

TECHNICAL REPORT

Visualizing the qualitative: making sense of written comments from an evaluative satisfaction survey

Keith V. Bletzer*

National Community Health Partners, Tucson, AZ; School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA

Abstract

Purpose: Satisfaction surveys are common in the field of health education, as a means of assisting organizations to improve the appropriateness of training materials and the effectiveness of facilitation-presentation. Data can be qualitative of which analysis often become specialized. This technical article aims to reveal whether qualitative survey results can be visualized by presenting them as a Word Cloud. **Methods:** Qualitative materials in the form of written comments on an agency-specific satisfaction survey were coded and quantified. The resulting quantitative data were used to convert comments into “input terms” to generate Word Clouds to increase comprehension and accessibility through visualization of the written responses. **Results:** A three-tier display incorporated a Word Cloud at the top, followed by the corresponding frequency table, and a textual summary of the qualitative data represented by the Word Cloud imagery. This mixed format adheres to recognition that people vary in what format is most effective for assimilating new information. **Conclusion:** The combination of visual representation through Word Clouds complemented by quantified qualitative materials is one means of increasing comprehensibility for a range of stakeholders, who might not be familiar with numerical tables or statistical analyses.

Key Words: *Capacity building assistance; Qualitative analysis; Satisfaction survey; Word Clouds*

INTRODUCTION

Satisfaction surveys are used to evaluate the effectiveness of services from the viewpoint of recipients and consumers. Survey protocols can be administered at intervals to large-scale groups in commercial settings or one-time to recipients at the conclusion of training [1], educational services [2], or behavioral interventions [3], among other activities. Quantitative calculation is the common means of analysis, resulting in evaluative materials that can be applied to improving services delivery. Qualitative satisfaction survey data in contrast are included less often in monitoring and evaluation analyses. This article aims to quantify the written comments from an agen-

cy-specific satisfaction survey and to examine their conversion to the imagery of Word Clouds through the use of frequency counts.

METHODS

Data collection

Data were collected by a community agency, National Community Health Partners, funded by the Centers for Disease Control and Prevention (CDC) to provide Capacity Building Assistance (CBA) to AIDS service organizations, and state/local health departments, across the United States and its territories. Programs follow the model of “prevention with positives” that emphasizes active enrollment and retention in HIV care and services with the goal of undetectable viral loads [4]. Training participants included medical personnel and auxiliary staff, such as patient navigators, retention case managers, clinical nurses, intervention specialists, HIV testers-counsel-

*Corresponding email: keith.bletzer@asu.edu

Received: September 1, 2014; Accepted: April 12, 2015;

Published: April 16, 2015

This article is available from: <http://jeehp.org/>

ors, and behavioral therapists, among other providers. Although data on status were not collected, it is known that some participants may be, and some might not be, HIV-positive.

Quantitative and qualitative data were collected through a satisfaction survey administered by trainers at each CBA event that they organized and delivered as on-site training. Evaluation focused on CBA delivery, based on participant responses to fixed-choice questions framed by a five-point scale. Beyond monitoring and evaluation of training events, the evaluator collected additional program data by on-site training observations, post-training focus group interviews (teleconference), and online quarterly follow-up survey. Focusing on qualitative data from the satisfaction survey, this analysis covers the final year of a five-year grant cycle. As the agency was funded for another five years to provide CBA to community organizations, the use of past tense reflects the completed project.

To improve evaluative rigor and complement quantitative satisfaction data, three open questions [5] asked participants what was most and least effective in the training, and what they would recommend to improve the learning experience. Question 16 (Q16) and Question 17 (Q17) focused on training elements that were considered by participants as *most* and *least* effective, respectively:

Q16: What were the MOST effective parts of this CBA event for you?

Q17: What were the LEAST effective parts of this CBA event for you?

These two questions plus Question 18 (Q18) on recommendations to improve training requested comments, after participants had completed a dichotomous agreement-disagreement checklist of eleven questions on Content and Facilitation, and for certain trainings, another four questions on Skills-Building for a total of fifteen. Collecting written comments immediately after training assured no intervening events to interfere with recalling a memorable learning experience that was perceived as appreciated, ineffective, most/least liked, different from similar encounters, and other reactions, before being filtered from immediate memory soon thereafter [6].

The five-point scale permitted calculation of mean and median scores, standard deviation, and analysis of variance, among possible metrics. Both statistical analysis and qualitative comments were used to identify what went well and particular aspects that might require review and revision. For presentation to stakeholders, written comments entered as textual data in SPSS (version 21) were sorted for coding and quantification, and made ready to prepare “input terms” to create Word Clouds.

As there were more than a thousand participants each of the final four years, this analysis is limited to the fifth and final year of the grant cycle. This circumscribes satisfaction data to the same group of trainers, guided by prevention-with-posi-

tives funding priorities initiated the fourth year, managed under the tenure of the same project director. In short, the analyzed data were collected under similar conditions.

Data coding

The first task was identifying which comments fit each question. That is, were respective comments on the most and least effective parts provided for Q16 and Q17? If not, they were coded by type and/or intent of comment, such as reflection on future utilization with clients and community members, and those unrelated to content or facilitation, e.g., distance to the training or someone “new” to the topic. This review of the qualitative data provided a perspective into participant language that facilitated code generation beyond what was “most” and “least” effective in the training.

The next step was coding. Comments varied. Most named one element such as “Interaction among participants” and “Hands-on activities were most effective.” Several persons wrote paired comments, such as “Role-play and discussion” and “Discussion with facilitator and activities,” where pairing might reverse order of the same elements. A few participants identified three elements, succinctly, “Writing, collaboration, and teamwork,” or elaborately, “Round-Table discussion – Examining myself [refers to self-reflection activity] – How one can handle stigma language.” Hyphens and capitalization were conventions several respondents used to differentiate one training element from another. For all coding, each element was listed in order of appearance.

Coded comments were reviewed by sorting variables in SPSS. This step facilitated a cross-check on coding accuracy by alphabetizing common phrases, such as “All was...” (36 times in total), variation across terms such as “role playing” (95 times in Q16; 19 times in Q17), processes such as “learning” (96 times in Q16 but not a term in Q17), activities by “group” (66 times in Q16; three times in Q17). Recommendations to improve training (Q18) were few in number and often began with “Needs more...” (33 times) or “Too long/brief” (28 times).

A few respondents inverted the intent of Q16 by responding critically on one element or the entire training, as one person who wrote, “This one did not ‘click’ for me,” with no comments on Q17 or Q18. More respondents, however, inverted the intent of Q17 on what was least effective, such as “Nothing came to my mind,” “None, it was all great,” “I enjoyed and learned from every part,” “I felt all parts were effectively implemented.” Those inverting Q17 on what was least effective more often responded to the contrastive question on what-was-most-effective (previous question), in contrast to Q16 inverters that infrequently wrote comments for what-was-least-effective (next question).

Many respondents utilized common terminology in express-

ing their concerns. Some used terms that varied slightly (plural or singular), such as “role play” and “role plays,” sought precision in their descriptions, such as “small-group” vs. “group,” and/or favored stylistic variations. These differences however slight as writing are important, as they determine the outcome of inserting input terms to create a Word Cloud. Stylistic variations were too infrequent to be captured by technology that identifies and sorts common words for proportional graphic representation as a Word Cloud.

Technical Information

“Double entries” [7] were not relevant in this analysis. Each respondent had a unique alpha-numeric identifier within a training encounter. Should someone attend more than one CBA event, each experience differed with respect to content. No more than 148 persons among 1,286 participants were duplicated across training during the project’s fifth year; in addition, ten persons attended three, and five attended four training events (12.7% total repeaters).

Language variability required adjustment to prepare input terms for Word Cloud technology. Respondents in Q16 and Q17 varied most in identifying “facilitator,” using close synonyms such as “trainer,” or related term, “instructor,” or possessive (e.g., trainer’s skills), and in some instances, a personal name. No more than one of 143 persons referred critically to a facilitator in Q16, compared to eight of 14 in Q17. Other common terms included “activities,” “learning,” “role-play,” “discussion,” and “interactive,” usually within favorable comments in Q16; “activities” and “role playing” were common in Q17 with less than half in critical comments. Frequent phrases in Q17 referred to dimensions of learning, such as “all was okay” and “everything was effective” (alternative terms: useful, beneficial, and informative).

Presentation conventions in this analysis included lower case in Word Clouds for affective-supportive, and capitalization for critical-negative, comments. Too few of the latter in Q16 meant that none appeared in the corresponding Word Cloud. Some appeared in Q17 in proportions less than that for comments appreciative of the training. As one person wrote, “I can’t imagine any of it out” (Q17). Proportional appearance in each Word Cloud highlighted differences in the frequency of responses to the two qualitatively dichotomous questions. Hyphens hold compound terms together in a Word Cloud (e.g., “small-groups”). The ~ symbol between terms performs the same action without a trace (e.g., “need more time”). Non-response codes were removed from input text. The code “xxx” (no response), e.g., is treated by Word Cloud technology as a word; to have it not appear, these cells were left blank.

Statistical analysis

Descriptive statistics were utilized. Coding multiple elements became cumbersome in frequency tables. Initially separated, singular elements were merged with combination statements to generate a list of elements that caught participant attention, tallied by a frequency count of identical elements.

RESULTS

Some 1,096 responses were offered by 1,286 participants at training delivered Year 5 (Table 1). Two-hundred-twenty-eight persons left blank Q16 (eight were “N/A”); 692 left blank Q17 (238 were “N/A,” 15 were “don’t know”); 190 skipped both questions. By far, more respondents in Q16 expressed appreciation and/or offered an affective-supportive comment (654/1,286: 50.9%) and identified training strengths (371/1,286: 28.8%), than those in Q17 providing criticism (155/1,286: 12.1%) and/or commenting negatively (71/1,286: 5.5%). A larger proportion (270/1,286: 21.0%) inverted the intent of Q17 that asked them to identify what was least effective than those inverting what-was-most-effective in Q16 with criticism (10/1,286: 0.8%). Thus, comments for what-was-most-effective were more frequent in Q16 than training elements reported as least effective in Q17. Altogether, four-fifths of respondents offered comments to Q16 on what-was-most-effective in the training they had attended (1,035/1,286: 80.5%) vs. less than half that commented in Q17 on what they felt was least effective (560/1,286: 43.5%). Extraneous responses were minimal (23 in Q16; 34 in Q17, or 1.8% and 2.6%, respectively).

The rationale for constructing each Word Cloud was to illustrate written comments by participants. Alterations follow-

Table 1. Written responses from satisfaction survey, coded by theme and intent

Theme of response	Question 16		Question 17	
	No.	%	No.	%
General comments				
Strength (STR)	371	28.8	None	
Negative (NEG)	None		71	5.5
Content-specific	654	50.9	155	12.1
Inverts the question	10	0.8	270	21.0
Critical suggestion (CS)	None		64	5.0
Extraneous responses				
Arrangements (site)	None		16	1.2
Review/new material	23	1.8	18	1.4
Non-responsive				
No response (blank)	219	17.0	439	34.1
"Don't know"	None		15	1.2
"N/A"	9	0.7	238	18.5
Total	1,286	100	1,286	100

Table 2. Twenty terms of high frequency from Q16 with corresponding term conversions

QUESTION 16	Word Cloud	ORIG terms	Added to ORIG	Close synonym	Related terms	Other reason	Critical comment
Facilitator	143	77	+ 66	6	9	52	-1
Activities	131	118	+ 13	3	10	0	0
Learning	96	80	+ 16	0	16	0	0
Role-play	95	6	+ 89	0	87	2	0
Discussion	93	93	+ 0	0	0	0	0
Interactive	70	73	+ -3	0	0	0	-3
"MI" [training]	67	13	+ 54	0	52	2	0
Small-groups	66	4	+ 62	3	59	0	0
Effective	65	55	+ 10	0	3	7	0
Client	57	57	+ 0	0	0	0	0
Information	57	32	+ 25	6	19	0	0
Knowledge	54	56	+ -2	0	0	0	-2
Training	53	55	+ -2	0	0	0	-2
Everything	42	29	+ 13	0	13	0	0
Good	41	41	+ 0	0	0	0	0
Skills	39	39	+ 0	0	0	0	0
Exercises	38	2	+ 36	0	36	0	0
Different	36	36	+ 0	0	0	0	0
Helpful	31	31	+ 0	0	0	0	0
Group discussion	24	19	+ 5	0	5	0	0

Most terms were constructive with very few critical comments. Original terms (ORIG) were converted by adding close synonyms, related terms, and other reasons, whereas critical comments, which were less common, were subtracted. Row 7 with "MI" (Motivational Interviewing) represents a popular training, sometimes capitalized, sometimes written with other terms such as "motivation," hence the need to render each of these the same as Word Cloud input terms

ed the direction of each question (enumerated above). To prepare Word Cloud input terms certain words were combined in Q16, such as "role play" (singular and plural), "roll play" (misspelled) and "role playing" into *role-play*; "group work" into *small-groups*, and "exercises" and "break-out" into *activities*, among others. Each trainer was mentioned at least once by name (naming someone typically indicates appreciation for that person); these nominalizations with the synonyms, "instructor" and "trainer," were converted into the term, *facilitator*. Words were similarly combined in Q17. The altered appearance of certain synonyms and misspelled words assured appearance validity in the Word Cloud. Thus, proportion of terms within a Word Cloud occasionally approximated rather than replicated numerical counts of respondent terms. Nonetheless, they followed respondent intent in what they wrote. In a narrative story, repetition adds emphasis to an intense experience, signals "sequential fit" to previous commentary or "repair" within conversational turn-taking, and assists in "sourcing" effective components (Table 2).

Respondents simplified responses in Q17 (148 times) usually to "none" and "nothing," or rare comments, such as "Can't think of any," to invert the question intent to collect information on what was least effective. Less frequent terms, "no" and "nope," were merged with "none" and "nothing," respectively.

Truncated statements, such as "everything was effective" and "all was good" ("okay," "great") were similarly merged. To generate phrases of similar length, 'everything' and 'all' were switched to "everything was okay" and "all was effective" (Table 3).

DISCUSSION

Word Clouds are dependent on holistic perception. They reinforce word recognition with the mathematical principle of proportions. Laura Ahearn [8], e.g., used Word Clouds to analyze thematic output in a professional journal. She collated keywords from *American Ethnologist* and converted these to frequency tables accompanied by Word Clouds.

Describing her process for creating Word Clouds, Ahearn noted the need to disaggregate phrases into core terms that represent author-chosen keywords. Based on two years of articles from each of the four past decades, she found that keywords were repeated no more than seven times over eight years, except for the descriptive terms, "anthropology" 36 times, and "social" 31 times. The frequency of common keywords diminished from early to recent decades [8].

Training elements mentioned in quantitative questions on the satisfaction survey had minimal influence on qualitative responses. Two were specific phrases, "hands-on activities"

Table 3. Nineteen terms of high frequency from Q17 with corresponding term conversions

QUESTION 17	Word Cloud	ORIG terms	Total altered	Close synonym	Related terms	Other reason	Critical comment
none	112	64	+ 48	0	48	0	0
effective	62	53	+ 9	9	0	0	0
everything okay	43	23	+ 20	0	20	0	0
all was effective	36	22	+ 9	0	14	0	-5
nothing	36	23	+ 13	11	2	0	0
time	35	46	+ -11	0	0	0	-11
training	25	38	+ -13	0	0	0	-13
information	22	20	+ 2	0	2	0	0
need more time	19	1	+ 18	0	18	0	0
facilitator	14	6	+ -2	4	2	0	-8
relevant	6	6	+ 0	0	0	0	0
exercises	4	5	+ -1	0	0	0	-1
Too-Long	28	9	+ 19	0	19	0	0
Activities	20	22	+ -2	0	3	0	-5
Role playing	19	8	+ 11	15	0	0	-4
Presentation	10	11	+ -1	0	-2	0	3
Discussion	10	16	+ -6	0	2	0	-8
Power-point	7	10	+ -3	0	1	0	-4
Few-examples	8	2	+ 6	0	4	0	-2

Conversion procedures were applied when a comment identified *least effective* (capitalized, bottom rows) or inverted the question when someone did not feel that anything was *least effective* (small case, top rows). Three columns (Close Synonym; Related Terms; Other Reason) were affective-supportive comments added to Original (ORIG) and Critical Comment was negative (thus, subtracted). Time comments usually mentioned “Too-Long” or prolonged breaks/lunch.

(Q12) and “questions and answers” (Q3); three were general, “materials” (Q4), “content” (Q5), “skills” (Q13-Q14-Q15); and one was global, namely, “CBA event” (ten questions). A few terms that preceded qualitative questions re-appeared in the qualitative comments, such as “activities,” (140 times), “skills” (39 times), “hands on” (11 times), “materials” (9 times), whereas others seldom were re-used, such as “content” (twice) and “questions and answers” (once). Participants in their written comments on what was most and least effective went beyond this vocabulary by repeatedly using thirteen original terms in responding to Q16 and twelve to respond to Q17. Thus, respondents demonstrated knowledge of training elements through those they repeated and those appearing as original language in their written comments.

Training element frequencies were used to prepare input terms to generate Word Clouds. Based on review of the qualitative data, close synonyms and terms of related meaning were replaced with more common root terms having the same meaning. Otherwise, synonyms would have diluted the visual impact of frequently-mentioned training elements in the Word Cloud. As Jonathon Friedman, who created the technology for Word Clouds, explains, “The size of the word in the visualization is proportional to the number of times the word appears in the input text” (<http://Wordle.com>). The program clusters each word according to frequency. Capitalized and lower case

words and those spelled correctly and incorrectly, appear separately.

Generated patterns are automatic in Word Clouds at the discretion of the online program. They are guided not manipulated. Choice of script and letter size, color schemes, number of words, and placement (horizontal, vertical, in-between) are what generate visible para-textual variation in pictorial Word Clouds [8]. Rethinking the reason for the Word Cloud as a visual display of responses led to placing the label at the top, as the question answered by the Word Cloud (Appendix). Participants identified effective elements in Q16, and for Q17 the words left no doubt that they appreciated the training and were responding figuratively against the notion of “least effective”.

A data table below the Word Cloud followed by summary text adds second and third components. The number table with each Word Cloud centers the satisfaction survey as a data-gathering tool to remind the viewer that the Word Cloud displays quantitatively converted qualitative data. Visible terms in the Word Cloud above a data table initially draw viewer attention, as ordered space above/below catches attention quicker than left/right or front/back [9]. Making the survey words “visual” [10] reveals overall participant intent.

Abbreviated text at the bottom summarizes the table data, whereas visual imagery on top receives prominent consideration. Comments that summarize and/or clarify aspects of the

Word Cloud (e.g., totals and sub-totals) are embedded in a horizontal-vertical table of rows in two columns followed by an across-the-page textual summary. Word Cloud multiplicity, then, eases the task of accessibility of meaning in going from top to middle to bottom. The three-tier display format facilitates cognitive processing of the qualitative participant responses.

The project described herein, similar to that reported by Laura Ahearn, generated an individualistic display of visual multiplicity [8]. A Word Cloud brings us closer to the truth of what participants wished to communicate versus the reality of frequency counts. When responding to qualitative question-prompts, training participants who receive CBA services are likely to produce a multiplicity of written comments that parallels the authors in Ahearn's analysis of keywords published in professional articles from *American Ethnologist*. As social beings presented with opportunity for self-expression, cognizant of an indeterminate audience, people seek to engage others by communication with novel meanings and aesthetic forms [11]. An intensive training experience motivates learners to move words and ideas around in their heads, gain from the experience, and later place them together in meaningful comments. I confirmed the efforts of training participants by arranging their qualitative comments into Word Clouds, and through coding and counting the terms they used to verify that I constructed a reasonable approximation that is true to their intentions.

ORCID: Keith V. Bletzer: <http://orcid.org/0000-0002-1782-6074>

CONFLICT OF INTEREST

No potential conflict of interest relevant to the study was reported.

ACKNOWLEDGMENTS

I thank Ann Verdine, who was project director at the time the analyzed data were collected; Jie Pu who was my evaluation mentor; and Albert Moreno for his leadership as CEO of National Community Health Partners that managed the capacity building assistance project described in the article. Funding for the capacity building assistance projects was provided by two successive grants from the United States Centers for Disease Control and Prevention (PS09-906, funded 2009-2014,

and PS14-1403, funded 2014-2018), awarded to National Community Health Partners (Tucson, AZ).

REFERENCES

1. Evenson ME. Preparing for fieldwork: students' perceptions of their readiness to provide evidence-based practice. *Work*. 2013; 44:297-306. <http://dx.doi.org/10.3233/WOR-121506>
2. Kirdar Y. The role of public relations for image creating in health services: a sample patient satisfaction survey. *Health Mark Q*. 2007; 24:3-4,33-53. <http://dx.doi.org/10.1080/07359680802119017>
3. Bjørknes R, Manger T. Can parent training alter parent practice and reduce conduct problems in ethnic minority children? A randomized controlled trial. *Prev Sci*. 2013;14:52-63. <http://dx.doi.org/10.1007/s11121-012-0299-9>
4. Sigaloff KCE, Lange JMA, Montaner J. Global response to HIV: treatment as prevention, or treatment for treatment? *Clin Infect Dis*. 2014;59: S7-S11. <http://dx.doi.org/10.1093/cid/ciu267>
5. Hruschka DJ, Cummings B, St John DC, Moore J, Khumalo-Sakutukwa G, Carey JW. Fixed-choice and open-ended response formats: a comparison from HIV prevention research in Zimbabwe. *Field Methods*. 2004;16:184-202. <http://dx.doi.org/10.1177/1525822X03262663>
6. Baddeley A. 2006. Working memory: an overview. In: Pickering SJ, editor. *Working memory and education*. Amsterdam, Netherlands: Elsevier; 2006. p.1-31.
7. Bauermeister JA, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data quality in web-based HIV/AIDS research: handling invalid and suspicious. *Field Methods*. 2012; 24:272-291. <http://dx.doi.org/10.1177/1525822X12443097>
8. Ahearn LM. Detecting research patterns and paratextual features in AE word clouds, keywords, and titles. *Am Ethnol*. 2014;41:17-30. <http://dx.doi.org/10.1111/amet.12056>
9. She HC, Chen YZ. The impact of multi-media effect on science learning: evidence from eye movements. *Comp Educ*. 2009;53: 1297-1307. <http://dx.doi.org/10.1016/j.compedu.2009.06.012>
10. Ozcelik E, Arslan-Ari I, Cagiltay K. Why does signaling enhance multimedia learning? Evidence from eye movements. *Comp Hum Behav*. 2010;26:110-117. <http://dx.doi.org/10.1016/j.chb.2009.09.001>
11. Nakassis CV. Brand, citationality, performativity. *Am Anthropol*. 2012;114:624-638. <http://dx.doi.org/10.1111/j.1548-1433.2012.01511.x>

