



GLOBAL JOURNAL OF MEDICAL RESEARCH: κ
INTERDISCIPLINARY
Volume 23 Issue 4 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4618 & Print ISSN: 0975-5888

A Short Satisfaction Questionnaire for an Independent Medical Examination: A Pilot Study

By Ian Bruce Mc Phee

University of Queensland

Abstract- Objective: This study was conducted to develop and validate a questionnaire to assess client satisfaction with an Independent Medical Examination (IME).

Design: The questionnaire (IMESQ) was developed as a short assessment of six items covering six domains relevant to an IME. The sample was tested for internal consistency reliability. Construct validity of the questionnaire was via a parallel measurement analysis.

Participants: Data were collected from 76 adult respondents (male = 44.7 %); mean age (42.9 ± 13.3 years).

Results: The questionnaire had good correlation of paired items (range $r = 0.55$ to 0.80 , $M = 0.65$). The Cronbach's α value of the questionnaire was $.91$ and the split-halves reliability was $.90$ suggesting good internal consistency. There was strong correlation ($r_s = .82$) with a numerical satisfaction scale (NSS). Compared with a Surrogate score calculated from the NSS, the intrarater reliability was high (ICC $.75$).

Keywords: *pilot study; independent medical examination; satisfaction questionnaire; internal consistency; validation; intrarater reliability; discordance; bland-altman analysis.*

GJMR-K Classification: NLMC: W 84.1



ASHORTSATISFACTIONQUESTIONNAIREFORANINDEPENDENTMEDICALEXAMINATIONAPILOTSTUDY

Strictly as per the compliance and regulations of:



A Short Satisfaction Questionnaire for an Independent Medical Examination: A Pilot Study

Ian Bruce Mc Phee

Abstract- Objective: This study was conducted to develop and validate a questionnaire to assess client satisfaction with an Independent Medical Examination (IME).

Design: The questionnaire (IMESQ) was developed as a short assessment of six items covering six domains relevant to an IME. The sample was tested for internal consistency reliability. Construct validity of the questionnaire was via a parallel measurement analysis.

Participants: Data were collected from 76 adult respondents (male = 44.7 %); mean age (42.9 ± 13.3 years).

Results: The questionnaire had good correlation of paired items (range $r = 0.55$ to 0.80 , $M = 0.65$). The Cronbach's α value of the questionnaire was .91 and the split-halves reliability was .90 suggesting good internal consistency. There was strong correlation ($r_s = .82$) with a numerical satisfaction scale (NSS). Compared with a Surrogate score calculated from the NSS, the intrarater reliability was high (ICC .75). A Bland-Altman analysis identified 14 Outliers who underrated or overrated the responses to the questionnaire.

Conclusions: The IMESQ appears to be a valid tool to assess client satisfaction with an IME. Larger trials with more examiners are required. The questionnaire is a 360-degree audit that satisfies the requirements of the professional bodies for Continuing Professional Development.

Keywords: pilot study; independent medical examination; satisfaction questionnaire; internal consistency; validation; intrarater reliability; discordance; bland-altman analysis.

1. INTRODUCTION

The patient is a consumer of health services and thus has certain rights, putting an emphasis on the delivery of quality health care. Patient satisfaction is an important and commonly used indicator for measuring the quality in health care [1]. Communication and health-related behavior are key markers of patient satisfaction [2]. The three most important aspects in patient – doctor communication are creating a good interpersonal relationship, exchanging information, and involving the patient in making treatment-related decisions [3]. Treatment related decisions are not relevant when doing an independent medical examination (IME) for a third party.

A qualitative study identified three key interacting components; "Listening", "Asking for information", and "Giving information" as central and covering two-thirds of the identified interaction types [4]. A study of patient satisfaction in a disease specific setting identified four factors that displayed eigenvalues greater than 1.0: 'Interpersonal Skills,' 'Quality of Time,' 'Empathy,' and 'Information Exchange' [5].

There are numerous satisfaction surveys available. Consumer satisfaction surveys are used by different organisations and individuals for varying purposes. The patient satisfaction field has increasingly been moving towards disease or condition-specific patient satisfaction questionnaires focussing on decision making, treatment and outcome, tapping aspects of satisfaction that are unique to that disease [6]. Researchers have used modified versions of commonly used questionnaires which excluded subscales not applicable in the context of these studies [7].

Hawthorne demonstrated that a Short Assessment of Patient Satisfaction (SAPS) scale based on seven questions each covering a separate dimension could adequately assess patient satisfaction of health care [8]. The questionnaire was proven robust, with good internal consistency reliability and correlated well with several accepted longer questionnaires. However, the domains are unsuitable for an IME.

An IME is unique in medicine. It is an *objective medical examination on behalf of a third-party to confirm the extent of the subject's injuries*. The report is often used to confirm or deny benefits, coverage, or settlements, or provide steps for safe return to work. The IME doctor is not the patient's treating doctor or personal adviser, and diagnosis, treatment and outcome are not an issue. However, client satisfaction with the IME is not excluded. The circumstances sometimes considered adversarial by the client might heighten dissatisfaction.

A literature search using the PubMed data-base and the search terms "independent medical examination", "satisfaction", and "questionnaire" failed to identify any relevant publications. A random search found one paper reporting a satisfaction survey conducted on patients from a medico-legal consultation [9].

The aim of this study was to develop a short generic satisfaction questionnaire, with acceptable psychometric standards, suitable for use in the setting

Author: Retired senior surgeon, Department of Orthopaedic Surgery, Royal Brisbane and Children's Hospitals, Herston, Australia. Retired Associate Professor, Orthopaedic Surgery, University of Queensland, Australia. Mail: 33 Derby Lane, Enoggera, 4051, Australia. e-mail: drbmcphree@gmail.com

of an IME. The scale required sub-scales or domains relevant to the examination that more than likely would influence the patient's/client's satisfaction. The questionnaire needs to be self-administrated, brief, understandable and easy to complete by clients aged 18 years and older. Whether the responses were influenced by socio-demographic factors or health status needed exploring.

II. MATERIALS AND METHODS

a) Questionnaire Development

The domains were selected to address aspect of the consultation perceived to be the most likely factors to affect the client's satisfaction with the encounter, focussing on doctor-client relationship, communication, and the client's perception of the adequacy of the history-taking and examination. The five domains selected were: interpersonal manner, communication, technical skill (examination), perceived exchange of information, and time allocation. The outcome was assessed by a question on general satisfaction. The complete questionnaire (IMESQ) of six items is attached in Appendix 1. Each question addresses a single domain and each has a 5-point Likert scale response. The scales are ordinal, varied and not uniformly weighted. Four of the six questions were positively worded. The responses were scored from 1 (the worst) to 5 (best) giving a total score ranging from six to 30. The higher the total score the greater the satisfaction.

To test the construct validity and consistency of the questionnaire, the study included a 0-to-10 numerical scale of satisfaction (NSS) for comparison with the total score from the six items. A further question inquired as to whether the client would be "willing to undergo another examination with the doctor if required". This assumes that higher levels of satisfaction would positively relate to Willingness to be re-examined by the doctor. The choices of responses were: "yes", "no", and "maybe". At the end of the structured questionnaire, a text box for open-ended comments was included to accommodate the qualitative information provided by the respondents. However, commenting was optional.

The study included an additional five questions relating to items thought to be independent variables: gender, age (in four categories), whether the client was born in Australia, had previously undergone an IME, and whether working or not. The latter items were included to see if immigrants, previous experience of a similar examination, and not working influenced the client's perception of satisfaction.

The questionnaire was loaded into Google Forms which allows emailing of the questionnaire to the clients. Once completed, the forms are automatically returned to the sender's email address and

automatically downloaded. The program provides automatic collation and analysis of the total data presented in graphic form, and allows access to individual responses for in-depth analysis.

Some items perceived as probably independent including level of education achievement (Unqualified), domestic status (Partnered), and the EQ-5D 3L quality of life utility scores ($M = .55$, $SD = .16$, $Mdn = .52$) and were added from the client's records for the in-depth analysis.

b) Study Site

This was a cross-sectional study of clients/participants who attend the author for an IME between November 2021 and March 2022. Initially the questionnaire was emailed to the client after the examination. There was only one response from the first thirteen clients and emailing was abandoned because of the poor response. The questionnaire was administered following completion of the consultation before the client departed the office. This ensured a satisfactory number of responses. The participants completed and signed a written informed consent.

c) Statistical Analysis

Data was collated into Excel and DATA tab for analysis. Primary analysis with Shapiro-Wilks test found that all data was not normally distributed except for chronological age. The items of the IMESQ, the total IMESQ score and the NSS were distributed with an asymmetrical left, negative skewness, with normal-like, left tails. As the item scores were ordinal and not normally distributed, Spearman's rank-order correlation coefficient was used to analyse the correlations for ordinal and continuous data. The rank-biserial correlation was used for continuous and ordinal data with binary data: "Willingness of further review", gender, Australian born, working, partnered (domestic status), and unqualified (no trade or tertiary qualification). Chi-square statistic was used for comparison of dichotomous and nominal data. For the descriptive purposes, an adapted Dancey and Reidy classification of the strength of correlation was used [10]. A $p < 0.05$ was accepted as the level of significance.

Initial experience suggested that test-retest evaluation was unlikely to succeed. In lieu of test-retest validation, split-half correlation coefficient with Spearman-Brown split-halves reliability was performed comparing odd to even numbered items. The split-halves approach has the advantage that it is not dependent on memory, but is dependent on a high level of internal consistency.

Construct validity of the questionnaire was tested against alternate forms of measuring satisfaction. The total IMESQ score was compared with the NSS using Spearman's correlation. An alternative hypothesis was that satisfaction would correlate with a Willingness to be examined by the doctor again. The potential future

utilisation of the service was tested using rank-biserial correlation. The impact of demographic factors on the IMESQ was explored.

Paired inter-Item correlations, Spearman-Brown split-halves reliability and Cronbach's alpha coefficient were used to test internal consistency reliability to determine the extent to which all of the items in the model measure the different aspects of the satisfaction [11]. Although all items contained non-normally distributed data which violates the axioms for using Cronbach's alpha, studies have reported it as a robust measure [12]. Alpha coefficient 0.70 or greater was considered sufficiently reliable [13]. Before performing a Principal Components Analysis (PCA), the suitability of the data for that analysis was tested by the Kaiser-Meyer-Olkin values and Bartlett's Test of Sphericity.

To test the accuracy of the participants' responses, intrarater reliability was assessed using Intra Class Coefficient (ICC) based on the analysis of variance (ANOVA) models [14-17]. The Concordance Correlation Coefficient (CCC) was calculated. Unlike the calculation of the ICC which assumes normal

distribution of data, Lin's CCC is the correlation of point estimates in relation to the 45° line through the origin and is suitable for non-normal data [18,19]. A Bland-Altman analysis identified discordant responses (Outliers) due to possible confounding factors or client bias [20-22]. The impact of demographic factors on the discordant data was explored.

III. RESULTS

There were 76 respondents. All questionnaires were completed without missing information. Metric and ordinal data was assessed for normality of distribution using Shapiro Wilk statistic. Only chronological age ($M = 42.9$, $SD = 13.3$, Range [18, 73]) was normally distributed.

The independent variables and descriptive data defining the demographics of the cohort is summarised in Table 1. There was a statistically significant relationship between Male and Unqualified, $\chi^2(1) = 11.13$, $p = .001$. Males were more likely to be unqualified, Odds ratio = 5.4, 95% CI [1.9, 15.0].

Table 1: Details of the Independent Variables

	Frequency	% of Cases
Male	34	45%
<40yo	27	36%
Unqualified	27	36%
Partnered	53	70%
Australian born	50	66%
Working	48	63%
Previous IME	36	47%
EQ5D <0.50	13	17%

a) The Questionnaire

Total scores for the IMESQ ranged from 18 to 30 with 36% of the participants having the maximum score of 30 ($M = 29$, $SD = 3.17$; $Mdn = 29$, $Q1 = 25$). The ceiling for each Item ranged from 49% to 55% and floor effect from 4% to 11%. As expected, multiple logistic regression indicated that there was a very strong collective significant effect between all Items and each Item and the IMESQ, $\chi^2(9,76) = 90.27$, $p < .001$, $R^2_N = .96$. The power for the entire model was medium (.67). Spearman correlation showed that there was a low but significant positive correlation between chronological Age and IMESQ, $r(74) = 0.28$, $p = .015$. A significant, small negative correlation was also noted between Unqualified and IMESQ, $r(74) = -0.25$, $p = .033$. The IMESQ had no statistical association with the other independent variables.

b) Internal Consistency Reliability

Inter-item correlation was high for all items indicating that the questionnaire is possibly unidirectional and that each dimension relates well to all others (Table 2). Cronbach's alpha coefficient was $\alpha = .91$. The corrected item-total correlations show that each item was consistent with the other items taken together (Table 3). Cronbach's alpha when each item was sequentially deleted was no greater than Cronbach's alpha for the whole scale. The high alpha coefficient indicates a high level of internal consistency reliability of the questionnaire and the total score obtained by summing the responses across all items reflects the same construct and the possibility that the questionnaire may be unidimensional [23].

c) Exploratory Factor Analysis (EFA)

Kaiser-Meyer-Olkin Measure of sampling adequacy was 0.87. A KMO value between 0.7 and 0.8 is good and suitable for factor analysis [24]. Bartlett's Test of Sphericity was also significant ($\chi^2 (15) = 308.56, p < .001$) confirming at least one significant correlation between two of the items [25].

Two of the assumptions for EFA were not met; the sample does not have multivariate normality and the sample size is small. Factor analysis was considered suitable. The total variance across all dimensions is 6 (Table 4). The two principal components explain 79.3% of the variance within the model. Only the first component has an eigenvalue greater than 1; all but the sixth component having an eigenvalue greater than 0.3. A Rotated Component Matrix (Varimax) and biplot confirmed the questionnaire was unidirectional.

d) Split-Halves Correlation

The IMESQ was divided into odd and even numbered questions for purpose of a split-half correlation. A Spearman's correlation between variables Odds and Evens showed a high, positive correlation between the variables, $r (74) = 0.78, p = <.001$. Pearson's correlation showed a higher, positive correlation, $r (74) = 0.82, p = <.001$ and the Spearman-Brown-adjusted Pearson correlation for reliability was $r_{SB} = 0.90$.

e) Construct Validation Willingness

Sixty-two (82%) of the participants indicated a "Willingness to submit for another examination if required" while the remaining fourteen participants indicated "maybe". A rank-biserial correlation between variables Willingness and IMESQ showed a medium, positive correlation, $r (74) = 0.41, p = <.001$. Univariate analysis of correlations between the Items and Willingness showed moderate association of all Items ranging from $r (74) = .31$ to $r (74) = .45 (p < .01)$. The Kruskal-Wallis test showed that there was a significant difference in mean ranks between the Items and Willingness, $\chi^2 (6) = 248.91, p < .001$. The Dunn-Bonferroni test revealed that the pairwise group comparisons of all Items with Willingness were significantly different ($p < .01$). Logistic regression analysis showed that the model as a whole was significant at predicting a Willingness to further examination, ($\chi^2 (9,72) = 23.01, p = .006, R^2_N = .42$). Backward step-wise analysis found that at ease was the sole predictor, $\chi^2 (1) = 11.75, p \text{ value} = <.001$. Willingness had no statistically significant association with any confounder.

If a suboptimal level of satisfaction is set at scores where both the IMESQ and NSS are in the first quartile, that is an IMESQ <25 and NSS <8 , then the criteria applies to 10 participants (13% of the cohort). Of

these participants, 8 were unwilling to submit to further examination compared with 6 of the remaining cohort. Participants who had low scores for both the IMESQ and NSS were forty times more likely (OR = 40.0, 95% CI [6.9, 233]) to be unwilling to undergo further examination, $\chi^2 (1) = 29.06, p < .001$.

f) Numerical Satisfaction Scale

Scores on the Numerical Satisfaction Scale (NSS) ranged from 5 to 10 with 42% giving the maximum score ($M = 8.78, SD = 1.36, Mdn = 9, Q1 = 8$). The distribution of data is shown in Figure 1. There was a strong correlation between the IMESQ and the NSS ($r (74) = 0.82, p = <.001$). The priori power was .76. Neither measure contained outliers. There is a medium, positive correlation between Willingness and NSS, $r (74) = 0.45, p = <.001$. Individually, all questionnaire Items had a strong correlation with the NSS, ranging from $r (74) = .58$ to $r = .70, p < 0.001$. The results of backward step-wise logistics regression indicated that there was a strong collective significant effect between the Items and NSS, $\chi^2 (9,76) = 53.34, p < .001, R^2_N = 0.67$). The individual predictors were examined further and identified Examination and Satisfaction as the only significant predictors in the model. NSS had no association with any demographic factors.

g) Concordance

A Surrogate score was devised based on the "perfect" relationship between the NSS and the IMESQ being identical measures of participant's satisfaction. The NSS score ranges from zero to 10. The IMESQ ranges from a minimum of six to a maximum of 30. The equation for the Surrogate score was $5 + 2.5 * NSS$ giving a range of five to 30. The Surrogate score was calculated for each participant for comparison with the IMESQ for analysis of association and discordance (Figure 1).

A Spearman correlation was performed to reaffirm the correlation between variables IMESQ, Surrogate score and NSS. There correlation between Surrogate and NSS was $r (74) = 1$. The correlation between IMESQ and Surrogate score was consistent with the correlation between the IMESQ and NSS, $r (74) = 0.82, p = <.001$.

A Wilcoxon Signed-Rank test showed that there is a non-significant small difference between IMESQ ($Mdn = 29$) and Surrogate scores ($Mdn = 27.5$), $z = -1.7, p = .085, r = -0.25$). A paired t-test for dependent samples also showed that this difference was small and was not statistically significant, $t (75) = 1.41, p = .161, d = .23$. Only 37 (49%) of the 76 data pairs were concordant. The Kendall tau coefficient ($\tau (76) = 0.03$) was poor, equivalent to a random response.

The means and standard deviations of the IMESQ ($Mn = 27.3, SD = 3.16$) and Surrogate scale

($Mn = 26.9$, $SD = 3.41$) are similar. The distribution of both have similar asymmetrical, left/negative skew with long left tails (skew -1.15 , -0.88 respectively). The Shapiro Wilk analysis shows that both measures were non-normal distributions.

The Levene's test found the variances are not significantly different for the IMESQ and Surrogate score, satisfying the assumption of homogeneity of variance, $F(1,150) = 7.12$, $p = .398$. The Brown-Forsythe the F two-sample test for variance also found no significant difference, $F(75,75) = .87$, $p = .266$. The IMESQ and Surrogate scores have equivalent variances. The assumption for both tests was that the distribution of data was normal.

h) Intra-Rater Reliability

The rater consistency was determined by ICC reliability calculated from an Analysis of Variance by three methods [16,17]:

ICC (A, 1) $p = .750$ (.632 - .834), $F(75,76) = 7.01$, $p < .001$,

ICC (C, 3) $p = .753$ (.634 - .834), $F(75, 75) = 7.10$, $p < .001$

ICC (A, 3) $p = .751$ (.634 - .834), $F(75, 75) = 7.10$, $p < .001$.

According to Cichetti and Sparrow, the intrarater reliability is good [26]. The ICCs calculated from the data are approximately equal and any bias is likely to be small or negligible [17]. However, ANOVA is poor at detecting bias [27].

Lin's CCC is $pc = .75$, $F(1,75) = 6.31$, $p < .001$, which according to McBride, the concordance between the IMESQ and Surrogate scale is poor compared to a gold standard measurement [28]. Others would disagree, the threshold of acceptable reliability should vary with the circumstances [29, 30]. Irrespective of the quantitative measure of reliability, there is evidence of intrarater discordance.

i) Discordance - Bland-Altman analysis

Spearman correlation showed that there was a significant small to medium positive correlation between Mean and Difference, $r(74) = 0.32$, $p = .005$. Simple linear regression was used to test if Mean of the scores significantly predicted the Difference in scores. The fitted regression model was: Difference = $2.62 - 0.08 * \text{Mean}$. The constant is positive suggesting that the questionnaire may have a small positive bias. The overall regression was not statistically significant ($F(1, 74) = 0.91$, $p = .343$, $R^2 = .012$). The Mean of the two scores does not significantly predicted the Difference in scores ($\beta = -.08$, $p = .344$). The coefficient for the variable Mean in the population is not different from zero. This result is unlikely to reflect the real circumstances where some participants may intentionally or subconsciously underrate the responses

to the questionnaire as dissent bias and others who overrate the responses as acquiescence bias.

j) Outliers

Outliers were identified by a Bland-Altman analysis of the Mean score and the Difference between the IMESQ and the Surrogate score (Figure 2). Note that the assumption of a normal distribution of Difference was not met ($W(76) = .87$, $p < .001$). There were four data points outside the limits of agreement ($M + 1.96 * SD$) for the Difference in scores (< -4.15 , > 4.90); three below and one above. There were 14 Outliers (< -2.08 , > 2.45) calculated using the Tukey Fences ($k = 1.5$) algorithm representing 18% of the sample. Five were low outliers underrating the IMESQ relative to the NSS and Surrogate scores, and nine were over raters where the responses to the questions were exaggerated.

Rank-biserial correlation showed that there was no significant correlation between Difference and Outlier, $r(74) = 0.20$, $p = .088$. Rank-biserial correlation showed that there was a significant negative correlation between IMESQ and Outlier, $r(74) = -0.33$, $p < .001$. Univariate analysis showed that Items 1, 2 and 4 in the questionnaire had a small negative but significant correlation with Outliers (range $r(74) = -.29$, to $r = -.34$], $p < .05$). The Kruskal-Wallis test showed that there was a significant difference in mean ranks between the Items and Outliers, $\chi^2(6) = 248.91$, $p < .001$. The Dunn-Bonferroni test revealed that the pairwise group comparisons of all Items with Outliers were significantly different ($p < .001$). Logistic regression analysis to examine the influence of the Items on the variable, Outlier, to predict the value "1" showed that the model as a whole was not significant, $\chi^2(9,76) = 16.02$, $p = .067$, $R^2_N = .31$. Backward step-wise analysis found that At ease (Item 1) was the sole predictor, $\chi^2(1) = 8.56$, $p = .003$, $R^2_N = .15$. A Chi-square test found no statistically significant relationship between Outlier and Willingness, $\chi^2(1) = 0.33$, $p = .569$. No association was found between any of the independent variables and Outliers. When the Outliers were divided into Underraters (coded -1, Difference < -2.08) and Overraters (coded 1, Difference > 2.45), analysis identified a significant association between Under/Over raters and gender. Underraters were entirely male whereas females accounted for two-thirds of the overraters (Table 5). A Chi-square test was performed between Male and Under/Over Outlier. At least one of the expected cell frequencies were less than 5. Therefore, the assumptions for the test were not met. There was a statistically significant relationship between Male and Under/Over Outlier, $\chi^2(2) = 6.85$, $p = .033$. A Kruskal-Wallis test showed that there is a significant difference between the categories of the independent variable Under/Over-rater with respect to the dependent variable Male, $\chi^2(2) = 6.76$, $p = .034$. The Dunn-Bonferroni test showed that the pairwise group

comparison of 0 to -1 (Underrater) has an adjusted p -value of less than 0.05 (Table 6).

Further analysis showed that if 4 points difference between the IMESQ and the Surrogate score was arbitrarily the cut-off for potential bias, then all but one Outlier identified by analysis would still be Outliers. This single case would be considered false negative.

IV. DISCUSSION

This study has shown that the IMESQ is a valid, unidimensional measure of patient satisfaction following an independent medical examination. The short questionnaire has been shown to produce meaningful results in the assessment of client satisfaction. Although the questions had a high ceiling effect, the aim of the IMESQ is not to identify high achievement but to identify deficiencies which can be corrected.

Inter-item correlation is high for all pairs of Items and each dimension relates well to all others. The internal consistency reliability is excellent, greater than the commonly recommended minimum value for Cronbach's alpha (0.70) [13]. A strong split-halves reliability is further evidence of internal consistency.

Construct validity has been demonstrated. The parallel measure, NSS, correlates strongly with the IMESQ. Although not strictly a direct measurement of satisfaction, client Willingness correlates moderately with the IMESQ. Willingness may be subject to confounding factors such as indecision or client bias. In the absence of a correlation with Outlier, it cannot be concluded that the responses to Willingness in this sample were biased. Some client might have difficulty expressing dissatisfaction with the service while they were still in the office, which could result in overestimated satisfaction of care. Nonetheless, the findings indicate that the IMESQ can be used to obtain reliable and valid information of satisfaction with an IME.

At ease was a predictor of Willingness and Outlying scores. "Interpersonal manner" appears to be the dominant determinant of client satisfaction on factor analysis. Communication skills, putting the client at ease and being respectful were dominant determinants of client satisfaction.

With the exception of chronological age and being unqualified, scores on the IMESQ were not influenced by the other socio-demographic characteristics. The association between chronological age and the IMESQ was small indicating that the score was better with age but not specifically for clients over 40 years of age. Other researchers have noted that satisfaction increased with age [31, 32]. The IMESQ was negatively associated with having no educational qualifications and may reflect a pessimistic outlook on returning to work. Furthermore, the IMESQ was not influenced by quality of life/general health status. This contrasts with other researchers who found that

satisfaction correlated with the health status and emotional well-being [33-35].

The client who is angry and frustrated by the cause of the injury (blame), medical management, and/or processing of the claim may intentionally or subconsciously underrate the scores on the questionnaire and produce low IMESQ scores due to dissent bias. From the data, it has been shown that the responses can be discordant. A small group, all male, was shown to underrate their responses, and another group, predominantly female, to exaggerate their responses. The addition of suitably a worded question representing a "Dissatisfaction" domain directed to assessing dissatisfaction with treatment and management might yield useful information of circumstances beyond the scope of the IME that influence the responses to the questions. Given that Outliers had a difference of 4 between the IMESQ and Surrogate score calculated from the NSS, an additional 5-point Likert scale question with reverse scores could be considered to gauge the impact of dissatisfaction as a confounder on client responses. A reverse score for a Dissatisfaction item may not alter the scoring balance.

a) Limitations

The study has some fundamental limitations which probably do not invalidate the conclusions. The study is an underpowered, small cohort of clients and one examiner. The study needs to be expanded to include several examiners and translated into other non-English languages.

Parametric statistical methods are not appropriate when using Likert scales as they violate the assumptions of use. Studies have shown that parametric statistics accommodate the violations of these assumptions and can be used without the concern of answers being invalidated [12]. Criticism of the statistical analysis may be well founded.

The Likert responses for each item were not rated equally. Only two of the six items had a central neutral response. The inclusion of a neutral response may represent neutral bias, true indecision or uncertainty. The use of an equivalent rating scale to all questions including a neutral central response would make for approximately equivalent weighting of all scales.

The use of balanced six or seven point Likert scales with more extreme anchors could reduce negative skew and the ceiling effect which was 42% to 67% across all items in this model. However, the purpose of the questionnaire is not the identification of high achievement but marginal or poor performance areas which can be addressed and improved. There is probably no advantage beyond a five-point Likert scale with a central neutral response if the purpose of the questionnaire is to identify suboptimal performance. High ceiling effects indicate a limited instrument range,

possible response bias, and inadequate questionnaire performance [36]. The floor effect of less than 15% for all items is within the recommended range [37].

Additional domains could be considered to expand the scope of assessment of the IME process. A question about the initial "Introduction" which makes the client aware of the process and improves the interpersonal relationship has value. The inclusion of a question to measure the clarity and use of "Common language" has strong merit. Doctors often use medical language with which the client may not be familiar and the subsequent response to an inquiry may be inadvertently inappropriate.

The issue of bias frequently occurs when using Likert scales. The high ceiling in the individual item scales more than likely has an element of acquiescent bias. This study has shown evidence of both underrated and exaggerated responses. If suboptimal measures are defined as a total IMESQ score less than 25 and a NSS score less than 8 (these being the first quartile), 10 (13%) respondents met these criteria. This would seem a reasonable threshold level to raise concerns.

V. CONCLUSIONS

In conclusion, scores on the IMESQ questionnaire can be used as a measure of global satisfaction with an Independent Medical Examination. The questionnaire is valid and has good internal consistency, reliability and reproducibility. The questionnaire is not influenced by socio-demographic, general health or psychological distress. All six items represent six different domains. Where the results of the questionnaire do not meet the level of "satisfactory", the weaker aspects of the IME examination can be identified and addressed. The IMESQ fulfils the criteria of a 360-degree audit. It satisfies the requirements of the various professional bodies for Continuing Professional Development. It may, on occasions be useful evidence in medical defence of a claim relating to the IME, if the allegation is not concordant with the responses given on the questionnaire.

Abbreviations

IMESQ	Independent Medical Examination Satisfaction Questionnaire
Item	Question, domain
NSS	Numerical satisfaction scale

REFERENCES RÉFÉRENCES REFERENCIAS

- Prakash B. Patient satisfaction. *J Cutan Aesthet Surg*. 2010 Sep; 3 (3): 151-5. doi:10.4103/0974-2077.74491. PMID: 21430827; PMCID: PMC3047732. <https://doi.org/10.4103/0974-2077.74491>.
- Schoenfelder, T., Klewer, J., & Kugler, J. (2011). Determinants of patient satisfaction: a study among 39 hospitals in an in-patient setting in Germany. *International journal for quality in health care : journal of the International Society for Quality in Health Care*, 23(5), 503–509. <https://doi.org/10.1093/intqhc/mzr038>.
- Ong, L. M., de Haes, J. C., Hoos, A. M., & Lammes, F. B. (1995). Doctor-patient communication: a review of the literature. *Social science & medicine (1982)*, 40 (7), 903–918. [https://doi.org/10.1016/0277-9536\(94\)00155-m](https://doi.org/10.1016/0277-9536(94)00155-m).
- Nowak P. (2011). Synthesis of qualitative linguistic research--a pilot review integrating and generalizing findings on doctor-patient interaction. *Patient education and counseling*, 82 (3), 429–441. <https://doi.org/10.1016/j.pec.2011.01.022>
- Loblaw, D. A., Bezjak, A., & Bunston, T. (1999). Development and testing of a visit-specific patient satisfaction questionnaire: the Princess Margaret Hospital Satisfaction with Doctor Questionnaire. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 17 (6), 1931–1938. <https://doi.org/10.1200/JCO.1999.17.6.1931>.
- Loblaw, D. A., Bezjak, A., Singh, P. M., Gotowiec, A., Joubert, D., Mah, K., & Devins, G. M. (2004). Psychometric refinement of an outpatient, visit-specific satisfaction with doctor questionnaire. *Psycho-oncology*, 13 (4), 223–234. <https://doi.org/10.1002/pon.715>.
- Hagedoorn, M., Uijl, S. G., Van Sonderen, E., Ranchor, A. V., Grol, B. M., Otter, R., Krol, B., Van den Heuvel, W., & Sanderman, R. (2003). Structure and reliability of Ware's Patient Satisfaction Questionnaire III: patients' satisfaction with oncological care in the Netherlands. *Medical care*, 41(2), 254–263. <https://doi.org/10.1097/01.MLR.0000044904.70286.B4>.
- Hawthorne, G., Sansoni, J., Hayes, L., Marosszeky, N., & Sansoni, E. (2014). Measuring patient satisfaction with health care treatment using the Short Assessment of Patient Satisfaction measure delivered superior and robust satisfaction estimates. *Journal of clinical epidemiology*, 67(5), 527–537. <https://doi.org/10.1016/j.jclinepi.2013.12.010>.
- Romain-Glassey, N., Gut, M., Cathieni, F., Hofner, M. C., & Mangin, P. (2011). A satisfaction survey conducted on patients of a medico-legal consultation. *Journal of forensic and legal medicine*, 18 (4), 158–161. <https://doi.org/10.1016/j.jflm.2011.02.002>.
- Dancey C.P., Reidy J. Pearson Education; 2007. *Statistics without Maths for Psychology*. Eighth edition. Harlow, England; New York: Pearson, 2020. Includes bibliographical references and index. Identifiers: LCCN 2019052888 ISBN 9781292276434 (hardback) ISBN 9781292276458 (epub)
- Parmenter, K., Waller, J., & Wardle, J. (2000). Demographic variation in nutrition knowledge in

- England. *Health education research*, 15(2), 163–174. <https://doi.org/10.1093/her/15.2.163>.
12. Norman G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education: theory and practice*, 15 (5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>.
 13. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York, NY: McGraw-Hill; 1994. p. 736. Identifiers: ISBN-10:007047849X ISBN-13:978-0070478497.
 14. Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86 (2), 420–428. <https://doi.org/10.1037//0033-2909.86.2.420>.
 15. McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1 (1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>.
 16. Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, 15 (2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.
 17. Liljequist, D., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation - A discussion and demonstration of basic features. *PloS one*, 14(7), e0219854. <https://doi.org/10.1371/journal.pone.0219854>.
 18. Lin L.I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45 (1), 255–268.
 19. Lin, L., Hedayat, A.S., Sinha, B, and Yang, M. 2002. Statistical Methods in Assessing Agreement. *JASA*, 97(457), 257-270.
 20. Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England)*, 1 (8476), 307–310.
 21. Giavarina D. (2015). Understanding Bland Altman analysis. *Biochemiamedica*, 25 (2), 141–151. <https://doi.org/10.11613/BM.2015.015>.
 22. Kwicien, R., Kopp-Schneider, A., & Blettner, M. (2011). Concordance analysis: part 16 of a series on evaluation of scientific publications. *Deutsches Arzteblatt international*, 108(30), 515–521. <https://doi.org/10.3238/arztebl.2011.0515>.
 23. Gardner, P. L. Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Research in Science Education*. 1995, 25 (3), 283–289. doi:10.1007/bf02357402.
 24. Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34 (1), 111–117. <https://doi.org/10.1177/001316447403400115>.
 25. Tobias, S., & Carlson, J. E. (1969). BRIEF REPORT: BARTLETT'S TEST OF SPHERICITY AND CHANCE FINDINGS IN FACTOR ANALYSIS. *Multivariate behavioral research*, 4(3), 375–377. https://doi.org/10.1207/s15327906mbr0403_8.
 26. Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*, 86 (2), 127–137.
 27. Camilli, G., & Shepard, L. A. (1987). The Inadequacy of ANOVA for Detecting Test Bias. *Journal of Educational Statistics* 1987, 12(1): 87-99. <https://doi.org/10.2307/1164630>.
 28. McBride GB. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Report: HAM 2005–062. National Institute of Water & Atmospheric Research Ltd: Hamilton, New Zealand, May 2005. <https://www.medcalc.org/download/pdf/McBride2005.pdf>
 29. Martins, W. P., & Nastri, C. O. (2014). Interpreting reproducibility results for ultrasound measurements. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 43 (4), 479–480. <https://doi.org/10.1002/uog.13320>.
 30. Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M., & Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of clinical epidemiology*, 64 (1), 96–106. <https://doi.org/10.1016/j.jclinepi.2010.03.002>.
 31. Jenkinson C, Coulter A, Bruster S, Richards N, Chandola T. Patients' experiences and satisfaction with health care: results of a questionnaire study of specific aspects of care. *Qual Saf Health Care*. 2002 Dec; 11 (4): 335-9. doi: 10.1136/qhc .11.4.335. PMID: 12468693; PMCID: PMC1757991.
 32. Nguyen Thi, P. L., Briançon, S., Empereur, F., & Guillemin, F. (2002). Factors determining inpatient satisfaction with care. *Social science & medicine (1982)*, 54 (4), 493–504. [https://doi.org/10.1016/s0277-9536\(01\)00045-4](https://doi.org/10.1016/s0277-9536(01)00045-4).
 33. Andoh, M., Kobayashi, K., Kudoh, S., & Takeda, F. (1997). *Nihon Ika Daigakuzasshi*, 64 (6), 538–545. <https://doi.org/10.1272/jnms1923.64.538>.
 34. Kerr, E. A., Smith, D. M., Kaplan, S. H., & Hayward, R. A. (2003). The association between three different measures of health status and satisfaction among patients with diabetes. *Medical care research and review: MCRR*, 60 (2), 158–177. <https://doi.org/10.1177/1077558703060002002>.
 35. Marshall, G. N., Hays, R. D., & Mazel, R. (1996). Health status and satisfaction with health care: results from the medical outcomes study. *Journal of consulting and clinical psychology*, 64(2), 380–390. <https://doi.org/10.1037//0022-006x.64.2.380>.

36. Bernstein, D. N., Houck, J. R., & Hammert, W. C. (2019). A Comparison of PROMIS UE Versus PF: Correlation to PROMIS PI and Depression, Ceiling and Floor Effects, and Time to Completion. *The Journal of hand surgery*, 44 (10), 901.e1–901.e7. <https://doi.org/10.1016/j.jhsa.2018.12.006>.
37. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, 60 (1), 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>.
38. Doğan N. Ö. (2018). Bland-Altman analysis: A paradigm to understand correlation and agreement. *Turkish journal of emergency medicine*, 18 (4), 139–141. <https://doi.org/10.1016/j.tjem.2018.09.001>.
5. How good was this doctor at providing time for your consultation? (plenty of time, not rushed)
 - o Very good
 - o Good
 - o Satisfactory
 - o Poor
 - o Very poor
6. Overall, how satisfied are you with your consultation? (Please check only one box)
 - o Very satisfied
 - o Satisfied
 - o Neither satisfied or dissatisfied
 - o Dissatisfied
 - o Very dissatisfied
7. On a 0-to-10 NAS scale how satisfied are you with the consultation where 0 = absolutely dissatisfied and 10 = almost satisfied.

0 1 2 3 4 5 6 7 8 9 10

Verv dissatisfied Verv satisfied

Appendix

IMESQ

1. How good was this doctor at making you feel at ease? (polite, being friendly towards you, treating you with respect; not cold, abrupt or apparently disinterested)
 - o Very good
 - o Good
 - o Satisfactory
 - o Poor
 - o Very poor
2. How good was this doctor at letting you tell your "story" and listening? (paying attention, giving you time to fully describe your illness in your own words)
 - o Very good
 - o Good
 - o Satisfactory
 - o Poor
 - o Very poor
3. How good was the doctor at very carefully checking everything when examining you? (meticulous and thorough)
 - o Very good
 - o Good
 - o Satisfactory
 - o Poor
 - o Very poor
4. How confident are you that this doctor fully understanding your condition... (that he had accurately understood your concerns; not overlooking or dismissing anything)
 - o No, definitely not
 - o No, generally not
 - o Uncertain
 - o Yes, generally
 - o Yes, definitely
8. Would you be willing to be examined by this doctor again?
 - o No
 - o Maybe
 - o Yes

Table 1: Details of the Independent Variables

	Frequency	% of Cases
Partnered	53	69.74%
Australian born	50	65.79%
Working	48	63.16%
Previous IME	36	47.37%
Male	34	44.74%
<40yo	27	35.53%
Unqualified	27	35.53%

Table 2: Matrix of Inter-Item Correlations

	At ease	Tell story & listening	Examination	Information exchange	Time	Satisfaction
At ease	1	0.70	0.72	0.58	0.72	0.57
Tell story & listening	0.70	1	0.80	0.59	0.76	0.69
Examination	0.72	0.80	1	0.59	0.67	0.55
Information exchange	0.58	0.59	0.59	1	0.60	0.59
Time	0.72	0.76	0.67	0.60	1	0.61
Satisfaction	0.57	0.69	0.55	0.59	0.61	1

Table 3: The Cronbach's Alpha Analysis for the Six Item IMESQ

	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
At ease	0.8	0.89
Tell story & listening	0.81	0.89
Examination	0.73	0.90
Information exchange	0.68	0.91
Time	0.82	0.89
Satisfaction	0.72	0.90



Table 4: Data from the Principle Component Analysis

Component	Total	% of variance - Eigenvalue	Accumulated %	Extraction	Component matrix - Component 1	Rotated Component Matrix – Component 1
1	4.24	70.64	70.64	0.76	0.87	0.87
2	0.52	8.67	79.31	0.77	0.88	0.88
3	0.47	7.86	87.16	0.67	0.82	0.82
4	0.37	6.09	93.25	0.6	0.78	0.78
5	0.23	3.8	97.05	0.79	0.89	0.89
6	0.18	2.95	100	0.65	0.81	0.81

Table 5: Distribution of Outliers According to Gender

		Female	Male	Total
Under-rater	-1	0	5	5
	0	36	26	62
Over-rater	1	6	3	9
	Total	42	34	76

Table 6: Results of Dunn-Bonferroni Test for Under/Over Outliers (-1/1)

	Test Statistic	Std. Error	Std. Test Statistic	p	Adj. p
0 : -1	-22.06	8.84	-2.5	.013	.038
0 : 1	3.27	6.78	0.48	.63	1
-1 : 1	25.33	10.61	2.39	.017	.051

Adj. P: Values Adjusted with Bonferroni Correction



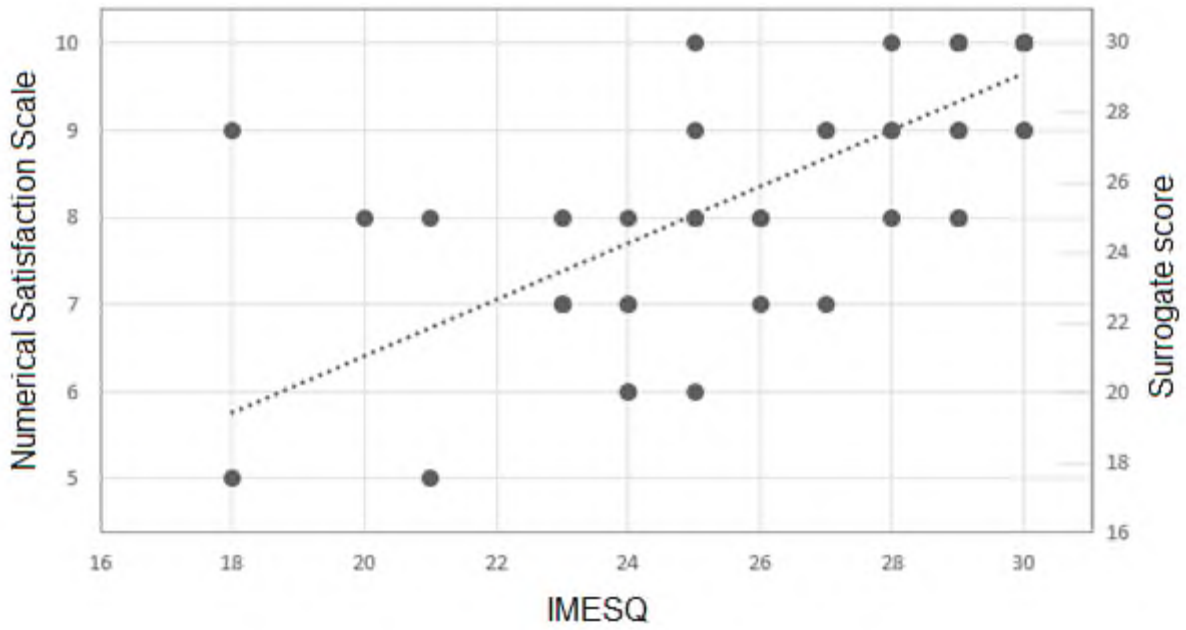


Figure 1: Composite Regression Plot of IMESQ and NSS / Surrogate Scores. For the Purpose of Interpreting Lin's CCC, the Trendline can be Compared with the "Gold Standard" (16,16 To 30,30)

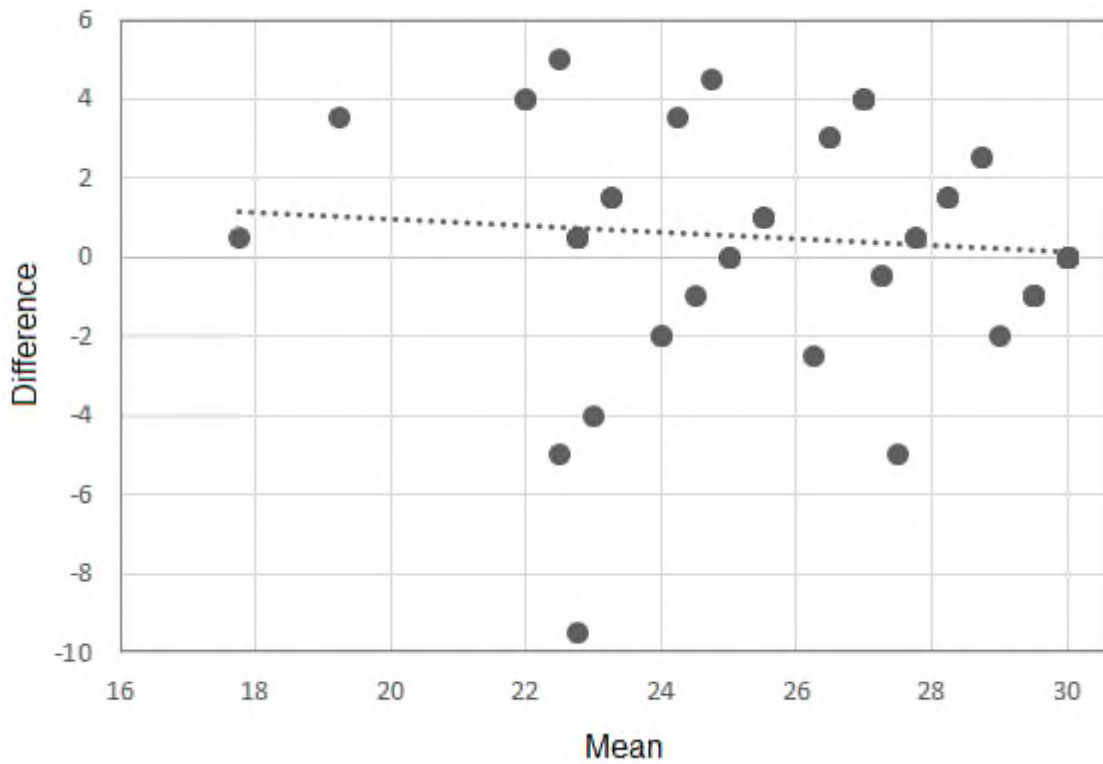


Figure 2: Bland-Altman Plot: Limits of Agreement $<-4.1, >4.1$; Outliers $<-2.4, >2.3$. The Trendline is an Index of Bias [38]