Cogent Mental Health

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/oamh20

Are we listening to every word? Using multiple analytic methods to examine qualitative data

Ryan L. Boyd, Nicholas R. Morrison, Sarah D. Horwitz, Rachel Maciag, Emma Travers-Hill & Youngsuk Kim

To cite this article: Ryan L. Boyd, Nicholas R. Morrison, Sarah D. Horwitz, Rachel Maciag, Emma Travers-Hill & Youngsuk Kim (2024) Are we listening to every word? Using multiple analytic methods to examine qualitative data, Cogent Mental Health, 3:1, 1-24, DOI: 10.1080/28324765.2024.2433791

To link to this article: https://doi.org/10.1080/28324765.2024.2433791

COGENT MENTAL HEALTH
2024, VOL. 3, NO. 1, 2433791
https://doi.org/10.1080/28324765.2024.2433791

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Are we listening to every word? Using multiple analytic methods to examine qualitative data

Ryan L. Boyd [a], Nicholas R. Morrison [b], Sarah D. Horwitz[c], Rachel Maciag[d], Emma Travers-Hill [e] and Youngsuk Kim [c,d]

[a]Department of Psychology, School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX, USA; [b]Department of Psychology, Westfield State University, Westfield, MA, USA; [c]Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA; [d]Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK; [e]School of Psychology, University of Kent, Canterbury, Kent, UK

ABSTRACT

Psychological researchers are increasingly striving to enhance methodological integrity, including in qualitative methods. Although computerized text analysis tools originally emerged as a potential replacement for manual coding approaches, recent studies have underscored the unique yet complementary value of employing several methods. The current study applies two text analysis methods across one qualitative dataset to explore whether each method yields information not clearly evidenced by the other, nor through traditional thematic analysis. Interviews exploring the experiences of paraprofessionals delivering Brief Psychological Interventions (BPIs) were analyzed through Linguistic Inquiry and Word Count (LIWC) and the Meaning Extraction Method (MEM). Results revealed LIWC, MEM, and thematic analysis to be complementary in nature, each providing unique insights that could be missed by implementing any one method alone. Moreover, text analyses can serve as a form of validation for more traditional qualitative approaches while also revealing otherwise indiscernible relationships and patterns within texts.

KEYWORDS qualitative methods; text analysis; thematic analysis; LIWC; Meaning Extraction Method

Until the introduction of grounded theory in 1967 (Glaser & Strauss, 1999), there was little work focusing on qualitative analytic methods (Adams & Preiss, 1960; Junker, 1960; Kahn & Cannell, 1957). Since then, qualitative approaches, with various epistemological underpinnings, have been increasingly employed in psychological research, with considerable emphasis in counseling and psychotherapy (Hays et al., 2016; Levitt, 2015; McLeod, 2000).

CONTACT Ryan L. Boyd boyd@utdallas.edu Department of Psychology, School of Behavioral and Brain Science, University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/28324765.2024.2433791

Despite paradigmatic differences between qualitative traditions, some researchers have made trans-epistemological efforts to advance the integrity of qualitative analysis (Leech & Onwuegbuzie, 2007; Whittemore et al., 2001). To date, approaches for enhancing trust in qualitative analysis include methods like data triangulation to reduce investigator bias, data auditing to 'check' primary raters, and rater reflexivity to balance participant meaning and researcher interpretation (Kisely & Kendall, 2011). More recently, the Society for Qualitative Inquiry in Psychology, a section of Division 5 (Quantitative and Qualitative Methods) of the American Psychological Association (APA), appointed a Task Force on Resources for the Publication of Qualitative Research. This initiative culminated in recommendations for designing, presenting, and reviewing qualitative research in psychology (Levitt et al., 2017). Among these recommendations, the Task Force included a variety of analytic considerations, including grounding study findings within the data, establishing coherence among the findings, generating meaningful contributions toward the project goal, and determining whether the researcher's perspective is managed to enhance fidelity. The Task Force was clear that methodological integrity is a matter of interpretation, and they were hesitant to provide procedural recommendations on how best to establish it (Levitt et al., 2017).

As qualitative researchers have continued to grapple with issues pertaining to methodological integrity, new approaches that are congruent with the APA Task Force recommendations continue to emerge and evolve. For example, while various data triangulation techniques have long existed (Denzin, 1978), more recent efforts involving data analysis triangulation can combine two or more methods of analyzing data, otherwise referred to as *intramethod triangulation* (Casey & Murphy, 2009; Renz et al., 2018). With the ever-increasing capacity for computer-assisted analytic approaches, there has been an increase in the use of text analysis as a triangulation method that can be combined with thematic analyses of qualitative data (e.g., Firmin et al., 2017; Whitney et al., 2005).

Whitney et al. (2005) utilized both qualitative thematic analysis and computerized text analysis on narratives written by carers. They found that while thematic analysis allowed for a comprehensive, unrestricted exploration of the data, it raised concerns about reliability and researcher bias. In contrast, computerized text analysis provided a more reliable and less biased assessment, though it was limited by rigid categories and a literal interpretation of the text, which restricted its ability to capture deeper meanings. Despite these differences, both methods produced similar findings, offering some validation of the results. Over a decade later, Firmin et al. (2017) conducted a similar comparison, analyzing the same qualitative data using both text analysis and traditional thematic analysis. Their findings echoed those of Whitney et al. 2005, revealing complementary strengths in each method.

Computerized text analysis was efficient in identifying patterns and managing large datasets but lacked the depth and contextual understanding offered by manual thematic analysis. Although thematic analysis required more researcher involvement, it provided richer insights into complex themes and participant experiences. Both studies concluded that a combined approach using computerized text analysis could enhance the rigor and depth of qualitative research.

Before discussing the research in this area on what needs to be compared and how, we provide here a brief background on computerized text analysis and methods.

Computers, text analysis, and psychometrics

Computerized text analysis methods were first introduced to psychological research in the 1960's (Stone & Hunt, 1963). They were initially treated as a potential replacement for manual scoring of texts, and have since evolved into a complementary approach that capitalizes on statistical regularities of verbal behavior (Boyd & Schwartz, 2021). Whereas human coders can 'understand' the meaning behind a text and account for things like context and speaker characteristics (i.e., the 'content' and 'context' of a sample of language), automated methods can pick up on incredibly subtle, but psychologically revealing, 'stylistic' variations in language—that is, differences in *how* a person says something, regardless of *what* they are saying. Such stylistic differences prove difficult for humans to ascertain, in part, because they are subtle and statistical (Ireland & Mehl, 2014). Ultimately, a major strength of automated methods is that they are able to provide objective clues about how a person is thinking that are difficult for human coders to ascertain (Chung & Pennebaker, 2019).

Broadly speaking, computerized methods for text analysis fall under two categories: 'top-down' and 'bottom-up' (Boyd, 2017). Top-down methods involve scanning texts and counting up the number of times various predetermined 'categories' of language appear (Gottschalk, 1997). This is also known as dictionary-based coding. A common example is that we may have a computer program scan texts for words like 'happy' and 'wonderful' as a way to quantify the degree to which a person is experiencing a positive emotional state. Such top-down, dictionary-based methods are designed to be probabilistic in nature: no serious researcher treats them as 'perfect' measures of some underlying construct—for example, the word 'happy' may be used in a sarcastic way, or a speaker may be saying that they cannot remember the last time that they were happy. However, given a large enough sample of text, the assumption generally holds true that greater rates of positive emotion words correspond to a greater experience of positive affect (e.g., Kahn et al., 2007; Saxbe et al., 2013; Smith, 2018); this rationale largely holds true across psychological constructs (see: Kennedy et al., 2022).

Bottom-up methods, on the other hand, are data-driven and are much more diffuse in terms of their goals and purpose. One of the more common families of bottom-up methods is known as 'topic modeling' (Currin McCulloch et al., 2021). Topic modeling can be used, generally speaking, as a way to identify the themes that characterize an entire body of text in 'broad strokes'. That is, topic modeling is an approach that is used to describe the major themes that emerge from a corpus of text in fairly general terms, but will not usually reveal (relatively) rare or nuanced themes. Most topic-modeling methods work by identifying words that tend to co-occur and 'cluster' together to form meaningful themes (Maier et al., 2018). For example, texts that use the word 'father' are also more likely to include words like 'mother' and 'child' than texts without it. By analyzing a collection of texts, these networks of word co-occurrences can be detected statistically, forming 'themes' or 'topics'. Bottom-up methods are, therefore, useful for identifying content/'categories' of meaning that characterize an entire collection of text. They are complementary to top-down approaches in that top-down approaches require an explicit definition of the categories that one is trying to capture/measure (Tong et al., 2020). Topic modeling, on the other hand, can be used to 'uncover' themes that are not already codified within an existing dictionary. Notably, a few studies such as Nikolenko et al. (2017) and Gregson et al. (2022) approached topic modeling and qualitative analysis from different perspectives and applications, while convergently demonstrating the importance of thematic analysis in understanding complex social and textual phenomena.

Analyzing qualitative data using text analyses have become incredibly popular due not only to their ability to accelerate the research process but because of their ability to reliably and objectively quantify certain aspects of human psychology which may be difficult for humans to detect (e.g., such as 'thinking styles' and other nuanced states). Critically, however, such methods are rarely paired with qualitative analyses in any formal manner. We suggest that a clear collaboration between qualitative and automated text analytic methods is important precisely because of their complementary nature in terms of relative strengths and weaknesses.

Current study

This study aims to build on pre-existing research (e.g., Firmin et al., 2017; Whitney et al., 2005) surrounding the use of multiple analytic methods for examining qualitative data. Namely, this is an exploratory study using text analyses on qualitative data to determine whether such analyses will yield information that was not evident from using a traditional qualitative method.

Thus, our exploratory research questions include 1) To what extent do computerized text analyses and qualitative analyses yield complementary insights into a specific phenomenon, in this case, the training and supervision experiences of paraprofessionals delivering Brief Psychological Interventions (BPIs); and 2) Do themes identified through traditional qualitative analysis align with patterns detected using computerized text analysis. To this end, we compared themes identified manually with those emerging from computerized text analysis to assess convergence/divergence and explored discrepancies to understand the unique contributions and limitations of each method, with the specific intention to elucidate for the reader methodological considerations that can be encountered when integrating qualitative methods and computerized text analyses.

Method

The study relied upon qualitative data that was previously gathered to explore the experiences of paraprofessionals who were trained and supervised to deliver Brief Psychological Interventions (BPIs).[1] BPIs were introduced in two secondary care mental health teams in Cambridge, UK, as a service initiative to improve access to psychological interventions. The three manualized interventions consist of 1) behavioral activation, 2) distress tolerance, and 3) anxiety management. They are case formulation-driven and developed using Cognitive Behavior Therapy and Dialectical Behavior Therapy frameworks. BPIs are delivered weekly for 6 to 12 weeks by trained paraprofessionals. Service evaluations have found significant improvements across all three BPIs in symptom reduction and improvement in well-being (Roberts et al., 2021; Wright et al., 2020). In addition to examining the effectiveness of the BPIs, a qualitative study was carried out to explore the experience of paraprofessionals delivering the BPIs (Maciag et al., 2023) and this data was used for the purpose of this study.

Data collection

The Quality Improvement of the local National Health Service Trust approved this study. Data was collected from October 2018 to March 2019. Semi-structured interviews were audio-recorded using digital recorders and transcribed verbatim. Steps were taken to preserve the anonymity and confidentiality of the interviewees.

Among the eleven participants, there were seven Support Time Recovery (STR) workers, two peer support workers (PSWs), and two assistant psychologists. STR workers are non-credentialed/non-licensed mental health staff with various responsibilities, including providing practical and social support. PSWs have lived experience with mental health conditions and have completed specialized training to support patients. Eight participants were female, and three were male. Academic or relevant training levels varied, ranging from no formal qualifications to a Master's degree in psychology. More specifically, four participants held psychology degrees, two had social care training or qualifications, and two were completing PSW training. Experience in mental health settings ranged from none to over 10 years, while the duration of delivering the BPI spanned from under three months to over three years. Caseloads varied: five participants handled fewer BPI cases than other duties, two had an even split, and four had more BPI cases than other responsibilities. All participants were interviewed individually by one of the authors (RM) and the interview was audio-recorded and transcribed. Further information on interviewees and data collection is detailed in the qualitative study (Maciag et al., 2023).

The interview began with background information before delving into 5 domains of inquiry (i.e., 'topics'):

(1) Training
(2) Manuals
(3) General experiences with delivery
(4) Supervision
(5) Views of BPIs

The specific prompts used to initiate discussion within each topic can be found in Supplementary Materials A. Three authors accessed the transcripts in either audio (RM, ETH) or de-identified text format (RM, ETH, RB). Only the output of the data analyses was shared with the other authors (YK, NM, SH).

The Framework Approach, as outlined by Braun and Clarke (2006), was used to capture participants' original accounts. This process involved familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the final report.

Three main themes emerged: 1) training, 2) supervision, and 3) general experience. Training subthemes consisted of self-directed learning, learning through doing, shadowing, and being shadowed. Supervision subthemes consisted of adherence and flexibility, peer support and supportive mechanisms. General experiences subthemes consisted of warmth and empathy in delivering staff, importance of boundaries in delivering paraprofessionals' role, and impact of delivering staff's professional identity.

Computerized text analysis procedures

The transcribed interview texts were analyzed using two separate methods, each with their own analytic considerations and preprocessing strategies.

Here, we provide a concise overview of each method used and the steps taken to create quantitative insights from the raw language data.

Initially, all transcribed texts were contained within Microsoft Word files formatted to differentiate language spoken by the interviewer and the interviewee; transcripts were additionally separated into topic areas denoted with topic-specific headings. A custom Python script was used to automatically identify and separate texts on a 'by speaker' basis, resulting in a prepared dataset that contained only the language spoken by interviewees, separated by topic.

Top-down content coding

To quantify markers of psychological states within each interviewee's responses, all interviewee texts were analyzed using the Linguistic Inquiry and Word Count (LIWC) 2015 version of the LIWC software and dictionary (Pennebaker et al., 2015), separately by topic. One of the primary strengths of LIWC is that it is largely context-insensitive, meaning that it can tell us a lot about the psychology behind a sample of text without needing much training (Pennebaker, 2011). In particular, it is one of the very few systems that explicitly quantifies function words (e.g., prepositions, conjunctions, etc.) as measures of psychological/thinking styles, which have been found to be particularly powerful clues to a person's mental state. Stylistic differences include, for example, 'We played all night long' versus 'The game lasted all night'. Both contain the same number of words, but one can 'feel' that the second example is more detached and less enthused.

Briefly described, LIWC2015 analyzes texts for words that belong to one of ~80 categories of language that have been demonstrated to reflect various thinking styles, attentional processes, and other psychological traits over thousands of empirical studies (see, e.g., Tausczik & Pennebaker, 2010). In general, scores generated by LIWC reflect the relative percentage of words that belong to each language category. For example, the phrase 'I adore watching old movies' would be scored as 20% first-person singular pronouns ('I'), 20% positive emotion words ('adore'), 20% perception words ('watching'), and so on. The current study focused on 5 LIWC categories that were determined *a priori* to be potentially meaningful: word count, positive emotions, negative emotions, discrepancies, and certainty. In Table 1, we provide a brief description and examples for each measure.

Importantly, LIWC2015 provides output for an additional 4 'summary measures' that reflect broader thinking styles that can be estimated from verbal behavior. These summary measures include: 1) analytic thinking, 2) clout, 3) authenticity, and 4) emotional tone. Unlike the above-mentioned measures, LIWC2015's summary measures reflect normative percentiles; that is, where a text scoring 50 is typical for that dimension, on average. Table 2 includes brief descriptions and more extended reading for each

Table 1. Overview of relevant LIWC measures, with example words

Measure	Description	Examples
Total Word Count (WC)	Total number of words spoken by interviewee.	N/A
Positive emotions (posemo)	Words suggestive of positive evaluation or emotional state	good, love, happy, hope
Negative emotions (negemo)	Words suggestive of negative evaluation or emotional state	bad, hate, hurt, tired
Discrepancies (discrep)	Words indicating a discrepancy between two states	would, should, could
Certainty (certain)	Words suggesting a confidence in one's evaluation or beliefs; a lack of questioning	sure, certain, absolutely, definitely

Table 2. Overview and further reading for LIWC "summary" measures

Measure	Description	Citations
Analytic Thinking	The extent to which a text is formal and logical, identifying connections between ideas (high scores) versus informal, narrative-driven, and personal (low scores)	Pennebaker et al. (2014); Jordan et al. (2019)
Clout	Reflects the speaker's confidence and self-perceived social standing relative to others	Kacewicz et al. (2014); Fox and Royne Stafford (2021)
Authenticity	The degree to which a text is produced in a spontaneous, unfiltered manner	Newman et al. (2003); Kalichman and Smyth (2021)
Emotional Tone	The overall balance between positive and negative emotions within a text	Cohn et al. (2004); Monzani et al. (2021)

metric. We opted to include LIWC's summary measures in our analyses given their potential to reveal important psychological nuance in interviewees' responses, such as the degree to which responses were confident (clout) or presented in a spontaneous, unfiltered manner (authenticity).

Bottom-up content coding

The LIWC method of scoring texts relies on predefined word-to-category mappings, allowing us to quantify the degree to which interviewees were responding to questions, stylistically speaking. However, we additionally sought to score all interviewees' responses for *emergent* themes—that is, common ideas and concepts that emerged naturally across interviewees in general (e.g., Brookes & McEnery, 2019). In order to explore those themes that were most prevalent, we performed the Meaning Extraction Method (MEM; Chung & Pennebaker, 2008), a topic-modeling procedure that identifies common word co-occurrence clusters.

The MEM procedure was conducted as a series of steps using the BUTTER text analysis platform (Boyd, 2020). All steps described below aligned with general recommendations provided elsewhere with regard to the extraction of themes from texts using the MEM (see Boyd, 2017; Markowitz, 2020).

(1) Responses were segmented into individual sentences, grouped by topic
(2) All texts were "tokenized" into individual words, then lemmatized to convert words to their most basic forms (e.g., the words "works", "worked", and "working" all get converted to their lemma: "work")
(3) All words with a raw frequency ≥5 were retained
(4) A document—term matrix (DTM) was created: a matrix that reflects which sentences contained which words

Once the interviewee responses had been converted to a DTM that was ready for topic-modeling to be conducted, we performed a Principal Components Analysis (PCA) with varimax rotation. Given the relatively small number of responses in our sample, we extracted 3 components per topic. An overview of our PCA results is provided in Supplementary Materials B; additional details of all analyses are presented in Supplementary Materials C.

Results

Traditionally, one of the primary goals of quantitative text analytic methods is the generation of data that will be used in some form of statistical modeling. However, even in the case of smaller samples that do not lend themselves to null hypothesis significance testing, we can learn much from a thoughtful interpretation of the results that are generated from computerized text analysis methods. Here, we provide a brief tour of the results from the LIWC and MEM approaches using the BPI qualitative data, highlighting general strategies for interpreting the results of these analytic methods in a manner that is complementary to qualitative analyses of the same data.

LIWC results

To get a general sense of how interviewee language varied for each LIWC measure, standard descriptive statistics were calculated (i.e., mean and standard deviations), by research topic; for all metrics; results are presented in Table 3. Interpretation of scores generated by dictionary-based coding requires two levels of analysis: 1) the psychological meaning of a particular dictionary category, and 2) concluding whether a difference is large enough to be meaningful given the context/research question. In truth, both levels of analyses are related to each other; a numerically large difference in a category like "positive emotion" language may be meaningful when comparing speeches given by the same orator, for example,

Table 3. Descriptive statistics of LIWC measures across interview topics

	Interview Topic				
LIWC Measure	Delivery M (SD)	Training M (SD)	Manuals M (SD)	Supervision M (SD)	Views of BPI M (SD)
Total Word Count	1833.73 (776.48)	1427.00 (1019.81)	1433.64 (780.59)	920.00 (472.88)	642.36 (275.01)
Positive Emotion	3.36 (0.85)	3.42 (0.91)	3.40 (0.50)	2.89 (0.54)	4.05 (0.86)
Negative Emotion	1.44 (0.57)	1.14 (0.62)	1.73 (0.64)	0.98 (0.65)	1.16 (0.84)
Discrepancies	1.77 (0.62)	1.52 (0.55)	1.64 (0.42)	2.25 (1.07)	2.11 (0.73)
Certainty	1.36 (0.51)	1.79 (0.70)	1.37 (0.47)	1.89 (0.92)	1.44 (0.88)
Analytic Thinking	18.58 (7.24)	19.24 (5.83)	22.33 (5.34)	15.03 (3.77)	18.61 (8.87)
Clout	43.79 (12.85)	33.60 (12.74)	40.48 (13.20)	32.70 (11.55)	43.93 (12.54)
Authenticity	60.78 (15.49)	72.59 (11.85)	67.44 (12.14)	75.37 (11.31)	61.90 (14.49)
Emotional Tone	59.85 (19.67)	65.43 (18.64)	56.80 (13.67)	60.79 (18.24)	76.39 (15.51)

but less so if one speech was delivered at a birthday party while the other was a funeral eulogy. In our case, we combine our knowledge of the *subject domain* for each LIWC score with an *a priori* understanding of what each category tends to reflect when used at relatively high-versus-low rates (see, e.g., Tausczik & Pennebaker, 2010). We revisit this important point in the discussion.

In looking for differences, we recommend thinking about category differences in terms of percent differences, or relative differences on a within-category basis. In some cases, the reason for this is clear: for example, word count (WC) and certainty words are measured on completely different scales (raw counts versus relative frequencies, respectively), such that a direct comparison across categories is not meaningful. For example, while often used primarily as a measure of data quantity to ensure that enough language data is present for reliable estimates, Word Count is an often-overlooked but highly revealing metric that can also reveal the degree of engagement a person has with a given topic. As any therapist will tell you: people often have more to say about a topic that is important to them. In our sample, we saw that interviewees had the least to say about their views of the pros/cons of BPIs ($M = 642$ words) but a considerable amount to say about their personal experiences with delivering BPIs—in fact, nearly three times as much ($M = 1833$ words).

Similarly, if we consider the degree to which interviewees used language indicative of certainty—that is, a concrete, preconceived notion of their own thoughts or beliefs about a topic—we can see that interviewees appeared to have relatively strong or well-developed ideas around training in BPIs ($M = 1.79$) relative to how they described their experiences with delivering BPIs (1.36) or the manuals used to deliver the intervention (1.37). In both cases, the use of certainty language was >30% higher when describing their training, potentially as a result of varied levels of experience in delivering BPIs or using BPI manuals.

Taking a step back, we also note the value in a more holistic perspective that considers multiple category differences at the same time to aid in the interpretation of interviewees' general feelings and mindset when discussing different topics. When looking at the average word count for the discussion topic, for example, one might be tempted to conclude that interviewees had little in the way of opinion or strong thoughts about supervision (*M* word count = 920) relative to most other topics (e.g., *M* word count = 1834 for "Delivery"). However, when considering the relative scores of multiple categories, a slightly different picture emerges. Note that emotional language, both positive and negative, scored considerably lower than that for other writing prompts, as did the "Clout" measure which tends to reflect one's confidence or social standing. Taken together, we might instead conclude that interviewees were hesitant to provide any type of judgment or evaluation of the supervisory components of the BPIs. Despite the fact that interviewees were assured that their responses were confidential and anonymous, they may have experienced a degree of restraint in sharing their "true" opinions on a topic that might be interpreted as reflecting poorly on their peers and supervisors, instead opting to share only factual information about which they were certain—in particular, focusing on discrepancies between ideal and actual features of supervision.

It is important to emphasize that many, or perhaps most, of the differences that we observed in interviewees' verbal behaviors across prompts would be difficult to see without the assistance of computerized methods. Indeed, in a subjective reading of the interview transcripts, the degree to which an interviewee's verbal behavioral profile reveals psychological states were often nuanced and not initially apparent on the surface. We found that such analyses helped us identify and further explore nuances of each interviewee's responses, highlighting how they may have been *thinking or feeling* about a question regardless of the *content* of their responses.

MEM results

In much (if not most) topic modeling research, the ultimate goal is to quantify the degree to which any given text reflects or is comprised of each theme. For example, one might seek to classify a large number of documents by their main topic (e.g., politics, sports, technology) to make them easier to organize and search (e.g., Hingmire et al., 2013), or quantify the extent to which a person talks about specific personal interests online, such as reading or going to parties (Schwartz & Ungar, 2015). Such approaches often involve the construction of a topic model on a corpus of text, then a "reapplication" of the model to the data to automatically evaluate how likely it is that each theme is being used within any given text.

However, results from most topic modeling methods can also be explored at various stages of the analysis, and the ultimate goal may not be quantification of themes but, rather, the *discovery* of themes within a corpus of text. For example, what researchers may be most interested in is simply understanding what themes are present across an entire collection of texts—this stage is often useful in and of itself as a way to better understand the topography of a corpus's content, with less focus on measuring whether a *specific* theme is present in a *specific* text (see, e.g., Currin McCulloch et al., 2021; McCloskey et al., 2022). Put more simply: often, it is meaningful to explore what themes emerge from a topic modeling method, without going further into measuring whether any specific text has a theme about "family" or "politics", for instance.

The subjective inspection, and interpretation, of MEM themes can generally be accomplished in two different ways. First, one would typically look at the component loading table (i.e., the "pattern matrix" for readers familiar with dimension reduction methods) to determine what words statistically "clump together"—that is, which words tend to co-occur above some arbitrary threshold. In this vein, a simple way to understand a given MEM theme is to look at some of the words that most strongly loaded it, as is shown in Table 4 for themes extracted from the "Supervision" topic. A complete table showing high-loading words across each topic can be found in the Appendix A.

Often, reading through lists such as those presented in Table 4 is helpful in identifying common words and concepts that tend to cluster and can additionally be used to identify words that should be removed from further analyses or re-analyses due to a lack of clear meaning or relevance to the subject matter (e.g., words like "thing" or "yes"). In our example, it was easy to remove words that did not appear to contribute in deriving an overall theme. Furthermore, the remaining words appeared to be closely associated with the prompts which in turn helped to generate the themes. For instance, for prompts under Delivery, words clustered in a way three distinctly different themes emerged (i.e., awareness of difficulties, measuring change, and details of care). Words such as "anxiety", "understand", and "emotion" clustered together, forming a theme of awareness of difficulties and "measure", "outcome", and "improve" clustered in another to form a theme in measuring change. There was one anomaly word for a theme (see the Appendix under Views of BPIs, theme 2, the word "house"), which in such instances researchers would be required to revisit the subset of the data to generate a suitable theme. We hope that this demonstrates that overall the MEM output helps to efficiently elucidate themes.

Table 4. Themes and associated words extracted from responses to the "supervision" topic

Theme Number	Theme Name	Example Words
1	Process - how	Engage, feel, good, person, client, week, lot
2	Content - what	Experience, patient, working, talk, BPI, supervision
3	Wider context - roles, responsibilities	Support, worker, team, work, caseload, hear

Figure 1. Word clouds for each of the three themes extracted from the supervision topic. Each theme is presented as its own word cloud: Process (left), Content (middle), and Wider Context (right). Word size is a reflection of the strength of each term's loading onto a given theme.

A second, increasingly popular way to illustrate MEM themes is via the use of word clouds: a visualization technique that illustrates various words in a topic, with more "important" or "central" words being rendered in larger font sizes. Word clouds have rapidly emerged in the past 10–15 years as a popular method for summarizing the content of a body of language ranging from illness symptoms (Sellars et al., 2018) to student responses as a way to holistically gauge student learning (e.g., DePaolo & Wilkinson, 2014). One of the main strengths of word clouds is their ability to intuitively and effectively convey a holistic sense of a topic's content, particularly to non-experts who may prefer them to simply reading word lists and loading values to try to decode a theme's meaning and importance (see Kuo et al., 2007).

Today, there are several packages that can be used to create word clouds, ranging from easy-to-use websites (e.g., https://www.jasondavies.com/wordcloud) to powerful, advanced packages in programming languages like Python.[2] For the current research, we used the *wordcloud* package for the R scripting language (Fellows, 2018). Most commonly, word cloud packages use word frequency as the default property to determine the scale of each word in the cloud; such an approach makes sense when trying to visualize word frequencies in a corpus of text to get a general understanding of common versus uncommon words. In our case, however, word size was scaled as a function of component loadings such that larger words are those that showed a stronger relationship with that theme. Here, we highlight the three themes extracted from the *Supervision* topic (see Figure 1); a complete collection of word clouds is available in the online supplementary materials.

Complementation of qualitative analysis

The results from the present study illustrate that computerized methods can give insights into aspects of interviewees' experiences that were not visible or obvious during initial qualitative analyses. In addition to LIWC's measures of positive or negative words, we were able to ascertain interviewees' psychological states of certainty and authenticity in a way that is difficult to do, objectively, through qualitative analyses. As the chief example from our study: putting the findings from LIWC and MEM together, we found that, while interviewees talked about concepts that seem to pertain to the topic of supervision, they did so in a way that was somewhat guarded and less transparent than other topics. These analyses revealed what was not obvious in a qualitative reading of the interview texts: that the interviewees may not have felt as comfortable sharing more socially sensitive views about the supervision process. Previous thematic analysis had not captured the interviewees' reluctance and, importantly, we likely would have continued to miss this important aspect of the study.

Discussion

When considering the multitude of approaches that exist for analyzing data throughout the social sciences, it is critical for research teams to consider—and regularly revisit—the impact of *how* their data is analyzed. This is especially important when considering whether to adopt a more traditional qualitative analytic approach, the text analyses described in the present study, or a combination of the two. The epistemological lens through which a research team interprets the data will matter.

We recommend that researchers consider the inclusion of computerized text analysis methods to help reveal precisely those types of social, cognitive, and affective processes that can be revealed through subtle linguistic cues but are virtually invisible to human coders. Indeed, the meaning and interpretation of *what* an interviewee says may completely change as a function of *how* they say it, psychologically speaking. The additional layer of social information—such as interviewee authenticity or confidence—provides a powerful addition to the analysts' toolkit, helping to ensure a more full or well-rounded perspective on interviewee responses.

Although we have identified additional benefits of text analyses that complemented our initial qualitative analysis, we believe it is important to recognize that additional results may have been gleaned from alternative qualitative approaches that do not rely on text analysis. For example, the results described in the present study are exclusively dependent upon transcribed, text-based data. Qualitative approaches to the study of nonverbal communication, including the interpretation of gestures, facial expressions, body language, tone of voice, spatial positioning, and other forms of nonverbal communication, can be crucial to understanding meaning, social interactions, and/or psychological states. Obviously, text analyses will fall short in capturing what is left unsaid, and as Miles Davis is credited saying, "It's not the notes you play, it's the notes you don't play". Researchers would do well to remember this maxim when considering which method(s) to employ in the study of their phenomena of interest.

Do computerized text analyses converge with qualitative analysis?

In our study, we did indeed find convergence between the automated MEM approach and a subjective reading, reinforcing previous analyses of the texts by providing a more "detached" perspective. Those researchers who approach their work from objective rather than subjective paradigms can view many computerized methods as a supplementary, but powerful, form of data triangulation, building on the assumption that multiple analytic approaches help the researcher to narrow in on the "objective truth". This approach is sometimes described in the literature as "small q" qualitative analysis: the use of qualitative tools and techniques within a positivist, or objective, paradigm (Kidder & Fine, 1987).

Several studies have adopted this strategy by incorporating both qualitative analysis and LIWC. For example, Whitney et al. (2005) used qualitative analysis and computerized text analysis (LIWC) to analyze narratives written by parents as part of a family intervention. The authors noted that the multiple analytic methods produced similar findings, suggesting a degree of validation in the results. Firmin et al. (2017) also combined thematic analysis methods and LIWC to increase confidence in the validity of the thematic analysis codes. The authors remarked on LIWC's objective method of category generation to serve as a validity checker for subjective coding processes like thematic analysis. LIWC's ability to uncover surprising relationships was also noted, further supporting the need for multiple analytic methods. This suggests that perhaps there is a degree of independence which then allows for useful validation and confidence in one's results.

Fewer studies have explored topic modeling and qualitative analyses being used to bolster one another (Nikolenko et al., 2017), however, such studies do exist (e.g., Gregson et al., 2022). Results from our study reinforce the notion that automated analyses can, at a minimum, reassure qualitative scholars of their readings of collections of text and help to alleviate concerns of subjective biases. Importantly, such analyses can also help to give researchers a sense of the relative prominence of several themes and, in some cases, shed light on themes that may have been overlooked or under-evaluated during a qualitative analysis.

What issues arise when using text analyses?

While there have been calls to integrate multiple analytic methods available and draw on the strengths that accompany analytic integration (Johnson & Onwuegbuzie, 2004) assuming that methodological integrity is maintained (Levitt et al., 2017), there is no guarantee that multiple methods will always combine into a meaningful “bigger picture”. While our use-case shows what a fortuitous pairing of methods can look like, we caution that others may encounter difficulties at various stages of analysis, ranging from interpreting highly nuanced or specific measures to integrating knowledge across domains.

At a macro level, it is important for qualitative researchers to consider the ontological and epistemological suppositions underpinning their research approach. The rise of computerized methods in qualitative research has significantly influenced data analysis, providing researchers with tools that increase efficiency and enhance objectivity (O’Kane et al., 2021). However, despite these practical benefits, these methods can fall short of aligning with the deeper philosophical foundations essential to qualitative inquiry, especially if used as a “shortcut” or “quick fix” to analyzing data. Qualitative research, rooted in ontological and epistemological perspectives, traditionally emphasizes understanding human experience through subjective and context-dependent interpretations. Epistemology, which explores the nature of knowledge and its acquisition, is especially crucial in this context as it shapes the researcher’s relationship with the data and the methodological approach taken (Al-Ababneh, 2020).

Notably, constructivist epistemology, commonly employed in qualitative research, stresses the interpretative role of the researcher in constructing meaning through engagement with participants’ perspectives and experiences (Lee, 2012). In contrast, computerized methods tend to focus chiefly on data analysis, often reducing complex, subjective information to coded segments that may lack the in-depth interpretative engagement used in traditional qualitative inquiry. This focus on objectivity and efficiency can be at odds with qualitative research’s emphasis on immersive, human-centered approaches, particularly those that emphasize a constructivist approach. A valid critique is that computer-aided qualitative data analysis software has the potential to lead to an instrumental approach that emphasizes pattern detection over researcher interpretation. The concern is that by adhering too closely to software-driven analyses, researchers may lose sight of the constructivist or social constructionist paradigms essential for qualitative work. Qualitative paradigms often require not only coding and organizing data but actively constructing knowledge that resonates with the unique contexts of both participants and researchers (Sohn, 2017; Vignato et al., 2021).

Constructivist approaches, such as grounded theory (e.g., Charmaz, 2014) emphasize the co-construction of knowledge between researcher and participant, allowing the researcher to interpret data within the social and contextual fabric of each participant’s lived reality. This approach acknowledges the researcher’s subjectivity and places value on understanding complex, nuanced experiences rather than extracting repetitive patterns. By capturing unique, non-repetitive insights, constructivist approaches in qualitative research can promote a richer, more nuanced analysis that aligns more closely with epistemological goals (Soraya et al., 2019). Thus, while computerized methods offer the advantage of systematic data organization, they often lack the philosophical depth inherent in traditional qualitative research. To preserve the integrity of qualitative methods, researchers would benefit from using computerized approaches only after articulating a clear ontological and epistemological framework.

As a general recommendation, our primary suggestion to combat challenges is to build research teams with diverse, but overlapping, ranges of interests and expertise. When doing so, we encourage regular communication among team members regarding both the aims of the analyses and their implementation. This is important both at the outset of the analyses and throughout the work. For example, it can be helpful at the outset of the work for team members to openly discuss their epistemological or paradigmatic approaches to research and analysis. It would be important, for instance, for researchers on the team who value "small q" or objective approaches to qualitative analysis to reconcile their views with researchers on the team who value more traditional or subjective qualitative approaches. Notably, this did not happen for the present analysis, and it may be the case that both the qualitative process and output may have shifted if it had. That said, researchers with differing ontological and epistemological frameworks could still use text-based analyses, and the presentation of results would likely vary depending on the paradigms adopted in a similar way that the results of traditional qualitative interviews are currently reported in the literature.

Moreover, it can be helpful at each phase of data analysis to have the analyst on the team most familiar with the approach explain the findings to the other team members. This can help facilitate dialogue and uncover meanings in the data that had not been previously considered. For example, the interpretation of how the interviewees in this study may have been reluctant to discuss supervision openly given differences in power dynamics between supervisor and supervisee emerged after our team's expert in text analysis was able to break down the findings to the team and solicit feedback. These regular opportunities to discuss data analysis, both at the outset and across the analysis, were crucial to our team's interpretation of the findings, especially for those not well versed in text analytic approaches.

Future directions

Text-based analytic methods, particularly when used in conjunction with more traditional qualitative methods, are still in their infancy, and a variety of applications merit consideration in the future. For example, from a cultural perspective, might some combination of these approaches be more appropriate when interviewing participants in a second language? Future steps could involve determining these tools' adaptability and cross-cultural robustness for data collected across diverse languages and cultural settings. Additionally, future research might apply these methods to longitudinal qualitative datasets to determine how participant insights evolve over time and whether these methods differ in their sensitivity to temporal changes in participant narratives. Given the burgeoning role of artificial intelligence (AI) across all aspects of technology, including throughout academia, it will also be important to consider the role(s) for AI-driven text analysis models, including natural language processing tools like GPT-based models, in conjunction with text analysis methods like LIWC and MEM as the boundaries of qualitative data interpretation continue to expand.

Increasingly, we find ourselves with easy access to tools, methods, and ways of thinking that emerge from perspectives, backgrounds, and disciplines outside of our usual paths of inquiry. In the context of calls for mixed-methods research urge for scholars to stretch outside of their analytic routines and methodological comfort zones, there are fewer reasons for scholars, researchers, and practitioners to silo themselves within traditions that, by their very nature, can only provide partial perspectives on complicated—and fundamental—questions about human thought, feeling, and behavior. Blurring the boundaries between methodologies to create exciting new compositions is not without its challenges. However, such challenges are worthy of conquest. In overcoming challenges of creatively combining methods, there is nearly infinite uncharted space that lives between the gaps of existing pedagogy. Our study is but one illustration of how various methods can shine light on different, important aspects of qualitative data, and we invite others to do the same.

Notes

1. For more information, see Maciag et al. (2023).
2. See, for example, https://github.com/amueller/word_cloud

Author note

Anonymized OSF repository for peer review: https://osf.io/7djpr/?view_only=12cefee5d60b4286b8151f1714bf4cd7

Acknowledgments

The authors would like to acknowledge the general managers of Cambridgeshire and Peterborough NHS Foundation Trust, Kim Masson and Maggie Page, for their permission and support of a service audit that yielded the qualitative data for this article. We are grateful to the psychologists who carried out Brief Psychological Interventions training and supervision: Drs. Nina Brauner, late Lee Brosan, Lizzie Felter, Fiona Glenn, David Haggarty and Lindsay Ridgeon. We are indebted to Dr. Stephanie Casey, Katherine Parkin, and Dr. Nick Oliver. Last but not least, this study would not have been possible without the support workers, peer support workers and assistant psychologists who participated in the study.

Disclosure statement

Dr. Boyd is one of the co-creators of the Linguistic Inquiry and Word Count software.

Data availability statement

Due to the confidential nature of the responses provided by our participants, the raw language data cannot be shared publicly to maintain the privacy and confidentiality agreements made with the participants. However, the Linguistic Inquiry and Word Count (LIWC) scores derived from this data are available upon request. Interested researchers may contact the authors for access to these aggregated scores, subject to compliance with our data sharing policies and ethical guidelines.

ORCID

Ryan L. Boyd http://orcid.org/0000-0002-1876-6050
Nicholas R. Morrison http://orcid.org/0000-0002-7268-1170
Emma Travers-Hill http://orcid.org/0000-0001-7589-4043
Youngsuk Kim http://orcid.org/0000-0001-6481-2532

References

Adams, R. N., & Preiss, J. J. Eds. (1960). *Human organization research: Field relations and techniques*. Published for the Society for Applied Anthropology by the Dorsey Press

Al-Ababneh, M. (2020). Linking ontology, epistemology and research methodology. *Journal of Science & Philosophy*, *8*(1), 75–91. https://eiris.it/ojs/index.php/scienceandphilosophy/article/view/500/732

Boyd, R. L. (2017). Psychological text analysis in the digital humanities. In S. Hai-Jew (Ed.), *Data analytics in digital humanities* (pp. 161–189). Springer International Publishing. https://doi.org/10.1007/978-3-319-54499-1_7

Boyd, R. L. (2020). *BUTTER: Basic Unit-Transposable Text Experimentation Resource*. https://www.butter.tools

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language & Social Psychology*, *40*(1), 21–41. https://doi.org/10.1177/0261927X20967028

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brookes, G., & McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, *21*(1), 3–21. https://doi.org/10.1177/1461445618814032

Casey, D., & Murphy, K. (2009). Issues in using methodological triangulation in research: Dympna Casey and Kathy Murphy explore the advantages and disadvantages of using triangulation. *Nurse Researcher*, *16*(4), 40–55. https://doi.org/10.7748/nr2009.07.16.4.40.c7160

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Sage.

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, *42*(1), 96–132. https://doi.org/10.1016/j.jrp.2007.04.006

Chung, C. K., & Pennebaker, J. W. (2019). Textual analysis. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in social psychology* (pp. 153–173). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780429452925-7

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693. https://doi.org/10.1111/j.0956-7976.2004.00741.x

Currin McCulloch, J., Stanton, A., Boyd, R., Neaves, M., & Jones, B. (2021). Understanding breast cancer survivors' information-seeking behaviours and overall experiences: A comparison of themes derived from social media posts and focus groups. *Psychology & Health*, *36*(7), 810–827. https://doi.org/10.1080/08870446.2020.1792903

Denzin, N. K. (1978). *Sociological methods: A sourcebook* (2d ed.). McGraw-Hill.

DePaolo, C. A., & Wilkinson, K. (2014). Get your head into the clouds: Using word clouds for analyzing qualitative assessment data. *Technology Trends*, *58*(3), 38–44. https://doi.org/10.1007/s11528-014-0750-9

Fellows, I. (2018). *Wordcloud: Word clouds* 2.6. https://CRAN.R-project.org/package=wordcloud

Firmin, R. L., Bonfils, K. A., Luther, L., Minor, K. S., & Salyers, M. P. (2017). Using text-analysis computer software and thematic analysis on the same qualitative data: A case example. *Qualitative Psychology*, *4*(3), 201–210. https://doi.org/10.1037/qup0000050

Fox, A. K., & Royne Stafford, M. B. (2021). Olympians on twitter: A linguistic perspective of the role of authenticity, clout, and expertise in social media advertising. *Journal of Current Issues & Research in Advertising*, *42*(3), 294–309. https://doi.org/10.1080/10641734.2020.1763521

Glaser, B. G., & Strauss, A. L. (1999). *The discovery of grounded theory: Strategies for qualitative research* (1st ed.). Routledge.

Gottschalk, L. A. (1997). The unobtrusive measurement of psychological states and traits. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 117–129). Lawrence Erlbaum.

Gregson, R., Piazza, J., & Boyd, R. L. (2022). 'Against the cult of veganism': Unpacking the social psychology and ideology of anti-vegans. *Appetite*, *178*, 106143. https://doi.org/10.1016/j.appet.2022.106143

Hays, D. G., Wood, C., Dahl, H., & Kirk-Jenkins, A. (2016). Methodological rigor in journal of counseling & development qualitative research articles: A 15-year review. *Journal of Counseling & Development*, *94*(2), 172–183. https://doi.org/10.1002/jcad.12074

Hingmire, S., Chougule, S., Palshikar, G. K., & Chakraborti, S. (2013). Document classification by topic labeling. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 877–880. https://doi.org/10.1145/2484028.2484140

Ireland, M. E., & Mehl, M. R. (2014). Natural language Use as a Marker of personality. In T. M. Holtgraves (Ed.), *The oxford handbook of language and social psychology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.013.034

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. *Content Patterns from Large-Scale Language Data*PreprintPsyArXivhttps://doi.org/10.31234/osf.io/ra5ze. https://doi.org/10.3102/0013189X033007014

Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, *116*(9), 3476–3481. https://doi.org/10.1073/pnas.1811987116

Junker, B. H. (1960). *Field work: An introduction to the social sciences*. University of Chicago Press.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language & Social Psychology*, *33*(2), 125–143. https://doi.org/10.1177/0261927X13502654

Kahn, J. H., Tobin, R. M., Massey, A. E., Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word Count. *The American Journal of Psychology*, *120*(2), 263. https://doi.org/10.2307/20445398

Kahn, R. L., & Cannell, C. F. (1957). *The dynamics of interviewing; theory, technique, and cases*. John Wiley & Sons.

Kalichman, S. C., & Smyth, J. M. (2021). "And you don't like, don't like the way I talk": Authenticity in the language of bruce springsteen. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000402

Kennedy, B., Ashokkumar, A., Boyd, R. L., & Dehghani, M. (2022). Text analysis for psychology: Methods, principles, and practices. In M. Dehghani & R. L. Boyd (Eds.), *The handbook of language analysis in psychology* (pp. 3–62). Guilford Press.

Kidder, L. H., & Fine, M. (1987). Qualitative and quantitative methods: When stories converge. In M. M. Mark (Ed.), *New directions for program evaluation* (pp. 57–75). Jossey-Bass.

Kisely, S., & Kendall, E. (2011). Critically appraising qualitative research: A Guide for clinicians more familiar with quantitative techniques. *Australasian Psychiatry*, *19*(4), 364–367. https://doi.org/10.3109/10398562.2011.562508

Kuo, B. Y. L., Hentrich, T., Good, B. M., & Wilkinson, M. D. (2007). Tag clouds for summarizing web search results. *Proceedings of the 16th international conference on World Wide Web - WWW '07* (p. 1203). ACM Press.

Lee, C. G. (2012). Reconsidering constructivism in qualitative research. *Educational Philosophy and Theory*, *44*(4), 403–412. https://doi.org/10.1111/j.1469-5812.2010.00720.x

Leech, N. L., & Onwuegbuzie, A. J. (2007). An array of qualitative data analysis tools: A call for data analysis triangulation. *School Psychology Quarterly*, *22*(4), 557–584. https://doi.org/10.1037/1045-3830.22.4.557

Levitt, H. M. (2015). Qualitative psychotherapy research: The journey so far and future directions. *Psychotherapy Theory, Research, Practice, Training*, *52*(1), 31–37. https://doi.org/10.1037/a0037076

Levitt, H. M., Motulsky, S. L., Wertz, F. J., Morrow, S. L., & Ponterotto, J. G. (2017). Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. *Qualitative Psychology*, *4*(1), 2–22. https://doi.org/10.1037/qup0000082

Maciag, R., Travers-Hill, E., Morrison, N. R., & Kim, Y. (2023). Paraprofessionals delivering brief psychological interventions: Qualitative accounts of training and supervision. *Counselling and Psychotherapy Research*, *23*(1), 64–73. https://doi.org/10.1002/capr.12566

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Markowitz, D. M. (2020). *The meaning extraction method: An approach to evaluate content patterns from large-scale language data*. PsyArXiv. Preprint https://doi.org/10.31234/osf.io/ra5ze

McCloskey, K., Ramírez-Esparza, N., & Johnson, B. T. (2022). Strange new worlds: Social content in popular star trek fanfiction versus commercial novels. *Psychology of Popular Media*, *11*(2), 152–162. https://doi.org/10.1037/ppm0000395

McLeod, J. (2000). The contribution of qualitative research to evidence-based counselling and psychotherapy. In N. Rowland & S. Goss (Eds.), *Evidence-based counselling and psychological therapies: Research and applications* (pp. 112–126). Routledge.

Monzani, D., Vergani, L., Pizzoli, S. F. M., Marton, G., & Pravettoni, G. (2021). Emotional tone, analytical thinking, and somatosensory processes of a sample of Italian tweets during the first phases of the COVID-19 pandemic: Observational study. *Journal of Medical Internet Research*, *23*(10), e29820. https://doi.org/10.2196/29820

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality & Social Psychology Bulletin*, *29*(5), 665–675. https://doi.org/10.1177/0146167203029005010

Nikolenko, S. I., Koltcov, S., & Koltsova, O. (2017). Topic modelling for qualitative studies. *Journal of Information Science*, *43*(1), 88–102. https://doi.org/10.1177/0165551515617393

O'Kane, P., Smith, A., & Lerman, M. P. (2021). Building transparency and trustworthiness in inductive research through computer-aided qualitative data analysis software. *Organizational Research Methods*, *24*(1), 104–139. https://doi.org/10.1177/1094428119865016

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Pennebaker Conglomerates.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., Beaver, D. I., & Gong, Q. (2014). When small words foretell academic success: The case of college admissions essays. *PLOS ONE*, *9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Renz, S. M., Carrington, J. M., & Badger, T. A. (2018). Two strategies for qualitative content analysis: An intramethod approach to triangulation. *The Qualitative Health Research*, *28*(5), 824–831. https://doi.org/10.1177/1049732317753586

Roberts, K., Travers-Hill, E., Coker, S., Troup, J., Casey, S., Parkin, K., & Kim, Y. (2021). Brief psychological interventions for anxiety and depression in a secondary care adult mental health service: An evaluation. *The Cognitive Behaviour Therapist*, *14*, e29. https://doi.org/10.1017/S1754470X21000258

Saxbe, D. E., Yang, X.-F., Borofsky, L. A., & Immordino-Yang, M. H. (2013). The embodiment of emotion: Language use during the feeling of social emotions predicts cortical somatosensory activity. *Social Cognitive and Affective Neuroscience*, *8*(7), 806–812. https://doi.org/10.1093/scan/nss075

Schwartz, H. A., & Ungar, L. H. (2015). Data-driven content analysis of social Media: A systematic overview of automated methods. *ANNALS of the American Academy of Political and Social Science*, *659*(1), 78–94. https://doi.org/10.1177/0002716215569197

Sellars, B. B., Sherrod, D. R., & Chappel-Aiken, L. (2018). Using word clouds to analyze qualitative data in clinical settings. *Nursing Management*, *49*(10), 51–53. https://doi.org/10.1097/01.NUMA.0000546207.70574.c3

Smith, N. (2018). Relations between self-reported and linguistic monitoring assessments of affective experience in an extreme environment. *Wilderness & Environmental Medicine*, *29*(1), 61–65. https://doi.org/10.1016/j.wem.2017.08.023

Sohn, B. (2017). Phenomenology and qualitative data analysis software (QDAS): A careful reconciliation. *Forum: Qualitative Social Research*, *18*(1). https://doi.org/10.17169/fqs-18.1.2688

Soraya, S., Binti, L., & Kamal, A. (2019). Research paradigm and the philosophical foundations of a qualitative study. *International Journal of Social Sciences*, *4*(3), 1286–1394. https://doi.org/10.20319/pijss.2019.43.13861394

Stone, P. J., & Hunt, E. B. (1963). A computer approach to content analysis: Studies using the General inquirer system. *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference*, 241–256. https://doi.org/10.1145/1461551.1461583

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and Computerized text analysis methods. *Journal of Language & Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Tong, C., F, E., Wibowo, K. A., Makki, T. W., & Slatcher, R. B. (2020). Self-presentation and impressions of personality through text-based online dating profiles: A lens model analysis. *New Media & Society*, *22*(5), 875–895. https://doi.org/10.1177/1461444819872678

Vignato, J., Inman, M., Patsais, M., & Conley, V. (2021). Computer-assisted qualitative data analysis software, phenomenology, and Colaizzi's method. *Western Journal of Nursing Research*, *44*(12), 1117–1123. https://doi.org/10.1177/01939459211030335

Whitney, J., Murray, J., Gavan, K., Todd, G., Whitaker, W., & Treasure, J. (2005). Experience of caring for someone with anorexia nervosa: Qualitative study. *British Journal of Psychiatry*, *187*(5), 444–449. https://doi.org/10.1192/bjp.187.5.444

Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *The Qualitative Health Research*, *11*(4), 522–537. https://doi.org/10.1177/104973201129119299

Wright, I., Travers-Hill, E., Gracey, F., Troup, J., Parkin, K., Casey, S., & Kim, Y. (2020). Brief psychological intervention for distress tolerance in an adult secondary care community mental health service: An evaluation. *The Cognitive Behaviour Therapist*, *13*, e50. https://doi.org/10.1017/S1754470X20000513

Appendix A

High-loading Words for Meaning Extraction Method, All Topics

Prompt	Theme Number	Theme Name	Example Words
Delivery	1	Awareness of difficulties	Anxiety, understand, emotion, reablement, knowledge, people
Delivery	2	Measuring change	Measure, outcome, improve, fill, change, give
Delivery	3	Details of care	Day, disorder, hour, psychosis, client, year
Formal Training	1	Context—intervention within the team	Referral, refer, intervention, team, difficult, support
Formal Training	2	Self and learning tasks	Word, worker, str, read, felt, manual
Formal Training	3	Interpersonal reception	Nice, day, people, oc, helpful, person
Manuals	1	Case formulation language	Hierarchy, create, messy, cycle, anxiety, change
Manuals	2	Reflection	Learn, calm, bring, resource, feel, breathe
Manuals	3	Emotions	feeling, give, messy, normal, distress
Supervision	1	Therapeutic alliance	Engage, feel, good, person, client, week, clinical
Supervision	2	Working with patients	Experience, patient, working, talk, bpi, supervision
Supervision	3	Wider context of supervision	Support, worker, team, work, caseload, hear
Views of BPI	1	Who is who of BPI	Member, psychologist, intervention, deliver, term, staff
Views of BPI	2	Positive qualities of BPI	Easy, end, week, session, house, good
Views of BPI	3	Wider context of BPI	Team, feel, support, deliver, bpi, end