PSYCHOMETRIC ANALYSIS AND VALIDATION

OF THE

SOUTH AFRICAN ADDICTION TREATMENT ASSESSMENT

MEASUREMENT INSTRUMENT

(SAATSA)

FINAL REPORT

13 September 2012

Prof. Rajen Govender

KhyAna

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER1: INTRODUCTION AND OVERVIEW

This report details and discusses the psychometric analysis of the South African Addiction Treatment Service Assessment measurement instrument (SAATSA). The SAATSA was developed by the Service Quality Measurement (SQM) initiative, a project funded by the Centres for Disease Control and located at the Medical Research Council (MRC). The SQM and the SAATSA was initiated to address the glaring lack of proper metrics for the assessment of service quality in the treatment of substance abuse in South Africa. The initiative is based on similar efforts in the United States, and began its work by adoption of some of the learning and measurement protocols from the US work.

The SQM project structure is constituted of a National Steering Committee with three working groups: the SAATSA Working Group, an Administrative Data Working Group, and a Policy Working Group. The SAATSA Working Group undertook the primary work to develop the SAATSA, with input received from the Administrative Data Working Group and the members of the National Steering Committee and ordinary SQM members. Based on these deliberations, a questionnaire was developed and subject to cognitive testing to examine and correct for language, age, education and other respondent demographic effects. The outcome of the cognitive testing was a set of 26 evaluative items indicating attitudes towards various aspects of service quality in substance abuse treatment. These items were formulated into a questionnaire together with additional questions pertaining to respondent demographics and relevant information about their substance abuse profile and background (see Appendix A). This questionnaire constituted the survey instrument for the pilot study.

The SAATSA pilot study was conducted with the co-operation of various substance abuse treatment facilities and their respective administrative authorities in two provinces in South Africa. The purpose of the pilot was to assess the validity and reliability of the SAATSA, as a precursor to the formalisation of a measurement instrument for potential deployment at facilities nationwide. To this end, various initiatives relating to policy and legislative imperatives and requirements are already underway in the SQM. The finalisation of a proper measurement instrument is deemed essential to facilitating discussion and negotiations with relevant roleplayers and stakeholders in the governmental, non-governmental and private sectors.

As the likely audience for this report is quite diverse, every attempt has been made to report the analyses and corresponding outputs in the most non-technical manner possible, including the use of graphic representations of outputs and obtained psychometric properties (though due consideration was given to the specific reporting requirements of the different quantitative analyses). Additionally, and where necessary, detailed explanations of relevant statistical terminology, theoretical foundations and analytic techniques are supplied, and the reader may elect at their discretion to study or overlook these.

Various recommendations are provided throughout the report for refinement and improvement of the SAATSA. Some of these relate to research design and methodology (sampling and fieldwork administration) and others relate to measurement design and construction (questionnaire wording, format, routing, etc.). These recommendations are intended to be addressed both jointly and severally as they all have import for the SAATSA measurement instrument and the protocols necessary in future data collection and data processing.

This report is structured as follows:

Chapter 2: Sampling and Fieldwork – this chapter provides an overview of the pilot study research design, including sampling, fieldwork administration and data processing.

Chapter 3: Data Validation and Univariate Analysis – this chapter raises and discusses various issues relevant to validation and vetting of the data, with a specific focus on missing data and the analysis of the univariate distributions of the SAATSA items.

Chapter 4: Sample Characteristics – this chapter provides a brief overview of the pilot sample in terms of key demographic and other features, and is descriptive and not analytic in nature.

Chapter 5: Methodology for the Psychometric Analysis - this chapter provides an overview of the methodology employed for the psychometric analysis of the data, with particular emphasis on the measurement theories employed and the techniques that emanate thereof.

Chapter 6: Psychometric Analysis of the SAATSA - this chapter provides the detailed psychometric analysis of the SAATSA in terms of its hypothesized domains, dimensions and the individual items.

Chapter 7: Conclusion - this chapter concludes the report and highlights and addresses the various recommendations for improvement of the SAATSA measurement instrument.

CHAPTER 2: SAMPLING AND FIELDWORK

This chapter outlines and discusses the sampling frame employed for the research and various issues relevant to fieldwork and survey administration. All fieldwork began in September 2011 and was concluded in July 2012.

Sampling Criteria

To ensure good coverage of the various types of substance abuse treatment facilities in the country, the sample was stratified according to the following criteria:

1. Facility Administration – Both state owned and operated and non-state owned and operated facilities were selected for the study. This is principally because of the prevalence of both these types of facilities in the treatment of substance abuse in the country.
2. Admission Type – To account for the differential nature and characteristics of substance abuse presenting symptoms and treatment programmes in the inpatient and outpatient admission of patients, the sample was structured to capture data for both admission types.
3. Patient Demographics – The sample of facilities needed to account for the variable demographic characteristics of the national population – such as ethnic group, education and language – and the spatial distribution of these characteristics.

Sampling Design

The research was conducted using purposive sampling. Following from the sampling frame, the sampling of treatment facilities was based on the access afforded to the research team by relevant organisational/institutional administration and facility staff. Access to the various facilities was negotiated and obtained by the Medical Research Council either by way of the relevant provincial authorities in the case of government owned/administered facilities or the relevant organisational administrations in the case of non-governmental facilities and/or facilities owned by the state but operated on an agency basis. The primary non-governmental organisation selected for provision of access to their facilities was the South Africa National Council on Alcoholism and Drug Dependence (SANCA), an organisation established for over 50 years and with a footprint in all nine provinces in South Africa. As will be seen later, the long lead time to securing the requisite permission and access for some facilities resulted in considerable variability in the number of completed surveys from individual facilities.

The specific sampling strategy was motivated by three primary considerations:

1. Access – systematic and rigorous service quality assessment in the treatment of substance abuse is a nascent development in South Africa. Consequently, there was a requirement for significant

effort by the SQM to ensure the co-operation and participation of administrators and practitioners and their respective facilities. To circumvent the protracted communication and media process that would be necessary to secure access at facilities generally, it was decided to engage with organisations and facilities where access could be more readily negotiated.

2. Budget - for reasons of a limited budget, the sampling was restricted to facilities located in two major metropolitan areas in two provinces in the country, these being the Durban and Pietermaritzburg metropolitan areas in KwaZulu-Natal and the Cape Town metropolitan area in the Western Cape.
3. Logistical Support - additionally, these locations were chosen because of the presence of the Medical Research Council in these areas, which ensured proper logistical support for the fieldwork process and research personnel.

The purposive sampling strategy was deemed acceptable as the objective of the research was to test and validate the psychometric properties of the SAATSA without the need to generalise these findings or any of the analyses to the wider relevant population. Accordingly, this data is in no way representative - nor presented as being representative - of the adult population of substance abuse treatment facility patients in either of these provinces or the country as a whole.

Fieldwork And Questionnaire Administration

Fieldwork was undertaken by personnel from the MRC, with assistance from staff at the various treatment facilities. Once access had been approved and provided, site visits were undertaken by research personnel to brief the facility staff on the research and the survey and to prepare the facility for the administration of the survey questionnaire, including where necessary the provision of training to staff.

All surveys were self-administered, and patients were identified and approached for the survey by facility staff and research personnel. For this reason, it was essential to establish protocols for determination of patient eligibility for the research. The principal requirement was that patients would have been in the facility for a sufficient period of time to make informed judgements about service quality. As the treatment regimens varied in duration at different facilities, it was decided that patients would be targeted only after they had completed the majority portion (70-80% or more) of their designated treatment programme. For those in 16 week programmes, this meant having completed at least 12 weeks, while those in 5 week programmes were required to have completed 4 weeks and those in three week programmes were required to have completed more than two weeks. Further, it was decided that all patients completing the survey would still be enrolled in their treatment programmes (either on an inpatient or outpatient basis) and would not have completed the programme. This was considered important as the assessment of service quality can and does vary considerably based upon whether or not patients are still enrolled or have already left the programme or the facility. Treatment facilities furnished all completed questionnaires immediately to the MRC or stored them on their premises for later submission to the MRC. The MRC captured all questionnaires electronically and developed the dataset for analysis of the pilot study.

CHAPTER 3: DATA VALIDATION AND UNIVARIATE ANALYSIS

Data validation or vetting is a process whereby the data is examined to verify that it is free of errors. The overarching objective is to ensure that the data, or more specifically the distribution of relevant variables measured by the data, is unaffected by values which are not a true reflection of the measurement of each of the variables. To do this requires a combination of various techniques, both conceptual and analytic. The specific purpose of these various techniques is to examine the data and to establish if the observed variable distributions are as expected, whether or not they are unduly affected by any known or unknown systematic bias, the potential reasons for such bias and their implications for variable measurement. Further, if such biases or errors are present, the process addresses how to resolve them so that the bivariate and multivariate analysis required for psychometric validation of the instrument are immunised from such effects.

Errors may present themselves in various forms in datasets. For instance, errors may be considered as data that is inadmissible for the relevant variable, that is, the particular values presented for some of the respondents on a particular variable may be inadmissible in terms of the a priori configuration of that variable measurement. The SAATSA items were all formulated as attitudinal questions with a Likert type response scale ranging from 1 (Disagree) to 4 (Strongly Agree) (see Appendix A). If a value outside this range is present in the dataset – such as 7 - then it is inadmissible in terms of the analyses. The most common reason for such errors are entry and coding mistakes, wherein data processors mistakenly enter the wrong value, and unless the data entry method has been programmed to detect and reject such inadmissible values, they can easily be punched into the dataset. Inadmissible values are clearly problematic because they reflect values on a variable for which there is no substantive meaning.

Similarly, missing values may also present problems to data, though in different ways. Missing values are typically values for which data is not present because the respondent missed or overlooked the item. Hence there is no response to code in the dataset. Alternatively, missing values are present when obtained values are inadmissible and consequently are recoded to appear as missing in the dataset. Regardless of origin, all missing values share the common feature of being excluded from any analyses conducted. In some instances missing values are perfectly legitimate and hence pose no problems in the data. For instance, if the questionnaire featured question routing to account for the fact that some questions are not relevant or applicable to certain respondents (questions pertaining to pregnancy are for women and not men, questions on salaries are applicable only to those earning an income, etc.), then their missing values for these particular questions are valid and regarded as such. However, when such reasons do not exist, then missing values are problematic. In the case of the SAATSA, all questions were applicable to all respondents and hence responses were required on every question. Despite this, and as will be seen later on, there were a number of missing values for individual items in the SAATSA.

All research assumes some degree of risk in the return of data. It is thus expected that some quota of missing data will be present in the dataset purely by chance factors alone. In such cases the distribution of these missing values is generally also randomly distributed. When values are missing at random (MAR), the effect on the univariate distribution of variables is normalised and thus the impact on the variable distribution is minimal. In such instances the missing values are accepted and those cases for which they are applicable are legitimately excluded from any analysis. The only critical issue here is whether or not the quantum of missing values is sufficiently large so as to compromise sub-sample analysis by depleting the number of cases available for analysis.

However, when missing values are not MAR, they require proper attention as the lack or absence of randomness implies that there is an undetected systematic effect which is affecting the data. There could be various reasons for this. Firstly, the item itself could be problematic (poor wording or phrasing, too much ambiguity, etc.), and hence respondents could not provide a response to it. Alternatively, the item could only be understood by respondents with particular level of education, hence it indicates a systematic effect for respondent education. Additionally, the item may unknowingly or unintentionally be relevant to only some respondents and not to others (requesting information on condom use for respondents who are in monogamous relationships) and this produces missing data for some respondents based on specific non-random characteristics. Regardless of origin, the presence of non-MAR missing values indicates the need for investigation and corrective action prior to the conducting for substantive analyses, and this approach was adopted for the SAATSA.

In addition to establishing errors and inadmissible values, or generally after one has done so, the variable distributions need to be examined to establish if they display univariate normality. Univariate normality is a precursor to multivariate normality, an essential requirement for the latent trait analysis and factor analyses as was conducted in this study. The importance of univariate normality is best exemplified by alluding to the variance of a distribution. Normal distributions have good variance properties, and since all multivariate statistics are focussed on the analysis of variance, the greater the pool or variance available for analysis, the better the power to detect significant effects. Conversely, the more shrunken a distribution, the more diminished the variance and the lesser the power to detect effects. Establishing univariate normality is thus the best first line defence against potential problems that might be encountered during multivariate analysis.

Assessments of univariate normality are generally undertaken by examining distribution characteristics such as skewness and kurtosis. Skewness refers to the extent to which the values of the distribution are clustered towards one or the other end of the distribution, while kurtosis refers to the height of the distribution, that is, the extent to which the frequency of values are spread out (a flat distribution) or bunched together (a tall distribution). In the real world study variables rarely achieve perfect univariate normality. Fortunately, normality as a basic assumption of parametric statistics is easier to achieve that is imagined, thanks largely to the robust nature of available

techniques that analyse variance. In other words, the techniques are forgiving of variations from true normality, and thus variable distributions may show skewness and kurtosis and yet be accepted as being not significantly different from normality. The critical consideration is thus not the deviation from normality but rather the degree of such deviation, and skewness and kurtosis provide good indicators of this.

Missing Values Analysis

For logical reasons, missing values analysis must necessarily precede analysis of univariate distribution characteristics, as values that are missing and thus inadmissible have a direct impact on the nature of these distributions. In the case of the SAATSA, this analysis was applicable to all the SAATSA scale items as well as critical demographic items which are relevant to the psychometric analysis of these items, such as gender, ethnic group, etc. To give effect to the missing values analysis it was decided to employ a generally accepted cut-off criterion: items which have missing values for 5% or more of respondents were deserving of proper consideration, while proportions of missing values less than 5% were regarded as being acceptable and not necessarily deserving of further attention. In all instances, however, the missing values needed to be normally distributed, that is, missing at random, though this requirement is optional for proportions well below 5%.

Demographic Variables

Examination of the obtained data for demographic variables revealed low levels of missing values for age (0%), gender (1.3%) and ethnic group/race (1.3%). In contrast, the missing values for education stood at 17.0%, and this merited some attention, particularly so as the level of education of respondents may be an important differentiator in the comprehension of the SAATSA items. To identify whether or not the missing values for respondent education were MAR, an analysis was undertaken cross referencing this variable on the other demographic variables. This was done to establish if the missing responses were pertinent to one of the other of the categories of these demographic variables, that is, respondents failed to provide their education level as a result of their gender, age, and so forth.

Examination of gender effects in the missing values for education revealed a slight but statistically non-significant deviation from the rest of the sample ($\chi^2 = 1.76$, $p>0.05$). Similarly, the distribution of missing values for education according to ethnic group was also consistent with the distribution of these ethnic groups in the sample ($\chi^2 = 2.3$, $p>0.05$). In other words, in both instances the missing values were as likely to occur as were the specific demographic categories, meaning that they were not systematic and actually missing at random. Or phrased differently, male and female respondents and those from all ethnic groups were equally likely to report missing values for their response to the education variable. Further, the distribution of missing values by age also revealed non-

systematic effects. In sum, then, while the proportion of missing values for education was found to be quite high (almost one in five respondents) there was little to suggest that these missing values were not randomly distributed amongst respondents in terms of their age, gender and ethnic group. This enabled the use of the education variable and the other demographic variables in the psychometric analysis, the only qualification being that one-fifth of the sample became unavailable for analysis in the instance where the education variable was included.

In addition to the base demographic variables, it may be argued that there are other variables which are highly relevant to the SAATSA items and thus needed to be considered in terms of the missing values for themselves as well as in relation to the education variable. The most relevant example of this in this research was that of admission type (inpatients versus outpatients). Examination of this variable revealed 0% missing values. Examination of the distribution of missing values for education across the admission types revealed a non-significant variation (χ^2 = 1.2, p>0.05), indicating that those admitted as outpatients were as likely to report missing values for education as those admitted as inpatients. Likewise, though probably less relevant than admission type, receipt of prior treatment was also examined for similar effects. The results once again indicated acceptable proportion of missing values for prior treatment (0.3%) and non-significant deviation for missing values for education amongst those who received treatment previously as compared to those who had not (χ^2 = 2.9, p>0.05). Hence both these variables were cleared for further analysis.

Furthermore, various other base variables were present in the dataset, such as date of admission, primary substance abused, and payment method. Missing values for the latter two variables were within acceptable proportions, while for date of admission it was much higher at 21.1%. The high rate of failure to recall the admission date is probably due to the patient's psychological state at the time of admission and/or a consequence of the abuse of specific substances which degrade memory effects. However, as these variables were included as background information and were not essential to the psychometric validation of the SAATSA it was not necessary to give further attention to them. For future application it is advisable to collect this information directly from the facility treatment records rather than relying on self-reporting by patients as such administrative records are a far more complete and reliable source for this type of data than are patients themselves.

Key recommendation: Data on patient admission, substance for which problematic use necessitated admission, date of admission and length of treatment, as well as any background information relating to prior treatment should be collected directly from the administrative records of the facility rather than from patients themselves as these are far more complete and reliable.

SAATSA Items

Examination of the missing values for the 26 SAATSA items revealed no item with a proportion of missing values exceeding 5% of the total response for that item (see Appendix B). Most items contained about 2-3% missing values, with only a minority containing missing values around 4%. In

these latter instances it may be argued that the missing value could be accepted as a valid response - such as for *Item 15: I am more able to practice safe sex* - an item which is arguably applicable to those respondents who were actually sexually active at the time of the survey. Overall, though, the quantum of missing values for all of the SAATSA items was reassuringly well below the critical threshold.

As indicated previously, missing values in a traditional sense refer to variable categories for which data is simply not present at all, either through omission of that item by the respondent and/or through failure to capture the value during data processing. However, values can be considered as missing in the instance when they reflect response categories for which there is no evident or discernible substantive meaning. A typical instance of this is in the case of the Not Applicable category in response formats. In instances where the category reflects a conceptually valid response, the value is accepted and submitted for relevant analysis. However, in the instance where such categories do not indicate conceptually valid responses, they must be reconsidered for transformation and/or exclusion as their inclusion in further analysis will clearly confound the psychometric properties of that variable.

In the case of the SAATSA, all items were measured using a 4 point Likert response scale, with numeric values and applicable Likert response categories as follows: 1 – Disagree; 2 – Somewhat Agree; 3 – Agree; 4 – Strongly Agree; and 5 – Not Applicable. The lattermost category is of note here. Indicating a Not Applicable response is conceptually valid and thus a perfectly acceptable response for items such as *Item 15: I am more able to practice safe sex* for the reasons discussed earlier on. Similarly, the response is valid for *Item2: I can afford my treatment*, as the patient may have been treated at the expense of the state, family or employer, and hence the costs of treatment are of little or no consideration to them. The same is also true for *Item 5: I have a say in deciding about my substance abuse treatment that I am receiving here*, as treatment facilities are not generally in the habit of allowing patients to determine the nature of their treatment.

In contrast to these items, the SAATSA contains various other variables for which the Not Applicable response category could not be considered as being conceptually valid. For instance, it is difficult to see why the response category is valid for *Item 9: The staff at the treatment centre are sensitive to my background,* as this variable seeks to assess the cultural sensitivity and reflexivity of staff at the facility to people from multi-ethnic and multicultural backgrounds, patently important in a country such as South Africa with its diverse population. Similarly, and given that all respondents were patients admitted for alcohol and substance abuse problems, it is difficult to understand why *Item 13: I am less likely to use alcohol and other drugs,* would merit a Not Applicable response. Moreover, even in the unlikely event that these responses, and similarly those to a number of other items, could be argued to have some conceptual validity, there is insufficient information in the survey to be able to propose and substantiate such a claim. In other words, given the available information it is impossible to establish 1) how and why the item is not relevant to the respondent and 2), how and

why the Not Applicable response provided represented some valid conceptual alternative to that provided by the other response categories of the Likert scale. Given this, these responses could only be regarded as being inadmissible, and consequently were treated as such.

Key Recommendation: Future SAATSA versions must exclude the "Not Applicable" option from the Likert response categories for all items where such an option contains no substantive meaning and/or where choosing such an option does not provide a conceptually valid response to the specific item for which it is applied.

Following from this reasoning, the 26 items of the SAATSA were revisited and all Not Applicable responses were recoded as missing. This changed the complexion of the missing values table as reported previously (in Appendix B). The new proportions of missing values reflecting the Not Applicable as missing is provided in Table 1.

TABLE 1: Proportions of Missing Values for SAATSA Items: Original and Recoded (to include Not Applicable)

SAATSA ITEMS	INITIAL MISSING VALUES	RECODED MISSING VALUES
1. The amount of time I had to wait to get services was acceptable to me	1.5%	5.0%
2. I can afford my treatment	3.6%	14.0%
3. The staff at this treatment centre treat me with respect	1.3%	2.3%
4. The people I went to for treatment services spent enough time with me	1.8%	3.1%
5. I have a say in deciding about my substance abuse treatment I am receiving here	2.8%	7.5%
6. The staff told me about services that will help me stay off drugs and alcohol	1.8%	3.6%
7. I am given a choice of services in this treatment centre	3.1%	7.8%
8. This treatment centre teaches me how to avoid getting HIV	2.8%	12.1%
9. The staff at this treatment centre are sensitive to my background	2.3%	4.7%
10. My general health is improving	3.4%	3.6%
11. I am better able to cope when things go wrong	3.9%	4.5%
12. I am better able to accomplish the things I want to do	4.2%	5.0%
13. I am less likely to use alcohol or other drugs	3.9%	6.1%
14. I expect to do better at work/finding work or at school	3.6%	6.1%
15. I am more likely to practice safe sex	4.5%	11.5%
16. There is some-one who cares about whether I am doing better	3.6%	5.0%
17. I have some-one who will help me when I have a problem	3.6%	4.7%
18. I have people in my life who are positive influence	3.6%	5.3%
19. The people who care about me are supportive of my treatment	4.5%	5.6%
20. My friends and family are more able to count on me	3.6%	5.0%
21. I have friends who are not using alcohol or drugs	4.2%	7.5%
22. I have some-one who will listen to me when I need to talk	2.3%	2.3%
23. I know that using alcohol and drugs is a problem for me	2.0%	2.3%
24. I need to work on my problems with alcohol and/or drugs	2.6%	3.4%
25. The treatment centre is helping me to recover from using drugs and alcohol	2.0%	2.0%
26. I would recommend this treatment centre to a friend	2.3%	3.4%

As is evident from Table 1, a number of SAATSA items were seen to have proportions of missing values that exceeded the 5% critical upper limit. As discussed earlier, these responses may be considered as legitimate for some of the variables, such as Item 2 (treatment is paid for by another party hence cost is not an issue for the patient) Item 15 (patient is not sexually active) and Item 8 (patient is not sexually active and/or active but in a committed monogamous relationship). Reassuringly, these were the only items for which the proportion of missing values was more than double the critical upper limit. Of the remaining 23 items, 13 had missing values for less that 5% of the total number of responses, and the remaining 10 items had missing value proportions which range from 5 to 10%.

In considering how to address the (recoded) missing values for the SAATSA items where such proportions exceeded 5%, it was decided that any further analysis and necessary data transformations would be guided by the fact that each of the 26 items represented a premium for the research and that their exclusion should be guided exclusively by the psychometric validation process. This is to say that any data transformations undertaken at this stage would have to be done in a manner that minimised the possibility of excluding items from the scale, so that all items were present for the latent trait analysis stage. To achieve this, it was decided that the missing value analysis would be approached from a different perspective, that is, by examining the distribution of missing values not in terms of item characteristics but in terms of respondent characteristics. This was based on the hypothesis that missing values may reflect problems related to specific respondents rather than the SAATSA items themselves. Such an approach permitted adjustments to the data by the dropping of cases rather than items, thereby allowing all SAATSA items to be submitted to the subsequent psychometric validation tests (though caution was necessary to prevent excessive depletion of the sample available for analysis)

Using this approach, an analysis was conducted using the Missing Values Analysis module of SPSS, with outputs set to examine for respondent effects. The summarised distribution of missing values for SAATSA items by respondents is reflected in Table 2.

TABLE 2: Distribution of Missing Values for SAATSA Items by Proportions of Respondents

Number of SAATSA Items with Missing Values	Proportion of Items	Proportion of Cases
No Items	0	59.1%
One Item	3.8%	19.5%
Two Items	7.7%	7.8%
Three Items	11.5%	4.7%
Four to Nine Items	15% - 35%	4.5%
More than Nine Items	35% -100%	4.4%
TOTAL		**100%**

As is evident from Table 2, for the vast majority of respondents (78.6%) there were either no missing values (59.1%) or missing values for only one SAATSA item (19.5%), thereby permitting use of these cases without adjustment. A further 7.8% of cases reported missing values for two items, which was also regarded as being acceptable given that some of these represented values recoded from Not Applicable to missing. A similar case was made for the 4.7% of cases which reported missing values for three SAATSA items, as many of these missing values were loaded onto the same items. These cases were thus deemed fit for inclusion in the further analysis without any transformations.

Looking at the remainder of the sample, it was found that almost one in ten respondents (8.9%) returned missing values for 15% or more of the SAATSA items (4 or more items). In half of these cases there were missing values for more than one third of the SAATSA items (9 items) and in some instances there were missing values for the majority or even all of the SAATSA items. In these latter instances the missing values were not principally that which had been recoded from the Not Applicable category, but rather items for which no data was present at all, that is, the respondent has not responded to a single SAATSA item. It is difficult to understand why this data was processed when there were clearly very few or even no valid responses available, and suggests the failure or absence of proper protocols for backchecking and validation. In any event, it was decided that the missing value quota for these respondents was simply too high to permit their inclusion in further analysis, and all cases which reported missing values for more than 15% of SAATSA items (35 cases in total) were permanently removed from the pilot sample.

Key recommendation: Further iterations of the SAATSA require proper and stringent protocols for backchecks and validation before and during data processing. Backchecks should be undertaken for at least 25% of all received surveys (the generally accepted minimum critical threshold for backchecks) and data validation during and after data processing must identify and account for cases which present large proportions of missing values as they clearly indicate problems in the administration of the survey.

Having resolved the respondent effects for missing values, a further missing values analysis was conducted to investigate once more for item effects. This analysis confirmed the validity of the approach focussing on respondent effects. The exclusion of the 8.9% of the sample which constituted the problematic cases resulted in only four SAATSA items registered missing values in excess of 5%, these being *Item 2: I can afford my treatment* (8.8%), *Item 7: I am given a choice of services in this treatment centre* (5.2%), *Item 8: This treatment centre teaches me how to avoid getting HIV* (7.4%) and *Item 15: I am more likely to practice safe sex* (5.5%). These proportions were considerably more acceptable than previously, and did not suggest any need for transformation of this data. Further, as discussed previously, there are entirely valid reasons why these items would not be relevant to some of the respondents and hence their missing responses can be accepted as being legitimate (although these cases were necessarily excluded from the subsequent analysis).

Treatment of these items in the future would require better routing in the questionnaire to properly account for divergent respondent attributes.

Key Recommendation: Future SAATSA iterations must incorporate proper question routing to ensure that items which are genuinely not applicable to certain respondents and/or facilities are regarded as such in the survey. This covers content areas such as cost of treatment, sexual behaviour, etc.

The removal of the problematic cases concluded the preparatory work on the data in terms of resolving issues related to missing values. The dataset was thus deemed suitable for further use in terms of the minimal requirement of being "clean".

A final note: as indicated previously, missing values are an inevitable aspect of research, and invariably appear in datasets. The inclusion of such cases compromises the integrity of obtained data and any analyses undertaken using such data, hence the need for proper missing values analysis. However, it is far more effective to address the matter in a preventative (a priori) rather than a curative (post hoc) basis. The extent to which missing values are present and the magnitude of impact they have on the validity of the data can and should be properly accounted for upfront by application of the correct protocols and procedures in measurement design and construction (encompassing question construction, construction of response scales and categories thereof, and questionnaire sequencing and routing) as well as in terms of fieldwork in general and survey administration in particular. In the latter case it is instructive to note that while the excluded cases were distributed proportionately across all treatment facilities, there was a significant difference in the distribution of these cases according to admission type (outpatient versus inpatient) with almost two thirds of these "problem" cases (63%) being respondents who were undergoing treatment as outpatients. Outpatients are by definition further removed from the facility than are inpatients, and hence are likely to need more convincing and/or closer supervision during the administration of the survey to ensure that they return fully completed surveys. Alternatively, or additionally, the rate of backchecks needs to be higher for outpatients than it is for inpatients to establish and correct for incomplete surveys.

Key Recommendation: Outpatients require improved survey administration protocols and a higher rate of backchecks to ensure better returns on returned surveys (completion rate).

Univariate Distributions

Once the problems posed by missing values had been resolved by removal of the problematic cases, it was possible to proceed to the analysis of the univariate distributions of the SAATSA items. The purpose of this analysis was to examine for any observations that indicated specific problems with the distribution of responses for these items. As discussed earlier, univariate normality is a necessary

though not a sufficient condition for the multivariate analyses, and thus essential prior to the psychometric analyses. Of particular importance for this stage of the analyses were the following:

1. The likely ceiling effects that was present for items, which might have posed a threat in terms of diminished variance for multivariate analysis.
2. The extent of deviation from normality of the univariate distributions, which might have rendered items unsuitable for inclusion in multivariate analysis.

Ceiling Effects

To begin with, the ceiling effects were examined. Ceiling effects refer to the phenomenon whereby the distribution of a variable is skewed towards certain response categories. In the case of the SAATSA the individual items are worded such that higher scores reflect greater agreement and accordingly more positive evaluations. Hence high concentrations of responses in categories such as Strongly Agree indicate uniformly positive evaluations. Should these responses be in disproportionate proportions relative to the other response categories, it indicates that respondents were unduly biased towards positive evaluations of the specific aspect of service quality being measured by the item.

Before proceeding, it must be noted that ceiling effects are not uncommon in consumer service quality assessment surveys, and more so for surveys which are conducted while the service is still being received, as in the case of the SAATSA. Accordingly, it was to be expected that respondents would provide far more positive than negative evaluations, and the only concern was the extent of such skewness towards these positive evaluations (the ceiling effects). To mitigate these anticipated potential ceiling effects the SAATSA items were developed using the learning obtained from the testing of similar survey items in the United States. In the initial round of the US survey the service quality items were scored using a five point Likert scale which ranged from Strongly Disagree to Strongly Agree, with neutral as the midpoint. However, Rasch modelling on that data revealed that the response scale could be reduced to four categories as few, if any, respondents indicated responses for the Strongly Disagree category. Inclusion of the Strongly Disagree category compelled respondents further towards the Strongly Agree side of the scale, showing up distributions as being very skewed in this (positive) direction. For this reason, the SAATSA response scale was confined to the four categories employed. The principal objective of doing so, given the naturally occurring bias towards the positive end of the scale, was to ensure that the four categories utilised provided sufficient differentiation in evaluation and hence sufficient (or more) variance for multivariate analysis.

Examination of the distribution of individual SAATSA items (see Appendix C) confirmed the generally positive bias in respondent evaluations of the various aspects of service quality. However, it is debatable whether or not the elimination of the extreme negative evaluation category (Strongly Disagree) fully curbed the likely ceiling effects, as most responses for the SAATSA items were either

3 (Agree) or 4 (Strongly Agree). This reflects a phase shift by respondents in the new response scale. That is, when the extreme negative response was eliminated, the net result was simply to push the bias across all existing categories, thereby still maintaining the general positive bias and likely ceiling effects. Nevertheless, and despite the apparent ceiling effects observed, all SAATSA items appeared to possess sufficient variation in the response categories, thereby facilitating their inclusion in the psychometric analysis.

Measures of Normality

The assessment of normality for the individual SAATSA items was conducted within the context of the preceding analysis, which clearly demonstrated the inherent positive bias in respondent responses on each of these items (positive evaluations of the service quality aspect). That is, based on the research conducted in the US as well as from this initial analysis it was already known and expected that these distributions would deviate from normality. The question posed in this part analysis was thus not whether or not such deviation was present, but rather the degree of deviation, and following from this, the assessment of whether or not this extent of deviation rendered the data in its raw form unusable in the later analysis.

Assessment of deviation from normality entails application of a number of different tests and techniques, each of which generates different estimates and output statistics. In approaching such analysis it is important to bear in mind that there are no absolute tests of such deviation. Different tests produce different results which attest to different aspects of normality testing, and they may contradict each other even when applied to the same data. Moreover, there are no absolute criteria for interpretation of some of these statistics, as such interpretation is subject to other considerations such as sample size, measurement scales, etc. Despite this variation in estimates and interpretation, these techniques do offer a more rigorous assessment of normality than does the visual inspection of histogram plots such as those provided in Appendix C.

For the purpose of testing the SAATSA items, it was decided that the test of normality would be conducted by examination of the properties of skewness and kurtosis. Skewness refers to how biased the scores are towards one or the other end of the distribution (the horizontal property – the SAATSA items can be seen to be skewed towards the right), while kurtosis refers to the height of the distribution (the vertical property – how flat or peaky the distribution is). The analysis of skewness and kurtosis statistics is available through SPSS, which provides values for the estimates for each of these as well as the standard errors for these estimates. In both instances the true (absolute value) is zero, and hence any deviation in the positive or negative direction away from zero indicates deviation from normality either or both horizontally and vertically.

While the skewness and kurtosis values are generally more informative when converted to standardised scores, in large samples such as this one (n>300) it is more useful to look at the actual values rather than their standardised counterparts as the increase in sample size typically reduces the standard error of the estimate and thus amplifies the size of the skewness and kurtosis estimates. As a general rule, the raw scores for skewness and kurtosis should be between -2 and +2 for one to conclude that the distributions do not deviate substantially from normality. Table 3 presents the skewness and kurtosis estimates for the SAATSA items, with the measures of central tendency also supplied for reference purposes.

TABLE 3: Estimates of Skewness and Kurtosis and Measures of Central Tendency for SAATSA Items

SAATSA ITEMS	Mean	Median	Skewness	Kurtosis
Item 1: The amount of time I had to wait to get services was acceptable to me	3.19	3.00	-1.013	1.583
Item 2: I can afford my treatment	2.77	3.00	-0.617	-0.620
Item 3: The staff at this treatment centre treat me with respect	3.45	4.00	-0.952	1.261
Item 4: The people I went to for treatment services spent enough time with me	3.43	3.00	-0.642	0.385
Item 5: I have a say in deciding about my substance abuse treatment that I am receiving here	3.16	3.00	-0.992	1.149
Item 6: The staff at this treatment centre told me about services in my area that will help me stay off drugs and alcohol	3.34	3.00	-1.057	1.344
Item 7: I am given a choice of services in this treatment centre	3.11	3.00	-1.028	0.735
Item 8: This treatment centre teaches me how to avoid getting HIV	3.17	3.00	-1.104	0.858
Item 9: The staff at this treatment centre are sensitive to my background	3.25	3.00	-0.982	1.214
Item 10: My general health is improving	3.45	3.00	-0.897	1.354
Item 11: I am better able to cope when things go wrong	3.28	3.00	-1.057	1.524
Item 12: I am better able to accomplish the things I want to do	3.39	3.00	-0.652	0.689
Item 13: I am less likely to use alcohol or other drugs	3.34	3.00	-1.298	1.757
Item 14: I expect to do better at work/finding work or at school	3.48	4.00	-0.807	0.550
Item 15: I am more likely to practice safe sex	3.48	4.00	-1.098	1.429
Item 16: There is some-one who cares about whether I am doing better	3.54	4.00	-0.771	0.122
Item 17: I have some-one who will help me when I have a problem	3.46	3.00	-0.681	0.480
Item :18 I have people in my life who are positive influence	3.45	4.00	-1.245	2.208
Item 19: The people who care about me are supportive of my treatment	3.53	4.00	-0.993	0.843
Item 20: My friends and family are more able to count on me	3.34	3.00	-1.045	1.187
Item 21: I have friends who are not using alcohol or drugs	3.11	3.00	-0.984	-0.006
Item 22: I have some-one who will listen to me when I need to talk	3.41	3.00	-1.206	2.173
Item 23: I know that using alcohol and drugs is a problem for me	3.61	4.00	-1.395	2.437
Item 24: I need to work on my problems with alcohol and/or drugs	3.41	4.00	-1.491	2.575
Item 25: The treatment centre is helping me to recover from using drugs and alcohol	3.59	4.00	-0.733	-0.673
Item 26: I would recommend this treatment centre to a friend	3.63	4.00	-1.455	2.884

As indicated by Table 3, while there was notable variation in the skewness estimates for the individual SAATSA items, none of these estimates exceeded the specified range of -2 to +2. Hence, and despite the observed positive bias in respondent evaluations as seen earlier on, we could conclude that the distributions of all of these items did not deviate significantly from normality.

Turning to the estimates of kurtosis, it was found that the majority of SAATSA items had kurtosis estimates that fell within the acceptable range. The exceptions were Items 18, 22, 23, 24 and 26, with Items 18 and 22 being somewhat marginal and hence acceptable. Inspection of Items 23, 24, and 26 confirmed the interpretation of the kurtosis estimates as all responses were almost entirely at the top end of these distributions, resulting in these distributions being taller than expected. However, as the skewness scores for these items do not show deviation outside the acceptable range, it was decided that these items could be still used in their current configuration as they contained enough variance to permit multivariate analysis. Nevertheless, a note was added to consider these items more intensively when the latent trait analysis was undertaken.

Overall, then, the univariate distributions appeared to be as anticipated given the nature of the SAATSA items, with most items recording positive evaluations and hence their distributions reflecting some degree of skewness and kurtosis. However, it was concluded that there was not sufficient deviation from normality for any of the distributions to warrant transformation of the data (such as log transformations) to achieve such normality. Hence all SAATSA items were cleared for psychometric analyses.

CHAPTER 4: SAMPLE CHARACTERISTICS

The initial sample for the pilot comprised 400 patients admitted to and receiving treatment at various substance abuse treatment facilities in the provinces of KwaZulu-Natal and the Western Cape in South Africa. Given the legal requirement to secure parental/guardian permission for the participation of minors (under 18 years of age), and the associated difficulty in securing this permission particularly in the case of substance abuse where familial relations are often strained and problematic, the pilot sample was restricted to adults. Nevertheless, one respondent indicated their age as 17, and this case was immediately excluded from the analysis. Furthermore, based on the missing values analysis reported in the preceding chapter, a further 35 cases were excluded from the dataset, realising a final pilot sample of 364 respondents. This final sample provided a more than adequate platform for the required psychometric analysis.

Geographic Location, Treatment Facility And Admission Type

As discussed previously, the pilot employed purposive sampling, hence no conclusions about representivity can be drawn from the obtained data, and the outputs are supplied purely for illustrative (descriptive) purposes. The distribution of respondents according to province and specific treatment facility is reflected in Table 4.

TABLE 4: Distribution Of Respondents Across Treatment Facilities

KWAZULU-NATAL		WESTERN CAPE	
SANCA Pietermaritzburg	46 (12.6%)	Kensington Treatment Centre	57 (15.7%)
SANCA Durban	88 (24.2%)	Toevlug	62 (17.0%)
		SANCA Tygerberg	41 (11.3%)
		SANCA Athlone	45 (12.4%)
		Tafelsig	7 (1.9%)
		Delft	4 (1.1%)
		Kenliworth Clinic	14 (3.8%)
TOTAL	**134 (36.8%)**	**TOTAL**	**230 (63.2%)**

The small number of respondents from Tafelsig, Delft and Kenilworth Clinic reflects the late approval granted for access to these facilities. Given the long lead time for this pilot, it was decided to forego any potential further respondents that may accrue from these facilities. The (small) number of these additional respondents did not justify the necessary delay in finalising the fieldwork and initiating data analysis, particularly so as progression to the policy and institutional implementation aspects of the project was contingent on validation and finalisation of the survey instrument.

The distribution of respondents by facility and admission types is reflected in Table 5, showing a sample slightly biased towards inpatients.

TABLE 5: Distribution Of Sample By Facility And Admission Type

FACILITY	ADMISSION TYPE	
	INPATIENT	OUTPATIENT
SANCA Pietermaritzburg	0 (0%)	46 (12.6%)
SANCA Durban	69 (19.0%)	19 (5.2%)
Kensington Treatment Centre	57 (15.7%)	0 (0.0%)
Toevlug	61 (16.8%)	1 (0.3%)
SANCA Tygerberg	0 (0%)	41 (11.3%)
SANCA Athlone	0 (0%)	45 (12.4%)
Tafelsig	5 (1.4%)	2 (0.5%)
Delft	0 (0%)	4 (1.1%)
Kenliworth Clinic	14 (3.8%)	0 (0%)
TOTAL	**206 (56.6%)**	**158 (43.4%)**

Sample Demographics

The realised sample ranged in age from 18 to 74 years, with the mean age being 32.5 years and the median age being 30 years. The sample was predominantly male (76%), predictably so given the well established higher preponderance of substance abuse amongst males (though it could also reflect the fact that males are more likely to present themselves at such facilities for treatment than are women). This male bias is relatively consistent at each of the different treatment facilities (see Table 6).

TABLE 6: Distribution of Males and Females Across Treatment Facilities

FACILITY	MALE	FEMALE
SANCA Pietermaritzburg	68.4%	31.6%
SANCA Durban	83.2%	16.8%
Kensington Treatment Centre	68.3%	31.7%
Toevlug	77.0%	23.0%
SANCA Tygerberg	75.6%	24.4%
SANCA Athlone	72.1%	27.9%
Tafelsig	85.7%	14.3%
Delft	75.0%	25.0%
Kenilworth Clinic	57.1%	42.9%

In terms of ethnic group, the largest proportion of respondents were Coloured (50.0%), followed by African respondents (27.2%) and then White and Indian respondents (13.3% and 9.4% respectively). The distribution of patients according to their ethnic group is reflective of the spatial distribution and location of these ethnic groups in the country, with the high proportion of Coloureds being

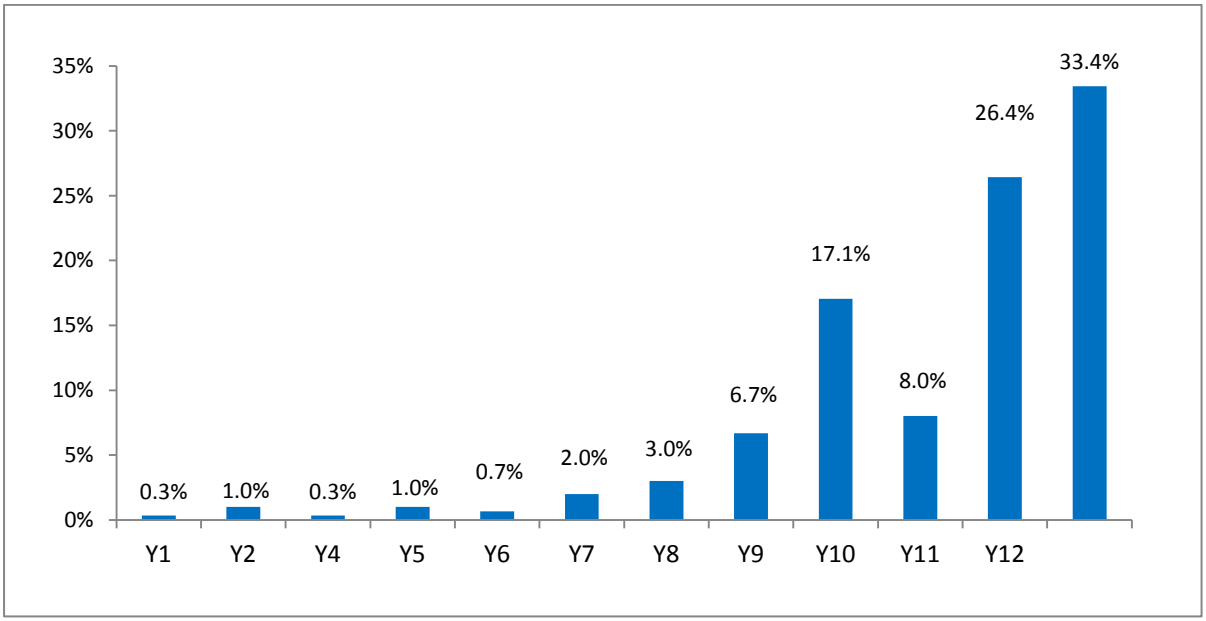
consistent with the large proportion of the sample from the Western Cape, a province where this ethnic group is in the majority.

The education level of the respondents was measured in terms of the number of years of education attended and/or completed up to graduation from high school, and a further single category measuring attendance and/or completion of post high school education/qualification. This measurement was deemed problematic at a conceptual level as it compels examination of education for those who attended and/or completed high school separately from those who proceeded further beyond high school (there being differentiation in the former and none in the latter). While this might have been considered useful considering the lexile level at which the SAATSA items were formulated, there is no reason why the variable cannot be improved without in any way compromising the ability to determine effects for lexile levels that fall at less than 12 years of education.

Key recommendation: The variable measuring education should be reformulated to reflect sufficient differentiation in terms of substantive measurements of different levels of educational achievement/attainment at all levels. In this regard it is recommended that the measurement as adopted by official statistics surveys be employed as it provides well differentiated categories of achievement which are extensively accepted and utilised in the country. This should also assist in reducing the high rate of missing values for this variable as seen in the pilot data.

Of those 82.1% of respondents who reported their education, two thirds (66.6%) had attended education up to and including the final year of high school while the remaining one third had progressed to some post-matric education and/or qualification(see Figure 1).

FIGURE 1: Respondent Years Of Education

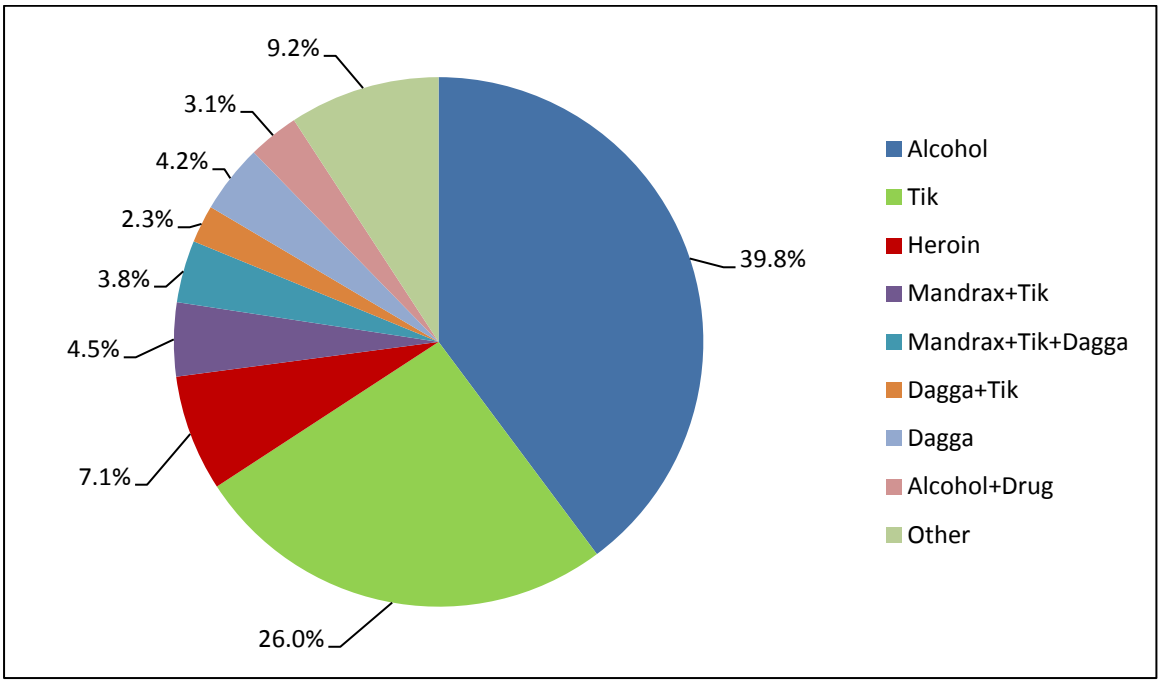


Amongst those respondents who has progressed only up to the final year of school (year 12), the mean number of years of education was 10.3 years and the median 11 years, indicating an even split between those who completed only ten years of school education and those who had progressed to the remaining two years. A minority of the total number of respondents (5.4%) had only completed primary school. For reasons already articulated the one third of respondents who progressed beyond high school could not be further differentiated.

The majority of the individual facilities selected for participation in the study are relatively specialised in terms of the treatment of specific types of substance abuse, and the distribution of presenting symptoms per facility reflects this. Respondents abusing Methamphetamine/Tik were concentrated primarily at SANCA Athlone, Toevlug, Kensington Treatment Centre, and to a lesser extent, SANCA Tygerberg, while respondents abusing alcohol were concentrated mainly at SANCA Durban and SANCA Pietermaritzburg. Once again, as there is no information available to weight the realised sample against facility records, these results cannot be generalised in any way.

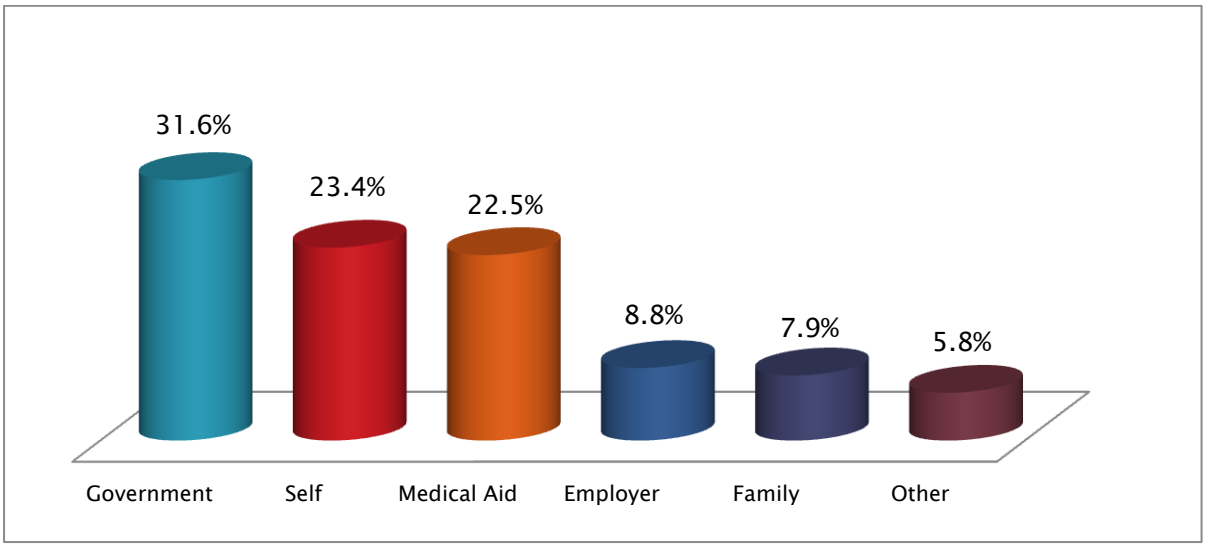
Overall, problematic use and abuse of alcohol was by far the most common presenting symptom across all facilities (see Figure 2), accounting for slightly under two-fifths of the sample (39.8%). This was followed by Methamphetamine/Tik (26.0%), though this latter prevalence is localised overwhelmingly in the Western Cape. One in 15 respondents reported treatment for abuse of heroin (7.1%), followed by slightly smaller numbers who reported abusing a combination of Tik and Mandrax (4.5%) and a combination of Tik, Mandrax and marijuana (4.0%).

FIGURE 2: Primary Substance For Which Respondents Reported Treatment



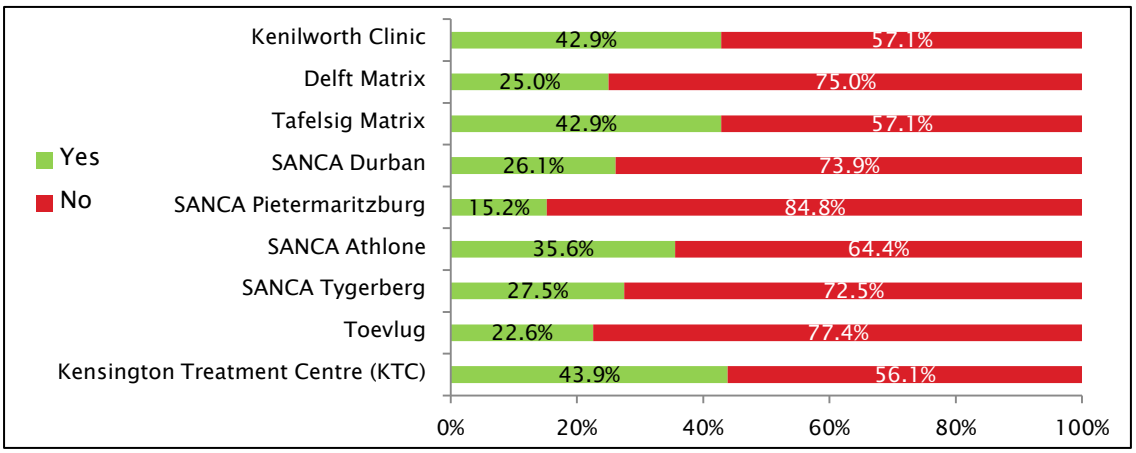
As discussed earlier, the SAATSA item measuring respondent views on their capacity to afford their treatment recorded a high incidence of missing values, the principal reason for this being that the items was not applicable to a sizeable proportion of respondents as their treatment was being paid for by someone other than themselves (see Figure 3).

FIGURE 3: Payment Of Respondent Treatment Costs



The biggest funder of treatment was the state, accounting for almost one in three respondents, followed by the respondents themselves, accounting for slightly less than one in four respondents, with a similar proportion being funded by their health insurance schemes. Less than one in ten respondents were dependent on their employer or family for payment for their treatment. Slightly over two thirds of the respondents were being treated for substance abuse for the first time (70.8%), though this figure varied depending on the specific facility in question (see Figure 3). At the Kensington Treatment Centre patients who had been in programmes previously accounted for two in every five respondents, while at SANCA Pietermaritzburg this figure dropped to around 15%.

FIGURE 4: Respondent Prior Treatment For Substance Abuse



To reiterate, the sampling design was purposive and hence no generalisations are possible to the population of adult substance abusers in the country. Moreover, given the absence of

supplementary data to properly weight the realised sample, no conclusions can and should be drawn about the prevalence of these categories in the enrolled populations of these treatment facilities. Nevertheless, it is useful to note that the sample did contain sufficient differentiation by way of critical demographic variables (gender, ethnicity, etc.) as well as in terms of other important variables (admission type) to permit proper psychometric assessment of the SAATSA items by application of these variables.

In closing, it should be noted that the demographic variables included in the pilot survey provided a good basis for determining respondent effects in the later psychometric analysis. However, one remiss in the pilot was the failure to include a variable measuring the primary language spoken by the respondent. While ethnic group can be used as a proxy for primary spoken language it is at best a flawed proxy because of linguistic differentiation within ethnic groups across different geographic locations. For instance, Coloureds in the Western Cape predominantly or even exclusively list Afrikaans as their primary language whereas for Coloureds in KwaZulu-Natal this could be either English or Afrikaans. Similarly, there is considerable variation amongst African respondents in different parts of the country in terms of their preference for English as compared to their home language. As home or primary spoken language plays a significant role in the learning and comprehension of English, should the SAATSA be administered in English in the future it would be necessary to record this variable to examine for language effects on the SATTSA scale items.

Key recommendation: A variable be added to capture information on the primary spoken language of the respondent. This will be invaluable for assessing any likely linguistic effects in the SAATSA, an important consideration if it to be administered in only one language and/or in a language that is other than the language spoken most often by the individual respondent.

CHAPTER 5: METHODOLOGY FOR THE PSYCHOMETRIC ANALYSIS

To begin with, it is instructive to note that the SAATSA items are intended to measure various underlying theoretical constructs relevant to service quality in the treatment of substance abuse. These constructs are latent in that they are not directly observed, but they can be measured by use of manifest (observed) variables, all of which are intended to be representative of the latent variable to varying degrees. The measurement of latent constructs by observed variables is inevitably prone to some degree of error as a result of the design, conceptualisation, and operationalisation of these measured variables. Additionally, such measurement is also vulnerable to the effects of sampling and sampling error. To examine the extent of such error and the overall competence of the observed variables in measuring the latent construct, it is necessary to investigate the psychometric properties of each of these individual measured variables in relation to the latent construct.

Two concepts are important in regard to item/scale psychometric properties: validity and reliability. Validity, the more important of the two, refers to the extent to which the observed variables are actually measuring the latent construct which they are intended to measure. Validity is typically differentiated in terms of face validity (measured variables on face value appear to measure the latent construct), content validity (measured variables provide coverage of important content domains relevant to the latent construct), convergent validity (measured variables correspond to some independent index in a predictable manner) and construct validity (measured variables capture and measure the conceptual meaning of the latent variable). In psychometric analysis it is construct validity which is of paramount importance, as achieving this typically ensures that the other three types of validity are also achieved. Reliability refers to the extent to which measured variables consistently measure whatever it is they are intended to measure. Reliability is a necessary but not a sufficient condition for validity: variables may consistently measure the same thing but this "same thing" may have little correspondence with the actual latent construct of interest. In contrast, construct validity is a fully sufficient condition for reliability: if measured items are highly indicative of some latent variable, they will be so consistently in repeated applications of the variables.

Measurement Theory

Investigation of the psychometric properties of scales and items is typically undertaken using the framework of measurement theory and all of the techniques and procedures than emanate from this. However, measurement theory is not monolithic, and encompasses a range of paradigms, each of which posits its own epistemology and supplies its own methodology. In the social sciences generally (public health, psychology, education, behaviourial medicine, etc) the dominant paradigm for the past 50 years or so has been that defined by Classical Test Theory (CTT). CTT itself originated from the Classical Model which was has existed for well over the past century. The basic premise of CTT is that there is a true score for every latent/measured variable under investigation (and for every person who is measured using these variables), and that this true score would be obtained if

there were no errors in the measurement process. However, as error is an inevitable part of measurement, any observed score is actually the true score with some degree of error. CTT further assumes that all measurement error is normally distributed, that is, it is equally distributed around the average. As this deviation occurs on both sides of the true score, the errors cancel each other out and the average becomes the true score. The principal objective of CTT is the examination of scale properties (a scale being a measure comprising several individual items), and even when item properties are investigated it is only in relation to these scale properties.

In contrast to the CTT emphasis on the scale itself, Modern Measurement Theory (MMT) focuses more directly and specifically on the properties of items, in and of themselves as well as in relation to other items. Additionally, MMT also provides a platform to examine these item effects from the perspective of respondent characteristics (such as age, ethnic group, gender, etc.), thereby enabling the investigation of the manner in which such demographic features can and may affect how respondents respond to any given individual item. The most well known form of MMT is that referred to as Item Response Theory (IRT). IRT provides the capability to examine not only individual items in total (do the items perform adequately in measuring the latent variable?) but also the separate response categories of these individual items (do the response categories for the items - the Likert scale in the instance of the SAATSA – perform adequately for that item?), and both of these analyses can also be subject to investigation in terms of critical respondent characteristics which might be assumed to impact on either the item as a whole and/or the individual response categories (Does the item perform adequately for all respondents regardless of their age, gender, ethnic group, education, etc?). Consequently, IRT is able to provide detailed psychometric properties for individual items rather than only for the scale, thereby permitting a higher resolution and considerably more detailed examination of these items as compared to CTT. The psychometric properties obtained via IRT are also not subject as in CTT to the specific sample being investigated, but are rather assumed to be invariant across all samples and populations. These are but a few of the much lauded advantages that IRT offers over CTT, advantages which clearly place it at the forefront in shaping the nature and direction of measurement theory for the next few decades, if not longer.

Despite is frequent characterisation as modern, IRT actually has its roots in work done by Thurstone early in the 20 century. The most significant contribution after this came from Georg Rasch, who proposed the central concept of IRT: specific objectivity, the idea that item and person parameters can be assessed independently but be comparable on the same metrics. Rasch's work subsequently led to the development of statistical techniques and models to enable application of IRT to the analysis of social sciences data. This has resulted in increased use and application of Rasch models in IRT in the past two decades, and such application has inevitably lead to comparisons with, and challenges against, the hitherto dominant CTT. The debate about the comparative and relative merits of CTT and MMT has grown more pronounced of late, with proponents from both perspectives citing various arguments in support of their respective positions. The vigorous nature of these arguments suggests that it is likely to continue for some time into the future. In approaching this debate, it is important that the differences between the two approaches not obscure the fact

that both of them are intended for the same purpose, this being the assessment of the measurement of latent variables by observed variables. In this regard, the different perspectives can be considered as different lenses through which the same data may be viewed. Importantly, if these approaches are applied properly and robustly to the same data, there should be - and typically is - a very high correspondence between the outputs of their respective analyses.

Notwithstanding the arguably clear advantages that IRT offers over CTT, particularly in terms of analysing items and respondent effects, there are currently some limitations to its widespread use. Principal amongst these is the relative novelty of the approach. IRT remains an emergent paradigm yet to properly penetrate graduate training curricula and find widespread application in research in education, health and the social sciences. Additionally, and unlike for CTT, there are relatively few software programmes available for application of the techniques, an essential requirement for an approach that is math intensive. The net effect of all of this is that IRT has yet to find sufficient currency amongst educators, researchers and practitioners in their own work. Additionally, and not insignificantly, the use of IRT is often constrained by the capacity of the audience to assimilate its outputs. This was an important consideration in the analysis of the SAATSA data, especially so as the readers of this report are likely to be quite diverse - academics, practitioners, administrators, researchers, policy makers, etc. - all with significantly variable exposure to and expertise in measurement theory. To enable the widest possible dissemination and consumption of this report, it was necessary to find a good balance between what was or might already be known to readers and new material which they could potentially assimilate from a careful reading of the report.

In pursuit of such an enhanced understanding of this report, the analysis of the SAATSA data employed a combination of IRT and CTT techniques. IRT was employed at the first stage of the psychometric analysis to examine the characteristics of the individual items and their response categories and to test for respondent effects, and CTT techniques were applied to examine the scale characteristics and to formally establish their psychometric properties. It is hoped that this complementary approach will afford readers who are familiar with CTT and not IRT the opportunity to understand the psychometric analysis as well as gain a useful introduction to its newer measurement theory stable mate. Where necessary and possible, detailed explanations are supplied to assist in understanding thereof, particularly for IRT but also for CTT techniques and outputs. This paired approach was applied for each separate analysis of the individual SAATSA scales.

IRT Analysis

The IRT analysis of the SAATSA capitalised on the advantages provided by this approach in examining for item and response category characteristics and for respondent effects. The analysis was conducted by use of the R statistical software package. The R package contains two modules applicable to IRT analysis: the latent trait modelling (ltm) and the extended Rasch Modelling (eRm)

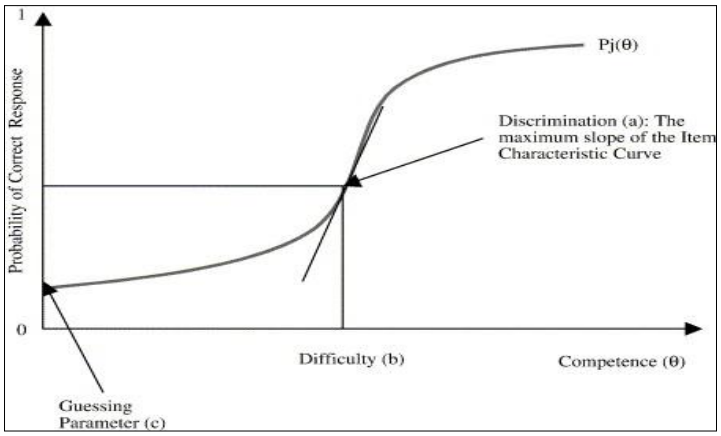
modules. The ltm is suitable for dichotomous data (data from items with only two response categories) and hence was not appropriate for the SAATSA items. The eRm module is appropriate to SAATSA as it addresses polytomous data (data from items with more than two response categories). Additionally, within the eRm module in the R package there are a number of Rasch models that may be run depending on the nature of the polytomous data. As the SAATSA items all use a Likert scale response format (with responses ranging from 1 to 4), the model most appropriate is the Graded Response Model (GRM), and this was employed for the IRT analysis.

As indicated earlier, IRT analysis provides a distinct advantage over CTT in that it permits analysis of the individual items and their response categories across individuals rather than for the entire scale. An added advantage is the graphic outputs which are derived from this analysis, and which provide a more readily assimilated overview of the specific properties of the items and the item response categories. Two such graphical outputs are of particular utility and importance and were applied in this analysis: Item Characteristic Curves (ICC) and Item Information Curves (IIC). The ICC provides information on the probability of a respondent selecting a given response category, the difficulty or easiness of the selected category, and the probability of a respondent merely guessing their responses to the item. These outputs are generally more relevant to tests of ability – which is where IRT has found much of its current application in education and psychology – but they are also applicable to evaluation of perceptual variables such as the SAATSA items (the probability of "success" or "correct response" should merely be interpreted as the probability of selecting a particular response category such as Strongly Agree). The figures that follow indicate and explain the information obtained from these two graphic outputs.

Item Curve Characteristics

When obtained from the analysis of perceptual data the ICC outputs provide important information on the probability of respondents selecting given response categories for that particular item. In the instance of the SAATSA they provide information on the likelihood that respondents will select any one of the four available responses categories: Disagree, Somewhat Agree, Agree, and Strongly Agree. As the objective of the analysis is to assess the extent to which the items measure a single underlying latent trait, it is to be expected that the probability of responding to one response category on an item should correspond to the probability of responding to the same response category on another item which is also measuring the same latent trait. That is, respondents are expected to be consistent in responding to different items from the same scale. Or stated differently, if a respondent indicated "Strongly Agree" for an item measuring some latent trait, they should respond in the same manner to another item which also measures the same latent trait. If these probabilities are different, it implies the items may not be measuring the same dimension as respondents are variable in how likely they are to respond to the different items purported to measure the same thing. The ICC reports these probabilities and other information in the graphic displays provided hereafter.

FIGURE 5: Properties of The Item Characteristics Curve in Rasch Models



Looking at Figure 5, we find the three item parameters referred to earlier on. Seen separately, the three parameters look as displayed in Figures 7, 8, and 9.

FIGURE 6: Item Easiness/Difficulty (Probability of Selecting A Given Response Category)

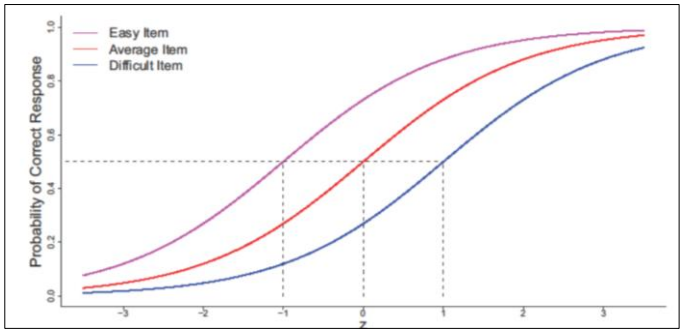


FIGURE 7: Item Discrimination Properties

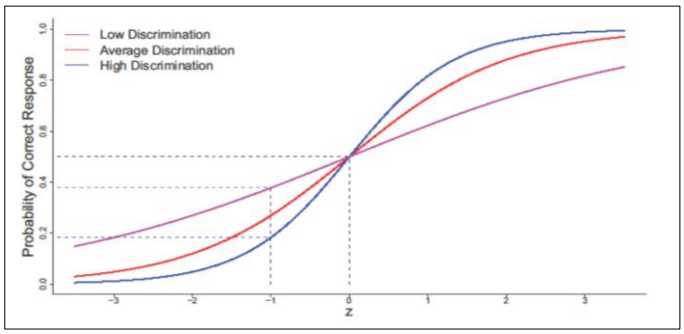
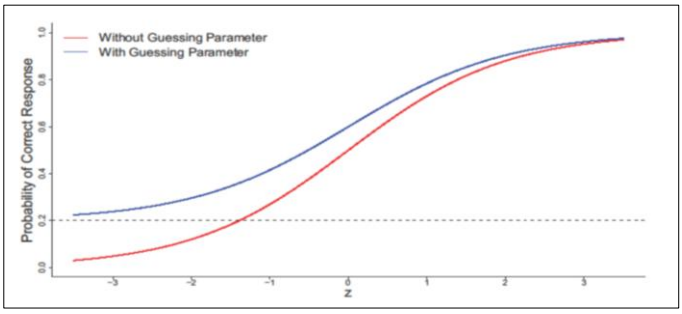


FIGURE 8: Item Guessing Properties

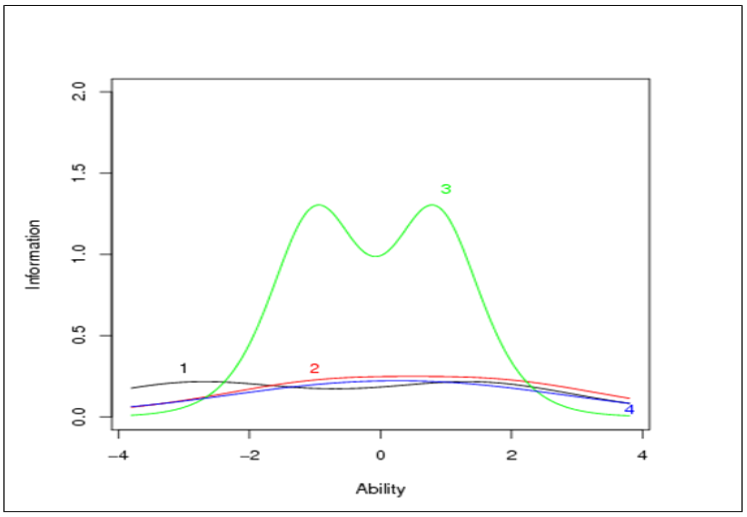


The latter three figures are slightly misleading as they imply different graphs for the different item properties. In reality, and as will be seen later in this report, all three pieces of information are captured in a single curve for each item. When applied in IRT, the curves for different items are compared to determine if their shape is different from one another. As all items are assumed to measure a single latent trait, we expect that the curves for each item should be quite similar to one another for all the response categories. If they are not, then the items are most likely quite variable at measuring the same latent dimension, which is not desirable, or they are in fact measuring different latent traits, which is also not desirable. A caveat in terms of this similarity in ICC is necessary: while we expect items to have similar curves, they should not be exactly identical, as this implies that the items are not really any different from each other and are actually interchangeable, meaning that there is redundancy in the scale. So item curve similarity with some difference is preferable.

Item Information Curves

The IIC in Rasch modelling supplies important information about how much information is provided by the individual items to the measurement of the latent trait and is considerably easier to interpret than the ICC. Items providing high levels of information are essential to the measurement of the latent trait and those providing low levels of information are less so and may be easily discarded (more technically, items with high information are those with low standard errors in measurement, and vice versa for items with low information). Figure 9 presents the IIC for four hypothetical items.

FIGURE 9: Item Information Curves



As is evident in Figure 9, the four items differ in terms of their information characteristics. Item 3 contributes the most information while Items 1, 2, and 4 contribute much less. In the IRT analysis conducted for the SAATSA, these levels of information must be interpreted as precision in measurement (hence the earlier reference to standard error of measurement): the higher the peak of the curve for an item the greater the precision with which that item measures the latent trait and vice versa.

Respondent Effects

A distinct advantage offered by IRT over CTT is the facility to examine for measurement effects due to respondent characteristics. This is important because respondents might respond differentially to the same SAATSA item based on critical demographic and other characteristics. In the case of the SAATSA, the following respondent characteristics were tested during the IRT analysis:

1. Language – linguistic proficiency plays a critical role in how well respondents understand and therefore respond to individual items. This is important in a multi-lingual country such as South Africa where there are 11 official languages, and especially so as the SAATSA was developed and applied only in English. Despite various controls being applied during the fieldwork to ensure greater consistency in language proficiency, it remained important to test for such language effects. As discussed earlier, the SAATSA dataset lacks information on the primary spoken language of the respondent, thereby complicating the testing for language effects. Accordingly, ethnic group/race was used as a proxy for language, though with the important caveat that it is an imperfect proxy for language for the reasons discussed previously.
2. Education – during the cognitive testing phase of the SAATSA careful attention was paid to the lexile level at which items were developed. In general, items were developed for the lexile level equivalent to the educational achievement of Grade 11, that is, eleven years of formal education. Consequently, this lexile level was used to separate the sample into two sub-samples: those with 11 years or less of education and those with more than 11 years of formal education. A new variable was formulated to reflect this distinction and applied in the IRT analysis.
3. Gender – research has shown that gender can and does play a role in how respondents respond to an item, with males and females responding similarly to others of their own gender and differentially from each other. This variable was thus also added into this part of the IRT analysis.
4. Admission type – the SAATSA is intended to measure attitudes towards the quality of service in substance abuse, and this service is often differentiated in nature and content depending on whether the patient is being treated as an inpatient or an outpatient. Consequently, it was deemed important to examine for effects due to the type of admission applicable to the respondent.

CTT Analysis

The CTT analysis was conducted using standard techniques for examining and establishing the validity and reliability of the SAATSA scales. Validity analysis was undertaken first, using exploratory factor analysis (EFA) to investigate the likely underlying dimensions measured by individual SAATSA scales. Unless otherwise stated, all EFA was conducted using Maximum Likelihood Estimation (ML) and Oblique Minimum Rotation (Oblimin). ML was deemed necessary as this provides the most robust test for real world data and Oblimin was employed to account for the fact that a single

SAATSA scale might actually comprise more than one latent dimension, hence these latent dimensions should be allowed to correlate with each other. In all EFA the standard criterion for selection of factors was an Eigen value greater than 1. The EFA was conducted using the SPSS software package.

Following the exploratory factor analysis of each scale, confirmatory factor analysis (CFA) was undertaken to confirm the latent structure observed in the EFA. This analysis was run using the AMOS module in SPSS. Where necessary, data imputation was executed to ensure that missing data was eliminated and thus modification indices could be produced to refine obtained latent models. As is well known in the testing of latent or structural equations models, model outputs are vulnerable to increased sample size, and this is particularly true for the χ^2 statistic, which becomes significant as the sample size increases past 200 (the final SAATSA sample is well over 350 respondents). Hence significant models may be rejected largely because of the sample size and despite being significant and conceptually valid. To overcome this effect, current approaches to latent model testing eschew the traditional over-reliance on the χ^2 statistic and instead employ a series of criteria to provide a much more comprehensive and detailed assessment of the adequacy of the obtained latent model. These criteria are as follows:

1. χ^2/df Ratio – The first criterion is the ratio of the χ^2 to the degrees of freedom for the model. A ratio of between 2 and 4 is deemed to be indicative of a very good fitting model, and anything less than 2 is indicative of a very good fit.
2. Comparative Fit Index (CFI) – the CFI ranges between 0 and 1, with higher values indicating better fit. A CFI of 0.90-0.94 indicates a good fitting model and a CFI of 0.95 and more indicates a very good fitting model.
3. Root Mean Square Error of Approximation (RMSEA) – The smaller the RMSEA (pronounced "Ramsey") value the better the fit of the model. A RMSEA value of 0.08 or less indicates a good fit while a value of less than 0.05 indicates a very good fitting model.

These criteria were applied during the CFA to assess the adequacy of the obtained latent models

Following the validity analyses, the reliability analyses were conducted for the SAATSA scales using the Cronbach alpha (α) statistic. The Cronbach alpha ranges between 0 and 1, with higher scores indicating greater levels of reliability and vice versa. The assessment of the returned reliability estimates was conducted using the following scale:

- Above 0.90 – excellent (but with possibility of some redundancy),
- Between 0.80 and 0.90 – very good,
- Between 0.70 and 0.80 – acceptable,
- Between 0.65 and 0.70 – marginally acceptable,
- Less than 0.65 – unacceptable.

Moreover, apart from obtaining the scale Cronbach α for the entire sample, analyses were conducted to examine the variation in this statistic based on the respondent demographic

characteristics identified earlier. The objective here was to examine whether or not the reliability statistic obtained for different sub-samples was within or outside the confidence intervals of the reliability statistic for the entire sample. Any sub-sample reliability statistic falling outside the sample confidence interval was considered indicative of respondent effects for that particular demographic variable. That is, it was indicative of the fact that the reliability of the scale differed significantly depending on whether respondents were male or female, were inpatients or outpatients, had a particular level of educational achievement, and as a result of their likely home language (as measured by proxy by their ethnic group). Ideally, all sub-sample reliability statistics should be within the same confidence interval, indicating that the scale was equally reliable for all respondents regardless of their demographic characteristics.

The SAATSA – Domains And Dimensions

As formulated, the 26 SAATSA items reflect a three tiered hierarchy of measurement. The first and highest tier is represented by two broad domains, one measuring Quality and the other measuring Effectiveness. Within these two domains, and at the second tier, are dimensions, each of which is believed to tap into a single latent construct. Within each of these dimensions, and at the lowest order in the hierarchy, are the observed variables represented by the individual SAATSA items. As the overall measurement was formulated using this conceptual framework, the psychometric analysis had to conform to this framework and provide a test of the sufficiency thereof. For this reason, it is instructive to review the conceptual framework underpinning the 26 SAATSA items.

The Quality and Effectiveness Domains

As reported earlier in the report, the SAATSA was developed based on the deliberations of the SAATSA Working Group and following on from the learning obtained from similar work done in the Unites States. The SAATSA Working Group eventually identified two primary domains for assessment of the service quality at substance abuse treatment facilities in South Africa. The first, Quality, measured respondent evaluations of the quality of health care which they were receiving at the various facilities. The second domain, Effectiveness, measured the extent to which respondents believed that the health care received was effective in them overcoming their substance abuse and returning to a life free of problematic drug and alcohol use. As reported previously, all individual SAATSA items were measured using the same 4-point Likert scale. All items were also worded and scored in the same direction, with higher scores indicating greater levels of agreement.

The SAATSA Dimensions and Items

Within each of the two SAATSA domains are resident various dimensions: Access, Quality, Person Centeredness, Quality of Life, Social Connectedness and Substance Abuse. Each of these six dimensions was assumed to be represented by a single latent construct, and each latent construct was assumed to be measured by the items making up that particular scale. For the purposes of

consistency, these dimensions will hereinafter be referred to as scales. The composition of the two domains and various scales according to the 26 SAATSA items is reflected in Table 7. Please note that the SAATSA contains two Quality latent constructs, one at the level of domain and the other at the level of the scale. These were kept intact for the purposes of the analysis.

TABLE 7: SAATSA Factors (Domains and Scales) And Items

QUALITY DOMAIN	
ACCESS	Item 1: The amount of time I had to wait to get services was acceptable to me Item 2: I can afford my treatment
QUALITY	Item 3: The staff at this treatment centre treat me with respect Item 4: The people I went to for treatment services spent enough time with me Item 6: The staff at this treatment centre told me about services in my area that will help me stay off drugs and alcohol Item 8: This treatment centre teaches me how to avoid getting HIV Item 26: I would recommend this treatment centre to a friend
PERSON CENTREDNESS	Item 5: I have a say in deciding about my substance abuse treatment that I am receiving here Item 7: I am given a choice of services in this treatment centre Item 9: The staff at this treatment centre are sensitive to my background
EFFECTIVENESS DOMAIN	
QUALITY OF LIFE	Item 10: My general health is improving Item 11: I am better able to cope when things go wrong Item 12: I am better able to accomplish the things I want to do Item 14: I expect to do better at work/finding work or at school Item 15: I am more likely to practice safe sex
SUBSTANCE ABUSE	Item 13: I am less likely to use alcohol or other drugs Item 23: I know that using alcohol and drugs is a problem for me Item 24: I need to work on my problems with alcohol and/or drugs Item 25: The treatment centre is helping me to recover from using drugs and alcohol
SOCIAL CONNECTEDNESS	Item 16: There is some-one who cares about whether I am doing better Item 17: I have some-one who will help me when I have a problem Item 18: I have people in my life who are positive influence Item 19: The people who care about me are supportive of my treatment Item 20: My friends and family are more able to count on me Item 21: I have friends who are not using alcohol or drugs Item 22: I have some-one who will listen to me when I need to talk

The putative latent structure of the SAATSA as reflected in Table 7 was tested using the IRT and CTT analysis. To satisfy the unidimensionality assumption for IRT modelling, each scale was examined separately. Following the analysis of the scales, the two domains were analysed to determine their psychometric properties.

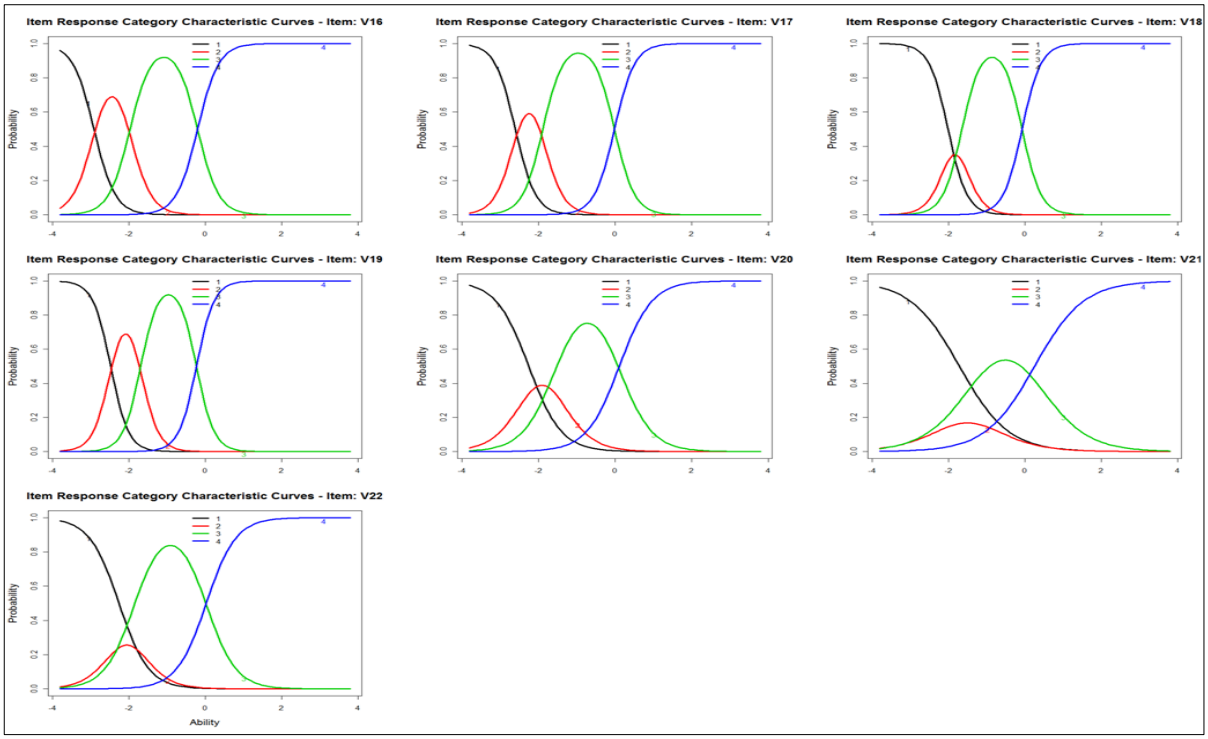
CHAPTER 6: PSYCHOMETRIC ANALYSIS OF THE SAATSA

All psychometric analyses were conducted using the analytic protocols and statistical techniques as specified in the previous chapter. Specifically, SAATSA items were analysed according to their hypothesized latent dimensions, meaning that the scales measuring these latent dimensions were analysed separately. The investigation of individual scales began with an assessment of item properties in IRT before proceeding to an assessment of construct validity using exploratory and confirmatory factor analyses. Following the validity test, the finalised scales were assessed for their reliability. Once the SAATSA dimension shad been analysed the domains of Quality and Effectiveness were investigated using the same protocols.

The Social Connectedness Scale

The Social Connectedness scale (SC) comprised seven individual SAATSA items: Item 16 to Item 22. These seven items were subject to a Rasch model test in R and the Item Characteristic Curves (ICC) are provided in Figure 10 (these indicate the probability of respondents selecting any of the four Likert response categories for each item).

FIGURE 10: Item Characteristic Curves For The Social Connectedness Items

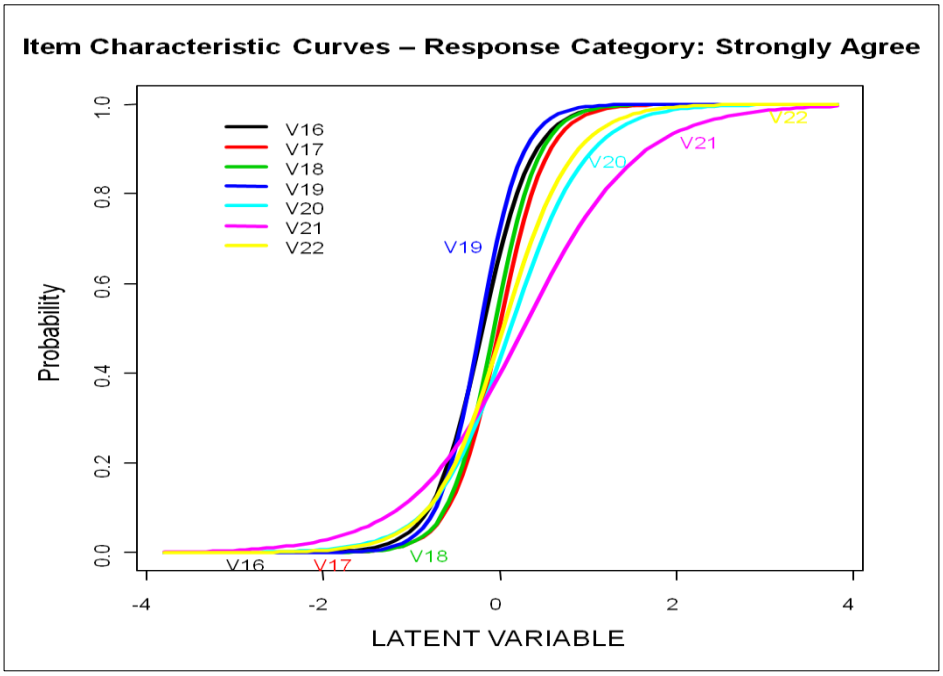


As can be seen in Figure 10, these probabilities are quite different for each response category on all of the items, and this is to be expected as we have already observed in the earlier analysis that there

is a preponderance of responses towards to the positive end of the response scale (Agree and Strongly Agree). Hence while this output is informative it is not especially useful and for the later analyses this particular output is not provided.

Of greater importance and relevance to the SAATSA analyses are the probabilities associated with a particular response category. In the case of the SAATSA, we are interested in how respondents evaluate the different service quality aspects, with such evaluation being measured by the degree of agreement with each of the individual items. Accordingly, it is the lattermost response category that is of primary significance to us, this being the Strongly Agree category. Using this category, we are able in the IRT analysis to assess how well the individual items work in terms of tapping into the hypothesized latent dimension of social connectedness. Figure 11 summarises the probabilities for this response category for all of the seven items in the SC scale and provides it in a single output.

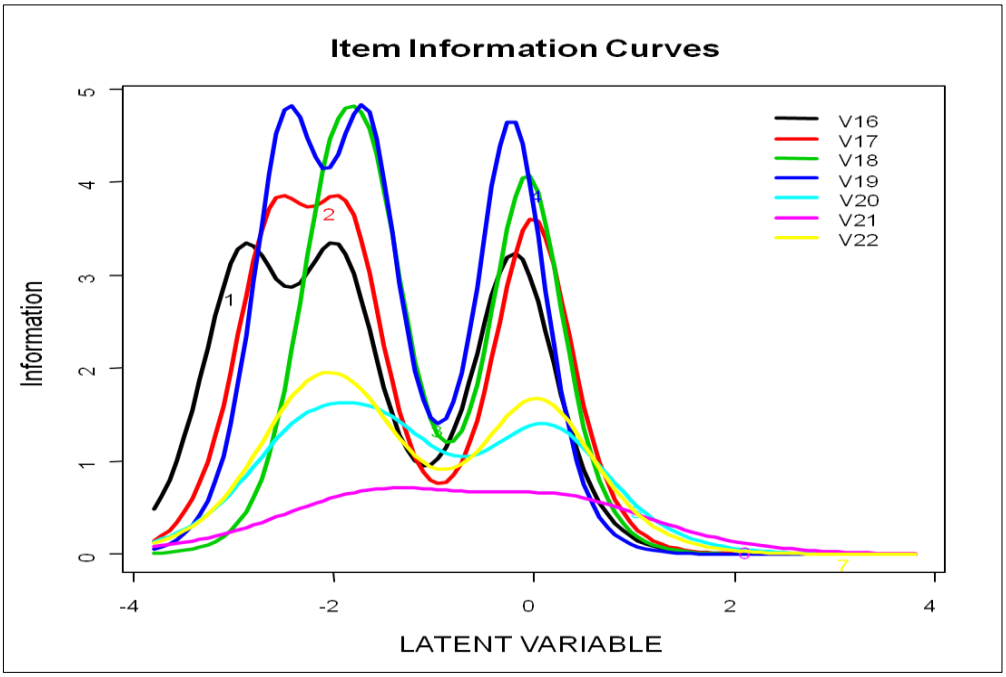
FIGURE 11: Item Characteristic Curves for the Strongly Agree Category on all Social Connectedness Items



The close grouping and similarity in curves for each of the first four items (Item 16 to Item 19) indicates consistency in how respondents responded to this response category on these items. In other words, the probability of respondents strongly agreeing to each of these items was similar and relatively consistent; hence these items are probably good indicators of the latent dimension. A note in interpreting these visual outputs: while it is desirable to have items with similar characteristic curves, too much similarity is not desirable as it indicates that the different items are actually all doing exactly the same thing, implying that there is high redundancy in measurement of the latent variable. In this instance the curves for these four items were close though not entirely overlapping, suggesting some differentiation amongst them, though this would only be determined more definitively during the factor analysis.

In contrast to the first four items, the remaining three items are not only dissimilar from the first four but also from each other, suggesting that they are working less efficiently at measuring the latent dimension of social connectedness. The extent of efficiency in measuring the latent variable can be ascertained by examining how much information the items provided to the measurement of the factor. Items which provide high levels of information are good indicators of the factor and those which do not are not so as they introduce too much error into the measurement of the latent variable. Figure 12 provides the graphical representation of the information supplied by each item to measuring the social connectedness factor. To recap, the higher the peak the greater the amount of information supplied by an item and vice versa.

FIGURE 12: Information Supplied By The Items In Measuring Social Connectedness



The graphic in Figure 12 confirms the earlier output as the item information curves for the first four items are considerably higher than that for the latter three items. This was also confirmed by the additional outputs from the Rasch modelling (difficulty scores and the discriminant function values). Taken together, they indicated that Items 16-19 were much better at measuring social connectedness than the remaining items. Additionally, the IRT analyses indicated that the item properties were invariant across the specified demographic categories of gender, education, admission type and ethnic group, meaning that there were no respondent effects which produced variable measurement for any of these items.

To test the latent structure of this scale, the seven items were subjected to an exploratory factor analysis (EFA) using the specifications indicated earlier on. The outputs of this EFA confirmed the results of the IRT: the analysis extracted a single factor (with an Eigen value greater than 1)

accounting for 61% of the total variance, with the first four items loading the most onto this factor (see Table 8).

TABLE 8: Factor Loadings For The Social Connectedness Scale Items

ITEM	FACTOR LOADING
Item 16: There is some-one who cares about whether I am doing better	0.612
Item 17: I have some-one who will help me when I have a problem	0.655
Item :18 I have people in my life who are positive influence	0.635
Item 19: The people who care about me are supportive of my treatment	0.676
Item 20: My friends and family are more able to count on me	0.511
Item 21: I have friends who are not using alcohol or drugs	0.256
Item 22: I have some-one who will listen to me when I need to talk	0.490

The factor loadings indicate a clear separation between the first four items and the latter three. Item 20 might also be considered to load onto the same latent factor as the first four items, but the relatively lower loadings for Items 21 and 22 suggested they would not. To validate the latent structure for this scale, a confirmatory factor analysis (CFA) was conducted using the structural equations modelling package AMOS. The CFA was structured to test three models: the first with all seven items as predictors of the latent variable, a second model testing for five items (Items 16, 17, 18, 19, and 20) and a third testing for four items (Items 16, 17, 18, and 19). The goodness of fit indices for the three models are presented in Table 9.

TABLE 9: Three Confirmatory Model Tests Of The Latent Structure For The Social Connectedness Scale

MODEL	χ^2/df Ratio	RMSEA	Comparative Fit Index
All seven items	2.86	0.072	0.979
Five: Items: 16-20	3.78	0.088	0.985
Four Items: 16-19	2.25	0.059	0.991

As Table 9 indicates, all three models provided a reasonably good fit to the data, with acceptable χ^2/df ratios and CFI values. Of the three, the final model comprising a single latent factor measured by only four items (Items 16, 17, 18 & 19) proved to be the most robust, especially when the RMSEA values are compared to that of the other two models. In sum, then, the model outputs confirmed the outputs of the earlier IRT analyses which indicated poorer discrimination and information values for Items 20, 21 and 22 as compared to the other four items. Based on this it was concluded that the social connectedness factor was best measured by use of only Items 16, 17, 18 and 19.

Once the validity of the latent structure of the SC scale was established, the scale was tested for reliability in both IRT and SPSS (to generate the required Cronbach alpha statistic as well as to test if

this reliability differed significantly as a result of the four identified demographic variables). The analyses returned the final reliability estimates reported in Table 10.

TABLE 10: Reliability Of The Finalised Social Connectedness Scale - Sample And Sub-Samples

SAMPLE	SUB-SAMPLE	CRONBACH α
FULL SAMPLE		0.88
GENDER	Male	0.82
	Female	0.89
EDUCATION	Grade 11 or less	0.88
	Grade 12 or more	0.88
ADMISSION TYPE	Inpatient	0.89
	Outpatient	0.86
ETHNIC GROUP	Coloured	0.88
	African	0.88
	White	0.86

The final reliability for the SC scale is very good and close to excellent, as was to be expected given the latent model analysis. Looking at the reliability estimates for each of the different subsamples we find that the most pronounced difference is between men and women, with the scale being less reliable for men than it is for women. Following this is the difference between inpatients and outpatients, though this is marginal, while the differences across education level and ethnic group are negligible. More importantly, all reliability estimates fell within the same confidence interval as that for the entire sample, meaning that they were not significantly different from each other and from the estimate for the entire sample. This confirmed that the scale was being equally reliable regardless of the demographic characteristics of the respondent.

The Quality Of Life Scale

The quality of life scale (QoL) was hypothesized to comprise five items: Item 10, Item 11, Item 12, Item 14 and Item 15. All five items were subject to an IRT analysis using the Rasch model module. The outputs of the analyses revealed some variability in the difficulty (probability) and discriminant function values for the items, with values being relatively equivalent for Items 10, 11, 12, and 14 and different for Item 15. This indicates that Item 15 was a poor indicator of the hypothesized latent variable and/or appeared to be measuring a separate latent variable other than the one being measured by the other items. The item characteristic curves for the Strongly Agree response category for all QoL items are presented in Figure 13. As discussed previously, the ICC for the other three response categories are not reported as they have minimal utility.

FIGURE 13: Item Characteristic Curves for the Strongly Agree Category on all Quality of Life Items

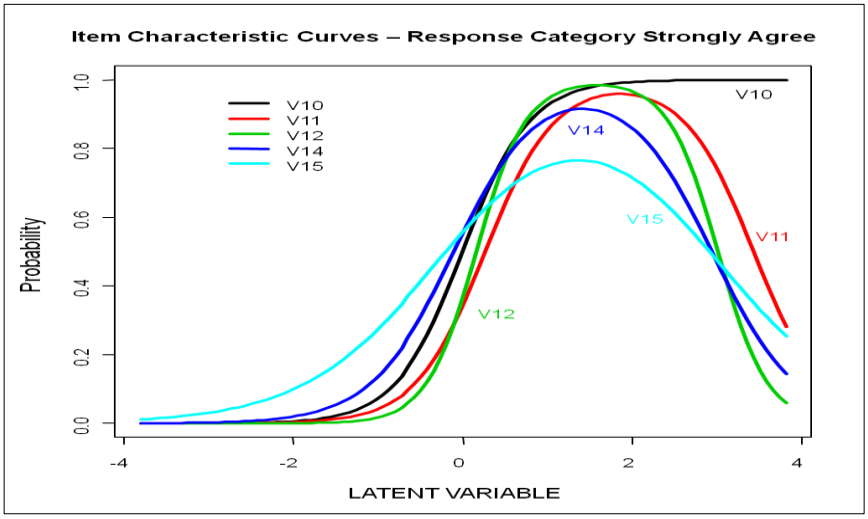
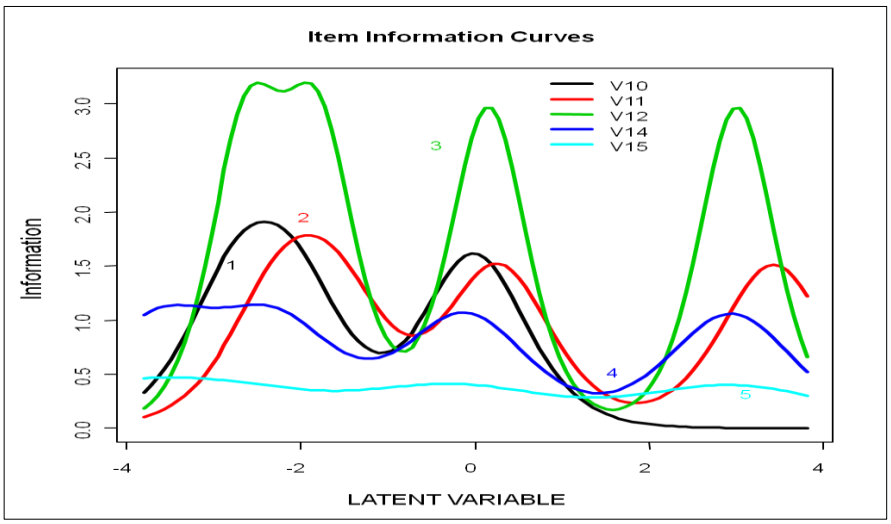


Figure 13 clearly indicates variability in the characteristics of the five items, with Item 10 conforming to the more traditional curve parameters while the others are much less so. The most notable of these latter four is that for Item 15, which is clearly functioning differently from the other items. Looking at the amount of information being provided by each item in the measurement of the latent variable Quality of Life (see Figure 14), the analyses confirms the earlier outputs in graphic form.

FIGURE 14: Information Supplied By The Items In Measuring Quality of Life



As seen in Figure 14, Item 12 provides the highest amount of information for measuring the Quality of Life, while Items 10, 11, and 14 provide less information, with Item 15 providing almost no information towards measurement of the latent variable.

An EFA conducted on all five items revealed a single factor with an Eigen value greater and 1 and accounting for 61.8% of the total variance. The factor scores for the five items are reported in Table 11.

TABLE 11: Factor Loadings For The Quality of Life Scale Items

ITEM	FACTOR LOADING
Item 10: My general health is improving	0.483
Item 11: I am better able to cope when things go wrong	0.450
Item 12: I am better able to accomplish the things I want to do	0.558
Item 14: I expect to do better at work/finding work or at school	0.410
Item 15: I am more likely to practice safe sex	0.316

The loadings for the five items confirmed the outputs seen previously from the IRT analysis, with Item 12 being the best indicator of the latent variable and Item 15 being the worst. This latter result was not unexpected, given the earlier discussion about the relative importance of safe sex and the practice thereof to different respondents. Depending on their demographics and personal preferences, respondents may or may not consider either sexual intercourse and/or the safe practice thereof to be an important part of their life. Accordingly, it will be assessed as having a variable impact on the determination of their quality of life.

Following from the EFA, a CFA was conducted to test for the fit of the latent model for the QoL scale. Two models were specified, one testing all five items and a second testing for four items excluding Item 15 (see Table 12). Both models provide a good fit to the data, but the second model excluding Item 15 is significantly better than the first model. This is the recommended model to be employed for the quality of life scale in the SAATSA.

TABLE 12: Three Confirmatory Model Tests Of The Latent Structure For The Quality of Life Scale

MODEL	χ^2/df Ratio	RMSEA	Comparative Fit Index
All five items	3.560	0.084	0.980
Four: Items: 10, 11, 12, 14	2.754	0.070	0.993

One note with regard to this four item QoL scale: Item 14 showed a lower loading onto the factor than did the remaining three items, all of which were relatively consistent in their loadings. This is possibly explained by the fact that while the first three items assess respondent views on the *current* state of their lives, Item 14 asks them to report on their *expectations* (I expect to do better at work/finding work and at school). It is likely that this phrasing confounds the present with the future, and thus shows some variance from the other three items. The available literature on the measurement of quality of life is clear about the need to keep separate assessments of current quality of life and assessments of future quality of life. Hence this variance for Item 14 is likely to be reduced or eliminated should the variable be reworded to ensure that it is consistent with the other three items in terms of assessing current rather than anticipated future states. However, for the purposes of this analysis the item was retained in the scale.

Key Recommendation: That Item 14 is rephrased to be more consistent with the temporal dimension of the other three Quality of Life items. More specifically, a phrasing such as " I am more likely to do better at work/finding work and at school" should perform somewhat better as it removes the expectation component which can easily be confused as referring to some future point and not the present time.

Having confirmed the construct validity of the QoL scale, the finalised scale of four items was assessed to determine its reliability (see Table 13). As previously, the Cronbach alpha was also computed for the different demographic categories to assess for any variability in reliability across these sub-samples.

TABLE 13: Reliability Of The Finalised Quality of Life Scale - Sample And Sub-Samples

SAMPLE	SUB-SAMPLE	CRONBACH α
FULL SAMPLE		0.83
GENDER	Male	0.83
	Female	0.82
EDUCATION	Grade 11 or less	0.81
	Grade 12 or more	0.83
ADMISSION TYPE	Inpatient	0.83
	Outpatient	0.83
ETHNIC GROUP	Coloured	0.83
	African	0.82
	White	0.85

As revealed by Table 13, the reliability of the scale for the total sample is very good. Moreover, while the reliability estimates for all of the sub-samples vary slightly, none of them deviate significantly from the reliability estimate for the full sample. It was thus concluded that the finalised quality of life scale is consistently reliable for all respondents regardless of their gender, ethnic group, education and admission type.

The Substance Abuse Scale

The substance abuse scale (SA) is the third hypothesized dimension of the Effectiveness domain, and comprises 4 items in the SAATSA measure: Item 13, Item 23, Item 24 and Item 25. Additionally, it was decided that Item 21 (I have friends who are not using drugs and alcohol) would be included in this analysis to assess its likely fit for this scale. The assumption here was that respondents who acknowledged they had a substance abuse problem and were keen to work on it would make changes in their social milieu that correspond to this acknowledgement and willingness to recover and consequently were more likely to associate with people who did not abuse drugs and alcohol. Should this be true, this item ought to load on a scale measuring attitudes towards substance abuse. The initial outputs from the Rasch model analysis are presented in Figure 15.

FIGURE 15: Item Characteristic Curves for the Strongly Agree Category on all Substance Abuse Items

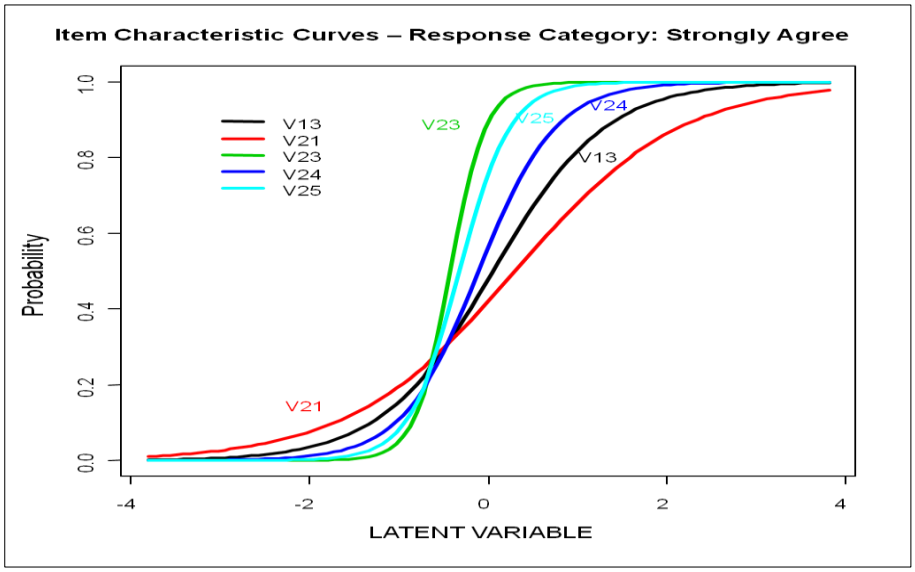
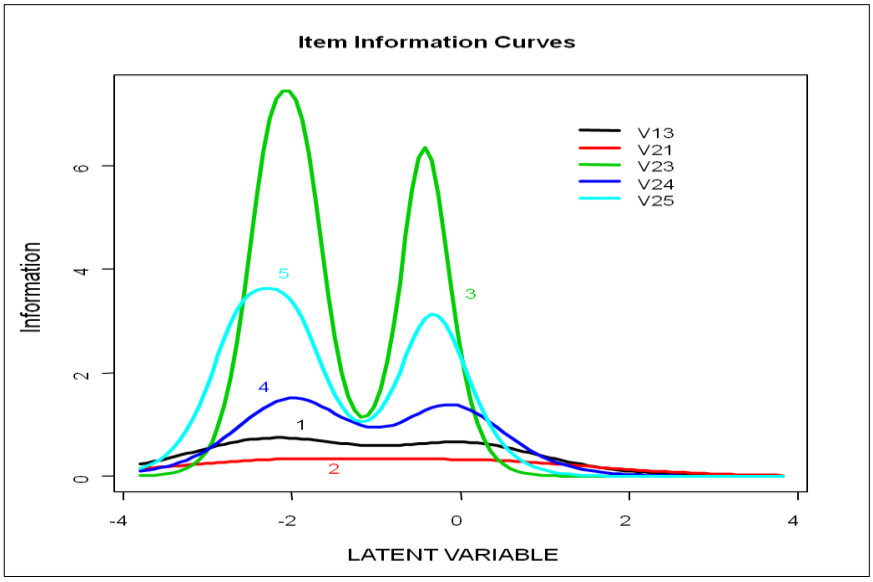


Figure 15 reveals a mixed bag of results for the five items purportedly measuring attitudes towards substance abuse. Items 23, 24 and 25 appear to have the greatest correspondence with one another, while Items 21 and 13 appear to be quite different from these and each other. This was confirmed by the various function values supplied by the Rasch model outputs and the IIC output (see Figure 16). Item 23 provides the most information and thus serves as the anchor for the latent variable, while Items 24 and 25 supply far less information but might nevertheless be considered useful in measuring the scale. In contrast, Items 13 and 21 supplied very little information and thus contributed negligibly to measuring the latent variable of substance abuse, with Item 21 in particular being very poor. The IRT analysis also revealed some variability in item psychometric properties based on respondent demographics, particularly education and ethnic group, but these were non-significantly and hence not in need of further consideration.

FIGURE 16: Information Supplied By The Items In Measuring Substance Abuse



The EFA for these five items produced a single factor with an Eigen value greater than 1 and accounting for 52% of the total variance. The loadings of the items on the factor as produced by this EFA are reported in Table 14.

TABLE 14: Factor Loadings For The Substance Abuse Scale Items

ITEM	FACTOR LOADING
Item 13: I am less likely to use alcohol or other drugs	0.281
Item 21: I have friends who are not using alcohol or drugs	0.192
Item 23: I know that using alcohol and drugs is a problem for me	0.501
Item 24: I need to work on my problems with alcohol and/or drugs	0.332
Item 25: The treatment centre is helping me to recover from using drugs and alcohol	0.440

As is clear from the table, Item 21 loads very weakly onto the factor, as does Item 13 though to a much lesser extent, while the remaining three items appear to load reasonably well. Moreover, and as seen previously, Item 23 provides the basis for the measurement of the substance abuse scale with the highest loading onto the factor. The follow-up CFA was conducted for the SA scale by testing for two models: the first model comprising all items and the second model excluding Item 21. The outputs for these two models are supplied in Table 15.

TABLE 15: Two Confirmatory Model Tests Of The Latent Structure For The Substance Abuse Scale

MODEL	χ^2/df Ratio	RMSEA	Comparative Fit Index
All five items	4.231	0.094	0.964
Four Items: 13, 23, 24, 25	0.965	0.001	0.999

The trimmed model comprising only four items is clearly significantly better than the five item model, with all model indices being excellent. It thus confirmed that the scale measuring substance abuse attitudes was best represented by the four items only. In considering Item 21 it is worth noting that this item suffers from poor formulation, as confirmed by its failure to fit two of the SAATSA scales. Over and above the obvious face validity issues, the failure to find a home in this scale is arguably due to the fact that while all of the four items refer to substance *abuse,* Item 21 refers specifically to substance *use* (*using* alcohol or drugs). While it is reasonable to expect that respondents might eschew the company of others who abuse drugs or alcohol, it is unreasonable to expect them to shun anyone who merely uses these substances as this could easily, given the prevalence of use, particularly for alcohol, in the general population, see them without any viable social circle at all. This item should remain excluded from the SAATSA until such time that it can be significantly reformulated to more directly focus on abuse and not merely use. However, as both the scales to which it might have belonged are sufficiently populated at present, its exclusion at the present time is without any risk.

The reliability for the four item SA scale for the entire sample as well as across the various demographic sub-samples is provided in Table 16.

Table 16: Reliability Of The Finalised Substance Abuse Scale - Sample And Sub-Samples

SAMPLE	SUB-SAMPLE	CRONBACH α
FULL SAMPLE		0.75
GENDER	Male	0.75
	Female	0.76
EDUCATION	Grade 11 or less	0.79
	Grade 12 or more	0.74
ADMISSION TYPE	Inpatient	0.78
	Outpatient	0.74
ETHNIC GROUP	Coloured	0.74
	African	0.72
	White	0.78

The reliability for the entire sample is respectable, as too are the reliability estimates for the various sub-samples. While these do not differ significantly from each other, it is notable that the scale is more variable in its reliability depending on the education of the respondent (more reliable for those with Grade 11 or less) and their ethnic group (more reliable for Whites than for Africans), though in the latter case it must be remembered that ethnic group was used here as a imperfect proxy for language. It is not immediately clear why the scale would be less reliable for those with higher levels of education, but it is recommended that future iterations of the SAATSA pay closer attention to this scale and/or any items that might be added to it.

The Quality Domain

The quality domain was meant to have been analysed using the same format as for the Effectiveness domain, that is, by separate investigation of each of the individual dimensions or scales that were believed to comprise this domain. However, upon closer examination of the data, various problems presented themselves for such an approach, and it is useful to address these problems before proceeding onto any investigation of the items from these scales. The specific issues relevant to each scale are as follows:

1. The Access Scale – the access scale as constructed in the SAATSA comprised only two items, and this immediately eliminated the possibility of any psychometric analysis for the scale as the number of items was below the minimum threshold for such analysis (scales should have at least three items – and ideally four - to enable and warrant proper psychometric testing). The problem was further compounded by the fact that one of the items (Item 2: I can afford my treatment) registered a high level of missing responses because of its non applicability to some respondents. As reported earlier on in the report, fewer than one in four respondents

(23.4%) reported that they were responsible for paying for their treatment themselves, with the remainder having their treatment paid for by employers, the state, family and others. Hence even in those instances when respondents did provide a response other than Not Applicable for this item, it has little direct meaning and hence minimal validity. For this reason, this item was removed from the analysis. This rendered the Access scale invalid as it comprised only a single item (Item 1: The amount of time I had to wait for services was acceptable for me), and this item was considered for inclusion elsewhere.

It is unclear why the development of the SAATSA did not properly account for a number of other issues relevant to access to treatment. While access is partly about cost, it is also about a range of other factors, such as convenience, time and safety. These issues are equally if not considerably more relevant and important in a country such as South Africa, where access to proper health services, particularly for the majority of the population, is well documented to be poor and has even resulted in successful challenges to the state at the level of the highest court in the country (the Constitutional Court). As indicated by various well validated studies and the Constitutional Court challenges, access to health care in South Africa is notably constrained by, *inter alia*:

- the spatial locations of these facilities, which renders them difficult to access in terms of the time required to travel there as well as the costs associated with such travel,
- the convenience of such locations in terms of the degree of effort required by patients to reach the facility, that is, whether or not transportation is available from their location to the facility, and the complexity of such travel in terms of modes of transport and transport schedules,
- the personal safety of the patient in travelling to and from the treatment centre and their home or workplace. This is an important issue in a country with a high rate of crime, and particularly so given the degree of violence associated with personal crime.

It is evident that these access issues are particularly relevant to outpatients who have to transit to and from the facility on a regular basis, but it could be argued that they are also applicable to inpatients as such patients do have to get to the facility in the first place to be admitted. Moreover, these issues are relevant for inpatients in terms of the access issues for family and friends who visit and provide the vital support structure for the patient. For these reasons, the issues identified can and do have significant import and impact in terms of access to substance abuse facilities and programmes, and accordingly, they must be properly accounted for in any measurement of access by the SAATSA.

Key recommendation: The SAATSA Working Group fully explore the relevant issues in terms of access and properly account for these in the measurement of the Access dimension in the SAATSA. Such investigation must give proper consideration to the full range of factors which are applicable in South Africa, be they systemic, institutional, or personal, and realise a proper Access scale accounting for these and comprising the requisite number of items for robust psychometric validation thereof.

2. The Person Centeredness Scale (PC) – the Person Centeredness Scale of the SAATSA was hypothesized to comprise three items, which is the absolute minimum to enable analysis for reliability and validity. However, closer inspection of the scale revealed that the items contained therein posed some issues as follows:
    - Item 7 – this item assessed the extent to which respondents were supplied with a choice in their treatment. This item may be deemed conceptually incomplete as it does not indicate the exact nature of these "services" (Is it the treatment service as a whole, or different components of the service for which choice is or is not supplied?). This ambiguity introduces some component of error into the measurement, but for the purposes of the analysis, the item was retained to ensure an empirical test via the psychometric analysis.
    - Item 3 – This item measured whether or not patients believed that they were being treated with respect at the facility. While this item may on face value relate to the person centeredness dimension, it can just as easily be considered an aspect of service quality. This is evident from the fact that the item has strong correspondence with Item 9 (The staff at the treatment centre are sensitive to my background): a respectful consideration of the patient is only possible with some measure of sensitivity to their background, and vice versa. Accordingly, this item was deemed to be better located in the Quality scale.

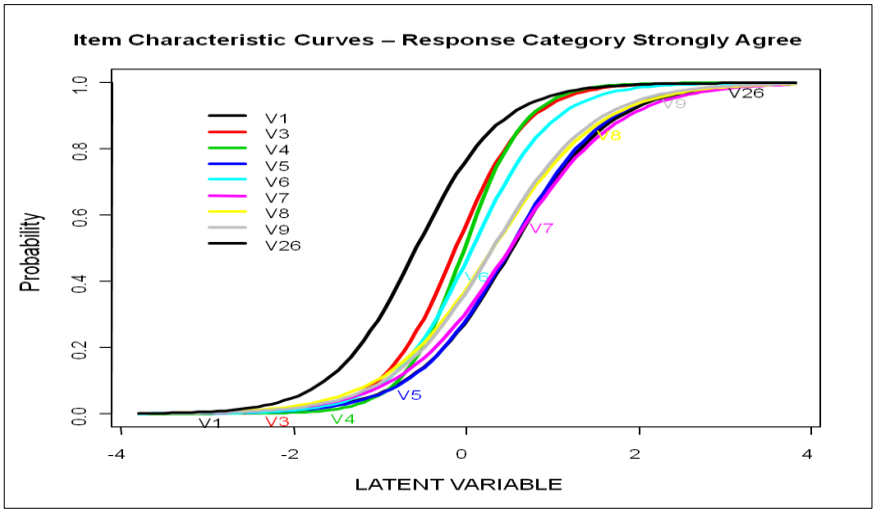
    The net effect of the removal of these two items was to render the Person Centeredness scale invalid as it constituted only one remaining item. Consequently, this scale was not tested.

Overall, then, the preliminary examination of the scales believed to comprise the Quality domain reduced the number of scales from three to one. As such, all items measuring these different scales were considered to be part of a single Quality scale and analysed accordingly. It must be noted that the collapsing of these scales into a single scale was also done when these items or items similar to these were tested in the United States. Further, as the scale in its revised form comprised the entirety of the Quality domain, the tests were conducted directly for the domain itself.

The Quality Scale (Domain)

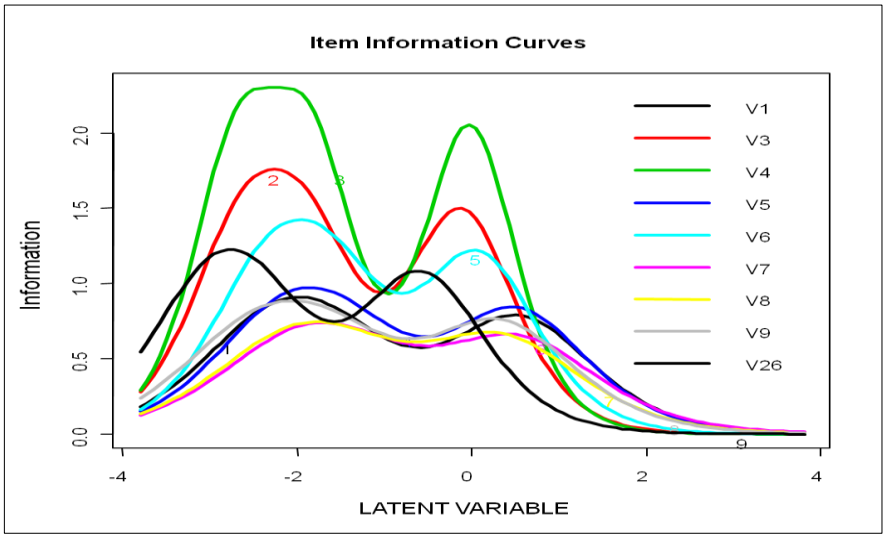
Having resolved some of the preliminary issues around the items and scales, and accounting for the preceding discussion of items considered either invalid or inappropriate located in terms of scales, a revised Quality scale comprising Items 1, 3, 4, 5, 6, 7, 8, 9 and 26 was formulated and subjected to a IRT analysis using Rasch modelling. The characteristic curves for the nine items in the scale are reflected in Figure 17.

FIGURE 17: Item Characteristic Curves for the Strongly Agree Category on all Quality Items



As Figure 17 indicates, while there is general consistency amongst the nine items measuring the Quality domain, there are some key differences as well. The probability of indicating strong agreement was highest for Item 26, indicating strong predisposition towards recommending the facility to others. In contrast, the probability of indicating strong agreement is lowest for Item 7 and Item 8. In the instance of Item 7 this may be attributable to the fact that treatment facilities require patients to conform to established treatment regimens which offer little choice for customisation, and in the instance of Item 8 it may be attributable to the low profile or even absence of safe sex counselling at the facility. These results were also somewhat predictable given the earlier analysis, and suggested weaker measurement of the latent variable Quality by these items. The remaining items all display relatively similar characteristic curves, suggesting better prediction of the Quality domain. This was confirmed by the item information curves (see Figure 18), which report the highest information being supplied to the measurement of Quality by Items 3, 4 and 6, while Items 7 and 8 report very low information levels, and Item 26 is somewhere in between these two extremes.

FIGURE 18: Information Supplied By The Items In Measuring The Quality Domain



To examine the scale in greater detail, an EFA was conducted entering all nine items into the analysis. The EFA realised two factors with an Eigen value greater than one and accounting for 53%

of the variance. The two factor solution confirmed the IRT analysis which suggested that some items were working consistently together while others were not, and hence the single factor solution may only be marginally correct.

To investigate the likelihood of this, CFA was performed specifying various latent variable models. In the first analysis two models were tested: the first model containing all nine items and the second model excluding Item 7 and containing eight items (see Table 17). As can be seen, the single factor solution comprising all nine items provided a poor fit to the data, with all model indices reflecting unacceptable values. The second model comprising eight items (excluding Item 7) fared significantly better, with all critical indices in the excellent range.

TABLE 17: Two Confirmatory Model Tests Of The Latent Structure For The Quality Scale (Domain)

MODEL	χ^2/df Ratio	RMSEA	Comparative Fit Index
All nine items	5.035	0.105	0.858
Eight items (excluding Item 7)	1.835	0.048	0.977

Once the eight item solution was confirmed a second analysis of the Quality scale was conducted specifying a two factor model. The results of this analysis indicated a very good model (χ^2/df Ratio = 1.933, RMSEA = 0.051, CFI=0.972). However, this model does not offer any significant improvement over the single factor model (its χ^2/df Ratio and RMSEA indices are in fact poorer than the single factor model), hence while it is equally plausible it is not of any appreciable additional value. For the purpose of the SAATSA, it would be better to regard the entire Quality scale as a single factor solution and to apply it as such in the future.

To finalise the Quality scale, a further IRT analysis was conducted to examine for differential item response as a function of respondent demographics. The results revealed consistent response characteristics on all items across the different demographic sub-samples with the exception of Item 8 (The treatment centre teaches me how to avoid getting HIV). Closer inspection of this item indicated that the differences were most pronounced for education and ethnic group. In the former case respondents with Grade 11 or less were significantly more likely to indicate strong agreement with the item than were respondents with Grade 12 or more education. In the latter case African and Coloured respondents were significantly more likely to indicate strong agreement than were white and Indian respondents. Taken together, the results indicate that the teaching of safe sex behaviour may be targeted more towards less educated African and Coloured patients than it towards better educated white and Indian patients. It is not clear whether or not this targeting is part of the policy of these facilities, but the fact that this pattern is consistent suggests some de facto targeting which is consistent with assumed prevalence rates in these sub-samples. If so, then there are implications for the item measurement properties. In the case of African and Coloured respondents, particularly those with lower levels of education, the item may be properly measuring *how well they received* their teaching on safe sex. In the case of white and Indian respondents,

particularly those with higher levels of education, the item may actually be measuring whether or not they *actually received* such teaching.

Despite the questions raised about the differential measurement by this item across the sub-samples, there are various reasons to retain it in the Quality scale. The first is that the testing of a latent model excluding this item did not produce any results that were significantly different from that obtained with the item included (χ^2/df Ratio = 2.056, RMSEA = 0.054, CFI=0.978). The second reason is the particular importance of the item for South Africa, with its high prevalence of HIV/AIDS as well as the well documented correlation between HIV/AIDS and substance abuse. Nevertheless, it is recommended that future iterations of the SAATSA establish the nature of such teaching on HIV/AIDS at the facility before soliciting patient opinions on the merits of such teaching. This would help significantly improve the psychometric properties of both this item and the Quality scale.

Key recommendation: Measurement of respondent views on the teaching received about HIV/AIDS should be properly contextualised by attendant measurement of the presence and nature of such programmes at the facility in question. That is, it must be established whether or not such programmes actually exist at the facility before patients are asked their views on these programmes.

Retaining Item 8, the Quality scale was analysed to obtain the scale reliability estimates for the entire sample and the estimates for the various demographic sub-samples (see Table 18).

TABLE 18: Reliability Of The Finalised Quality of Life Scale - Sample And Sub-Samples

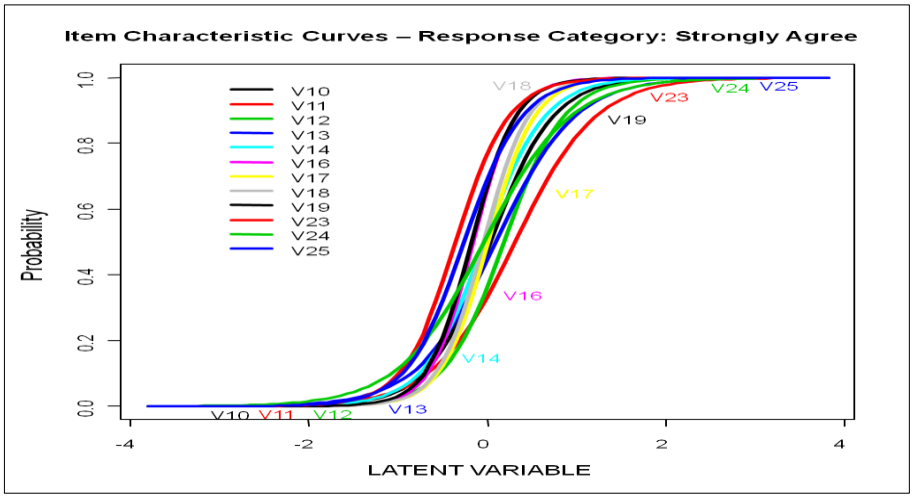
SAMPLE	SUB-SAMPLE	CRONBACH α
FULL SAMPLE		0.78
GENDER	Male	0.79
	Female	0.74
EDUCATION	Grade 11 or less	0.79
	Grade 12 or more	0.75
ADMISSION TYPE	Inpatient	0.78
	Outpatient	0.77
ETHNIC GROUP	Coloured	0.80
	African	0.77
	White	0.76

The Quality scale has very good reliability across the entire sample, and this reliability is relatively consistent across the demographic sub-samples. The most pronounced difference is between males and females, but both the estimates fall within the same confidence interval and hence are not significantly different from each other. Overall, then, the Quality scale works quite reliably for the entire sample, though it is likely that its reliability can be improved by consideration of the issues raised with regard to item 8.

The Effectiveness Domain

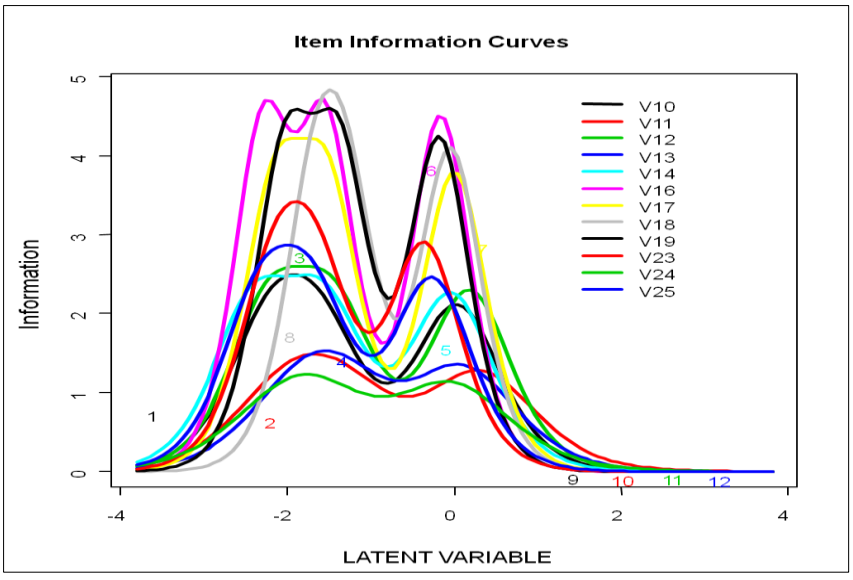
As indicated earlier, the Effectiveness domain is hypothesized to comprise three dimensions: quality of life, substance abuse and social connectedness. To test the psychometric properties of the domain, all items deemed during scale analyses to be a good fit for any of these three dimensions were collated to provide a single scale. This scale was submitted for an IRT analysis with the assumption of a single (higher order) latent dimension underlying all included items. The characteristic curves for each of the items in the Effectiveness domain are reported in Figure 19.

FIGURE 19: Item Characteristic Curves for the Strongly Agree Category on all Effectiveness Items



As Figure 19 reveals, the curve characteristics for the 12 items were relatively similar, suggesting somewhat equivalent probability and discriminant values for the individual items. As the items were so closed grouped together it was difficult to assess the extent to which they differed from one another. To clarify the differential contributions to the latent variable, the item information curves were plotted for each item, and the results of this are displayed in Figure 20.

FIGURE 20: Information Supplied By The Items In Measuring The Effectiveness Domain



As is seen in Figure 20, the IIC provides a better indication of how the different items grouped together in their determination of the latent variable Effectiveness. Three distinct groupings were evident, the first recording the highest amount of information (Items 16, 17, 18 and 19), the second providing a moderate amount of information (Items 10, 23, 24 and 25), and the third providing a relatively low level of information Items 11, 12, 13 and 14). These groupings suggest the presence of separate latent variables (predictable as the items were shown previously to comprise individual scales) though the assumption of a single latent variable at a higher order was still valid and held true for the IRT analysis.

To further investigate the validity of multiple latent variables with a higher order single latent variable, an EFA was conducted entering all items as is. The EFA extracted two factors from the twelve items, the first factor accounting for 53% of the total variance and the second accounting for 9%, realising a total of 63% of explained variance. The loadings for the items on the two factors are indicated in Table 19.

As more than one factor was extracted, the pattern matrix is provided. The pattern matrix output of the EFA provides loadings which are equivalent to standardised regression coefficients, that is, they indicate the unique contribution of an item to predicting the factor while holding the contributions of all other items constant. As expected, the two factors extracted during the analysis were significantly correlated ($r=0.740$, $p<0.01$).

TABLE 19: Factor Loadings For All Items Measuring The Effectiveness Domain

ITEM	Factor 1	Factor 2
Item 10: My general health is improving		0.637
Item 11: I am better able to cope when things go wrong		0.772
Item 12: I am better able to accomplish the things I want to do		0.894
Item 13: I am less likely to use alcohol or other drugs		0.504
Item 14: I expect to do better at work/finding work or at school		0.467
Item 16: There is some-one who cares about whether I am doing better	0.841	
Item 17: I have some-one who will help me when I have a problem	0.851	
Item:18 I have people in my life who are positive influence	0.733	
Item 19: The people who care about me are supportive of my treatment	0.654	
Item 23: I know that using alcohol and drugs is a problem for me	0.684	
Item 24: I need to work on my problems with alcohol and/or drugs	0.565	
Item 25: The treatment centre is helping me to recover from using drugs and alcohol	0.727	

As Table 19 reveals, the 12 Effectiveness items loaded neatly onto two factors, the first relating to quality of life and the second collapsing social connectedness and substance abuse into a single factor. The only misplaced item is Item 13, which was previously included in the substance abuse

scale but here loaded onto the quality of life factor. However, it may be argued that this loading is consistent with an interpretation of the item as a measure of the extent to which a respondent's life is free from drug and alcohol abuse: the lesser the prevalence of alcohol and drugs in their lives, the greater will be their quality of life. Accordingly, this result was considered acceptable.

The convergence of the social connectedness and substance abuse scales could be explained by the heightened importance of others in the lives of the respondents (partners, family, friends, etc.) and the impact of alcohol and substance abuse on the lives of these significant others. It is well established that self-realisation about the perils of alcohol and substance abuse are most pronounced when such abuse impacts the interpersonal relationships which are held most dear by the individual, meaning that the propensity to change is highest when it is motivated by an acknowledgement of the deterioration of such relationships and the consequent intention and desire to restore these to a former or new state of health. From this perspective, it is predictable that respondent views about their substance/alcohol problems would be highly correlated with their views about their social relations with others in their immediate familial and social milieu.

To confirm the latent structure for the Effectiveness domain, a CFA was conducted specifying three different models: the first comprising a single factor, the second comprising two factors as indicated by the EFA, and a three factor solution as indicated by the previous test of the individual scales comprising this domain. The latter two models were specified with factors correlated. The results of these three model tests are indicated in Table 20.

TABLE 20: Three Confirmatory Model Tests Of The Latent Structure For The Effectiveness Scale (Domain)

MODEL	χ^2/df Ratio	RMSEA	Comparative Fit Index
Single Factor Solution: Model 1	7.123	0.130	0.855
Two Factor Solution: Model 2	4.817	0.101	0.912
Three Factor Solution: Model 3	3.706	0.086	0.940

The three factor solution provides the best fit to the data, with all model indices significantly better than that for the other two models. However, the indices for Model 3 are still somewhat marginal, and to assess how the model could be improved the modification indices were examined. These specified the greatest parameter changes for correlated error terms for Item 13, which was anticipated at it loaded onto two scales. Specifying these correlated errors realised a model with the following indices: χ^2/df Ratio = 2.205, RMSEA = 0.059 and CFI=0.975. This model provides the best possible fit to the data accounting for the loading of Item 13 on more than one factor. All three factors were significantly correlated with each other.

Overall, then, the CFA confirmed the validity of the three dimensions as well as the correspondence amongst the three scales in their determination of the Effectiveness domain. It might be argued that the domain is better served by the removal of Item 13 but this is not recommended as the Item does provide a significant contribution to the Substance Abuse scale and eliminating it would prejudice this scale. Following the CFA, the reliability estimates were calculated for the Effectiveness domain for the entire sample as well as for the various demographic sub-samples (see Table 22).

TABLE 21: Reliability Of The Finalised Effectiveness Domain - Sample And Sub-Samples

SAMPLE	SUB-SAMPLE	CRONBACH α
FULL SAMPLE		0.91
GENDER	Male	0.91
	Female	0.89
EDUCATION	Grade 11 or less	0.91
	Grade 12 or more	0.91
ADMISSION TYPE	Inpatient	0.92
	Outpatient	0.90
ETHNIC GROUP	Coloured	0.91
	African	0.91
	White	0.91

The results of the reliability analyses revealed excellent reliability for the Effectiveness domain for the entire sample, with minor non-significant variations amongst the estimates for the subsamples. This confirmed the consistency with which the items were measuring the domain regardless of respondent demographics. These items should thus be retained intact for this domain.

CHAPTER 7: CONCLUSION

The purpose of this analysis was to examine the psychometric properties of the 26 SAATSA items and to confirm the dimensions and domains hypothesized to be measured by these items. The finalised results indicate a fairly robust instrument in some respects, with respectable validity for the individual scales measuring the dimensions and good correspondence between dimensions assumed to belong to the same domain. The measures were also found to be highly reliable, with reliability estimates ranging from good to very good to excellent for the different scales and domains.

Additionally, these reliability estimates were found to be relatively consistent across the different demographic sub-samples, and where differences were observed, they were found to be non-significant. This indicates considerable consistency in the measures regardless of respondent gender, education level, admission type and ethnic group (as a proxy for language). The lack of significance for respondent effects confirms the utility and validity of the cognitive testing undertaken during the development of the SAATSA.

A number of qualifications with respect to individual SAATSA items and the overall SAATSA questionnaire were noted and discussed. These qualifications have been addressed in the report by way of the various key recommendations, and any improvement of the instrument must give due consideration to each of these recommendations. In summary, the following are considered the most critical areas for improvement of the SAATSA questionnaire:

1. **Demographic Information** – The absence of measurement of home language or language spoken most often is a significant remiss and must be rectified immediately. There are various standard questionnaire items employed in the country for this purpose and one of these can easily be inserted as is into the SAATSA. Additionally, the education variable needs revision to provide better differentiation at discrete educational levels. Again, a standardised format is common currency in the country, and can be used intact for the SAATSA.

2. **Respondent Background Information** – apart from demographics, the SAATSA needs to ensure collection of key respondent background information such as admission data, primary substance abused, and so forth. This data must be collected directly from the facility and not from respondents. The Administrative Data Working Group has proposed a considered array of vectors to be collected from the facility. This data needs to be collected at individual level so that correspondence between returned questionnaires and respondent key background information is properly maintained (with due consideration of anonymity and confidentiality requirements).

3. **Item Response Format** – the analysis indicated that the Likert response format as employed from past experience elsewhere is robust and should be retained as is, i.e., a 4-point scale comprised of the following categories: Disagree, Somewhat Agree, Agree and Strongly Agree. However, the Not Applicable category as applied in this pilot study demands sharp attention. As recommended, this response category must be removed from all items SAATSA except for those items for which this category is conceptually valid and thus of some utility for analysis. Given the current configuration of the SAATSA (post-analysis) it is unlikely that this response category would be required at all.

4. **Question Routing** – there is a need for proper routing in the questionnaire to account for items that may not be relevant to certain respondents and/or facilities. One example of this is the teaching of safe sex at facilities. If such teaching is not present, respondents cannot be expected to provide their evaluation thereof.

5. **Fieldwork and Data Processing** – proper protocols are required for backchecking of returned data to ensure valid and vetted data for analysis. This is all that more critical when the SAATSA is implemented as the onus will be on facilities themselves to collect such data. These protocols need to ensure equally valid data across facilities and different respondent admission types.

6. **SAATSA items** – various recommendations have been proposed with regard to the individual SAATSA items, including modifications to existing items and/or deletion of certain items. These recommendations should be considered and addressed individually. Ideally, all recommendations should be implemented to build up the SAATSA into a robust measurement instrument.

7. **The Access Scale** – a substantive oversight in the SAATSA is the absence of a well developed scale measuring respondent access to substance abuse treatment. As argued previously, this is of paramount importance in a country where access to health services, both public and private, is well documented to be low and skewed depending on education, income and spatial location. Ideally, a full set of items should be developed to build a scale measuring access. Further, and as necessary, the measurement of access should account for the differential considerations applicable to inpatients and outpatients as the access issues across these two admission types are not always clear cut or even identical.

8. **The SAATSA Domains and Dimensions** – the psychometric analysis provided a detailed examination of the SAATSA in terms of the domains and dimensions, and the items within the dimensions. The finalised SAATSA measure is reflected in Table 22 (The Access and

Person Centeredness dimensions in the Quality domain are provided for illustrative purposes and are not intended to represent scales).

TABLE 22: Finalised SAATSA Domains, Dimensions (Scales) And Items

QUALITY DOMAIN	
ACCESS	Item 1: The amount of time I had to wait to get services was acceptable to me
QUALITY	Item 3: The staff at this treatment centre treat me with respect Item 4: The people I went to for treatment services spent enough time with me Item 5: I have a say in deciding about the substance abuse treatment that I am receiving here Item 6: The staff at this treatment centre told me about services in my area that will help me stay off drugs and alcohol Item 8: This treatment centre teaches me how to avoid getting HIV Item 26: I would recommend this treatment centre to a friend
PERSON CENTREDNESS	Item 9: The staff at this treatment centre are sensitive to my background
EFFECTIVENESS DOMAIN	
QUALITY OF LIFE	Item 10: My general health is improving Item 11: I am better able to cope when things go wrong Item 12: I am better able to accomplish the things I want to do Item 14: I expect to do better at work/finding work or at school
SUBSTANCE ABUSE	Item 13: I am less likely to use alcohol or other drugs Item 23: I know that using alcohol and drugs is a problem for me Item 24: I need to work on my problems with alcohol and/or drugs Item 25: The treatment centre is helping me to recover from using drugs and alcohol
SOCIAL CONNECTEDNESS	Item 16: There is some-one who cares about whether I am doing better Item 17: I have some-one who will help me when I have a problem Item 18: I have people in my life who are positive influence Item 19: The people who care about me are supportive of my treatment

REFERENCES

Baker, F.B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation.

Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Multivariate Software Inc., Encino, CA.

Bollen, K. and Stine, R. (1992). Bootstrapping Goodness-of-Fit Measures in Structural Equation Models. *Sociological Methods & Research,* 21.

Bond, T. And Fox, C. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* New Jersey, Lawrence Erlbaum Associates.

Cronbach, L.J. (1951). *Coefficient Alpha and the internal structure of tests*. Psychometrika, 16 (3).

Cronbach, L. J. and Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64(3).

DeVellis, R.F. (1991). *Scale Development*. Newbury Park, NJ: Sage Publications.

Doucette, A. (2008). *Modular Survey: Addressing the Need to Measure Quality*. The George Washington University.

Embretson, S.E. and Reise, S.P. (2000). *Item Response Theory for Psychologists*. New Jersey, Lawrence Erlbaum Associates.

Fields, A. (2009). *Discovering Statistics Through SPSS*. London, Sage Publications.

Mair, P. Hatzinger, R. And Maier, M.J. (2012). *Package 'eRm': Extended Rasch Modelling*. http://r-forge.r-project.org/projects/erm.

Myers**,** B., Louw J. and Pasche, S. (2010). Inequitable access to substance abuse treatment services in Cape Town, South Africa*. Substance Abuse Treatment, Prevention and Policy*, 5(28).

Myers B., Parry C.D.H. (2005). Access to substance abuse treatment services for black South Africans: findings from audits of specialist treatment facilities in Cape Town and Gauteng. *South African Psychiatry Review* 2005, 8.

Mair, P. And Hatzinger, R. (2007) Extended Rasch Modelling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, 20(9).

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Austria. http://www.R-project.org.

Rasch, G. (1977). On Specific Objectivity: An Attempt at Formalising the Request for Generality and Validity of Scientific Statements. *Danish Yearbook of Philosophy*, 14.

Reeve, B.P. (2008). *An Introduction to Modern Measurement Theory*. Division of Cancer Control and Population Sciences, National Cancer Institute.

Revelle, W. (2011). *The "New Psychometrics" - Item Response Theory.* www.personality-project.org/r/book/Chapter8.pdf.

Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software*, 17(5).

Venables, W.M., Smith D.M. and the R Core Team. (2012). Notes on R: A Programming Environment for Data Analysis and Graphics, Version 2.15.1.

APPENDIX A: THE SAATSA PILOT QUESTIONNAIRE

South African Addiction Treatment Services Assessment (SAATSA)

BASIC AND DEMOGRAPHIC INFORMATION:

1. Age	
2. Race	African / Coloured / Asian / White
3. Gender	Male / Female
4. What is your highest level of education?	Grade / Any tertiary
5. Name of city	
6. Name of this treatment centre?	
7. Is this an in- or outpatient centre?	inpatient / outpatient
8. When did you start the current treatment?	day / month / year
9. What is the primary substance for which you are receiving treatment now?	
10. How are you paying for your treatment?	
11. Have you received substance abuse treatment before?	yes / no
11.1. If 'Yes', when? (please provide the date/s)	
11.2. From which treatment centre/s?	

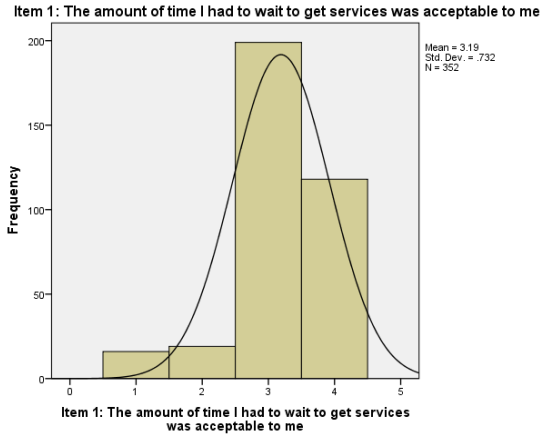
Please read each statement below and think about the rehab services you have received from the staff at this treatment center. Fill in the circle that best describes how you feel about services.

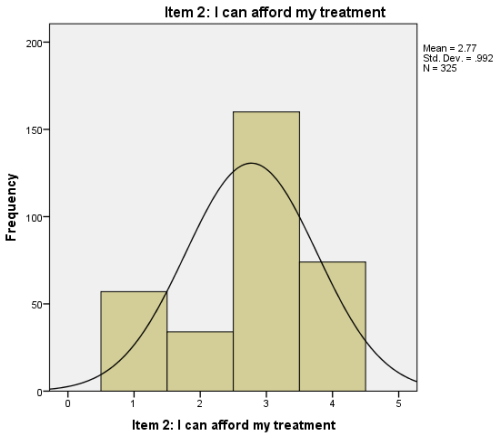
		Disagree	Somewhat Agree	Agree	Strongly Agree	Does Not Apply
For the statement below, please indicate the extent to which you agree…						
1	The amount of time I had to wait to get services was acceptable to me.					
2	I can afford my treatment.					
3	The staff at this treatment centre treat me with respect.					
4	The people I went to for treatment services spent enough time with me.					
5	I have a say in deciding about my substance abuse treatment that I am receiving here.					
6	The staff at this treatment centre told me about services in my area that will help me stay off drugs and alcohol.					
7	I am given a choice of services in this treatment centre.					
8	This treatment centre teaches me how to avoid getting HIV.					
9	The staff at this treatment centre are sensitive to my background.					
As a result of the services (treatment) I am receiving . . .						
10	My general health is improving.					
11	I am better able to cope when things go wrong.					
12	I am better able to accomplish the things I want to do.					
13	I am less likely to use alcohol or other drugs.					
14	I expect to do better at work/finding work or at school.					
15	I am more likely to practice safe sex.					
16	There is someone who cares about whether I am doing better.					
17	I have someone who will help me when I have a problem.					
18	I have people in my life who are a positive influence.					
19	The people who care about me are supportive of my treatment.					
20	My friends and family are more able to count on me.					
21	I have friends who are not using alcohol or drugs.					
22	I have someone who will listen to me when I need to talk.					
23	I now know that using alcohol and drugs is a problem for me.					
24	I need to work on my problems with alcohol and/or drugs.					
25	The treatment centre is helping me to recover from using drugs and alcohol.					
26	I would recommend this treatment centre to a friend.					

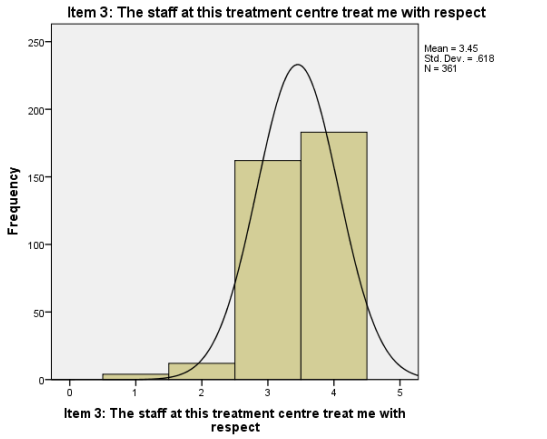
APPENDIX B: DISTRIBUTION OF MISSING VALUES FOR SAATSA ITEMS

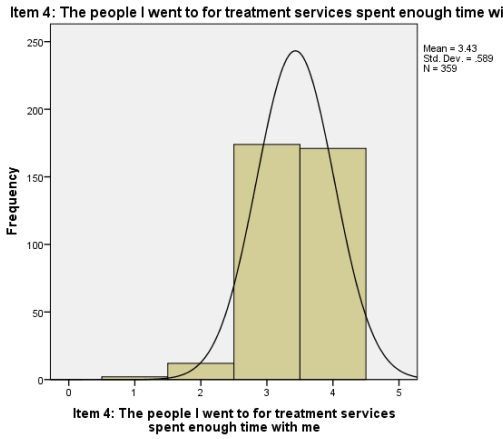
SAATSA ITEMS	VALID RESPONSES	MISSING FREQUENCY	MISSING PERCENTAGE
1. The amount of time I had to wait to get services was acceptable to me	393	6	1.5%
2. I can afford my treatment	385	14	3.6%
3. The staff at this treatment centre treat me with respect	394	5	1.3%
4. The people I went to for treatment services spent enough time with me	392	7	1.8%
5. I have a say in deciding about my substance abuse treatment that I am receiving here	388	11	2.8%
6. The staff at this treatment centre told me about services in my area that will help me stay off drugs and alcohol	392	7	1.8%
7. I am given a choice of services in this treatment centre	387	12	3.1%
8. This treatment centre teaches me how to avoid getting HIV	388	11	2.8%
9. The staff at this treatment centre are sensitive to my background	390	9	2.3%
10. My general health is improving	386	13	3.4%
11. I am better able to cope when things go wrong	384	15	3.9%
12. I am better able to accomplish the things I want to do	383	16	4.2%
13. I am less likely to use alcohol or other drugs	384	15	3.9%
14. I expect to do better at work/finding work or at school	385	14	3.6%
15. I am more likely to practice safe sex	382	17	4.5%
16. There is some-one who cares about whether I am doing better	385	14	3.6%
17. I have some-one who will help me when I have a problem	385	14	3.6%
18. I have people in my life who are positive influence	385	14	3.6%
19. The people who care about me are supportive of my treatment	382	17	4.5%
20. My friends and family are more able to count on me	385	14	3.6%
21. I have friends who are not using alcohol or drugs	383	16	4.2%
22. I have some-one who will listen to me when I need to talk	390	9	2.3%
23. I know that using alcohol and drugs is a problem for me	391	8	2.0%
24. I need to work on my problems with alcohol and/or drugs	389	10	2.6%
25. The treatment centre is helping me to recover from using drugs and alcohol	391	8	2.0%
26. I would recommend this treatment centre to a friend	390	9	2.3%

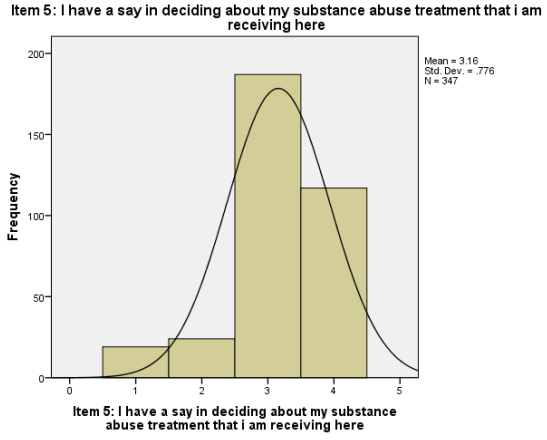
APPENDIX C: UNIVARIATE DISTRIBUTIONS OF SAATSA ITEMS

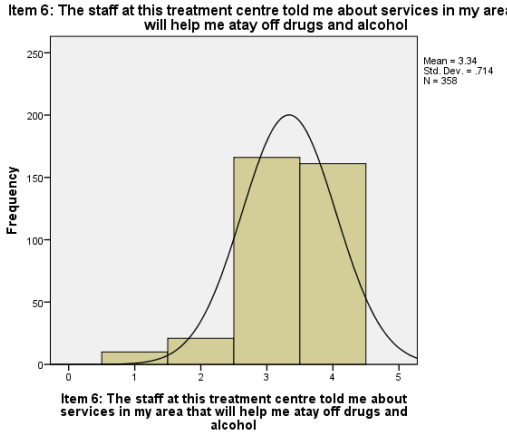


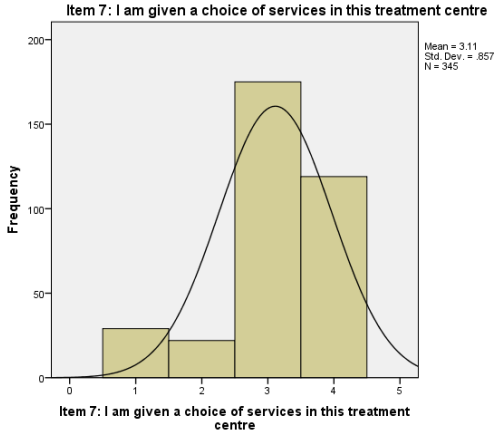


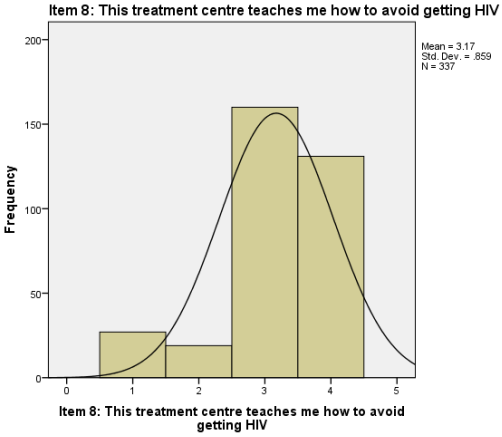


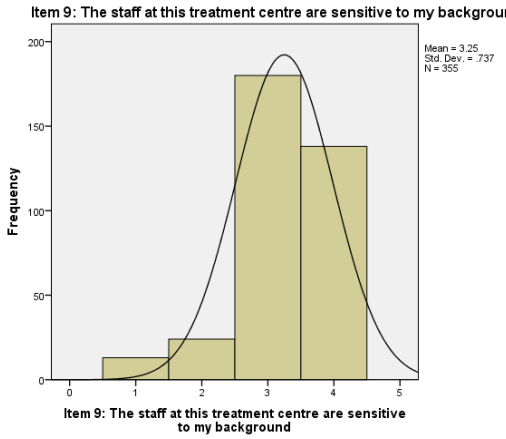


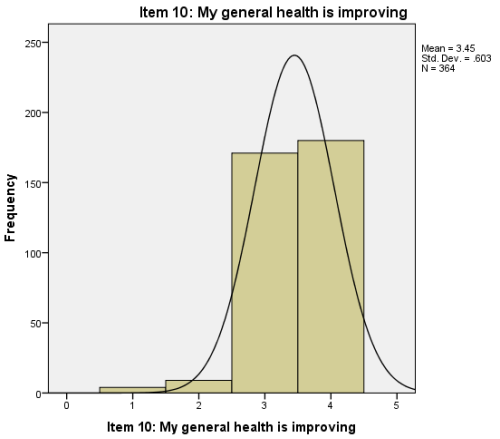




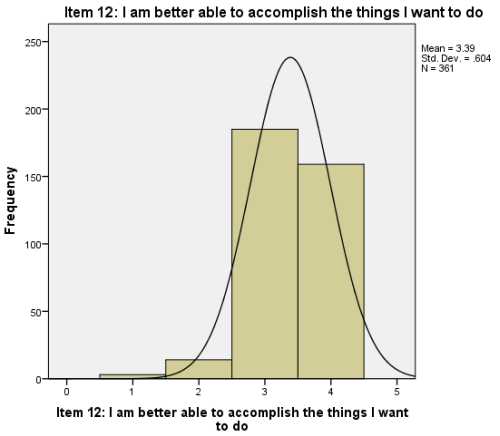

Item 7: I am given a choice of services in this treatment centre
Mean = 3.11
Std. Dev. = .857
N = 345
Frequency
Item 7: I am given a choice of services in this treatment centre

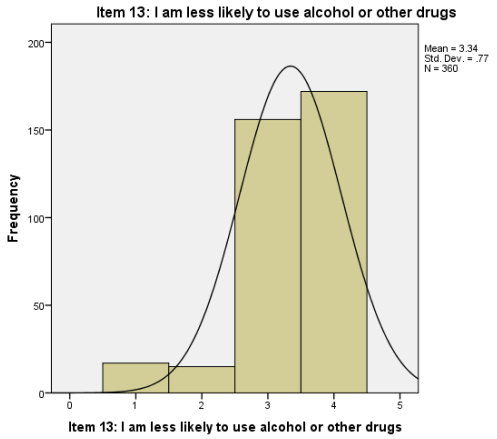

Item 8: This treatment centre teaches me how to avoid getting HIV
Mean = 3.17
Std. Dev. = .859
N = 337
Frequency
Item 8: This treatment centre teaches me how to avoid getting HIV

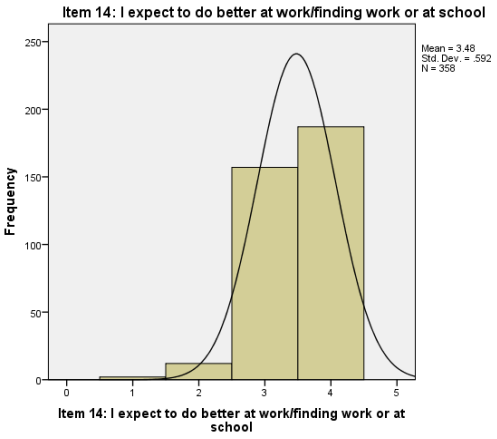

Item 9: The staff at this treatment centre are sensitive to my background
Mean = 3.25
Std. Dev. = .737
N = 355
Frequency
Item 9: The staff at this treatment centre are sensitive to my background

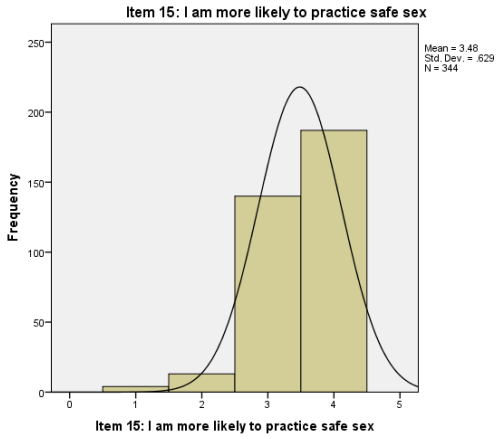

Item 10: My general health is improving
Mean = 3.45
Std. Dev. = .603
N = 364
Frequency
Item 10: My general health is improving

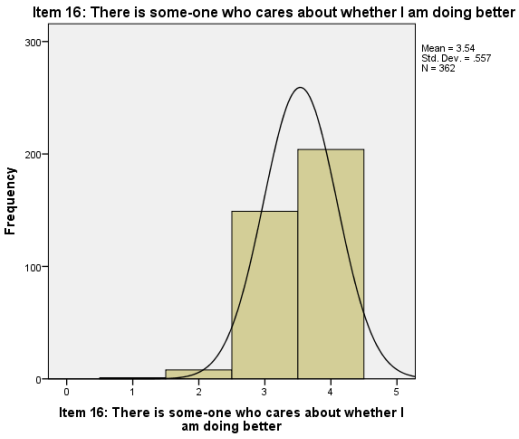

Item 11: I am better able to cope when things go wrong
Mean = 3.28
Std. Dev. = .724
N = 362
Frequency
Item 11: I am better able to cope when things go wrong

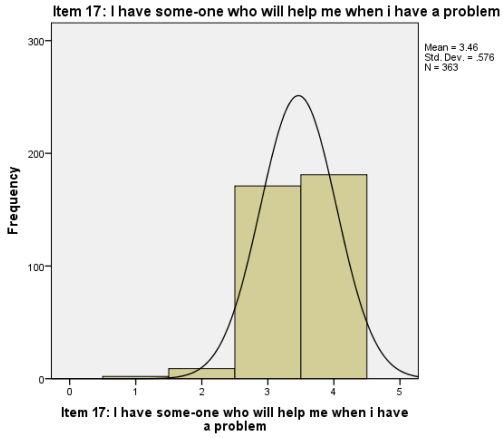

Item 12: I am better able to accomplish the things I want to do
Mean = 3.39
Std. Dev. = .604
N = 361
Frequency
Item 12: I am better able to accomplish the things I want to do

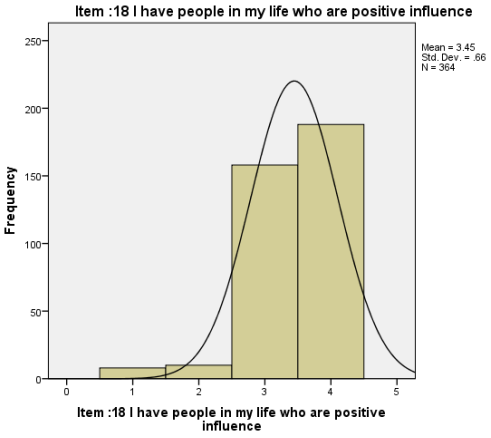

Item 13: I am less likely to use alcohol or other drugs
Mean = 3.34
Std. Dev. = .77
N = 360
Frequency
Item 13: I am less likely to use alcohol or other drugs

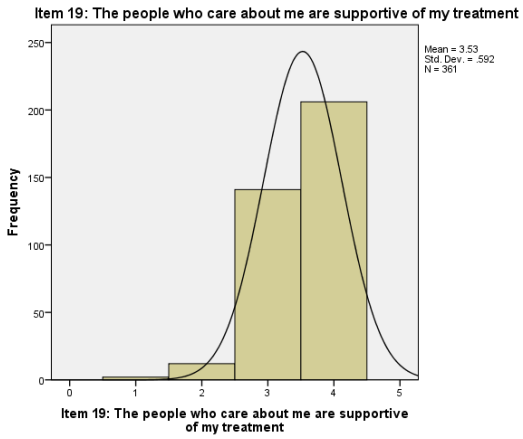

Item 14: I expect to do better at work/finding work or at school
Mean = 3.48
Std. Dev. = .592
N = 358
Frequency
Item 14: I expect to do better at work/finding work or at school

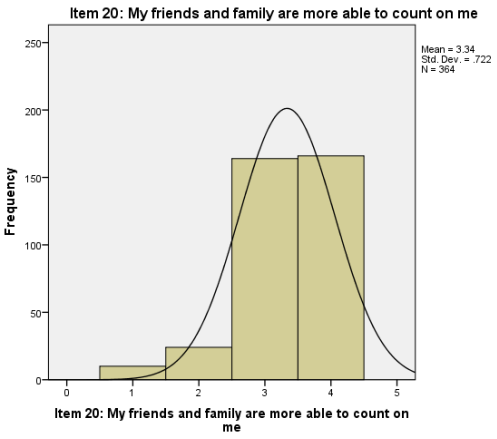

Item 15: I am more likely to practice safe sex
Mean = 3.48
Std. Dev. = .629
N = 344
Frequency
Item 15: I am more likely to practice safe sex

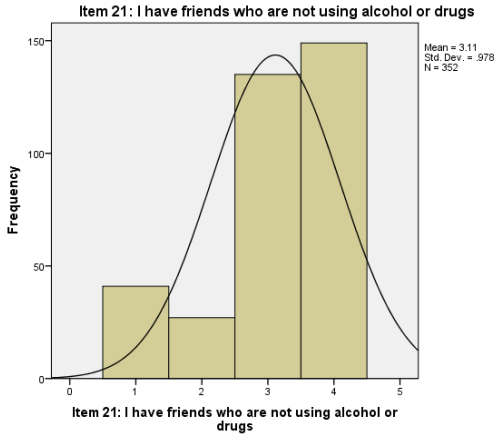

Item 16: There is some-one who cares about whether I am doing better
Mean = 3.54
Std. Dev. = .557
N = 362
Frequency
Item 16: There is some-one who cares about whether I am doing better

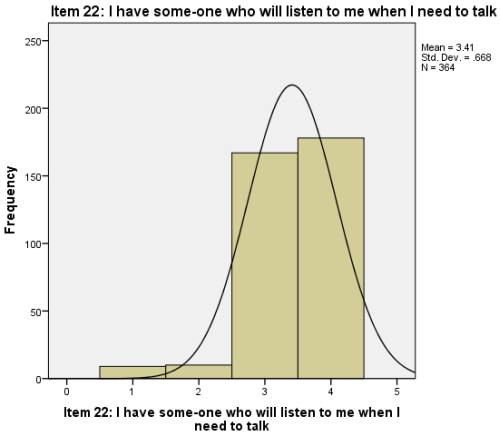

Item 17: I have some-one who will help me when i have a problem
Mean = 3.46
Std. Dev. = .576
N = 363
Frequency
Item 17: I have some-one who will help me when i have a problem

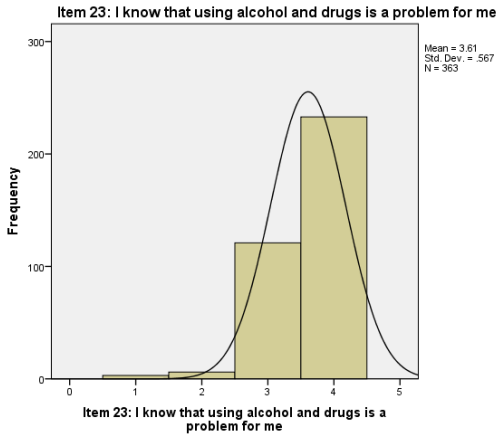

Item :18 I have people in my life who are positive influence
Mean = 3.45
Std. Dev. = .66
N = 364
Frequency
Item :18 I have people in my life who are positive influence

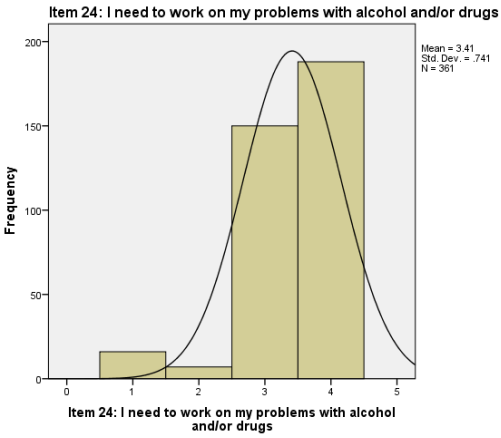

Item 19: The people who care about me are supportive of my treatment
Mean = 3.53
Std. Dev. = .592
N = 361
Frequency
Item 19: The people who care about me are supportive of my treatment

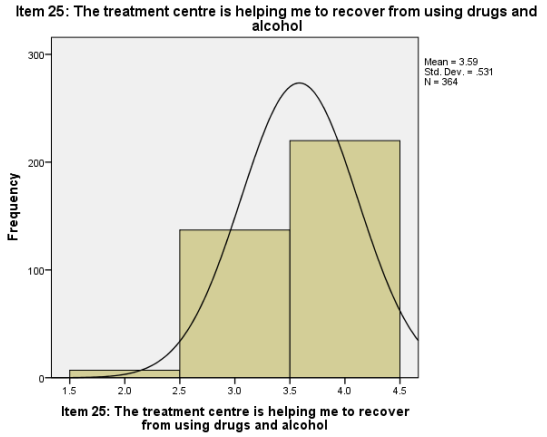

Item 20: My friends and family are more able to count on me
Mean = 3.34
Std. Dev. = .722
N = 364
Frequency
Item 20: My friends and family are more able to count on me

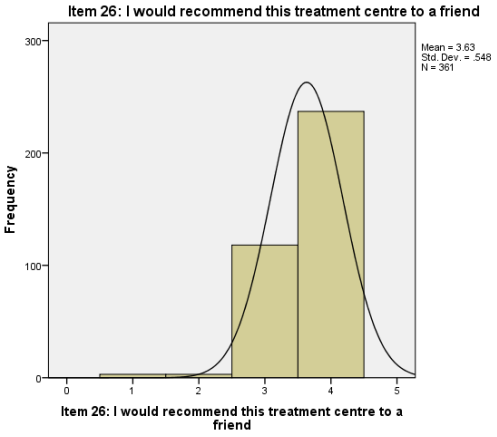

Item 21: I have friends who are not using alcohol or drugs
Mean = 3.11
Std. Dev. = .978
N = 352
Frequency
Item 21: I have friends who are not using alcohol or drugs


Item 22: I have some-one who will listen to me when I need to talk
Mean = 3.41
Std. Dev. = .668
N = 364
Frequency
Item 22: I have some-one who will listen to me when I need to talk


Item 23: I know that using alcohol and drugs is a problem for me
Mean = 3.61
Std. Dev. = .567
N = 363
Frequency
Item 23: I know that using alcohol and drugs is a problem for me


Item 24: I need to work on my problems with alcohol and/or drugs
Mean = 3.41
Std. Dev. = .741
N = 361
Frequency
Item 24: I need to work on my problems with alcohol and/or drugs


Item 25: The treatment centre is helping me to recover from using drugs and alcohol
Mean = 3.59
Std. Dev. = .531
N = 364
Frequency
300
200
100
0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
Item 25: The treatment centre is helping me to recover from using drugs and alcohol


Item 26: I would recommend this treatment centre to a friend
Mean = 3.63
Std. Dev. = .548
N = 361
Frequency
300
200
100
0
0
1
2
3
4
5
Item 26: I would recommend this treatment centre to a friend