# An evaluation of the assessment tool used for extensive mini-dissertations in the Master's Degree in Family Medicine, University of the Free State

**Brits H,** MBChB, MFamMed, FCFP, MHPE, Principal Specialist, Department of Family Medicine
**Bezuidenhout J**, BA, BA(Hon), MEd(Psychology), DTechEducation, Office of the Dean, Division Health Sciences Education
**Steinberg WJ,** MBBCh, DTM+H, DPH, DipObst, MFamMed, FCFP, Principal Specialist, Department of Family Medicine
**Joubert G,** BA, MSc, Head of Department, Department of Biostatistics
Faculty of Health Sciences, University of the Free State
**Correspondence to:** Hanneke Brits, e-mail: britsh@ufs.ac.za
**Keywords:** mini-dissertation, assessment tool, family medicine, Master's Degree

## Abstract

**Background:** Family Medicine became a speciality in South Africa in 2007. Postgraduate studies in Family Medicine changed from part-time Master of Family Medicine (MFamMed) to a full-time Master of Medicine (Family Medicine) [MMed(Fam)] degree, with changes in the curriculum and assessment criteria. The overall goal of this study was to evaluate the current assessment tool for extensive mini-dissertations in the postgraduate programme for Family Medicine, at the University of the Free State, and if necessary, to produce a valid and reliable assessment tool that is user-friendly.

**Method:** An action research approach was used in this study, using mixed methods. Firstly, marks given by 15 assessors for four mini-dissertations using the current assessment tool were analysed quantitatively. In Phase 2, the regulation of the assessment bodies and the quantitative results of Phase 1 were discussed by assessors during a focus group interview, and data were analysed qualitatively. An adapted, improved assessment tool (Phase 3) was developed and re-evaluated in Phase 4.

**Results:** The current assessment tool complied with the regulations of the assessment bodies. The scores allocated to specific categories varied with a median coefficient of variation of more than 15% in four of the possible 12 assessment categories. During the focus group interview, reasons for this were identified and the assessment tool adapted accordingly. During reassessment of the tool, individual assessors were identified as the reason for poor reliability.

**Conclusion:** The current assessment tool was found to be valid, but was not reliable for all assessment categories. The adapted assessment tool addressed these areas, but identified lack of training and experience in the assessment of extensive mini-dissertations by certain assessors as the main reason for unreliable assessment.

Ⓟ **Peer reviewed. (Submitted: 2013-03-06. Accepted: 2013-04-28.)**  **S Afr Fam Pract 2014;56(2)125-133**

## Introduction

The postgraduate programme in Family Medicine in South Africa changed from a three-year, part-time degree Master of Family Medicine (MFamMed) to a four-year, full-time speciality Master of Medicine (Family Medicine) [MMed(Fam)] in 2007. This change in the programme brought about a new curriculum, regulations and assessment criteria. A mini-dissertation was always part of the postgraduate Family Medicine programme, but was assessed differently, by different stakeholders.

Together with these changes, the new concept of a single-exit examination for all universities was introduced. Universities now use the same outcomes and are assessed accordingly. Therefore, the assessment criteria for the MMed(Fam) programme must now meet the regulations of the accredited programme at the university, as well as that of the single-exit examination body, currently the Colleges of Medicine of South Africa (CMSA).

The first objective of this study was to evaluate whether or not the assessment tool currently in use to assess an extensive mini-dissertation for a Master's Degree in Family Medicine at the University of the Free State is in line with the regulations for the requirements for an extensive mini-dissertation for a Master's Degree at the University of the Free State.

The second objective was to evaluate whether or not this assessment tool met the regulations of the current examining body's criteria, namely those of the CMSA.

The third objective was to adjust the current assessment tool if it was not found to be valid and reliable, and produce a valid and reliable assessment tool that was also user-friendly.

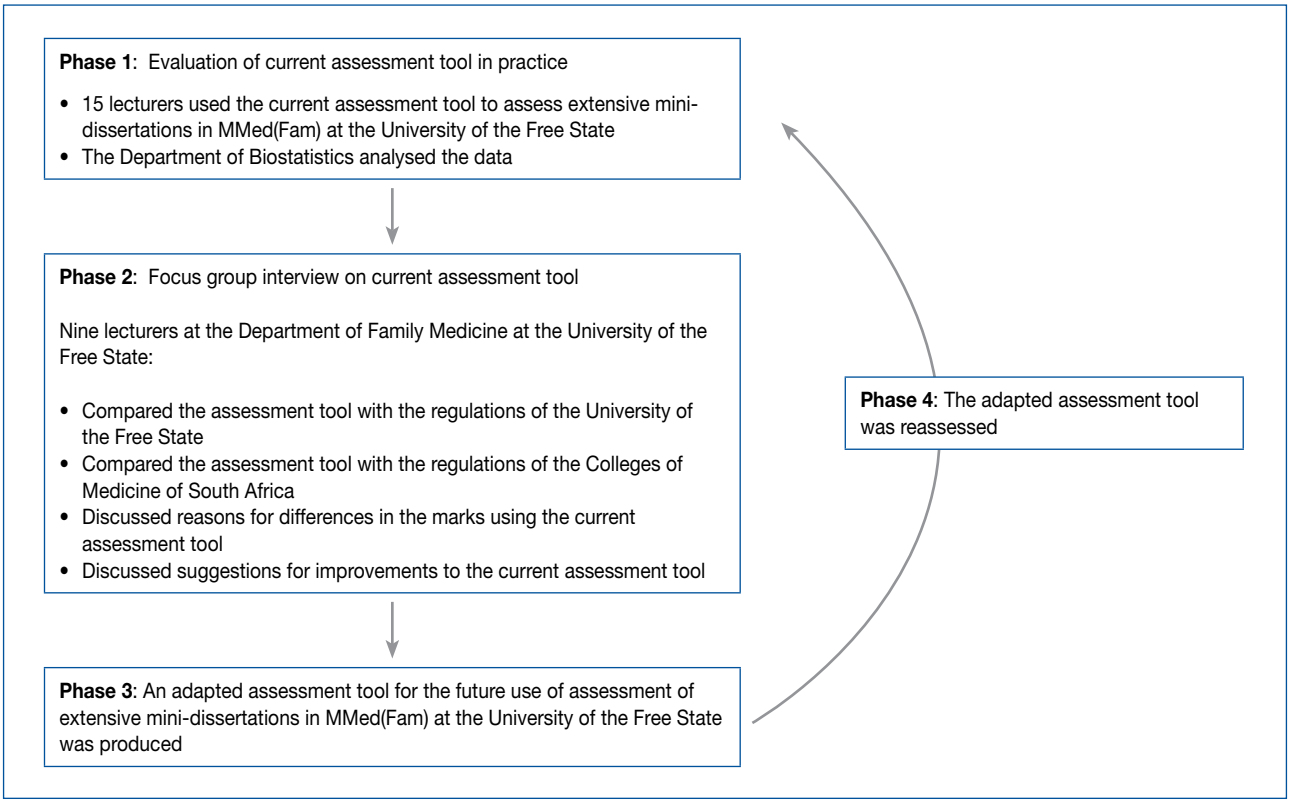


**Figure 1:** Flow diagram of action research steps followed in the study

The fourth objective was to replace the old assessment tool with the new assessment tool, and to re-evaluate the adjusted assessment tool in 2012.

Kurt Lewin,[1] a German psychologist, first described and practised action research. Creswell[2] describes action research as a process which aims to provide solutions to practical problems, where the main focus is to bring about change. Action research is a cyclic process, which starts by identifying a problem, planning an intervention through data gathering and open discussion, and then implementing the intervention and evaluating the outcome.[3]

## Method

### The research design in this study

An action research approach that used mixed methods was followed. This project consisted of four phases, as illustrated in Figure 1. In the first phase, the strategy for data collection was the assessment of extensive mini-dissertations with the current assessment rubrics, and data were analysed quantitatively. Phase 2 consisted of data gathering during a focus group interview, and the data were analysed qualitatively. In Phase 3, a new assessment tool was developed, and replaced the current tool. In Phase 4, the adopted tool was re-evaluated using the same methods as those used in Phase 1.

### Phase 1: Use of the assessment rubric and quantitative data collection

The target population for this study consisted of all Family Medicine lecturers involved in the assessment of extensive mini-dissertations for postgraduate studies in Family Medicine at the University of the Free State. This included internal and external assessors.

The entire target population, namely all of the lecturers in the Department of Family Medicine at the University of the Free State, was selected for the internal assessors. The departmental database of available assessors was used to select the external assessors. Because two of the mini-dissertations were in Afrikaans, only assessors who were proficient in Afrikaans could be selected to conduct the assessments. Therefore, the heads of the Family Medicine Department of each of these four medical schools with Afrikaans-speaking lecturers were asked to nominate one assessor from their department to participate in the study. The sample consisted of 11 internal assessors from the University of the Free State and four external assessors, one each from the University of Cape Town, Stellenbosch University, the University of Pretoria and the University of the Witwatersrand.

A pilot study was not carried out on the assessment rubric as the tool that was evaluated had been in use for many years. The assessment rubric in use is a single-page document. There are spaces for the student's name and the title of the extensive mini-dissertation. The assessment

**Table I:** Questions asked during the focus group interview

| | |
|---|---|
| Question 1 | Does the assessment tool currently in use to assess an extensive mini-dissertation for an MMed(Fam) at the University of the Free State meet the regulations for the requirements of an extensive mini-dissertation for a Master's Degree at the University of the Free State?<br>• The latest regulations regarding extensive mini-dissertations for a Masters Degree were used as a reference.[7-9] |
| Question 2 | Does the assessment tool currently in use for the assessment of extensive mini-dissertations for an MMed(Fam) meet the regulations of the current examining body's criteria, namely that of the Colleges of Medicine of South Africa?<br>• The latest regulations regarding the requirements for the assessment of mini-dissertations were downloaded from the website of the Colleges of Medicine of South Africa,[10] and used as a reference. |
| Question 3 | Why did different assessors score the same mini-dissertations differently when they used the same assessment tool?<br>• The results of the quantitative study and specifically areas where vast differences in mark allocation occurred were discussed in order to identify the reasons for these discrepancies.<br>• Copies of the mini-dissertations were available in order to refresh memories, if necessary. |
| Question 4 | How can we improve the current assessment tool in order for it to be valid and reliable, as well as user-friendly?<br>• The layout of the assessment tool, clarity on what was expected under each component of the tool, and weighting of marks, were discussed, as well as suggestions on how to improve the tool.<br>• Remarks and suggestions from external assessors were also mentioned and discussed during the focus group interview. |

tool consists of 12 assessment categories, each with a different weight allocated to it. There is a brief description of what is expected under that specific criterion below each assessment category. The total score of the rubric is 120, and marks are converted to a percentage. There is space for general remarks at the bottom of the page, as well as a tick box where the assessor must indicate if the student passes, passes with a distinction, needs improvement or fails the extensive mini-dissertation.

Each assessor received the total number of mini-dissertations in 2011 in the form of hard copies, as well as the assessment rubric, after he or she had signed an informed consent form to participate in the research project. Each assessor was given six weeks in which to complete the assessments, and received an honorarium per assessment on completion thereof.

Data from the assessment rubrics were entered on an Excel® spreadsheet after being checked for correctness. There were no missing data on the assessment forms. The assessors were not identified by name, but grouped into different categories, namely internal assessors who had completed the formal Short Programme on Assessment and Learning in Higher Education (SPALHE) course at the university, and internal assessors who had not undertaken the SPALHE course, and external assessors. The Department of Biostatistics used SAS®[4] to analyse the data quantitatively.

Data were interpreted and summarised by examining the combined results of all of the assessments. The mean and standard deviation of the assessors was calculated for each of the four mini-dissertations, and also for each of the 12 assessment criteria. From this, the coefficient of variation (CV) was calculated for each report, for each item. The CV "expresses the standard deviation as a percentage of the mean."[5] Owing to the difference in weight between the different assessment criteria, the CV was used to compare variations between the different criteria.

### Phase 2: The focus group interview and qualitative data analysis

The method for a focus group interview, as described by Boddy, was used in this phase.[6] The difference between a focus group interview and a focus group discussion is that in the former, the main objective is to obtain answers to specific questions, while in a focus group discussion, the interaction between the group and the group dynamics is as important as the information gathered.[6]

Owing to logistical reasons, only the internal assessors participated in the focus group interview. However, remarks from external assessors were included using the assessment tools, and were also discussed.

The sample consisted of nine of a possible 11 consultants from the Department of Family Medicine, University of the Free State. The tenth consultant was on sabbatical leave, and was not available for the focus group interview. The researcher, the eleventh member of the University of the Free State Department of Family Medicine, acted as the facilitator for the interview.

After the process was explained to the participants, they signed informed consent. The researcher conducted the focus group interview in English, and asked specific, focused questions in order for the research questions to be answered. The group decided that if at least seven of the nine assessors agreed on a specific answer to questions 1 and 2, and that no more discussion would convince them otherwise, it would be considered to constitute agreement. The four discussed questions are displayed in Table I. The focus group interview took 70 minutes, and agreement was reached on all of the answers.

The focus group interview was video- and audiotaped, and the researcher and an administrative assistant took notes during the interview. All of the responses were transcribed by the researcher from the audio recording and verified from the video recording when necessary. These transcriptions were given to the assessors to check for accuracy before being analysed.

The data gathered were grouped for the four questions, and analysed with the qualitative content analysis approach, as described by Hsieh and Shannon.[11]

### Phase 3: Adjustment of current tool and implementation of the adapted tool

The current tool was adjusted according to the recommendations of the assessment panel. The adapted tool was distributed among the assessors who participated in the focus group interview to ensure that it accurately reflected what was agreed. After agreement, the current tool was replaced by the adapted tool at the beginning of 2012.

### Phase 4: Reassessment of the current tool

In Phase 4, exactly the same steps were followed as those in Phase 1, but only two mini-dissertations were available. Nine internal and one external assessor evaluated each of the extensive mini-dissertations.

### Ethical considerations

The protocol was approved by the Ethics Committee of the Faculty of Health Science of the University of the Free State (ETOVS Number 38/2011).

Each assessor or lecturer from the Department of Family Medicine at the University of the Free State, as well as the four external assessors signed informed consent forms before participating in the study.

The names of the candidates were not revealed to the assessors. Information was handled confidentially. Mutual respect for one other formed the basis of the focus group interviews. Although it was possible to identify assessors, their assessments were handled anonymously. If desired, the results of their personal assessments in comparison with those of the group, could be obtained. This proved to be a good learning opportunity for the assessors.

## Results

### Phase 1: Use of the assessment rubric and quantitative data collection

All 11 consultants from the Department of Family Medicine at the University of the Free State, as well as all four external assessors, completed the assessment rubrics. This provided a total of 56 completed assessment rubrics for evaluation, as the two English-speaking internal assessors were unable to assess the two Afrikaans dissertations. The response rate was 100%.

The group of assessors consisted of seven females and eight males. Different race groups were represented (black, white and Asian) and their experience as assessors varied from three months to 28 years. The assessors were also divided into three subgroups, namely internal assessors with SPALHE, internal assessors without SPALHE and external assessors. The combination of all of the assessors will be referred to as the assessment panel henceforth.

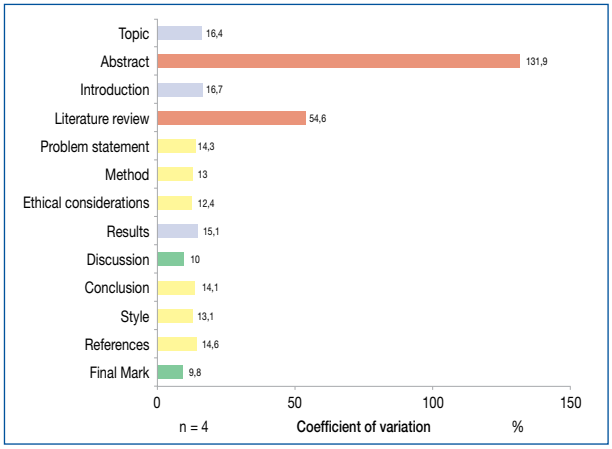


**Figure 2:** The median coefficient of variation for the different assessment categories

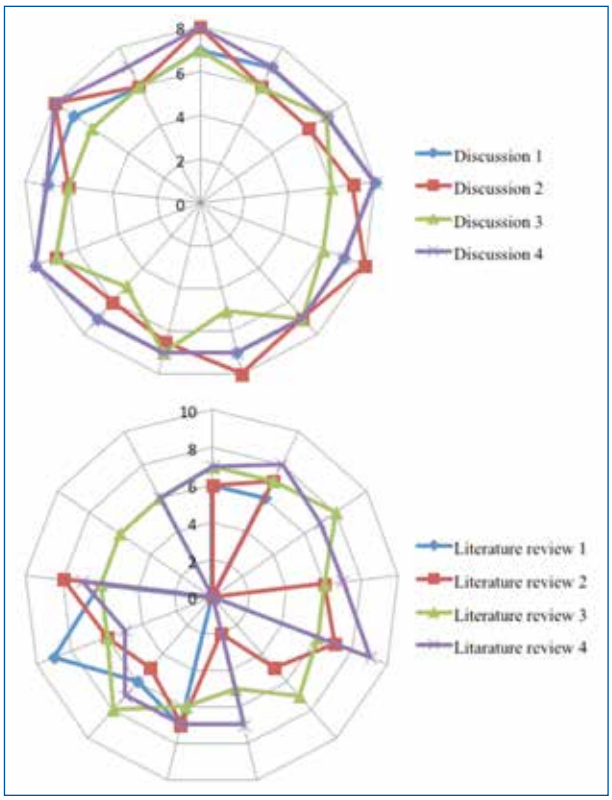


**Figure 3:** The difference between the "constant" marks allocated for the "Discussion" and the "scattered" marks allocated for the "Literature review"

The final marks allocated to each mini-dissertation were consistent with a median CV of only 9.8%. However, when each assessment category was analysed, the median CV was between 10% and 131.9%. Two of the assessment categories, namely the "Abstract" and the "Literature review" varied the most; by 131.9% and 54.5%, respectively. The rest varied by between 10% and 16.7%. The median CV for each assessment category is indicated in Figure 2.

Owing to the large CV in the "Abstract" and "Literature review", when compared to the other assessment categories, a separate radar graph (Figure 3) has been used to display the visual variance between the marks allocated

**Table II:** Coefficient of variation for the different assessment categories between the subgroups of assessors

| Assessment category | CVs of internal assessors with SPALHE (n = 5) | CVs of internal assessors without SPALHE (n = 4) | External assessors (n = 4) |
|---|---|---|---|
| Topic | 8.1 | 13.3 | 14.9 |
| Abstract | 118.6 | 107.7 | 18.3 |
| Introduction | 15 | 13.3 | 21 |
| Literature review | 39.6 | 66.9 | 93.4 |
| Problem statement | 9.2 | 9.5 | 17.6 |
| Method | 9.9 | 6.1 | 12.1 |
| Ethical consideration | 13.5 | 14.4 | 13.6 |
| Data management | 14 | 10.5 | 14.9 |
| Discussion | 9.6 | 10.6 | 10.2 |
| Conclusion | 13.7 | 8.2 | 12.4 |
| Style and layout | 13.8 | 7.9 | 10.2 |
| References | 14.9 | 7.4 | 12.1 |
| **Total marks** | **7.7** | **8.3** | **9.2** |

CV: coefficient of variation, SPALHE: Short Programme on Assessment and Learning in Higher Education

< 10% variation
10-15% variation
16-50% variation
> 50% variation

to an assessment category that scored "constant" marks and an assessment category that scored "scattered" marks.

In the bottom part of Figure 3, the marks allocated for the "Literature review" varied from 0-9 for mini-dissertations 1 and 4, 0-8 for mini-dissertation 2, and 5-8 for mini-dissertation 3 out of a possible 10, while the marks allocated for the "Discussion" varied from 5-7 for mini-dissertation 3, and 6-8 for the other three dissertations, out of a possible 10 in the top part of Figure 3.

The median CV was calculated per assessment category for each dissertation, after which individual and subgroups were compared with the gold standard, the combined average median marks of the assessment panel. No specific individual could be identified to have scored considerably differently from the others. The median coefficient of variation in each subgroup is displayed for each assessment category in Table II. Although differences of more than 100% occurred between the different assessment categories, all three subgroups of assessors gave a final mark within 10% of the median CV. A variation of < 10% was considered to be good, and ≥ 16% to be poor. As no specific value was available from the literature, values were decided on by consensus among the assessors.

The final marks allocated by the three subgroups of assessors for each of the dissertations showed a difference between the different dissertations, but a small variation between the different groups of assessors.

**Phase 2: The focus group interview and qualitative data analysis**

### Question 1

Does the assessment tool currently in use to assess an extensive mini-dissertation for an MMed(Fam) at the University of the Free State meet the regulations for the requirements of an extensive mini-dissertation for a Master's Degree at the University of the Free State?

To start the discussion, the first question posed to the participants was: "What do you think of these regulations?"

Answers like: "It's informative enough", "The guidelines are clear and easy to understand", "I don't see any problems with the guidelines", were the initial responses. "I'm not sure if we are doing number (h), the summary. Are we counting the words, and is it in Afrikaans and English?" followed. Number (h) stipulates: "A summary in Afrikaans and English of no more than 600 words each must be included at the back of the dissertation, together with approximately 10 key terms in English describing the subject of the dissertation" (University of the Free State, 2001: 45).

After the discussion, the group concluded that it would prefer an abstract instead of a summary. It also interpreted the guidelines as being that Afrikaans dissertations must have an English summary, but that English dissertations probably did not need an Afrikaans summary. The words in the abstract were not counted when the dissertation was assessed.

In conclusion, agreement was reached that there is compliance with the regulations of the university with regard to the assessment of mini-dissertations, and that there were no contradictions between the assessment tool and the regulations for postgraduate studies at the University of the Free State.

### Question 2

Does the assessment tool currently in use for the assessment of extensive mini-dissertations for an MMed(Fam) meet the

regulations of the current examining body's criteria, namely that of the CMSA?

Responses were: "Yes, and more. We included everything in their list. Maybe we used different words in some instances", and "Yes, I think that we included more in the current assessment tool. The grouping of information is a little different, but I'm happy that we comply with the regulations". The trend of all of the answers was along these lines.

Participants agreed that the assessment tool complied with the regulations of the CMSA with regard to the regulations for mini-dissertations in Family Medicine.

### Question 3

Why did different assessors score the same mini-dissertations differently when they used the same assessment tool?

Overall, remarks and clarification of terms and results were encouraged by the facilitator. It was explained that the assessment results of all of the assessments pertaining to each dissertation were compared for a variation in marks.

"How can you have such a big difference in marks? Are you sure you calculated it correctly?" was asked. This comment specifically pertained to the marks allocated to the "Abstract" and the "Literature review". The facilitator clarified that there were no abstracts in two of the dissertations, and that some assessors had allocated marks to this assessment category. After some amusement, this was taken seriously. A summary or abstract is a requirement of the university and of the CMSA. Therefore, it must be part of the extensive mini-dissertation. Everyone was in agreement that marks cannot be allocated if an abstract is absent.

Other comments included: "Although there was not a specific heading for 'Literature review' (in the mini-dissertation), I evaluated the literature used throughout the dissertation and allocated marks to it", "No, if there was not a specific 'Literature review', I scored them zero for that category", "I deducted some marks if a heading was not present, but still gave them marks", "Should you not rather deduct the marks for the category of 'Style and layout', and give them marks for the 'Literature review?'" and *"Dis goed dat ons daaroor praat"* ("It's good that we talk about it").

After these and other remarks, it was agreed that there was disagreement on how to allocate marks for the literature review. Some discussion followed, and after an anonymous vote, eight of the nine participants indicated that there should be a separate heading for "Literature review". A discussion also followed on whether or not the number of words to use in the literature review or numbers of references should be stipulated to the students. The conclusion was that the components indicated on the assessment rubric, namely: "Relevant, current, well interpreted and connected to study" were all important, and should all be addressed when allocating marks, rather than the number of words or references.

These were some of the responses to marks being allocated to the "Method" category: "I think, my own opinion is that I don't know enough about all the different methods to assess the 'Method' properly", "I would say that it is an assessor problem, rather than an assessment tool problem", "The assessors need training!" "Not even SPALHE taught us how to assess dissertations," "Supervision can also help to improve our knowledge and experience of assessment of dissertations," and "It boils down to one thing: we all need to refresh ourselves". In conclusion, all of the participants agreed that they needed more training on research methods, and more experience as supervisors to improve their own knowledge. The formal assessment course at the university, SPALHE, does not address the assessment of dissertations.

In the category, "Display of data" or "Results", these remarks were made during the discussion: "For me, proper display and good graphs will ensure good marks", "Most people are visual, and would like to see pictures or graphs," and "I think the display should be transparent and appropriate to what you want to achieve in your study, and not a rewriting of all the results that you have". Consensus was reached that the display of data or results should be "clear and accurate", as indicated by the assessment tool, but that the part "answer the research questions" should be changed to "in context with the research questions".

### Question 4

How can we improve the current assessment tool in order for it to be valid and reliable, as well as user-friendly?

A key comment was: "I think that the first part of the question was answered in questions 1 and 2, where we determined what we needed or wanted to assess, and if we assessed that. According to the regulations, we assessed what we are supposed to assess, and therefore the tool is valid in my opinion". Everybody agreed that the assessment tool was valid.

The current assessment tool was not reliable for four of the 12 assessment categories in producing the same results when different assessors used the same tool to evaluate the same dissertations. Therefore, steps are necessary to address this.

Two comments from the external examiners were: "Should you not change the total score to 100, rather than the 120 marks currently in use?" and "It's nice that the marks are not out of 100". Some of the remarks from participants on these suggestions were: "That will be easier to calculate", "I tend to calculate all the time when it is out of 100, and I don't think it is good," and "If you change the total score, you will need to change the weighting of each assessment category, and I think it's fine." Consensus was reached that the score out of 120 should be retained, and to keep the weighting the way it is presently.

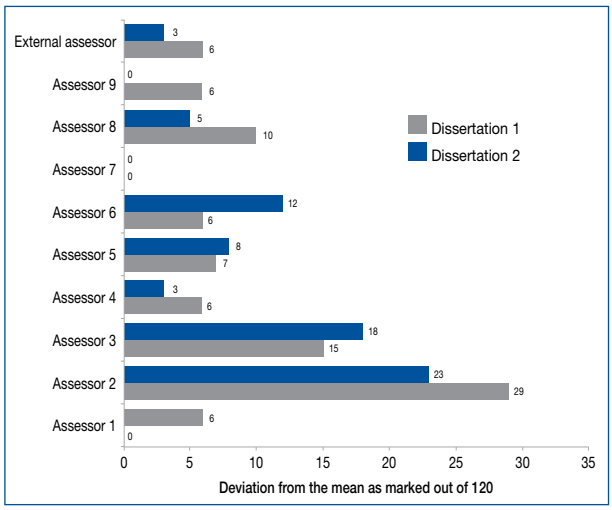


**Figure 4:** Deviation from the mean of the final marks of assessors

All of the participants agreed with the following statements: "I would like to suggest that we add a remarks column next to each assessment category as part of feedback to the students, and to be able to justify our marks", "I would encourage the use of decimals when allocating marks, and give marks like 6.5, rather than only 6 or 7", and "Global assessment should stay part of the assessment".

After discussion, the following was agreed regarding the assessment tool:

- It should be kept as a one-page document.
- The layout of the tool should be retained in its current format.
- The use of decimals or half marks should be employed, when appropriate.
- The global assessment should be retained as part of the assessment tool.
- The words in the "Data management" category should be changed (as discussed earlier).
- A column for comments should be added next to each assessment category.

Important identified areas from the discussion were the lack of training of assessors regarding:

- Use of the assessment tool.
- General research methods.
- Expertise in the supervision of the research projects.

**Phase 4: Reassessment of the new tool**

Nine internal assessors and one external assessor assessed two extensive mini-dissertations each. The final allocated marks varied from 50-73%, with a mean of 65% for dissertation 1; and from 58-88%, with a mean of 65% for dissertation 2. The CVs for different assessment categories did not vary markedly, and no specific assessment category could be identified as problematic. However, the final mark by individual assessors varied from 0-29% of the mean. Two individual assessors allocated marks for both dissertations that were more than 13% above or below the average marks allocated for the individual dissertations. In both cases, the one mark was above the average and the other mark was below it, while the other eight assessors all allocated marks within 10% of the average for both dissertations. The deviations from the mean of the final marks of the individual assessors are displayed (Figure 4).

## Discussion

Although only four dissertations were available for assessment, the 100% response rate of the assessors and the number of assessors ensured that 56 assessments could be compared with the current assessment tool. The assessors also represented a full spectrum with regard to gender, race, training and experience.

Two categories, namely the "Abstract" and the "Literature review", were responsible for these big differences. A study by Pathirage, Haigh, Amaratunga and Baldry[12] on the consistency of marks allocated to undergraduate dissertations also showed this difference in marks allocated per assessment category, as well as the overall marks. In their study, they compared the original marks allocated to a dissertation with those of assessors participating in their study. The marks allocated for "Referencing" varied from 39-95%, and the final marks for a specific dissertation were from 56-87% of the original mark allocated to the dissertation. These dissertations were marked by personnel assessing student dissertations as part of their work. In the study by Bettany-Saltikov, Kilinc and Stow[13] on the assessment of Master's dissertations, only the final marks were compared (not specific assessment criteria), and they varied by as much as 11%. Therefore, the current assessment tool is not reliable as different assessors allocated different marks (CV ≥ 16%) in four of a possible 12 assessment categories when they assessed the same mini-dissertations with the same assessment tool.

When the different subgroups, namely the the internal assessors with formal assessment training (SPALHE), the internal assessors without formal assessment training, and the external assessors, were compared, the same differences occurred for the different assessment categories. The final marks allocated were consistent for each dissertation, but varied between the different dissertations. The marks allocated by the three subgroups correlated with each other, although the subgroup without SPALHE training gave the highest marks to three of the four dissertations. The formal assessment course for lecturers at the University of the Free State (SPALHE) does not include a section on the assessment of dissertations.

The general arrangements for the focus group interview worked well. Participants contributed nicely and nobody dominated the discussions. Participants were open and admitted their own shortcomings. The discussions took place in a relaxed atmosphere, and finished within the allocated time.

Questions 1 and 2 addressed the validity of the assessment tool. After studying the regulations of the University of the Free State and the CMSA, participants were satisfied that the current assessment tool is valid. In a paper by Matthews[14] on *Dissertation: issues in guidance, supervision and assessment,* he emphasised the importance of complying with the regulations of a specific university to ensure the awarding of the degree.

Question 3 specifically addressed issues around reliability of the assessment tool. The quantitative assessment already demonstrated that the assessment tool was not reliable for four of a possible 12 criteria. The participants identified a problem with the assessors rather than with the assessment tool as the reason for the difference in marks for each assessment category. As an example, marks were allocated for abstracts that were not part of the dissertation. However, this is not unique to this study, for in a study by Pathirage, Haigh, Amaratunga and Baldry,[12] assessors also allocated marks for students' independence and initiative when they marked unidentified (anonymous) dissertations, and were not provided information on the students' independence or own initiative.

Confusion existed around the "Literature review", and whether or not it should be a separate heading in the mini-dissertation. Eventually, consensus was reached that it should be a separate part of the mini-dissertation. According to the literature,[15] a "Literature review" follows the "Introduction", and forms the basis of a good research project as it forms the theoretical basis of the study.

Lack of formal training and inexperience were identified as major factors with regard to differences in the assessment results. Participants expressed their desire for further training. A study by Webster, Pepper and Jenkins[16] on the assessment of undergraduate dissertations also identified lack of training and a variable understanding of terms as factors that could seriously influence fair assessments. Pathirage, Haigh, Amaratunga and Baldry,[12] in their study on *Enhancing the quality and consistency of undergraduate dissertation assessment,* identified the need for assessor training. After attending training workshops, the deviation between assessments decreased from 3.89 to 0.51. The importance of training on assessment tools and staff development was also emphasised by Bettanny-Saltokov, Kilinc and Stow,[13] Hand and Clewes[17] and Matthews.[14]

Question 4 focused on the improvement of the current assessment tool. Although major changes were not suggested, the addition of a comments column next to each assessment category and the use of decimal marks were encouraged. Although no study has proved that the allocation of decimal marks improves assessment, many arguments by academics support this practice, e.g. if the assessment is out of five marks, and the student scores border line, only 40% or 60% can be awarded. Minor change was also suggested to the "Data management" category.

A study by Webster, Pepper and Jenkins[16] demonstrated the difference in overall opinion (global score) and marks allocated to a research project. Bettany-Saltikov, Kilinc and Stow[13] concluded that it is best to use analytic and holistic marking combined, when assessing dissertations. Thus, the use of assessment category marks, as well as a global assessment, as proposed during the focus group interview, is supported by the literature.

Consensus was reached by assessors that marks should be within 10% of the average to be regarded as acceptable. Therefore, the difference in marks allocated by two individual assessors of more than 13% above or below the average is out of line with this agreement. The difference between the assessors, rather than a difference per specific assessment category, indicated that the problem resided with the individuals, rather than the adapted assessment tool.

## Limitations of the study

Only four extensive mini-dissertations were available for assessment during the study, as only four students completed their studies. Therefore, the study sample of assessments was small.

Two of the extensive mini-dissertations were in Afrikaans, which limited possible external assessors to those who were proficient in Afrikaans, and also excluded two of consultants in the Department of Family Medicine who were not able to assess those two dissertations.

## Conclusion

It can be concluded that the use of the current assessment tool for extensive mini-dissertations in postgraduate Family Medicine at the University of the Free State is not reliable for all assessment criteria, as it could not prove reproducibility when measured with CVs.

The focus group interview proved to be valuable in gaining information on how to improve the assessment tool, as well as identifying assessor problems, and motivating assessors to improve their own skills. The current assessment tool was shown to be valid in assessimg extensive mini-dissertations for postgraduate study in Family Medicine.

The adapted assessment tool identified assessor factors, rather than the assessment tool as the reason for differences in marks allocated when the tool was used.

## Recommendations

The first recommendation from this study is for the adapted assessment tool to continue to be used, and re-evaluated for reliability and validity.

The second recommendation is that the new assessment tool should be included in the module guide for extensive mini-dissertations in MMed(Fam).

Other recommendations are:

- To add an information leaflet to the assessment tool in order to provide assessors with clear instructions regarding use of the new assessment tool.
- To train consultants in the Department of Family Medicine on the new assessment tool, refresh knowledge on research methods, and encourage them to assess mini-dissertations and conduct the research themselves.

## References


1. Lewin K. Action research and minority problems. Journal of Social Issue. 1946;2:34-46.
2. Creswell JW. Educational research: planning, conducting and evaluating quantitative and qualitative research. Upper Saddle River: Pearson Prentice Hall, 2005; p. 281.
3. Cohen L, Manion K, Morrison K. Research methods in education. 6th ed. New York: Routhledge, Taylor and Francis Group; 2007.
4. SAS Institute. SAS/STAT® 9.1 User's guide. Cary: SAS Institute; 2004.
5. Joubert G, Ehrlich R. Epidemiology: a research manual for South Africa. 2nd ed. Oxford: Oxford University Press; 2007.
6. Boddy C. A rose by any other name may small as sweet but "group discussion" is not another name for a "focus group" nor should it be. Qualitative Market Research. 2005;8(3):248-255.
7. University of the Free State. Assessment policy. UFS [homepage on the Internet]. 2006. c2011. Available from: http://www.ufs.ac.za/documents/ufs facts/policies/assessmentpolicy
8. University of the Free State. General institutional rules and regulations for advanced and postgraduate qualifications. UFS [homepage on the Internet]. 2011. Available from: http:www.ufs.ac.za/dl/userfiles/Documents/00000/48_eng.pdf
9. University of the Free State. 2011. Faculty of Health Science: postgraduate degrees and diplomas yearbook. UFS [homepage on the Internet]. 2011. Available from: http://apps.ufs.ac.za/dl/yearbooks/151_yearbook_eng.pdf
10. Colleges of Medicine of South Africa. Fellow of the College of Family Physicians regulations. CMSA [homepage on the Internet]. c2011. Available from: http://www.collegemedsa.ac.za/view _ exam.aspx? examid=102
11. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005;15(9):1277-1288.
12. Pathirage C, Haigh R, Amaratunga D, Baldry D. Enhancing the quality and consistency of undergraduate dissertation assessment. Quality Assurance in Education. 2007;15(3):271-286.
13. Bettany-Saltikov J, Kilinc S, Stow K. Bones, boys, bombs and booze: an exploratory study of the reliability of marking dissertations across disciplines. Assessment and Evaluation in Higher Education. 2009;34(6):621-639.
14. Matthews L. Dissertations: issues in guidance, supervision and assessment. Newcastle: British Library, Northumbria University, 2007; p. 1-29.
15. Levy Y, Ellis TJ. A systems approach to conduct an effective literature review in support of information systems research. Information Science Journal. 2006;9:181-211.
16. Webster F, Pepper D, Jenkins A. 2000. Assessing the undergraduate dissertation. 2000. Assessment and Evaluation in Higher Education. 2000;25(1):71-80.
17. Hand L, Clewes D. Marking the difference: an investigation of the criteria used for assessing undergraduate dissertations in a business school. Assessment and Evaluation in Higher Education. 2000;25(1):5-21.


**Appendix 1: Assessment Tool and Quality Indicator sheet**

**Name of student**……..

**Title**……

| Component | Marks | Comments |
|---|---|---|
| **1. Topic and title** Accurate, concise, relevant to the field of Family Medicine | /5 | |
| **2. Abstract** All aspects covered, clear and accurate | /10 | |
| **3. Introduction/Background** Problem stated well and justification for the study is clear | /5 | |
| **4. Literature review** Relevant, current, well interpreted and connected to study | /10 | |
| **5. Problem statement, Aim and objectives** Clearly formulated, focused, feasible and logical | /10 | |
| **6. Method** Study design, sampling, measurement, errors and pilot study well described, logic and justified | /20 | |
| **7. Ethical considerations** Informed consent, approval and confidentiality addressed | /5 | |
| **8. Data analysis, management and presentation** Clear and accurate presentation, in context with research question/s? | /15 | |
| **9. Discussion** Accurate interpretation and supported by literature | /10 | |
| **10. Conclusion and recommendations** Accurate and justified from findings, logic, limitations described | /10 | |
| **11. Style, language and layout** Grammatically correct, consistent style and numbering, neat presentation | /10 | |
| **12. References** Good sources, Correctly referenced, up to date | /10 | |
| **Total** | /120 | |

**Global rating**

Fail ☐
Needs improvement ☐
Pass ☐
Distinction ☐

%

**General remarks** ……
……
……
……

**Quality Indicator sheet**

| Quality indicators | Description of indicators |
|---|---|
| **Unacceptable 0-39%** 0-1.5/5 0-3.5/10 0-5.5/15 0-7/20 | No or very poor understanding. No application of knowledge. Missing information. No effort to execute task appropriately. Unprofessional language or skills displayed. Fail |
| **Weak 40-49%** 2/5 4-4.5/10 6-7/15 7.5-9.5/20 | An attempt to understand concepts, poor application of knowledge. Some contradictory information and applications. Distractions are evident. Lack of professional language and skills displayed. Needs moderate improvements. |
| **Acceptable 50-60%** 2.5-3/5 5-6/10 7.5-9/15 10-12/20 | Some evidence of understanding of concepts. Some application. Task executed most of the time. No evidence of creativity. No evidence of extra effort. Acceptable language and professional skills. Pass |
| **Proficient 61-74%** 3.5/5 6.5-7/10 9.5-11/15 12.5-15/20 | Evidence of understanding of concepts and application of knowledge. Appropriate and correct task execution throughout. Good evidence of creativity and professional skills. Above average – Good work |
| **Exceptional 75-100%** 4-5/5 7.5-10/10 11.5-15/15 15.5-20/20 | Consistent evidence of understanding of concepts and application of knowledge. Excellent task execution throughout. Clear evidence of creativity and professional skills. Way above expectations. Distinction. |