"NovoPsych helps mental health services use psychometric science to improve client outcomes"

The Client Satisfaction Survey. Measuring Patient's Experiences in Private Psychology Practices.

Ben Buchanan[1 2] and David Hegarty[1]

2023

[1] NovoPsych Pty Ltd, Australia

[2] Monash University, Faculty of Medicine, School of Psychological Sciences

Corresponding author email: david@novopsych.com

Abstract

Objective: Measuring client satisfaction among users of mental health services is of critical importance to quality assurance and service improvement, yet few private outpatient clinics collect such data. This research sought to characterise the psychometric qualities of the Client Satisfaction Survey (CSS).
Method: Using 5,336 responses from private psychology clinics in Australia, a sentiment analysis and an exploratory factor analysis (EFA) was performed on the responses from the CSS.
Results: Response sentiment was overwhelmingly positive with trust being the most associated emotions. Four factors of client satisfaction were found as a result of the EFA: (1) practitioner, (2) reception staff, (3), scheduling, (4), quality of life improvements. Responses were highly skewed, with 43% of respondents giving the maximum satisfaction rating, indicating ceiling effects and/or a high level of satisfaction among respondents. Normative data was computed, giving clinicians valuable context for interpretation of the CSS. It was concluded that 88% of respondents were satisfied with the service they received, defined as an average score of "agree" or higher on the CSS.
Conclusions: The client experiences in private psychology practices in Australia are overwhelmingly positive. The CSS, which measures four factors of the client satisfaction experience, is an assessment that can be used by private practitioners to measure client satisfaction. Methods for interpreting collated satisfaction data at a practice level were discussed in light of the benchmarks provided.

Key Words: Client satisfaction; Independent private practice; psychology.

MANUSCRIPT DATE: 23rd August 2023

In mental health and psychology clinics, understanding and measuring client satisfaction with services received can help service providers to improve the quality of care and tailor their approach to better meet the needs and preferences of their clients. Despite this, few practices systematically measure client satisfaction. Client satisfaction is defined as the fulfilment of expectations and gratification experienced by the client/patient in relation to the service as a whole, including with their practitioner (e.g. therapeutic alliance) and the practice's administrative processes.

Research has shown that client satisfaction is associated with a range of positive treatment outcomes, including improved symptom reduction, increased functioning, and decreased dropout rates (Gibbons et al., 2013; Hsieh & Wang, 2017). A positive therapeutic alliance, characterised by a strong, collaborative relationship between therapist and client, is a key factor in promoting client satisfaction (Flückiger et al., 2018).

However, measuring client satisfaction can be challenging. Some clients may be reluctant to express dissatisfaction and therefore engage in socially desirable responding. As a result, many Patient Reported Experience Measures (PREMS) or client satisfaction scales suffer from significant ceiling effects, where scores cluster toward the top of the range for the total scale score (Meier, 2022; Meier & Feeley, 2021; Paap et al., 2019).

Existing PREMS like the Your Experience of Service Survey (2015) are designed for public mental health services and not necessarily appropriate for private outpatient mental health services such as psychology or psychiatry clinics.

Despite these challenges, the importance of measuring client satisfaction remains valuable for practitioners and practice managers for quality assurance and in an effort to continually improve the service based on feedback. Understanding and addressing clients' satisfaction is essential for building and maintaining a strong therapeutic relationship and promoting positive treatment outcomes. To assist in this endeavour, NovoPsych has developed the Client Satisfaction Survey (CSS) which has been in use since 2014. The aim of this study is to:

1. Characterise the typical level of satisfaction among private practice psychology practices in Australia, and creating normative data.
2. Assess the psychometric properties of the CSS using a factor analysis.

Method

Procedure. Data for the Client Satisfaction Survey (CSS) in NovoPsych, an online psychometrics platform, where users had given permission for anonymous data to be used for benchmarking purposes, was downloaded. After initial data tidying, exploratory data analysis was performed to determine the typical level of satisfaction for clients. Sentiment analysis was then completed on the two open-text responses of the CSS to determine the pervasive opinions of clients towards the clinician and/or their practice. Subsequently, an exploratory factor analysis (EFA) was performed to determine the factor structure of the first eight questions of the CSS. The EFA was completed to see if there were any applicable factors that may enhance the interpretation of the CSS for clinicians. Whether there were any relevant factors or not, a norming process was then completed to gather percentiles for available data so that NovoPsych users could more easily interpret their own results on the CSS in comparison to a normative group.

Participants. There were 5,710 individual responses in NovoPsych for the CSS that were gathered between 24 June 2014 and 9 June 2023. Data was tidied, whereby dummy data entered by clinicians was removed (where a computer script deleted client names such as “Test”, “Dummy”, “Generic”, which are automatically generated dummy names within the NovoPsych platform.). After data tidying, there were 5,336 responses from 637 different practices. The top 5 practices with most responses had 1,556 (29.1%) of the responses and all of these practices were located in Australia. Responses were from 26 different countries, but 5,087 (95.3%) of all responses were from Australia with 5,243 (98.3%) responses from Australia, United Kingdom, or New Zealand.

For one of the practices managed by co-author Ben Buchanan where feedback protocols and data was available, the response rate was analysed. This private practice with 13 psychologists located in Melbourne, Victoria, sent the CSS to clients over age 18 via email after the client had not had an appointment for 60 days. This email was emailed by reception on 653 occasions between Feb 2021 and Feb 2023, with 131 responses, making the response rate 20%.

Measure. The Client Satisfaction Survey (CSS) is a 10-item NovoPsych developed scale to measure overall client satisfaction. The CSS measures client satisfaction using eight Likert scale questions (1 of which is optional) and two open-ended questions. Questions 1 to 8 vary in their focus, with practitioner preparation (question 1), understanding of concerns (question 2), providing sufficient information (question 3), improved quality of life (question

4), scheduling availability (question 5), consistent information (question 6), friendly reception staff (question 7 - optional), and general satisfaction (question 8) measured. The open-ended questions focus upon other areas of strength (question 9) and suggestions for improvement (question 10). Questions 1 to 8 can be summed to provide a total score, with a maximum score of 40 (if all items are completed) or a maximum score of 35 (if question 7, which is optional, is left blank). The higher the score, the more a client is satisfied with the therapeutic experience.

Results

Data analysis was performed in RStudio (Version 2022.07.2+576; RStudio Team, 2022) using R (Version 4.2.0; R Core Team, 2022). After data tidying, data was separated into three sets: (i) all the data combined (n = 5,336); (ii) those who had completed all questions (n = 4,954), and; (iii) those who had left the optional question 7 blank (n = 382).

Exploratory Data Analysis. Consistent with many other similar measures (Meier, 2022; Meier & Feeley, 2021; Paap et al., 2019), as can be seen in the density plots in Figures 1 and 2, the data was significantly skewed to the left / had a negative skew (a ceiling effect).

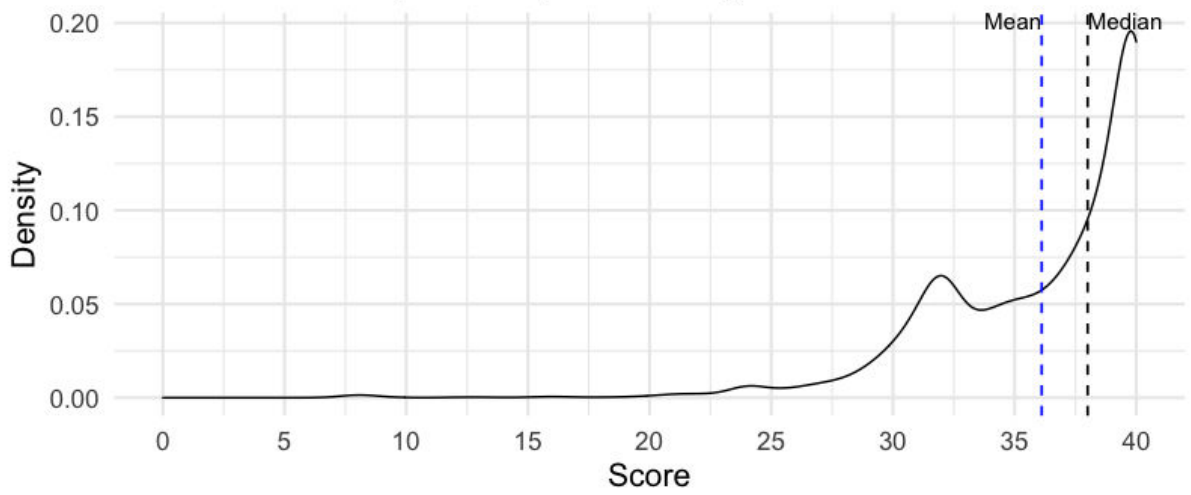

Figure 1. Density plot for total score for participants who completed all questions (1-8). The maximum total score was 40. N = 4,350.

Between the group of participants who did complete question 7 (set ii) and the group who left question 7 blank (set iii), there was no difference on the total score for the other questions (out of a maximum score of 35), Welch's t(449.9) = 0.28, p = 0.78, d = 0.01, $CI_{95\%}$[-0.09, 0.12].

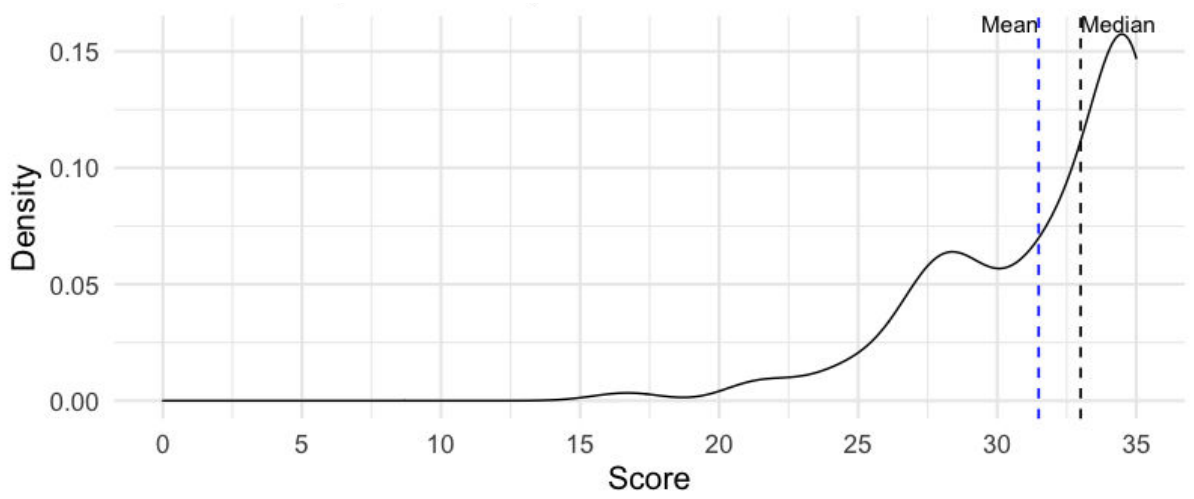

Figure 2. Density plot for total score for participants who left optional question 7 blank. The maximum total score was 35. N = 294.

Sentiment Analysis. An analysis of the open-text questions in the CSS (questions 9 & 10) was performed using the syuzhet package (Version 1.0.6; Jockers, 2015). The sentiment analysis matches client responses to the NRC emotion lexicon (Mohammad & Turney, 2013), which results in the responses' associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). Question 9 was assessing "Other areas of strength" for the practitioner / practice and it would be expected that these would be mostly positive aspects that would be outlined. There were 1,106 text responses to this optional question. As seen in Figure 3, the most associated emotion was 'trust' (35.6%) and the sentiment was overwhelmingly positive (88.9%; see Figure 4).

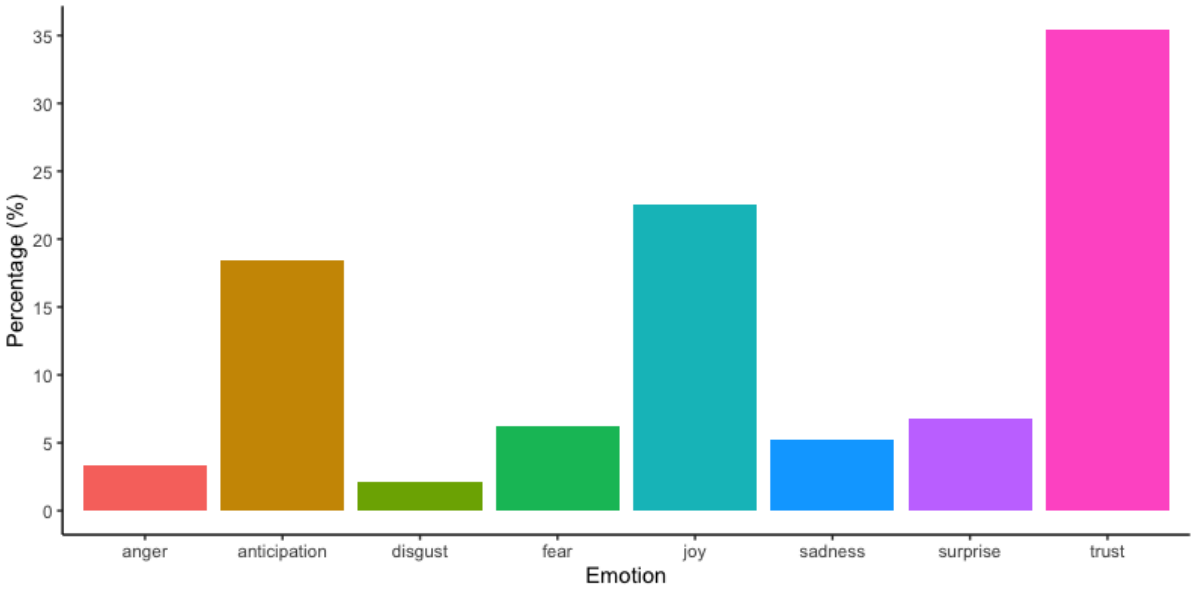

Figure 3. Percentage of NRC emotion lexicon (Mohammad & Turney, 2013) identified emotions within the open-text optional question 9 ("Other areas of strength").

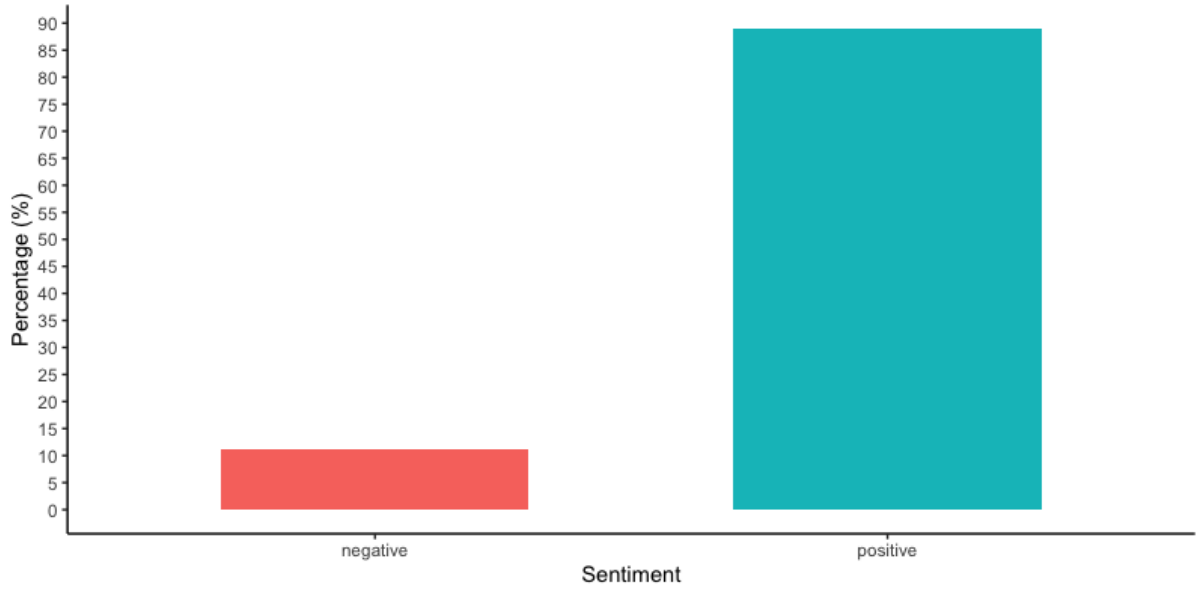

Figure 4. Percentage of NRC emotion lexicon (Mohammad & Turney, 2013) identified sentiment within the open-text optional question 9 (“Other areas of strength”).

According to OpenAI’s ChatGPT (OpenAI, 2023), the three top themes in question 9 were::

1. Positive feedback about the therapist's competence and effectiveness in helping the client. This includes comments about the therapist's ability to navigate complex mental health issues, their knowledge and expertise, and the effectiveness of the strategies and tools provided.
2. Comments about the client's comfort and satisfaction with the therapy experience, including feelings of respect, ease of booking, and the comfortable and accommodating environment provided.
3. Comments about the client's personal circumstances and needs, including the ability to receive support for specific health conditions, the need for multiple therapists, and the flexibility and responsiveness of the therapy process to individual needs.

In contrast, question 10 was assessing for “Suggestions for improvements” and it might be expected that the sentiment might be more negative. There were 741 text responses to this optional question. As seen in Figure 8, the most associated emotion was again ‘trust’ (27.7%), but closely followed by anticipation (23.9%), and the sentiment was again overwhelmingly positive (75.7%; Figure 9), but slightly less than what it was for question 9.

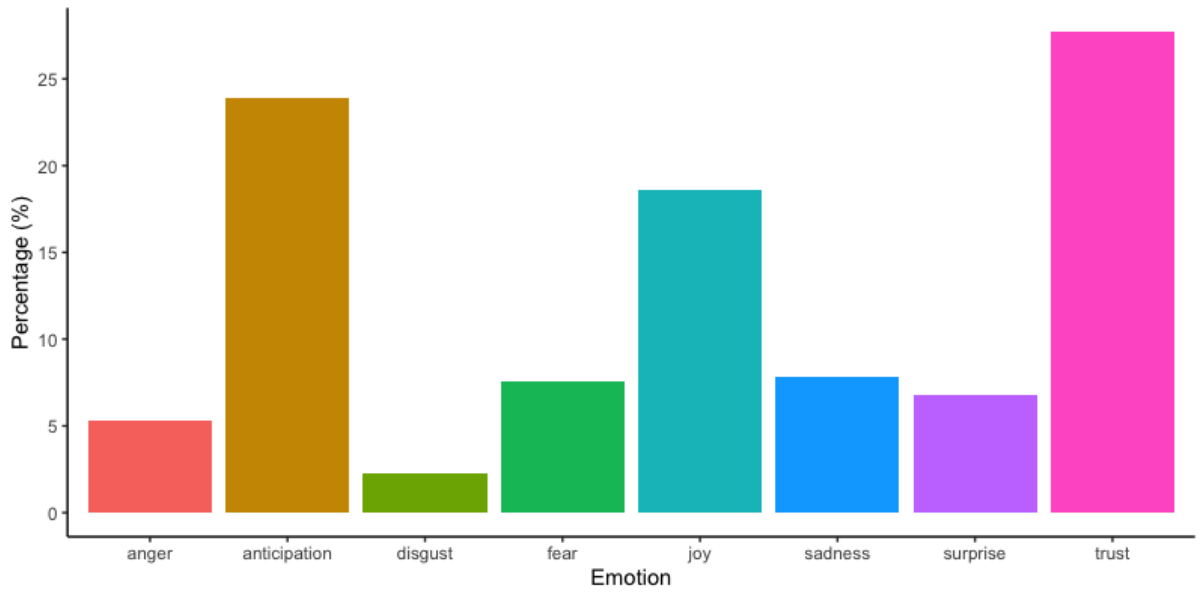

Figure 5. Percentage of NRC emotion lexicon (Mohammad & Turney, 2013) identified emotions within the open-text optional question 10 (“Suggestions for improvements”).

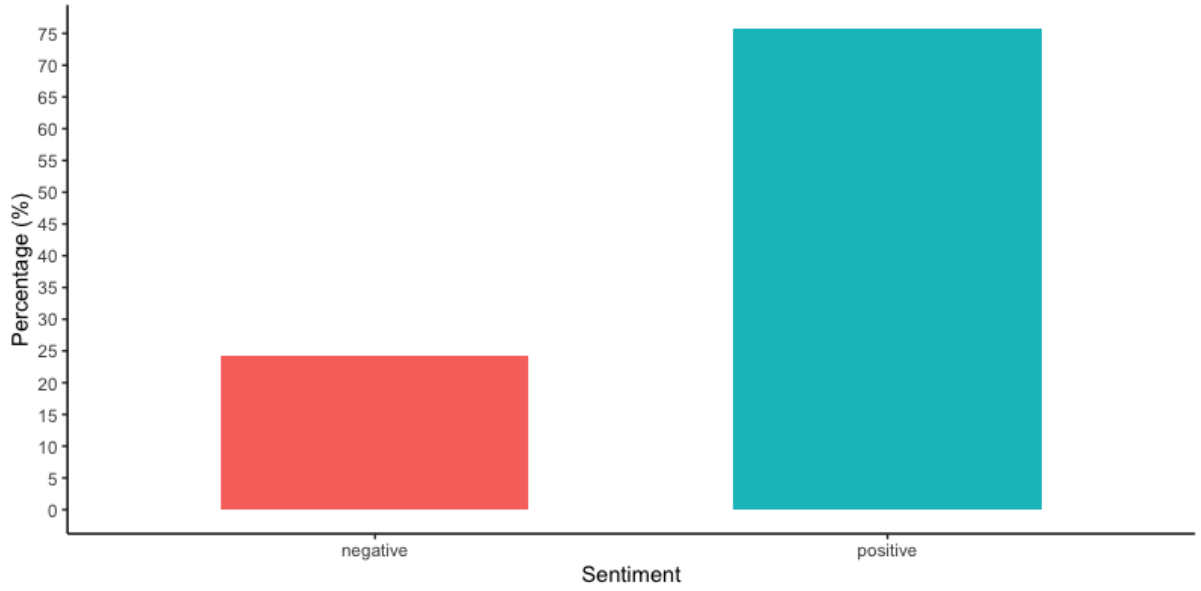

Figure 6. Percentage of NRC emotion lexicon (Mohammad & Turney, 2013) identified sentiment within the open-text optional question 10 (“Suggestions for improvements”).

So, although the identified negative sentiment more than doubled between question 9 and question 10, this would be expected given the nature of question 10. According to OpenAI’s ChatGPT (OpenAI, 2023), the three top themes in question 10 were:

1. Comments about the quality of the therapy experience, including issues with booking appointments, and concerns about the effectiveness of online therapy compared to in-person sessions.

2. Feedback about the therapist's competence and effectiveness in helping the client, including positive comments about specific therapists who were perceived to be helpful and effective.
3. Suggestions for improving the therapy process, including requests for more resources and support, suggestions for better communication and accessibility, and ideas for additional check-ins or support outside of regular appointments.

Factor Analysis. To determine if a factor analysis was appropriate, Bartlett's test of sphericity was performed on the complete data (set ii; n = 4,954) using the EFAtools package (Version 0.4.4; Steiner & Grider, 2020). Bartlett's test of sphericity, which tests the overall significance of all the correlations within the correlation matrix, was significant ($X^2(28) = 24{,}366.3$, $p < 0.001$), indicating that it was appropriate to use the factor analytic model on this set of data. The Kaiser-Meyer-Olkin measure of sampling adequacy indicated that the strength of the relationships among variables was high (KMO = .93), thus it was acceptable to proceed with the analysis. Using the corrplot package (Version 0.92, Wei & Simko, 2021), as seen in Figure 7, the correlation amongst each of the questions was moderate to strong.

Figure 7. A correlation plot amongst questions 1 to 8 of the CSS.

To initially determine the number of factors, a parallel analysis using 1,000 simulated random data sets was performed. Using means and EFA-determined eigenvalues, this determined that there were 4 factors (see Figure 8).

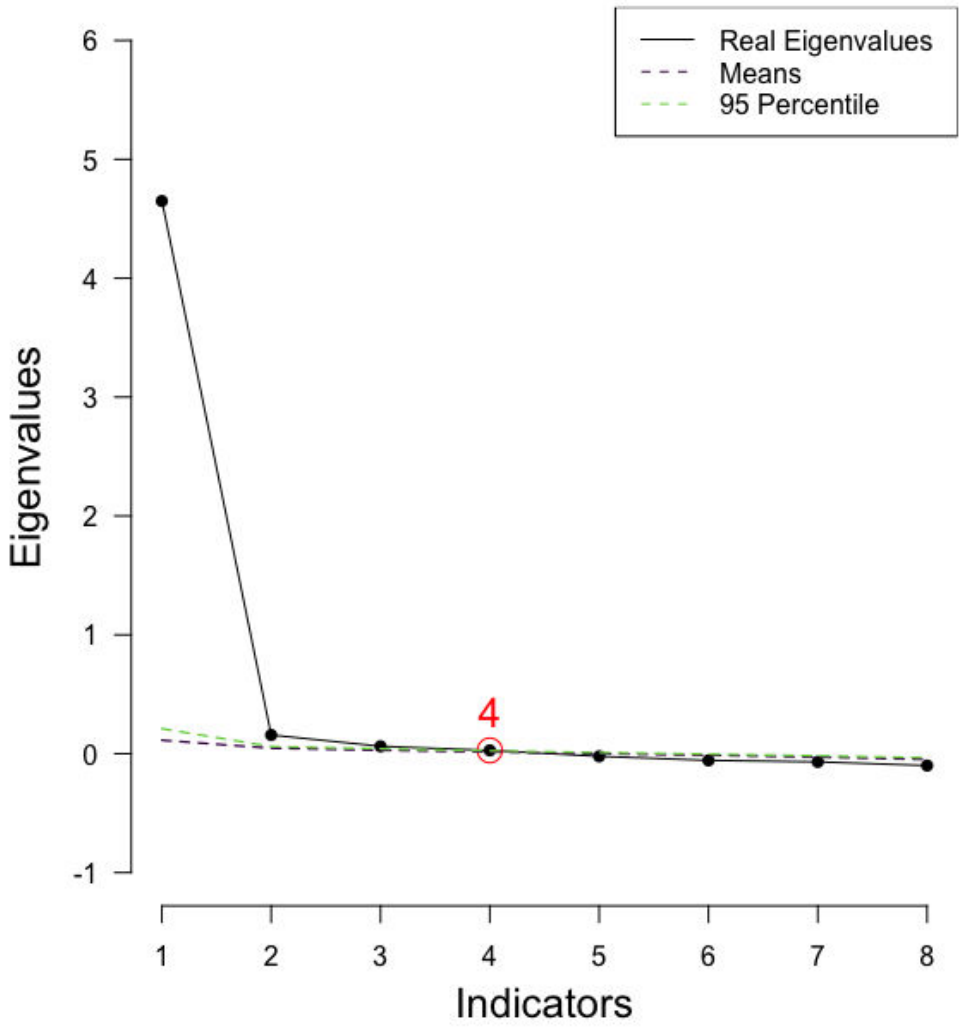

Figure 8. The number of factors to retain according to EFA determined eigenvalues.

EFA was completed in R using the psych package (Version 2.2.9; Revelle, 2022). The protocol adopted here for EFA was to use default settings initially (Principal Axis Factor - PAF) and to rotate the matrix of loadings to obtain orthogonal (independent) factors (Varimax rotation). The prime goal of factor analysis is to identify simple factors (items loadings >0.50 on only one factor) that are interpretable, assuming that items are factorable.

Table 1.
Pattern matrix with standardised loadings on each of the factors, percentage of variance explained by each factor and each factor's eigenvalue.

Scale Items	Factors			
	F1	**F2**	**F3**	**F4**
Question 1	**0.72**	0.21	0.22	0.25
Question 2	**0.79**	0.20	0.19	0.32
Question 3	**0.62**	0.20	0.22	0.51
Question 4	0.36	0.17	0.19	**0.61**
Question 5	0.26	0.15	**0.89**	0.22
Question 6	**0.51**	0.28	0.33	0.49
Question 7	0.26	**0.92**	0.16	0.20
Question 8	**0.58**	0.27	0.28	0.46
Percentage of Variance	30%	15%	15%	17%
Eigenvalue	2.39	1.17	1.17	1.34

Note. F1 - F4 = four identified factors. Standardised loadings in bold are the items that load onto each factor.

Table 1 shows the pattern matrix with the standardised loadings on each identified factor, the percentage of variance explained by each factor, and factor eigenvalues. The first factor (F1) consisted of questions 1, 2, 3, 6, and 8, whereas the other three factors consisted of individual questions. The model fit statistics for this 4-factor model were good (TLI = 0.982; RMSEA = 0.056) and the resultant path model can be seen in Figure 9.

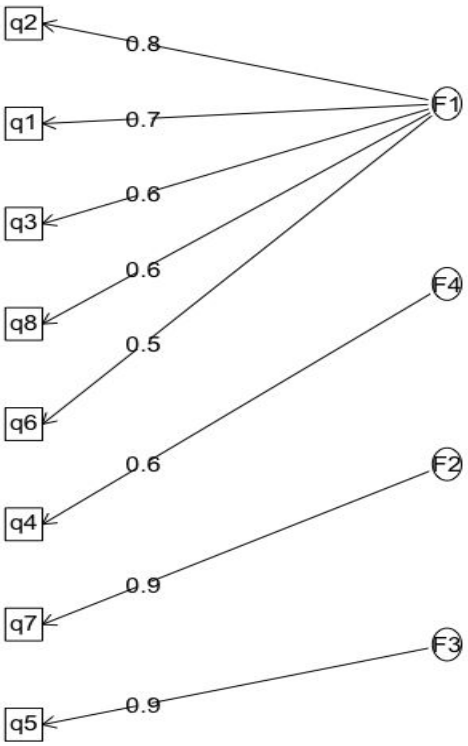

Figure 9. Path model of four factor EFA solution. q = questions 1 - 8. F1 - F4 = four identified factors.

As a result of this EFA, factor one (questions 1, 2, 3, 6, 8) was proposed to be a measure of the clients' satisfaction with the 'Practitioner' (e.g., preparation, empathy / understanding, information provision, consistency, general satisfaction). All questions in factor one except for question 8 explicitly refer to the practitioner, whereas question 8 asks about overall satisfaction with the service, which is primarily provided by the practitioner (All things considered, I am satisfied with the service I received).

Three other factors all had one question each. Factor two (question 7) was proposed to be a measure of the clients' satisfaction with 'Reception Staff', factor three (question 5) was a measure of satisfaction with 'Scheduling', and factor four (question 4) was a measure of client satisfaction with 'QoL (quality of life) Improvement'.

Norming / Percentiles. Using the four factors identified by the EFA process, percentiles for each factor were calculated using the cNORM package (Version 3.0.2; Lenhard & Lenhard, 2021). This method of norming estimates percentiles on the basis of the raw data without requiring assumptions about the distribution of the raw data. This method minimises bias arising from sampling and measurement error, while handling marked deviations from

normality, addressing bottom or ceiling effects and capturing almost all of the variance in the original norm data sample (Lenhard & Lenhard, 2021).

Table 2.
Percentile table for 'Practitioner' factor (Factor 1).

Raw Score	Percentile
14	< 1
15	1.5
16	2
17	3
18	6
19	10
20	16
21	23
22	31
23	38
24	47
25	> 57

Obviously, given the extreme ceiling effects in the CSS data, the percentiles generated as part of the norming process have large gaps at the upper end. For example, Table 2 shows the percentile table generated for the 'practitioner' factor (Factor 1). As seen in Table 2, a raw score of 25 (the maximum possible score) corresponds to a percentile of > 57. This indicates that a large percentage of clients (43%) scored 25, the maximum satisfaction score. Tables 3 - 5 show the percentile tables for the other three factors.

Table 3.
Percentile table for 'Reception Staff' factor (Factor 2).

Raw Score	Percentile
1	< 0.3
2	1
3	6
4	24
5	> 53

Table 4.
Percentile table for 'Scheduling' factor (Factor 3).

Raw Score	Percentile
1	< 0.4
2	2
3	6
4	26
5	> 71

Table 5.
Percentile table for 'QoL Improvement' factor (Factor 4).

Raw Score	Percentile
1	< 0.4
2	2
3	9
4	32
5	> 74

In addition to factor scores, percentiles were calculated for the total score for the complete data (Table 6; set ii; n = 4,954) and for when optional question 7 was not completed (Table 7, set iii; n = 382).

Table 6.
Percentile table for total score (when all items completed).

Total Score	Percentile
22	< 1
23	1.2
24	1.5
25	2
26	2.5
27	3
28	5
29	7
30	9
31	12
32	17
33	22
34	27
35	33
36	40
37	46
38	53
39	61
40	> 71

Note. Maximum score = 40.

Table 7.
Percentile table for total score (when optional question 7 is incomplete).

Total Score	Percentile
20	< 1
21	2
22	2.5
23	4
24	5
25	7
26	10
27	14
28	18
29	24
30	30
31	37
32	45
33	53
34	63
35	> 75

Note. Maximum score = 35.

Discussion

This study aimed to characterise the psychometric properties of the Client Satisfaction Survey (CSS) and measure typical levels of satisfaction among clients of psychology clinics. Clients who responded to the CSS typically report a very high level of satisfaction with services. For instance, 31.2 percent of respondents “Strongly Agreed” with all satisfaction statements and 86.7 percent were in general agreement with the statements (i.e. scores 32 or more).

The distribution of responses was significantly skewed and suffered from ceiling effects, which is not unusual for client satisfaction measures (Meier, 2022; Meier & Feeley, 2021; Paap et al., 2019). The results of the EFA demonstrated that there were four factors present, representing clients' satisfaction with the 'Practitioner' (e.g., preparation, empathy / understanding, information provision, consistency, general satisfaction), 'Reception Staff', 'Scheduling', and 'QoL (quality of life) Improvement'. Although the last three factors are represented by only one question each, the validity of single item factors / scales have been verified statistically and clinically previously (e.g., McKenzie & Marks, 1999). In our sample, respondents typically were most satisfied with reception staff (question 7).

The percentiles this study developed for each of the four factors had large gaps at the maximum scores, due to the extensive ceiling effects. Nevertheless, they are intended to provide clinicians with a better understanding of their clients' satisfaction in reference to a normative group, and are therefore of value.

Given the ceiling effects observed in this measure, slight deviations from maximum scores are likely to be meaningful. In particular, scores of 19 or below on the practitioner factor is likely to indicate client dissatisfaction and represents scores at the 10th percentile or less. This may be counterintuitive for practitioners reviewing a single client's results given that a score of 19 probably corresponds to answers mostly in agreement with the five questions.

As well as the CSS being useful for collecting feedback from individual clients, it may be useful for a practice to aggregate data to assess their service as a whole or consider practitioner factors related to patient satisfaction. If collating data, a key metric for practices to consider is the proportion of respondents who "Agreed" on all the satisfaction statements. (e.g. said at least Agree to all questions). For ease of calculation, five metrics are suggested for collated data.

(1) The percentage of respondents who had an average score of 4 or more (total ≥ 32 if question 7 is completed), indicating the proportion of people who were satisfied with their interaction with the service. Our data shows that typically 88% of clients have an average score of 4 or above. Therefore, a practice that deviates significantly from this percentage has patterns of responding, and therefore client experiences, that are different from typical practices.

(2) The percentage of the Practitioner factor average score 4 (total = 20) or above, indicating the proportion of people who reported a positive experience with the practitioner. Our data shows that typically 90% of clients score 20 or above.

(3) The percentage of people scoring Agree (four) or more on question 4 (Therapy has helped me improve the quality of my life). Our data shows that typically 91% of clients score 4 or above.

(4) The percentage of people scoring neutral (three) or less on question 5 (I was able to schedule appointments at times that suited me). Typically 94% of clients score above 4.

(5) The percentage of people scoring neutral (three) or less on question 7 (The reception staff were friendly and helpful). Typically 94% of clients score 4 or above.

Overall, the distribution of scores on the CSS had significant ceiling effects that are typical of client satisfaction measures. This effect may be due to sampling bias (whereby satisfied clients are more likely to respond), socially desirable responding, or that most clients are genuinely satisfied with services. A suggested direction for future research is developing a measure where responses are more normally distributed.

In any case, the use of normative data (skewed or otherwise) can provide important context for interpreting satisfaction scores in general, and CSS scores in particular. Benchmarking scores against patterns of typical responding helps service managers and clinicians better assess the quality of services provided and respond appropriately to feedback.

References

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy, 55(4), 316-340. https://doi.org/10.1037/pst0000172

Gibbons, C. J., Morris, Z. S., & de Leval, M. R. (2013). Waiting for elective surgery: Effect on health-related quality of life and psychological distress. Journal of Health Psychology, 18(10), 1352-1368. https://doi.org/10.1177/1359105312465781

Hsieh, C. M., & Wang, H. H. (2017). Patient satisfaction with and recommendation of a primary care provider: Associations of perceived quality and patient education. International Journal for Quality in Health Care, 29(5), 600-605. https://doi.org/10.1093/intqhc/mzx080

Lenhard, W., & Lenhard, A. (2021). Improvement of Norm Score Quality via Regression-Based Continuous Norming. Educational and Psychological Measurement, 81(2), 229–261. https://doi.org/10.1177/0013164420928457

McKenzie, N., & Marks, I. (1999). Quick rating of depressed mood in patients with anxiety disorders. British Journal of Psychiatry, 174, 266 - 269.

Meier, S.T. (2022). Investigation of causes of ceiling effects on working alliance measures. Frontiers in Psychology, 13:949326. https://doi.org/10.3389/fpsyg.2022.949326

Meier, S. T., and Feeley, T. H. (2021). Ceiling effects suggest a threshold structure for working alliance. Journal of Counseling Psychology, 69, 235–245. https://pubmed.ncbi.nlm.nih.gov/34292029/

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. Computational intelligence, 29(3), 436-465. https://doi.org/10.1111/j.1467-8640.2012.00460.x

OpenAI (2023). ChatGPT. Accessed at https://chat.openai.com/chat

Paap, D., Schepers, M., & Dijkstra, P. U. (2020). Reducing ceiling effects in the Working Alliance Inventory Rehabilitation Dutch Version. Disability and Rehabilitation, 42(20), 2944-2950. https://doi.org/10.1080/09638288.2018.1563833

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Revelle, W. (2022). psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych

RStudio Team (2022). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Steiner, M.D., & Grieder, S.G. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. Journal of Open Source Software, 5(53), 2521. https://doi.org/10.21105/joss.02521

Wei, T. & Simko V. (2021). R package 'corrplot': Visualization of a Correlation Matrix (Version 0.92). Available from https://github.com/taiyun/corrplot

Your Experience of Service. Australia's National Mental Health Consumer Experience of Care Survey (2015), retrieved August 22 2023: https://www.aihw.gov.au/getmedia/b49af1ae-cb32-4e15-9ad0-00c28341ea1e/YES-Guidance.pdf.aspx