “trust” (n=1,796), “joy” (n=1,250) and “anticipation” (n=1.161) were also the most common.

Figure 15. Proportions of emotions per body region for all reflections (n=1,365). The proportions were calculated as a ratio of the sum of occurrences of all words in reflections about a particular body region associated with an emotion to the sum of occurrences of all words in reflections about the body region related to any out of eight emotions.

Figure 16. Comparison of emotions’ proportions per body region between self- (n=688) and donor-related (n=677) reflections. The proportions were calculated as a ratio of the sum of occurrences of all words in reflections, about self or a donor, regarding a particular body region

associated with an emotion to the sum of occurrences of all words in reflections about the body region related to any out of eight emotions.

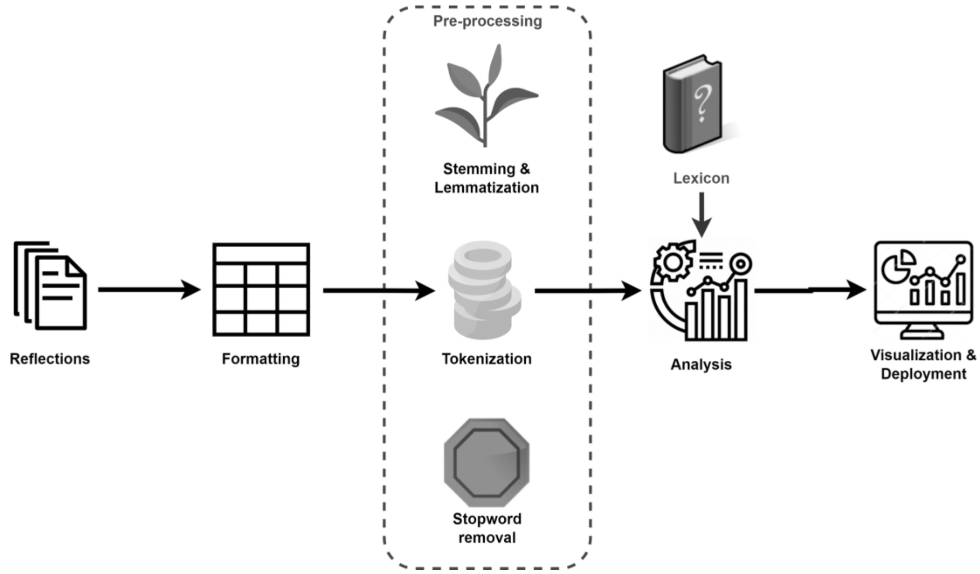

Figure 1. Reflections processing pipeline. Student reflections were formatted and pre-preprocessed by stemming and lemmatization, tokenization, and stop word removal. Lexicon-based analysis was performed to visualize the results.

175x98mm (300 x 300 DPI)

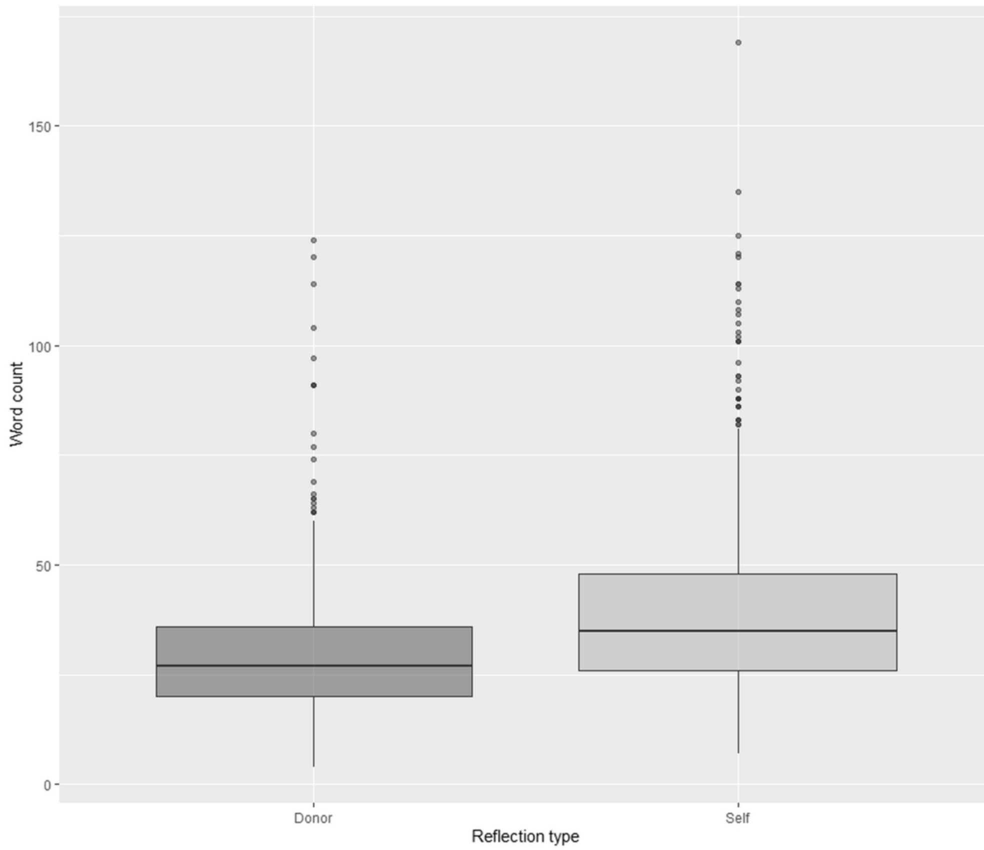

Figure 2. Comparison of the word count per reflection type (donor vs. self) excluding stop words. Reflections about the donor (n=677) averaged 29.4 [SD=14.9] words, whereas self-reflections (n=688) averaged 39.6 [SD=20.7] words.

175x149mm (300 x 300 DPI)

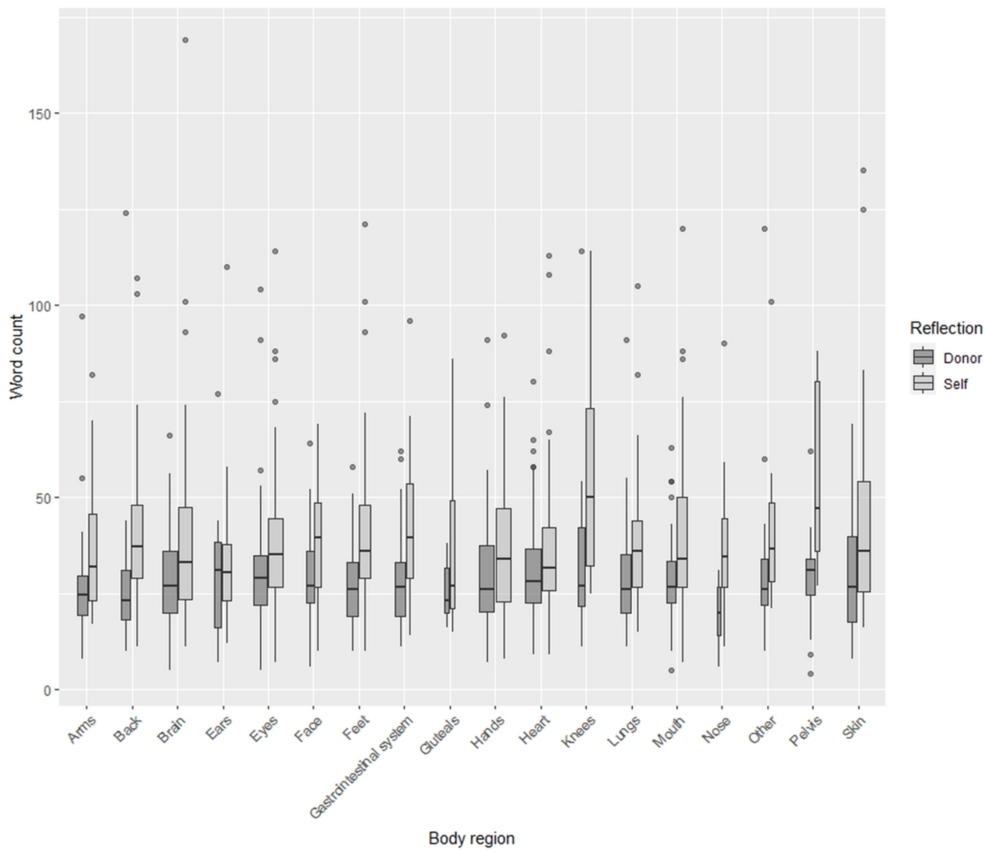

Figure 3. Comparison of word count per body region per reflection type (donor (n=677) vs. self (n=688)) excluding stop words.The width of the boxplots corresponds to the number of reflections about the particular body region. Reflections about donors' noses were the least elaborated. In contrast, reflections about students' knees were the most elaborated based on median word count.

175x149mm (300 x 300 DPI)

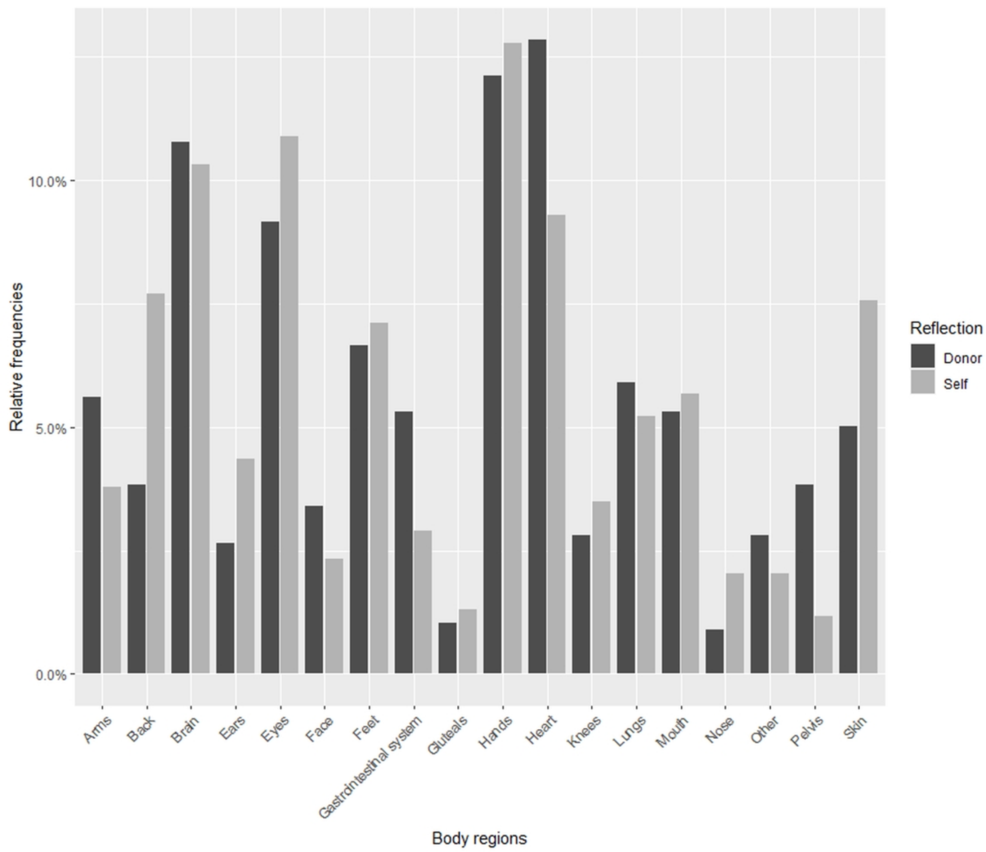

Figure 4. Comparison of relative frequencies of body regions students (n=132) chose to reflect on in reflections about donors (n=677) and self (n=688). Frequencies for each body region were calculated by dividing the number of reflections the students chose to reflect on by the number of reflections of the particular type. While self-reflecting, students chose to reflect mostly on hands (n=88), in contrast to reflections about the heart (n=87) when reflecting about the donor.

175x149mm (300 x 300 DPI)

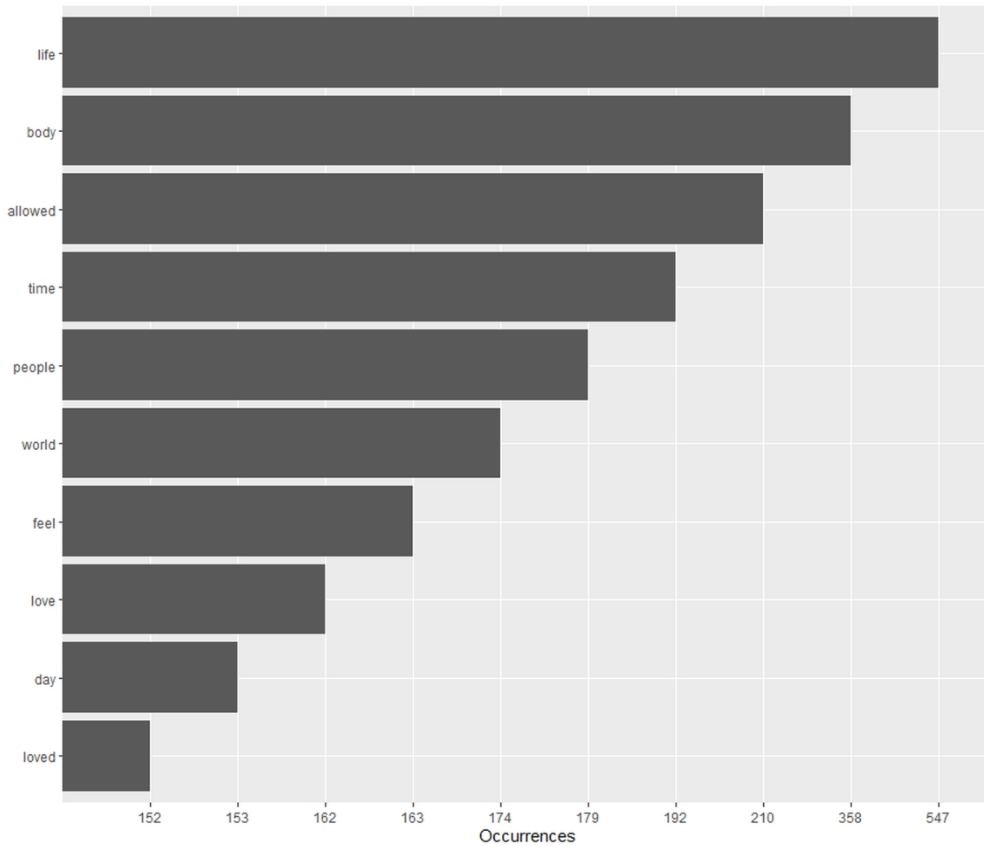

Figure 5. Top 10 most common words in all reflections (n=1,365) excluding stop words and words in the custom dictionary (n=21).

175x149mm (300 x 300 DPI)

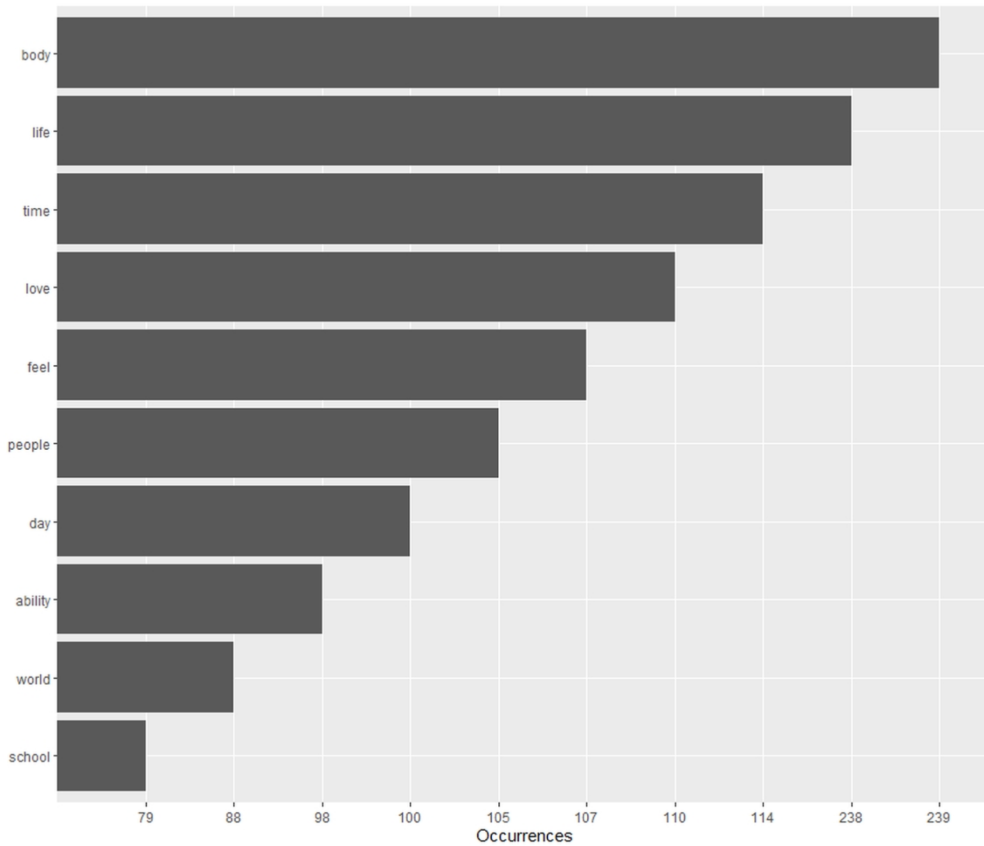

Figure 6. Top 10 most common words in self-reflections (n=688) excluding stop words and words in the custom dictionary (n=21).

175x149mm (300 x 300 DPI)

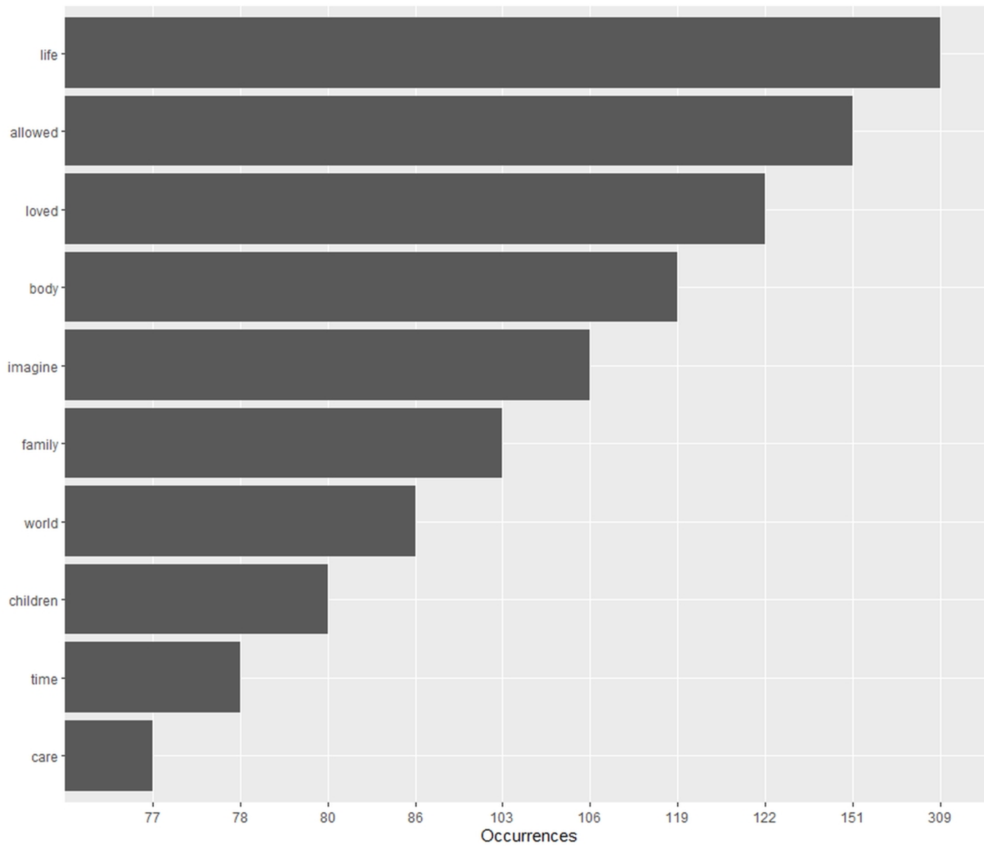

Figure 7. Top 10 most common words in donor-related reflections (n=677) excluding stop words and words in the custom dictionary (n=21).

175x149mm (300 x 300 DPI)

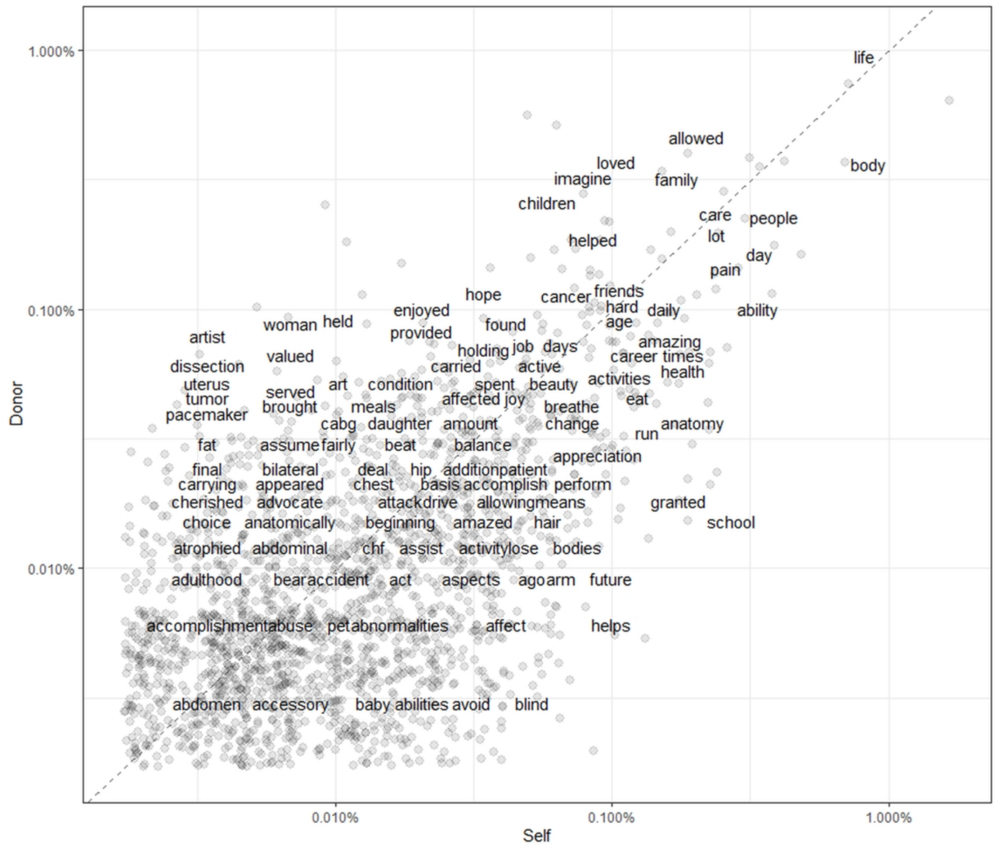

Figure 8. Comparison of word frequency in reflections about self (n=688) and donors (n=677) in the log-log scale. Only words appearing in both types of reflections were plotted (n=1,952).

175x149mm (300 x 300 DPI)

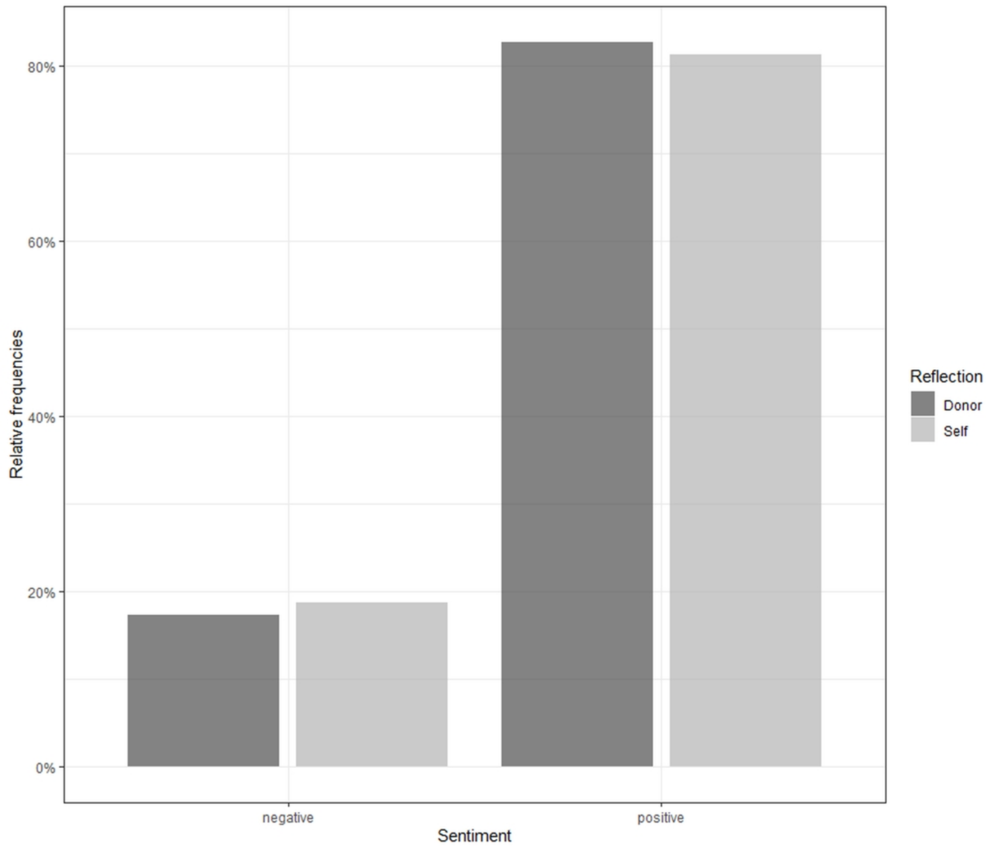

Figure 9. Comparison of two-level sentiment between donor-related (n=677) and self-reflections (n=688). The relative frequency of positive and negative sentiment reflections was calculated as a ratio between the number of reflections with particular sentiment and the number of reflections of a particular type (self vs. donor). For both reflection types, the majority of reflections had positive polarity.

175x149mm (300 x 300 DPI)

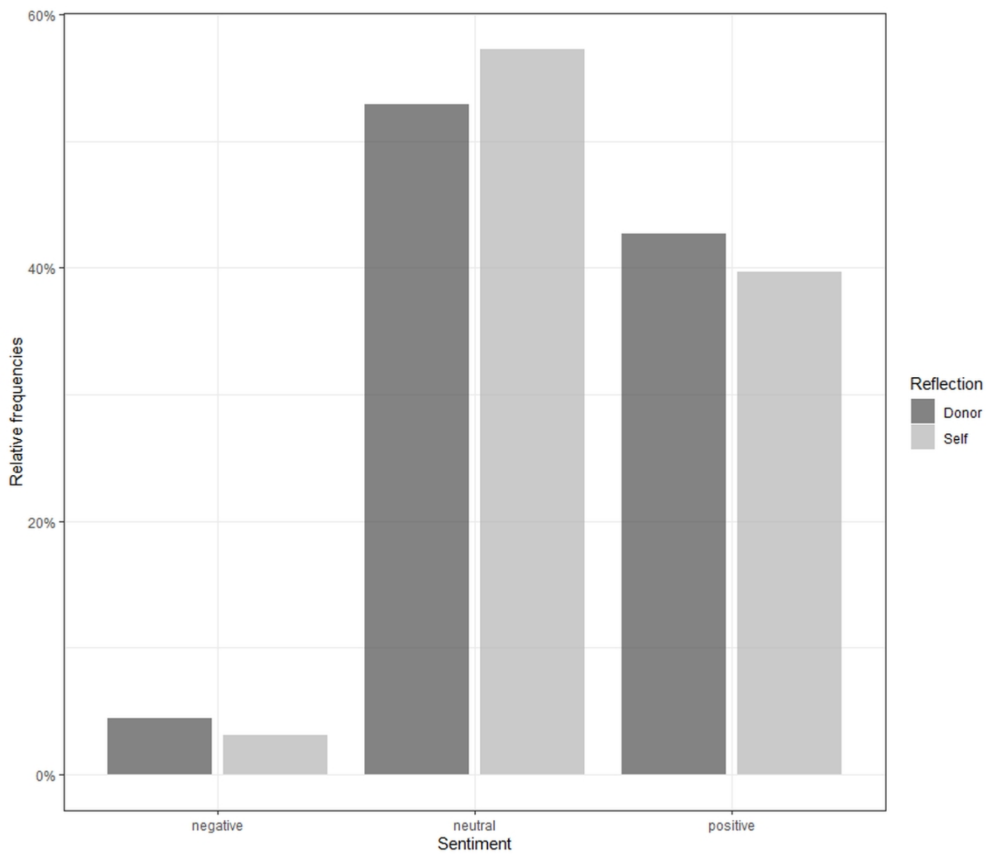

Figure 10. Comparison of three-level sentiment between donor-related (n=677) and self-reflections (n=688). The relative frequency of positive, neutral and negative sentiment reflections was calculated as a ratio between the number of reflections with particular sentiment and the number of reflections of a particular type (self vs. donor). For both reflection types, the majority of reflections had neutral polarity.

175x149mm (300 x 300 DPI)

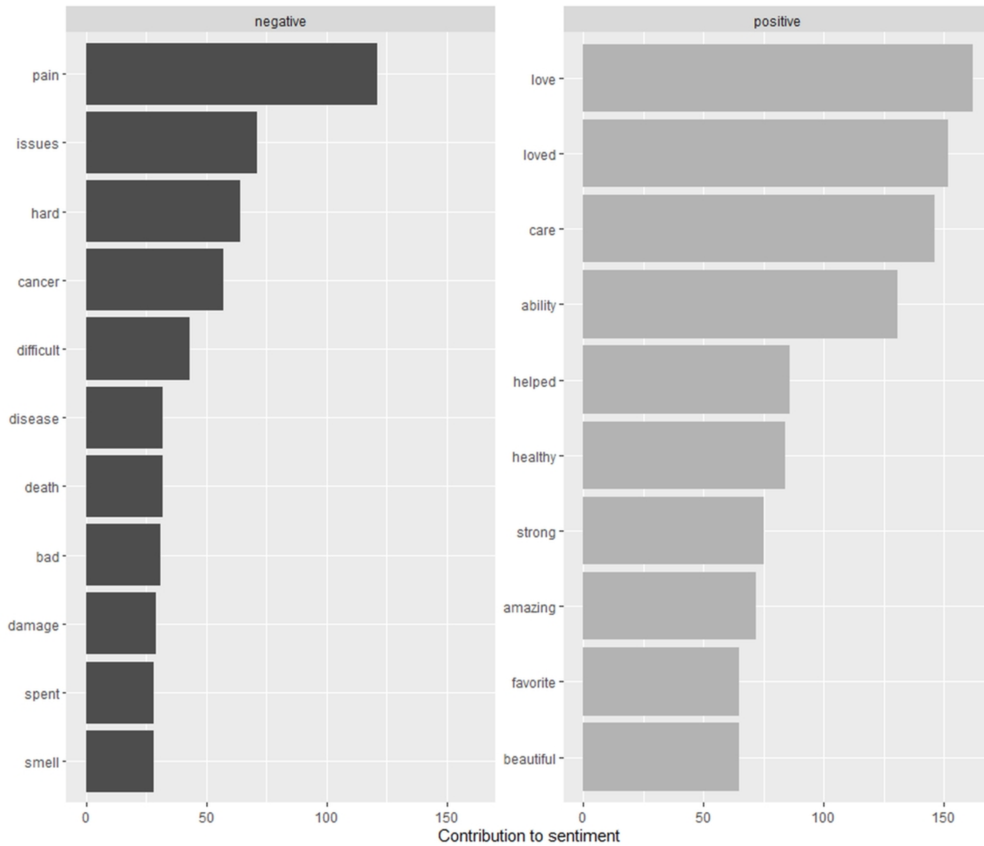

Figure 11. Top 10 words contributing to negative and positive sentiment of all reflections. Out of the total number of polarized words (n=1,786), "pain" occurred the most (n=121) amongst the negative words (n=837), whereas "love" occurred the most (n=162) amongst the positive words (n=949).

175x149mm (300 x 300 DPI)

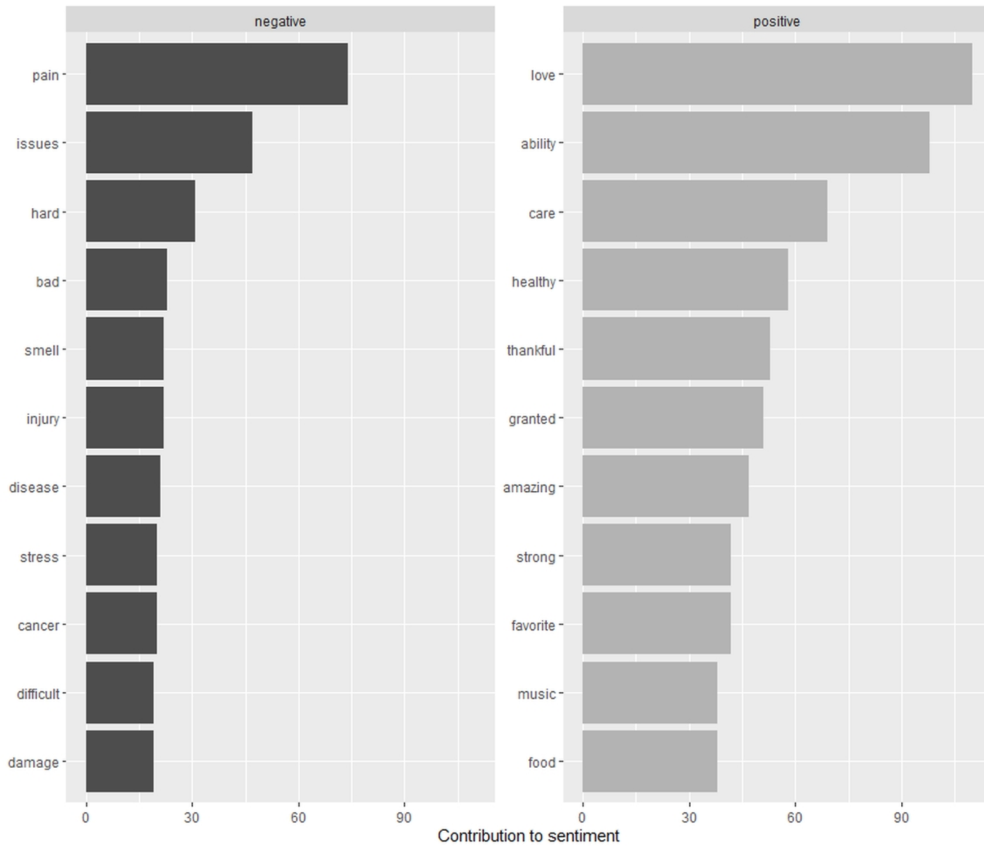

Figure 12. Top 10 words contributing to negative and positive sentiment of self-reflections. Out of the total number of polarized words (n=1,404), "pain" occurred the most (n=74) amongst the negative words (n=662), whereas "love" occurred the most (n=110) amongst the positive words (n=742).

175x149mm (300 x 300 DPI)

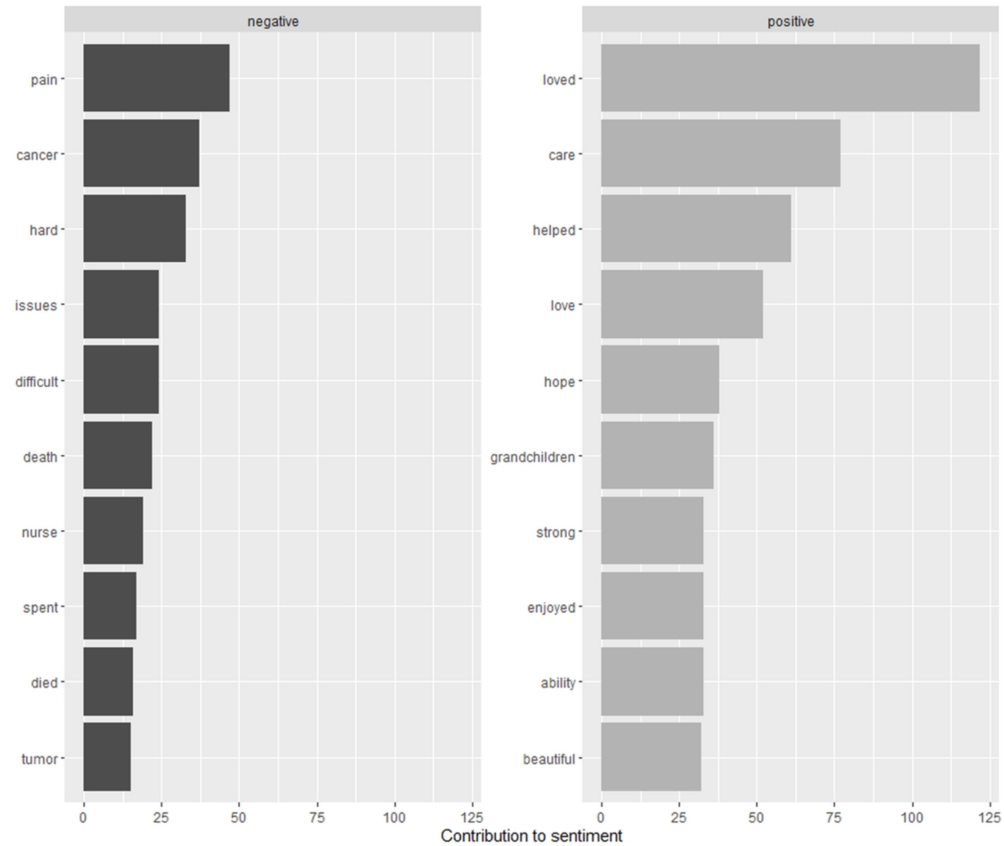

Figure 13. Top 10 words contributing to negative and positive sentiment of donor-related reflections. Out of the total number of polarized words (n=1,034), "pain" occurred the most (n=47) amongst the negative words (n=662), whereas "loved" occurred the most (n=127) amongst the positive words (n=742).

175x149mm (300 x 300 DPI)

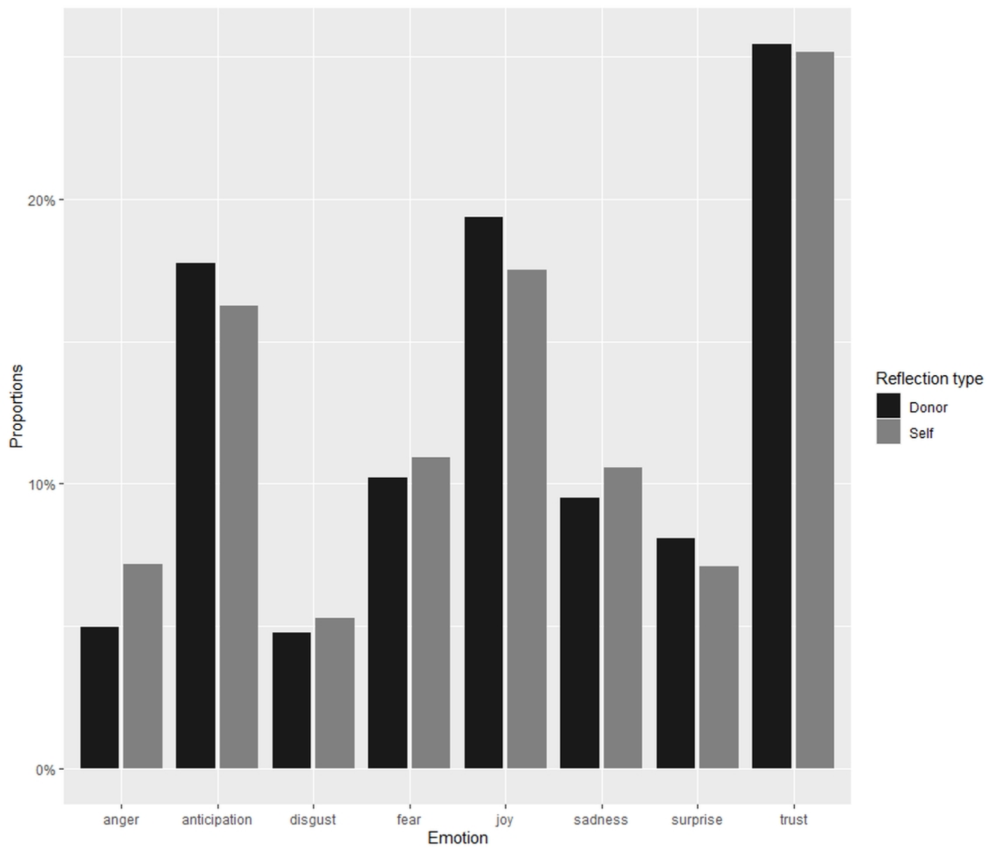

Figure 14. Proportions of emotions per reflection type. Proportions of emotions were a ratio of the sum of occurrences of words associated with a particular emotion to the sum of occurrences of words associated with any out of eight emotions. Out of all occurrences in reflections about a donor of words associated with an emotion (n=5,307), words related to "trust" (n=1,350), "joy" (n=1,027) and "anticipation" (n=941) were the most common. Out of all occurrences in reflections about self of words associated with an emotion (n=7,140), words related to "trust" (n=1,796), "joy" (n=1,250) and "anticipation" (n=1.161) were also the most common.

175x149mm (300 x 300 DPI)

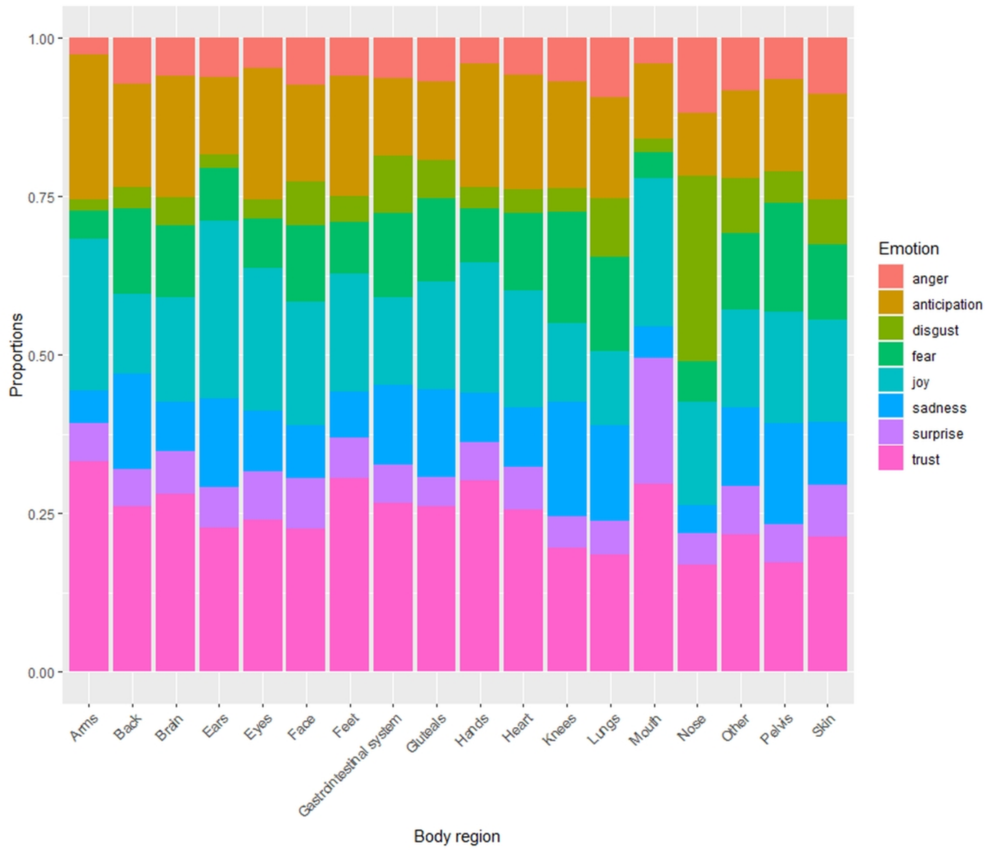

Figure 15. Proportions of emotions per body region for all reflections (n=1,365). The proportions were calculated as a ratio of the sum of occurrences of all words in reflections about a particular body region associated with an emotion to the sum of occurrences of all words in reflections about the body region related to any out of eight emotions.

175x149mm (300 x 300 DPI)

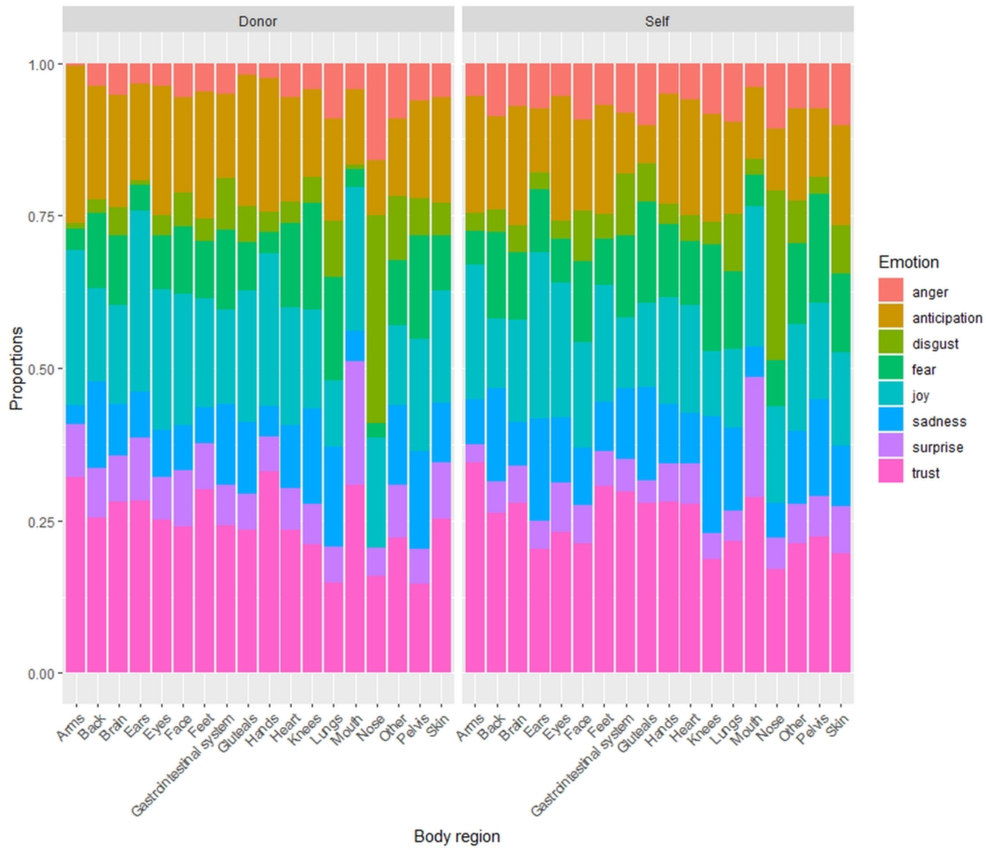

Figure 16. Comparison of emotions' proportions per body region between self- (n=688) and donor-related (n=677) reflections. The proportions were calculated as a ratio of the sum of occurrences of all words in reflections, about self or a donor, regarding a particular body region associated with an emotion to the sum of occurrences of all words in reflections about the body region related to any out of eight emotions.

175x149mm (300 x 300 DPI)