	checking for outliers compared to the average and standard deviations in the whole dataset. It should be noted that a value, which is identified as an outlier, does not necessarily indicate incorrect data.	
7.2.3	Data editing procedures should include cross-checks between related values, and ideally, should also include checks on summary totals for such related fields. *For example, this would include checking subtotals in Snap tables.*	6.2
7.2.4	For repeated surveys, editing procedures should compare values submitted previously for a specific school. Ranges are set for acceptable changes from one year to the next, and changes in values for individual schools are checked and verified against these ranges.	6.2
7.2.5	Editing procedures should include checks on the validity of values, such as checks against list tables for codes, values falling outside acceptable ranges, as well as required fields or sections that have not been completed.	6.2
7.2.6	Data cleaning procedures must be revised to take into account changes in the data collection instruments from one year to the next, which makes it impossible to successfully compare data from one year to the next. Such revisions must be documented for metadata purposes.	6.3
7.2.7	Methodologies used for editing should follow accepted standards, guidelines or good practices (national, international and peer-agreed).	7.3
7.2.8	Care should be taken if there are revisions to data *(for instance by a province that submitted an initial dataset and then submitted a revised dataset)*. If any data is submitted after publication of the results and such data is included in the national dataset, this additional data need to be tagged as such in the dataset.	7.5

2.3.2.8 STEP 8: Aggregation and analysis

Objectives to be reached during this step:

When the collected data is analysed, it is important to use sound methodologies for the analysis. The focus of the analysis should be on providing answers to key questions and not just to show all the results. Of further importance is to document all the procedures and methods used. Documentation of the procedures provides evidence of the acceptability and quality of the obtained results, and also ensures the repeatability of the results.

It should be noted that an analysis could involve providing descriptive statistics, such as aggregated totals and counts, but it could also comprise an in-depth (technical) analysis of the collected data. The guidelines cover general aspects that relate to both these kinds of analysis, as well as aspects that relate primarily to technical analysis.

QG8: Guidelines pertaining to actions carried out during this step:

	Action points: Aggregation and analysis	Related SASQAF indicator
8.1	**General**	
8.1.1	The methods planned for the analysis must be motivated by the objectives of the study and by statistical considerations, rather than by the data itself or by political interference. Therefore, the choice of the level of analysis, as well as the relevant analysis methods, should preferably be finalised before data collection takes place.	8.5
8.1.2	The analysis should, as far as possible, be aligned with the analysis plan that was drawn up during the planning and design steps, in order to ensure that the analysis provides answers that are in line with the objectives of the study.	
8.1.3	Electronic systems designed for data analysis must be designed so as to be compatible with the systems designed for data collection and capture. *If systems are developed by a third* party *(i.e. outside the Department of Basic Education), then this software product must link up with the Department of Basic Education capture and analysis tools and produce analysis outputs that are acceptable to the department.*	
8.1.4	The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated, and which was planned for the survey, should be adhered to during the analysis.	8.1

8.2	**Imputation of missing values**	
8.2.1	Imputation is the responsibility of national department. Provincial departments are not responsible for imputations.	
8.2.2	If imputation routines are used to fill in missing data, it is important to limit imputation to variables for which values are missing randomly, and not due to a systematic difference or bias. It is furthermore important to study and find an imputation method that is appropriate for a specific survey. Methods used must be reported in the metadata.	7.3

8.3	**Technical analysis**	
8.3.1	All assumptions underlying the analysis should be stated clearly, and appropriate methods for the analysis should be selected accordingly.	7.3
8.3.2	Assumptions, analysis methods and any limitations to the analysis should be documented for metadata purposes.	4.13
8.3.3	Measures of sampling errors for key variables must be calculated from the actual data that was collected, and these measures must be compared to the targets set for these values during the survey planning. As far as possible, results should be reported in accordance with the relevant standard errors or confidence intervals. This information must also be provided in the metadata that reports on data quality.	2.1, 4.13
8.3.4	Results should be reported in such a way that it supports appropriate conclusions from the analysis. Assumptions, limitations, data quality issues, response rates, potential biases and other important aspects that affect the analysis should be taken into account.	
8.3.5	If possible, measures of non-sampling errors must be calculated – such as errors picked up in the sample frame or during editing. It is also important to consider the effect of errors, such as deficiencies in sample frame or instruments, or other underlying biases to the analysis on the quality of the data and analysis. These errors and their impact should be documented for metadata purposes.	2.2
8.3.6	With regard to census surveys, it is important to report on non-response.	2.2

	The extent of non-response, possible reasons for the non-response and measures (if any) taken to address the non-response should be documented and reported.	
8.3.7	Documentation must be provided to describe record-matching methods and techniques used (for instance matching of collected data to the MLI), and other analysis and editing methods used. It is important to record what was done, what errors were found and how these errors were rectified.	2.7

2.3.2.9 STEP 9: Publication of results

Objectives to be reached during this step:

Publication is done in order to make results available to users in an accessible way. In order to ensure proper use of the results, metadata, explaining the contents of the data and the methods used to obtain and analyse it, should be provided with the results.

QG9: Guidelines pertaining to actions carried out during this step:

	Action points: Publication of results	Related SASQAF indicator
9.1	**Before disseminating the data**	
9.1.1	Determine whether there is a data dissemination policy relevant to the survey. *For instance, this would be a publication policy that governs the contents, the form in which it is distributed, to whom it is distributed, etc*[4].	4.10
9.1.2	Determine whether there are clearance procedures ensuring that the document has been approved by the unit producing it and whether there are guidelines on the issues to be addressed before publication could be approved. Ensure that these prescriptions are followed before publication of the results.	4.10
9.1.3	The Director-General: Basic Education will need to approve the contents before releasing it to the general public. Before information is referred to the DG: BE for approval, it must furthermore be approved by the relevant staff members internal to the department. These include the staff members responsible for conducting the survey, and also the EMIS Officer. Determine the requirements and processes in obtaining approval for release and ensure that these are correctly followed.	4.10
9.1.4	For qualitative data *(e.g. whole school evaluation)*, the publication should be in a format that would allow the data to be integrated and analysed.	4.10
9.1.5	Adding or removing results due to political pressure should be avoided.	8.5
9.1.6	Ministerial commentary when data is released should be identified as such, and should not be seen as part of the official statistics.	8.4

9.2	**Reporting on the quality of the results**	
9.2.1	Statistical measures of data quality should be reported with the results. It	2.2

[4] See document SC012, Standards for Publication of Education Statistics in this regard.

	should also be indicated how the obtained accuracy differed from the planned levels, including the response rate that was actually achieved. It should be noted that it could be important to indicate if a particularly bad response was obtained with regard to certain aspects, even if the overall response rate was satisfactory. *This helps the user to understand what the results can and cannot be used for.*	
9.2.2	The metadata collected throughout the survey should be collated. Issues recorded in the metadata that affect data quality or influence the use of the results should be published in conjunction with the results. *This could include objectives, methodology, data sources, accuracy, instruments, sampling plan, editing and imputation.*	4.13
9.2.3	There will be metadata that should be published with the data, metadata that may be released and metadata that will only be released if legally required to do so. Metadata for internal use carries a lot of detail, while metadata for publication carries less detail. Determine which portions of the recorded metadata should be released with the results. *A person who reads the report must understand how the study was conducted and what exceptions or unanticipated aspects were found in the data or the processes.*	4.13
9.2.4	Concepts, definitions and classifications underlying the data should be published in order to clarify the data. *This should form part of the reference section in all reports.*	5.1
9.2.5	Information on the scope, sampling, data sources and statistical techniques used should be published. Differences relating to accepted standards, guidelines or good practices should be pointed out and motivated.	5.1, 5.2
9.2.6	Some information must also be published in terms of the processing of the data. For internal use, metadata should include the actual information one would need to process the data, such as the statistical code used to run the analysis, and this code must be available if requested.	4.13, 5.2
9.2.7	All the statistical information published (tables, etc.) must be accompanied by an explanation of the contents, and clarification of the key findings. *There must be sufficient information to guide the user in the interpretation of the information.*	5.3
9.2.8	If there are revisions of data, or if data collection continues after publication of the results, information pointing this out must be included with the published results. *This must be done to warn users about differences between the published results and the survey data disseminated.*	7.4, 7.5, 7.6

9.3	**General issues**	
9.3.1	All activities should be monitored and controlled in order to ensure that publication deadlines are met.	3.2, 3.3, 3.5
9.3.2	When the results are published, it should be indicated whether this is a repeated survey and how often the survey has been repeated.	3.4
9.3.3	Ensure that the results will reach key users by employing the appropriate types of media/channels. Make sure that the correct level of confidentiality is maintained in the published results. *Do not release raw data or data that allows individuals to be identified.*	4.1, 4.4
9.3.4	Make sure that the release of data before publication adheres to the planned timelines and conditions.	8.2
9.3.5	Put systems in place to obtain feedback on the published results from key users. Furthermore, ensure that published results are provided to those who provided data for the survey.	1.5, 1.7

2.3.2.10 STEP 10: Handling queries

Objectives to be reached during this step:

The survey is only completed once all data has been finalised and is ready for release. Furthermore, once a study has been completed and results are published, there may be requests from users for the data attached to the survey. In clause 24 of the Education Information Policy, it is stated that the needs of researchers to analyse data pertaining to the education sector should be anticipated and that data should be prepared accordingly.

It is important to disseminate data from the study according to the same principles as the release of the publication itself, and therefore many of the requirements for release of the data are similar to those for publication purposes. However, in releasing data is it also important to ensure that the correct context is maintained. The context provided in terms of the entire published report should also be clearly applied to sections of the report released in the form of data or as individual summary tables.

QG10: Guidelines pertaining to actions carried out during this step:

	Action points: Handling queries	Related SASQAF indicator
10.1	**General**	
10.1.1	All the requirements in terms of clearance before release, accompanying metadata, confidentiality and comments on data quality listed in terms of publication of the results, should also be followed for the release of data. *For queries and for reports published on the Internet, make sure that the information retains the correct context by providing appropriate accompanying metadata, since extracts from the report may be misinterpreted if read out of context.*	2.2, 4.10, 4.13, 5.1, 5.2, 5.3, 8.5
10.1.2	The types of media/channels used for providing requested data should be adequate and should preserve confidentiality. *Also plan for a way to provide the relevant metadata in conjunction with the data.*	4.4
10.1.3	Catalogue systems to identify available information from the survey should be made available to users, and this catalogue should be regularly updated.	4.12
10.1.4	Data must be made accessible to users by ensuring that the data could be provided in a variety of formats that will satisfy the requirements of all users.	4.5

	Action points: Handling queries	Related SASQAF indicator
10.1.5	Make sure that data from more than one instance of a repeat survey is provided in a comparable format.	6.2, 6.3
10.1.6	Preliminary and revised data should also be identified in the data.	7.5
10.1.7	Data that is released from a specific survey should contain references to that survey. The date on which the survey was completed/finalised, as well as the date on which the data was supplied, must be indicated on the dataset.	
10.1.8	Publication policies should address more than just the published hard copy and should include policies pertaining to electronic copies, Internet versions of reports, queries and all output provided to users.	4.10
10.1.9	Make sure that part of the process for providing data includes a check on the methods or codes used to extract data in response to queries. *Someone must double-check queries used to extract data.*	
10.1.10	Put in place systems to obtain feedback on the published results from key users.	1.5, 1.7
10.1.11	Put in place systems to log all the queries in order to keep all queries responded to, on file.	

APPENDIX A: SASQAF Data Quality Indicators

Tables 2 to 10 that follow in this subsection provide the details on the SASQAF standard to be used to assess data quality, organises into the eight quality dimensions and the prerequisites for quality. Note that the numbering convention for the SASQAF indicators is to use the dimension as the first level of the number, so that the indicator numbered 3.2 refers to the second indicator under dimension 3. The table headings list the number of the dimension along with the name of the dimension.

Each table contains an indicator number, a description of the indicator, as well as the four possible assessment levels. When a certain statistic is measured in terms of an indicator, it is placed in one of the four levels, with level 4 indicating best and level 1 lowest quality.

Table 2 Prerequisites of quality (SASQAF dimension 0)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
0.1	The responsibility for producing statistics is clearly specified.	The responsibility for producing statistics is explicitly specified through a legal framework.	The responsibility for producing statistics is specified through a legal framework.	The responsibility for producing statistics is implied through a legal framework.	The responsibility for producing statistics is not specified.
0.2	Standards and policies are in place to promote consistency of methods and results.	All standards and policies are in place to promote consistency of methods and results, and are adhered to.	The majority of standards are in place to promote consistency of methods and results.	Some standards are in place to promote consistency of methods and results.	No standards are in place to promote consistency of methods and results.
0.3	Data sharing procedures and coordination among data-producing agencies are clearly specified and adhered to.	Data sharing procedures and coordination among data-producing agencies are explicitly specified through a legal framework.	Data sharing procedures and coordination among data-producing agencies are specified through a legal framework.	Data sharing procedures and coordination among data-producing agencies are implied through a legal framework.	Data sharing procedures and coordination among data-producing agencies are not specified.
		A data-sharing policy exists and is regularly updated and adhered to.	A data-sharing policy exists and for the most part is adhered to. It may not be up to date.	A data-sharing policy exists, but is rarely adhered to. It may not be up to date.	No data-sharing policy exists.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
0.4	Measures are in place to ensure that individual data are kept confidential, and used for statistical purposes only.	Measures (e.g. policies, documented procedures) exist and are fully enforced so that individual data are always kept confidential.	Measures exist and are partially enforced so that individual data are always kept confidential.	Measures exist, but are not enforced to always keep individual data confidential.	There are no measures that ensure confidentiality.
0.5	Resources are commensurate with the needs of statistical programmes (staff, facilities, computing resources, financing	All resources are completely commensurate with statistical programmes.	Resources are partially commensurate with statistical programmes.	Resources are inadequately commensurate with statistical programmes.	Resources are not commensurate with statistical programmes.
0.6	Measures to ensure efficient use of the above resources in 0.5 are implemented.	Measures (e.g. project plans and sign-off documentation) to ensure efficient use of resources are systematically implemented.	Measures to ensure efficient use of resources are often implemented.	Measures to ensure efficient use of resources are seldom implemented.	Measures to ensure efficient use of resources are not implemented.
0.7	Processes are in place to focus on, monitor and check quality.	Processes are consistently in place to focus on, monitor and check quality.	Processes are to some extent in place to focus on, monitor and check quality.	Processes are seldom in place to focus on, monitor and check quality.	Processes are not in place to focus on, monitor and check quality.

Table 3 Relevance (SASQAF dimension 1)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
1.1	Have both the internal and external users of the data been identified?	All users of the data have been identified with their most recent contact details.	All users of the data have been identified with some of the contact details not up to date.	Attempts have been made to create a user list.	No attempt has been made to create a user list.
1.2	Is there a process to identify user needs?	User needs are identified as a matter of course.	User needs are usually, but not always, identified.	User needs are identified on an ad hoc basis.	No attempt is made to identify user needs.
1.3	Are user needs and the usage of statistical information analysed?	User needs and the usage of statistical information are always analysed.	User needs and the usage of statistical information are often analysed.	User needs and the usage of statistical information are seldom analysed.	There is no effective interaction with users.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
1.4	Changes made as a result of user needs assessments.	The results of the assessment are always built into the corporate processes and influence decisions on the design of the survey/series.	The results of the assessment are often built into the corporate processes and influence decisions on the design of the survey/series.	The results of the assessment are seldom built into the corporate processes and influence decisions on the design of the survey/series.	No action taken to incorporate the results of assessments.
1.5	Is there a process to determine the satisfaction of users?	User satisfaction is measured and to a large extent has made an impact on the output.	User satisfaction is measured and to some extent has made an impact on the output.	User satisfaction is measured but has made no impact on the output.	User satisfaction is not measured.
1.6	To what extent are the primary data (e.g. administrative data and other data) appropriate for the statistical product produced?	The primary data are fully aligned to the statistical product released.	The primary data are mostly aligned to the statistical product released.	The primary data have limited relevance to the statistical product released.	The primary data are not at all relevant to the statistical product released.
1.7	Were special requests for estimates of statistical characteristics met?	All special requests were met.	Some special requests were met.	The majority of special requests were not considered.	No special requests were met.

Table 4 Accuracy (SASQAF dimension 2)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
2.1	Measures of sampling errors for key variables are calculated. Amongst others these are: Standard error, Coefficient of variation (CV), Confidence interval (CI), Mean square Error (MSE)	Sampling errors are calculated for the main variables and are available for the other variables on request, and fall within acceptable standards.	Sampling errors are calculated and made available for the main variables, and fall within acceptable standards.	Sampling errors are calculated but not made available, and fall outside the acceptable standards.	No sampling errors are calculated.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
2.2	Measures of non-sampling errors are calculated, viz.: Frame coverage errors (under-coverage errors, over-coverage errors), duplication in the frame / register used to conduct survey, number of statistical units out of scope (i.e. number of ineligible units), misclassification errors, systematic errors to determine the extent of bias introduced for both administrative records and surveys, measurement errors (questionnaire effects, data collection mode effects, interviewer effects, respondent effects), processing errors (data entry error rate, coding errors, editing failure rates, imputation rates), model assumption errors, non-response errors (overall response rate, item response rate, unit non-response, weighted and unweighted response rates)	Non-sampling errors are extensively described and analysed, and the measures fall within acceptable standards.	Non-sampling errors are described and analysed, and the measures are not far off from acceptable standards.	Non-sampling errors are described and analysed, and the measures are far off from acceptable standards.	Non-sampling errors are not described.