



# planning, monitoring & evaluation

---

Department:  
Planning, Monitoring and Evaluation  
**REPUBLIC OF SOUTH AFRICA**

## Report on the Assessment of Government Evaluations

Evaluation Title:

Northern Cape Provincial Transport Operations Grants  
(PTOG) Evaluation

Evaluation Number:

607

Start Date:

Sunday, 01 January 2017

Completion Date:

Friday, 31 March 2017

Period of Evaluation:

3 months

Submitted:

Tuesday, 30 January 2018 by Cara Hartley

Approved:

Wednesday, 31 January 2018 by Mike Leslie

## Evaluation Details

Evaluation Title:	Northern Cape Provincial Transport Operations Grants (PTOG) Evaluation
Evaluation Number:	607
Evaluation Completion Date:	Friday, 31 March 2017
Created:	Wednesday, 13 December 2017 by Cara Hartley
Submitted:	Tuesday, 30 January 2018 by Cara Hartley
Approved:	Wednesday, 31 January 2018 by Mike Leslie
Period of Evaluation:	3 months
Known Cost:	R 10400.00
Known Cost Type:	Estimate
Initiated By:	Northern Cape Department of Transport, Safety and Liaison
Initiated By Internal:	Yes
Service Providers:	N/A
Undertaken By Internal:	Yes

## Assessors

Cara Hartley	cara@pdg.co.za
--------------	----------------

## Assessment Documents

Document Name:	Document Type:	Added By:	Added On:
DTSL.PTOG Evaluation Report.pdf	Evaluation report	Cara Hartley	Wednesday, 13 December 2017
Northern Cape Department of Transport Evaluation Plan.docx	Any other relevant documentation pertaining to the evaluation process	Cara Hartley	Friday, 26 January 2018
Update on Evaluation.pptx	Progress reports	Cara Hartley	Friday, 26 January 2018
TERMS OF REFERENCE FOR THE NORTHERN CAPE DEPARTMENT OF TRANSPORT SAFETY AND LIAISON EVALUATION PLAN.docx	Any other relevant documentation pertaining to the evaluation process	Cara Hartley	Friday, 26 January 2018
PTOGs DATA COLLECTING QUESTIONS.docx	Assessment tools	Cara Hartley	Friday, 26 January 2018
Business Plan for PTOGs Evaluation Data Gathering (1).docx	Any other relevant documentation pertaining to the evaluation process	Cara Hartley	Friday, 26 January 2018
Assessment Report at Moderation.pdf	Assessment Report at Moderation	Mike Leslie	Monday, 29 January 2018

## Quality Assessment Summary

---

This evaluation was the first ever conducted by the Northern Cape Department of Transport, Safety and Liaison (DTSL) and it was done internally with no involvement of a supporting institution such as the Office of the Premier or a peer reviewer. It was done without the expectation that it would be considered within the National Evaluation System, and for this reason there are aspects of the "standard" evaluation process that were not considered or followed within the context.

The evaluation scores 2.59, meaning that its quality is below adequate in terms of the DPME standards for evaluation in government. Nevertheless, it was an important exercise for the department both in terms of yielding important insights about what is happening "on the ground" with the Public Transport Operations Grant, and as a demonstration to the department's senior leadership, of the potential value that evaluations can have for the department.

The planning phase was inadequate in terms of the standards, scoring 1.68. No Terms of Reference or internal evaluation proposal was developed and therefore there was limited consideration of evaluation questions, the theory of change, and analytical framework that would be applied to the programme.

The implementation phase scored 3.05, meaning it was just above adequate (3). An evaluation working group oversaw implementation, and it consisted of all the senior managers of the department. Their support and effective cooperation raised the score for this phase; so did the proper consideration given to ethics. However, there was no literature review, and limited piloting of the data collection instruments.

The reporting phase score is 2.25, which is relatively low. The report does not comprehensively cover the sections prescribed by the DPME's standards, and there are analytical errors in the report; moreover the analysis does not synthesise the findings but presents them in a relatively raw form. In the absence of evaluation questions and a theory of change, the findings and conclusions are structured not by theme or criterion but by original primary data source.

The follow-up, use and learning phase, however, is well above adequate, scoring 3.53. This is because of the clear value that the evaluation exercise had for the department - both in terms of learning about what is happening in the programme and shaping future practice; and in terms of demonstrating to management the general symbolic and conceptual value of evaluating a programme. Furthermore, the evaluation was completed within its minimal budget and three-month time frame.

Despite the shortcomings of the evaluation, the interviewed stakeholders see it as an important learning experience. This is reflected in the good rating (4) received for Participation and M&E Skills development. The way it has been received in the department bodes well for management support and engagement with evaluations in the future.

## Quality Assessment Scores

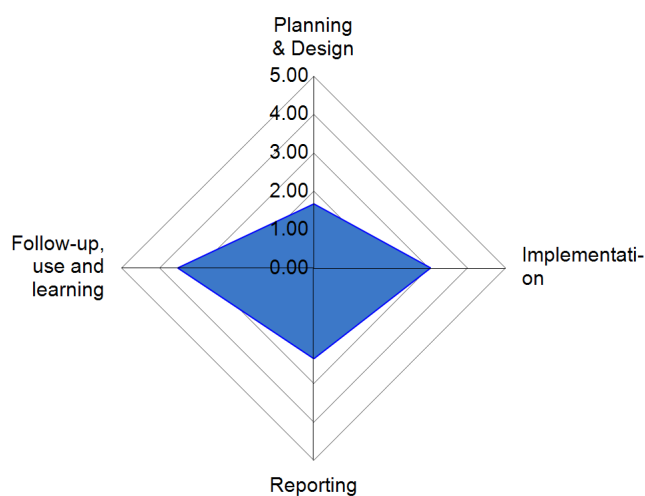
---

Phase of Evaluation	Score
Planning & Design	1.68
Implementation	3.05
Reporting	2.35
Follow-up, use and learning	3.53
Total	2.59

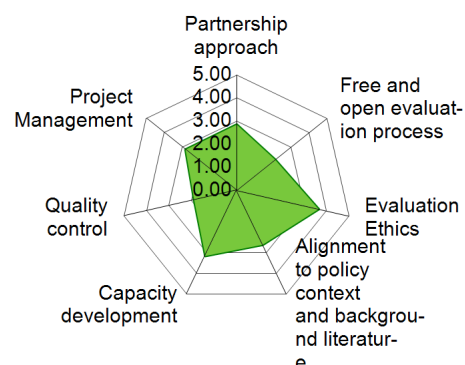
Overarching Consideration	Score
Partnership approach	2.88
Free and open evaluation process	2.17
Evaluation Ethics	3.70
Alignment to policy context and background literature	2.67
Capacity development	3.20
Quality control	1.92
Project Management	2.86

Total	2.59
-------	------

Scores: Phases of Evaluation



Scores: Overarching Considerations



Phase of Evaluation	Area of Evaluation	Score
Planning & Design	Quality of the TOR	1.00
Planning & Design	Adequacy of resourcing	2.00
Planning & Design	Appropriateness of the evaluation design and methodology	2.09
Planning & Design	Project management (Planning phase)	3.00
Implementation	Evaluation ethics and independence	3.50
Implementation	Participation and M&E skills development	4.00
Implementation	Methodological integrity	2.37
Implementation	Project management (Implementation phase)	3.50
Reporting	Completeness of the evaluation report	2.50
Reporting	Accessibility of content	2.50
Reporting	Robustness of findings	2.00
Reporting	Strength of conclusions	1.67
Reporting	Suitability of recommendations	2.50
Reporting	Acknowledgement of ethical considerations	4.00
Follow-up, use and learning	Resource utilisation	3.00
Follow-up, use and learning	Evaluation use	3.67
Total	Total	2.59

## Planning & Design

---

### Quality of the TOR

Standard:	The evaluation was guided by a well-structured and complete TOR or a well-structured and complete internal evaluation proposal (e.g. Background, Purpose, Evaluation Questions, Design & Methodology, Deliverables & Timeframes, Resource requirements, Intended Audience & Utilisation, etc).
Comment and Analysis:	No standalone terms of reference (TOR) or internal evaluation proposal was developed. This was the first evaluation undertaken as part of the DTSL's newly approved Departmental Evaluation Plan (DEP) and it was only briefly conceptualised as part of the larger plan. At the planning stage, although the DEP had articulated the broad purpose of DTSL evaluations, no documentation of the following aspects were developed for the PTOG evaluation (as opposed to future departmental evaluations in general): Background of the programme, purpose of the evaluation, evaluation questions, deliverables and timeframes, or resource requirements. The intended audience for all evaluations was implied not in a TOR specific to this evaluation, but in a terms of reference that was developed to guide the departmental evaluation working group (consisting of senior managers).
Rating:	1: There is no TOR or internal evaluation proposal available, or the TOR was never finalised
Standard:	The approach and type of evaluation was suited to the purpose and scope of the evaluation TOR (or an internal evaluation proposal)
Comment and Analysis:	There was no TOR or internal evaluation proposal stating the purpose and scope of the evaluation. The DEP, which noted the intention to conduct this PTOG evaluation, also did not specify the approach, type, purpose or scope of this evaluation.
Rating:	1: There is no TOR or internal evaluation proposal available, or the approach and type of evaluation was not stated in the TOR
Standard:	The TOR (or an internal evaluation proposal) identified the intended users of the evaluation and their information needs
Comment and Analysis:	There is no TOR or internal evaluation proposal. It is noted however that the departmental evaluation plan indicates that evaluation reports (in general) will be shared with all relevant stakeholders and the office of the Head of Department. It states that it this is anticipated to lead to resolution of general audit queries, and changes to some programmes. No documentation explains the users and information needs for this evaluation in particular.
Rating:	1: There is no TOR or internal evaluation proposal available, or the TOR made no mention of the users of the evaluation or their information needs
Standard:	Key stakeholders were involved in the scoping of the TOR and choosing the purpose of the evaluation
Comment and Analysis:	There is no TOR, therefore a score of 1 is given. However, to provide background as to the participation in the planning: the evaluation was conceptualised by the Chief Director: Policy and Planning and a senior researcher in the Unit: Evaluation and Research. It was then proposed to and discussed with the senior management committee (ManCo) of the department, including the Senior Manager: Transport Operations, who is responsible for the PTOG. Thus, key decisionmakers within the department were consulted. Neither the Office of the Premier nor any external stakeholders of the PTOG (see below) were involved in planning the evaluation. The interviewees explained that they saw the evaluation as an internal process.
Rating:	1: There is no TOR or internal evaluation proposal available, or there is no evidence of any consultation amongst evaluation stakeholders in the development of the TOR

## Adequacy of resourcing

Standard:	The evaluation was adequately resourced in terms of time and budget allocated
Comment and Analysis:	There was no budget for the evaluation and the entire project was intended to be conducted in three months. The scope of work was accordingly kept relatively small (no literature and document review, no theory of change development process, no paid peer reviewer, a brief primary data collection exercise) but the interviewees nevertheless found the narrow time frame challenging.
Rating:	2: The evaluation was resourced with tight timeframes and budget which were challenging from the outset
Standard:	The team conducting the evaluation was adequately resourced in terms of staffing and skills sets
Comment and Analysis:	<p>The leader of the evaluation has a PhD and some of his honours research was related to transport. He thus had sector expertise. He also had plenty of research experience. He did not have any evaluation experience. The Senior Manager: Policy and Planning who also played an important role in overseeing and giving input on the evaluation also has research experience, and in addition has considerable experience as a manager in the public sector. They had recently studied the National Evaluation System and compiled the DEP, but had no first-hand experience of implementing evaluations.</p> <p>Unfortunately, the limited evaluation expertise does appear to have hampered the quality of the evaluation, as key components were absent (for instance, clarity on the evaluation type; evaluation questions; and conclusions that explicitly judge the programme against a transparent framework or set of criteria). The evaluation served as a significant hands-on training experience for them and has doubtless paved the way for future evaluations.</p>
Rating:	2: The evaluation was under-staffed or lacked some skills sets appropriate for the type and sector of the evaluation

## Appropriateness of the evaluation design and methodology

Standard:	There was explicit reference to the intervention logic or the theory of change of the evaluand in the planning of the evaluation
Comment and Analysis:	The interviewed stakeholders indicated that theory of change thinking was central to their approach, guiding them in picturing the causal pathways that the programme relies on and making sure to develop data collection instruments that would assess those causal pathways. However, there was no documented reference in any of the evaluation documents to the theory of change, nor was there any systematic documentation of the intervention logic. It therefore appears that theory of change thinking was implicit in their approach rather than ever explicitly documented.
Rating:	1: There was no reference to the intervention logic or the theory of change in the TOR or the Inception Report

Standard:	The planned methodology was appropriate to the questions being asked
Comment and Analysis:	<p>The report does not include any evaluation questions, and the interviewees confirmed that no questions were formulated. The general areas of enquiry, as mentioned by the interviewees, were as follows: (1) the efficiency of grant implementation (appropriate application of grant funds); (2) the extent to which the PTOG provides impact/benefits (mostly conceptualised at the immediate outcome level) to the public, the subsidised transport companies, and the department; and (3) the extent of alignment of the programme as currently implemented, to the department's strategic plan.</p> <p>If these are taken as the implicit evaluation questions, then firstly, part of the methodology would need to be the articulation of the PTOG intervention logic / theory of change, so that it is clear what type of efficiency, impact and alignment is sought. This was not done, leaving the evaluation vague about what constitutes success. In terms of specific data, area 1 would be suitably addressed through the analysis of expenditure data and trends and some of the data collected from the transport companies, but data from programme staff (who administer the grant) would also be important and was not provided for; area 2 by interviewing and surveying the public, bus drivers and transport company managers; but area 3 was not addressed by any particular part of the methodology. One would expect a document review of the strategic plan itself, and possibly a review of the larger national grant framework for PTOG.</p> <p>Overall therefore, the planned methodology did not cover all the areas of enquiry sufficiently, although it still had the potential to provide considerable insight on the programme's implementation.</p>
Rating:	2: The planned methodology was not entirely appropriate for addressing all of the questions being asked
Standard:	The sampling planned was appropriate and adequate given the focus and purpose of evaluation
Comment and Analysis:	<p>A Data Gathering Business Plan details the planned sampling. It specifies six geographic areas in which fieldwork will be done on specific bus routes, and indicates that it will cover 6 bus companies (there are 8, but the remaining 2 have much smaller passenger volumes). Thus, the sample would include representatives from the most important bus companies. The plan also states that passengers will be randomly sampled, but will not be preselected. The plan does not elaborate on how many drivers and passengers will be sampled. It also mentioned that traffic enforcement agents will be engaged in the data collection, but does not explain how they will be selected. If one accepts that the samples would not be representative in aggregate, but provide a sense of the variety and spread of stakeholder experience in a variety of geographic areas and across the main companies, then this sample is adequate. Still, the omission of DTSL staff from the data collection sample is an important one considering that DTSL would ideally use the evaluation to improve on areas within its control (which DTSL staff would be well placed to describe).</p>
Rating:	3: The sampling planned was appropriate and adequate given focus and purpose of evaluation

## Project management (Planning phase)

Standard:	The inception phase was used to develop a common agreement on how the evaluation would be implemented
Comment and Analysis:	<p>There was no standalone inception phase, but there appears to have been adequate communication among the Evaluation Working Group members, and the members were reportedly eager and enthusiastic to engage with this first departmental evaluation. According to the interviewees, there was good cooperation between the Policy and Planning unit and the Transport Operations units, (the two units which were the most closely involved in carrying out the evaluation). According to the interviewees, everyone was certainly on the same page once the plans had been discussed in the Working Group. Planning was certainly adequate to secure buy-in and support for the exercise. It is noted, however, that it was only discovered later that no road traffic enforcement agents could comment on the PTOG because of staff shortages, and this was not picked up in the inception phase despite this reported high level of engagement between the units.</p>
Rating:	3: The inception phase was used to develop a common agreement on how the evaluation would be implemented

## Implementation

---

### Evaluation ethics and independence

Standard:	Where data was gathered in contexts where ethical sensitivity is high, informed consent, assurances of confidentiality and appropriate clearance were achieved; e.g. through an ethics review board, in evaluation involving minors, institutions where access usually requires ethical or bureaucratic clearance
Comment and Analysis:	No aspect of the evaluation was highly ethically sensitive. Where individuals were interviewed and surveyed, their names were not recorded and they were assured of confidentiality (as evidenced by the introductory paragraph at the start of the data collection questions). Furthermore, the questions asked of them were not very personal. The lead evaluator noted the possible risk of putting the bus companies in a bad light in the report and indicated that the questionnaire was therefore developed to study the PTOG but not expose the companies unnecessarily. The report does not document how data notes are treated.
Rating:	3: There was clear evidence that ethical protocols were observed for some data collection instances including: informed consent agreements; confidentiality; documenting and storing data notes, recordings or transcripts; and ethics review board approvals where appropriate
Standard:	Where external, the evaluation team was able to work without significant interference and given access to existing data and information sources
Comment and Analysis:	The evaluation team was internal, but this does not automatically eliminate the risk of interference or withholding of information. Fortunately in this case, the evaluation team had good support from the DTSL senior management. They reported no challenges in terms of interference; and noted that an effort was made to provide them with all data they requested.
Rating:	4: The evaluation team was able to work freely without interference and was given access to all sought data and information sources

### Participation and M&E skills development

Standard:	Key stakeholders were involved in the evaluation through a formalised mechanism or institutional arrangement
Comment and Analysis:	As already alluded to, the evaluation Working Group is the senior management committee (ManCo) of the department. This Working Group was involved from approval of the proposal/request through to the end. Interviewees report that senior managers were generally quite interested in the evaluation and engaged with the deliverables. A score of 4 is awarded to reflect their regular, active involvement, but it should be noted that it appears that the Working Group did not go so far as to provide in-depth written comments on the deliverables (data collection plan and draft report). It should also be noted that since the ManCo is the Working Group, it included only managers of the department and not any external role players.
Rating:	4: Key stakeholders were regularly, actively involved in the evaluation and contributed through a formalised mechanism or institutional arrangement (e.g. a steering committee or reference group)



Standard:	Where appropriate, an element of capacity building of partners responsible for the evaluand and evaluators was incorporated into the evaluation process
Comment and Analysis:	As part of this first departmental evaluation, those conducting the evaluation introduced everyone on the Working Group to relevant M&E frameworks and to the National Evaluation System as part of their presentations and progress updates. Furthermore, the data collection teams were drawn from EPWP workers; they underwent a day of training on how to conduct the data collection, including general principles of field research such as body language and how to treat respondents. These capacity building elements were highly valuable to those involved and were informally "structured" around the phases of the evaluation.
Rating:	4: Structured capacity building of evaluators and partners responsible for the evaluand was incorporated into the evaluation process

## Methodological integrity

Standard:	A literature review was developed which informed the analytical framework and findings of the evaluation
Comment and Analysis:	No literature review was developed. The report has only a paragraph noting the legal framework for PTOG and the need to audit and evaluate the programme.
Rating:	1: A literature review was not undertaken or was very poor
Standard:	The methods employed in the process of the evaluation were consistent with those planned and implemented adequately
Comment and Analysis:	<p>Data collection deviated somewhat from the plan and was not entirely systematic. The Business Plan for PTOG Data Gathering stated that 6 bus companies would be surveyed in 6 geographic areas. In the final report 8 companies and their passengers were surveyed - the 2 that were added are much smaller in terms of monthly passenger volumes. According to interviewees, this is all the companies receiving PTOG subsidies from the Northern Cape DTSL. The increase in the bus company sample was a beneficial deviation from the plan. It is not clear whether the drivers' and passengers' surveys were done in the 6 geographic areas that had been planned or in different / more areas.</p> <p>The data collection plan indicated that passengers would be randomly selected because it would be difficult to identify them in advance. In the interview the evaluation leader explained that surveyors boarded buses on their daily routes - the fact that some bus companies transport far more passengers than others was not specifically taken into account, and it appears that the small bus companies' passengers were over-sampled - but fortunately the results are presented by company with an indication of the passenger sample size before being aggregated. Data collectors interviewed whatever passengers on the bus they saw fit. Some enthusiastic data collectors also decided to collect data by approaching passengers waiting at bus stops, and some even did home visits. These responses were included into the total data collected and it is unclear which responses were collected in these alternative settings, or whether this skews the data. Still, even though responses were not systematically collected, the data collected appears informative as a first snapshot of the variety of experiences across the different companies, which is adequate for the evaluation.</p> <p>The evaluation also found that road traffic enforcement agents are so short-staffed that none of them monitor subsidised transport vehicles. The data collection therefore deviated from what was planned in this regard.</p> <p>As for the analysis of secondary data, as intended, basic descriptive statistics were generated based on this. The quality of this analysis is discussed further below.</p>
Rating:	2: The methods employed in the process of the evaluation deviated somewhat from those planned or implementation was inadequate

Standard:	A pilot of basic data collection instrumentation occurred prior to undertaking data collection and it was used to inform the research process
Comment and Analysis:	An informal pilot took place with the data collection team as part of their training. It involved them asking the questions of each other.
Rating:	2: A pilot of data collection instrumentation occurred but not in a way that could meaningfully test or improve upon instrumentation
Standard:	Data was collected from key stakeholders (e.g. implementers, governance structures, indirectly affected stakeholders) as data sources
Comment and Analysis:	Data was collected from bus companies, drivers, and passengers. Programme staff in the DTSL were not formally interviewed, nor did the evaluation draw on any sources e.g. quarterly reports on indicators, project reports etc. to supplement the expenditure data. This is an important omission, and helps to explain why the evaluation report did not present much analysis / tracing of the operational causes of challenges observed on the ground. Despite this omission, the bus companies, drivers and passengers constitute the harder to reach stakeholders and it is very valuable that the evaluation engaged them. The programme staff are arguably the easiest to involve later in analysing the findings and improving the programme.
Rating:	3: Data was collected from key stakeholders (e.g. Implementers, governance structures, indirectly affected stakeholders) as data sources
Standard:	The methodology included engaging beneficiaries appropriately as a key source of data and information
Comment and Analysis:	Passengers are the main intended direct beneficiaries of the PTOG. The (non-systematic) random sample of them was surveyed one-on-one by the surveyors. This is an appropriate way to obtain their views on the public transport that they use. The survey was appropriately short and focused on their transport experience, and it included a limited number of open-ended questions that elicited richer information from them. It was no doubt an important gesture for the survey respondents that such research was done, and the views that they shared certainly contributed significantly to the value of the report.
Rating:	4: The methodology included meaningfully engaging beneficiaries as a primary source of data and information (or if based on secondary data, includes data from beneficiaries and beneficiaries consulted on emerging findings)

## Project management (Implementation phase)

Standard:	The steering committee, technical working group and service provider worked together adequately to facilitate achievement of the objectives of the evaluation
Comment and Analysis:	The evaluation Working Group was supportive of the evaluation and the Transport Operations unit reportedly assisted wherever necessary to facilitate achievement of the evaluation objectives. The lead evaluator reportedly prepared well for presentations and communicated well about the project's progress. The interviewees noted no challenges regarding the cooperation between these stakeholders. If anything, the Working Group may have been too quick to approve the deliverables and may not have applied a sufficient level of scrutiny to the final report, as some errors slipped through as discussed below.
Rating:	3: The steering committee, technical working group and service provider worked together adequately to facilitate achievement of the objectives of the evaluation

Standard:	Support provided by the evaluation secretariat (e.g. the administrators responsible for the evaluation) facilitated achievement of the objectives of the evaluation (eg turnaround times, addressing problems, preparation for meetings etc)
Comment and Analysis:	The Policy and Planning unit reportedly performed the secretariat function well, for instance assisting with the logistics of the data collection and keeping track of the project progress. The senior manager: Policy and Planning worked constructively to support the evaluation, for instance, she came up with the idea of compensating for the lack of a peer reviewer (because of cost constraints) by sending the report to a peer department in the Western Cape for comment.
Rating:	4: Good support was provided by the evaluation secretariat and facilitates timely and constructive achievement of the objectives of the evaluation

## Reporting

---

### Completeness of the evaluation report

Standard:	The first draft evaluation report was of a sufficient quality to go to stakeholders and did not require major changes
Comment and Analysis:	The Working Group expressed their satisfaction with the first draft, so only unsubstantial further editing (e.g. typing errors, grammar errors) were required thereafter. Despite their satisfaction, a higher rating is not assigned because arguably, some weaknesses in the report slipped through the Working Group's review.
Rating:	3: A first draft of the evaluation report was of a sufficient quality to go to stakeholders and did not require major changes prior to sharing

Standard:	The final evaluation report is well-structured and complete in terms of the following: executive summary; context of the development evaluation; evaluation purpose, questions and scope; methodology; findings and analysis; conclusions and recommendations
Comment and Analysis:	The final evaluation report is not complete in terms of the components listed above. It includes a brief background / context paragraph, purpose, findings and analysis, and recommendations. It has no executive summary, evaluation questions and scope, and the methodology is not discussed upfront but rather briefly in the text where new data is presented (focusing on data collection methodology; no discussion of analysis methodology or limitations). There are no conclusions as pertains the findings; the "interpretations" section discusses some of the possible causes of the findings while the "summary" section at the end reminds the reader of the context of the study and speaks broadly about the need to use the findings. It is noted that the report in fact follows the DTSL's DEP report outline and the quality assessment thus demonstrates that following the DEP report outline may result in a report that does not align to the DPME standards.
Rating:	2: The final evaluation report is poorly structured and does not address all of the following components: executive summary; background/context of the evaluation; evaluation purpose, questions and scope; methodology; findings and analysis; conclusions and recommendations

### Accessibility of content

Standard:	The final evaluation report is user-friendly, written in accessible language and adequate for publication (e.g. adequate layout and consistent formatting; complete sentences and no widespread grammatical or typographical errors; consistency of style and writing conventions; levels of formality; references complete and consistent with cited references in reference list and vice versa; etc.)
Comment and Analysis:	The report is accessibly written - the writing is concise and to the point. There are no notable errors in language or layout of the main report, but two of the appendices are illegible in the copy that was sent to DPME. The report could have been made easier for the user to navigate if it was stated more clearly up front how the different sections contribute to the whole, and if the main chapters were better distinguished e.g. with a larger heading, or starting on a new page.
Rating:	3: The final evaluation report is user-friendly, written in accessible language and adequate for sharing (e.g. some spelling, grammar or formatting mistakes but these do not seriously detract from the report)

Standard:	Figures, tables and appropriate conventions are used in presentation of data (e.g. use of appropriate statistical language; reporting of p-values where appropriate; not reporting statistically insignificant findings as significant; clarifying disaggregation categories in constructing percentages; not using quantitative language in reporting qualitative data, etc.) and are readily discernible to a reader familiar with data presentation conventions
Comment and Analysis:	<p>The report makes somewhat inappropriate use of the term "average" (in some places it is correctly used, but in other places it is conflated with the concept of "aggregate" or "sum"). The word "total" is also used inappropriately where the value presented is in fact an average (rows 4 and 5, page 5). Still, since limited analysis of the data was done, it is not difficult for the reader to understand the point.</p> <p>Tables are clearly presented, barring one or two isolated errors (e.g. the headings of the months on page 4). The detail in figures are sometimes hard to interpret, for instance in the bar charts (p.7-8) a red-yellow-green colour scheme is used but it is not clear what the colours mean; furthermore, the bar charts do not always match their headings (fig.3.1; fig 3.3); and the bars in the charts do not bear a clear relationship to the values presented above them. The figures and graphs could also have better supported the logical flow of the discussion if they had been interspersed in the report; instead, the entire collection of figures and graphs are presented in pages 4 to 9, and then afterwards, they are discussed in the text. Similarly, a detailed table with aggregate data from all the survey instruments is presented in tabular form (pages 14 to 24), and then in the next section, it is interpreted somewhat more (frequencies in bullet points) and discussed in text. While presentation of the entire dataset is very transparent, it could perhaps have been done in the appendix in order to support the flow of the report.</p>
Rating:	2: Some figures, tables and conventions are used in presentation of data but not entirely appropriately or consistently

## Robustness of findings

Standard:	Data analysis appears to have been executed to an adequate standard
Comment and Analysis:	<p>In some areas, the data analysis is (1) inadequately explained, (2) hampered by errors, and/or (3) limited in depth. For these reasons, the report is rated as inadequate in terms of this quality assessment standard. The data may nevertheless be very useful where no such data has been available to decision makers in the past.</p> <p>The secondary data (expenditure and activity reports) is analysed appropriately with trends and averages, though there is little discussion of what this means for the programme. For instance, sections 3.2.3 and 3.2.4 simply sum together the values presented previously in graphs and tables, without discussing the importance of the results. Where there is some qualitative explanation for a trend, e.g. the decline in the number of subsidised vehicles, it is not always clear where the explanation for the trends was obtained (e.g. departmental staff? there is no mention of interviews or written communication with them).</p> <p>For the primary data, basic analysis was done. There is however a heavy reliance on the frequency of a finding, rather than the identification of patterns, correlations or apparent contradictions. It is reported question by question, using frequencies (number of respondents who gave a certain answer) in bullet points. There is limited discussion of what the data means for the programme (e.g. how important is the finding that some bus drivers are sometimes delayed, by management and/or passengers, in terms of the stated PTOG intention of benefiting the public?), and how the answers to different sections relate to each other. In the discussion of the passenger survey, the report notes that "responses provided depended on the demographics of the areas" - a potentially very important finding that is not satisfactorily addressed in the discussion that follows. Similarly, "some" passengers (no quantity given) reported waiting 5 to 10 minutes for buses, while others wait 1-3 hours. This is a major difference but is not analysed further.</p> <p>While the data analysis appears free of errors in most respects, the average subsidy amount per passenger is erroneously added together for two financial years on page 11, and the term "on average" is used incorrectly, for instance on pages 25 and 26 where it should read "in total".</p>
Rating:	2: Data analysis was executed to an extent but it appears inadequate or significantly lacking for some datasets

Standard:	Findings are supported by evidence which is sufficiently and appropriately analysed to support the argument, integrating sources of data
Comment and Analysis:	There is plenty of evidence presented, but it is not sufficiently analysed to produce an organised set of findings. Most of the report is written in list format (bullets or numbered sentences), which does not lend itself to a nuanced discussion of issues and arguments. Section 6: Interpretations, mostly reiterates points made in the presentation of the primary and secondary data (and is therefore directly supported by primary evidence). But it also includes some further explanatory arguments that are added here for the first time (e.g. a new reason is given for the fluctuating number of subsidised routes; and it is argued that friction between drivers and passengers is caused by the lack of driver training). These seem like they may have originated from the steering committee or management, but the sources are not cited. An early section presents a SWOT analysis, the evidence base for which is not cited. The sources of data are kept separate all the way up to the end of the report; they are not integrated into a framework. Therefore, the report findings read as a list of observations about how different role players are experiencing the PTOG.
Rating:	2: The evidence gathered has been analysed to support the argument to an extent but this is not entirely sufficient or appropriate, and different data sources may be presented separately rather than integrated
Standard:	There is appropriate recognition and exploration of the possibility of alternative interpretations
Comment and Analysis:	The evaluation was not founded in a literature review, or other framework that could illuminate alternative points of view on the topic. The report for the most part simply describes the available data. Explanations, where they are put forward, are stated as fact. For instance, the lack of driver training is cited as the cause for friction with passengers, but it is not explained why driver training is the cause, as opposed to any of the other issues mentioned in the report that may affect driver-passenger relations. In short, the report does not explicitly weigh up the validity of different explanations for the findings it presents.
Rating:	2: There is an implicit or indirect recognition of alternative interpretations
Standard:	The report appears free of significant methodological and analytic flaws
Comment and Analysis:	It is important to note that the analysis undertaken in this report was relatively limited. If one accepts the limited extent of analysis undertaken, then the analysis is mostly free of flaws. The only analytic flaws picked up in a review of the report are an error in calculation of the average subsidy amount per passenger on page 11, and the fact that the term "on average" is used incorrectly, for instance on pages 25 and 26 where it should read "in total".  The methodology was not explicitly stated nor justified in the report. Any evaluation needs to render a judgment on the merit or worth of something. This report did so without articulating evaluation questions, articulating the programme logic / theory of change, or explaining the standard to which the programme was being held. In this sense the report does not have the basic methodological features that are expected of government evaluation.
Rating:	2: The report appears to include some minor methodological and analytic flaws, but these are not significant
Standard:	Limitations of all aspects of the methodology and findings are clearly articulated (e.g. limitations of scope or evaluation design, recommendation for additional research, data collection challenges, etc)
Comment and Analysis:	Limitations are not comprehensively discussed. The evaluation's resource limitations and reliance on untrained data collectors are noted as "threats" in the SWOT analysis on page 3. (However, this is phrased as if the evaluation has yet to take place. The report does not state how these threats were dealt with.)
Rating:	2: There is some acknowledgment of the limitations of the methodology and findings but these are not clear or exhaustive

## Strength of conclusions

Standard:	Conclusions are derived from evidence
Comment and Analysis:	Section 6: Interpretations presents a summary of the data, by type (secondary, followed by each primary data instrument individually). Most, but not all the findings in this section are supported by evidence. Some introduce new facts, apparently based on discussions with DTSL management: for instance, the fact that the number of subsidised routes fluctuates is attributed for the first time in this section to the fact that many routes are under the jurisdiction of SANRAL or the Northern Cape Provincial Roads and Public Works department. This section does not entirely fulfil the role of a Conclusions section, as it does not pull together all that was presented. Instead it remains somewhat fragmented, being presented by data source instead of by root cause, theme or issue.
Rating:	2: Conclusions are derived from some evidence but do not encompass all of what was presented

Standard:	Conclusions address the original evaluation purpose and questions
Comment and Analysis:	The Interpretations section does not specifically address the stated evaluation purpose - it does not clearly judge its alignment to the strategic plan nor does it set out the extent to which the grant benefits the public (passengers) or the department. Some of the points imply ways in which the programme benefits or doesn't benefit passengers, but these findings are not set out systematically and they are still presented as bullet-point lists by source data instead of by theme or question. The interpretations section also does not address the evaluation questions, since these were never articulated. On the whole the Interpretations section, and the summary section at the end of the report, address the main areas of enquiry only broadly.
Rating:	2: Conclusions address the original evaluation purpose and questions in implicit or indirect terms to an extent

Standard:	Conclusions are drawn with explicit reference to the intervention logic or theory of change
Comment and Analysis:	The evaluation leader indicated that theory of change thinking played an important role in how the programme was evaluated. However, in the report itself, only a passing reference to the programme logic / theory of change was presented in the introductory sections, and in the Interpretations and Summary sections there is no discussion at all of what the findings mean for the validity or appropriateness of the theory of change.
Rating:	1: Conclusions fail to make reference to the intervention logic or theory of change

## Suitability of recommendations

Standard:	Recommendations are made in consultation with relevant government officials, stakeholders and sectoral experts
Comment and Analysis:	The recommendations were drafted by the lead evaluator and then presented for discussion in the Working Group before being finalised. This was not an in-depth workshop but there was an opportunity for them to comment. The draft report, including recommendations, was also circulated to all the bus companies for input and they expressed their approval in individual follow-up telephone calls. Lastly, the report was shared with the Western Cape Department of Public Works, Roads and Transport for "peer review" (this consisted not of detailed written comments but of approving remarks made in a telephone call). Thus there was some consultation with relevant government officials, stakeholders (bus companies) and one could call the Western Cape DPWRT an institution with considerable expertise in the evaluation sector.
Rating:	3: Recommendations are made in consultation with relevant government officials, stakeholders and sectoral experts

Standard:	Recommendations are useful- they are relevant, specific, feasible, affordable and acceptable
Comment and Analysis:	There are 9 recommendations, and they vary in their usefulness. The fact that the Working Group of senior managers approved the report, and that it was subsequently signed off by the head of department, suggests that the recommendations are acceptable. However, as the lead evaluator pointed out, some recommendations may not be feasible to implement - for instance, one recommendation is that if DTSL subsidises some of a company's routes, it should subsidise all to avoid the company having to either set all tickets at the subsidised price (and make a loss) or set different ticket prices for different routes. Most are specific about what should be done and by whom, although the last two argue for the importance of interventions (regular contact with passengers, and introduction of traffic officers) without explaining who should take responsibility for this and how it should be done. The fact that these recommendations were accepted casts some doubt on whether the intention is there to implement them all. The recommendations are nevertheless valuable and the experience of working to address/implement these recommendations will likely lead to valuable learnings about how to craft recommendations in future evaluations.
Rating:	2: Recommendations are of limited use - they vary in the degree to which they are relevant, specific, feasible affordable and acceptable

## Acknowledgement of ethical considerations

Standard:	The full report documents procedures intended to ensure confidentiality and to secure informed consent where necessary (in some cases this is not needed - e.g. evaluation synthesis - in which case N/A should be recorded)
Comment and Analysis:	All the primary data collection instruments are included as an appendix. It starts with a paragraph that briefly informs the respondent about the evaluation, and assures them of their anonymity.
Rating:	4: The full report documents all procedures to ensure confidentiality and to secure informed consent and provides some examples in appendices
Standard:	There are no risks to participants or institutions in disseminating the evaluation report on a public website
Comment and Analysis:	The interviewees are not aware of any such risks, nor are any such risks apparent in the report.
Rating:	4: There are no risks to participants or institutions in disseminating the original full evaluation report on a public website



## Follow-up, use and learning

---

### Resource utilisation

Standard:	The evaluation was completed within the planned timeframes and budget
Comment and Analysis:	The evaluation was completed within time frames and within budget, insofar as any budget was available for the fieldwork. The whole project took just three months, meaning the report was produced under considerable pressure, but it was completed. After March 2017, the lead evaluator was seconded to a different department and was only able to attend a ManCo later in the year, to present the final results and get the head of department to sign off on it.
Rating:	3: The evaluation was completed within the planned timeframes and budget

### Evaluation use

Standard:	Results of the evaluation have been presented to relevant stakeholders
Comment and Analysis:	The final evaluation results have been circulated to an extent within government. It was of course presented to the ManCo (which includes the head of department) and subsequently to the MEC. Thereafter it was shared with the Office of the Premier (but not presented as such). The report in its near-final form had also been shared with the bus companies for comment, who are key stakeholders external to government.
Rating:	4: Results of the evaluation have been presented to all relevant stakeholders, inside and outside of government

Standard:	A reflective process has been undertaken by the steering committee with the service provider (if no steering committee exists then by the evaluation management team or the involved department officials) to reflect on what could be done to strengthen future evaluations
Comment and Analysis:	No formal reflective session took place, but the Senior Manager: Policy and pIPnning and the evaluation lead have reflected informally on the experience at various occasions. Furthermore, although not involving the steering committee, there was an element of reflection in the data collection, with a daily reflection between the lead evaluator and the data collectors to determine "whether the data collectors did learn or experience anything during the data collecting procedure".
Rating:	2: The steering committee undertook a meeting in which some form of reflection occurred, but not in a clear, reflective process

Standard:	The evaluation study is seen by interviewed stakeholders as having added significant symbolic value to the policy or programme (e.g. raised its profile)
Comment and Analysis:	<p>According to interviewed stakeholders, the evaluation gave an opportunity for the province to become aware of the PTOG, as something that is "active" in the province (an effort on the part of government to support transport). Given that the ManCo served as the Evaluation Working Group, the evaluation also increased the time that the entire senior management team spent discussing the strengths and challenges of this programme, i.e. it improved its visibility within the department.</p> <p>Not only did the evaluation raise the profile of the PTOG, but it also strongly raised the profile of the national and departmental policies and policy framework on government evaluations. Management's interest in evaluation reportedly has grown; they "value the evaluation process" to a greater extent now, because this evaluation was a positive experience and really underscored the value of a systematic study of a programme. Senior managers reportedly now want their own programmes to be evaluated and there is currently a process underway to conduct the next evaluation.</p> <p>Thus, the evaluation was of substantial symbolic value to government stakeholders, not just for the programme being evaluated, for government evaluation policy.</p>
Rating:	4: The evaluation study is seen by interviewed stakeholders as being of substantial symbolic value to the policy or programme and has noticeably raised its profile amongst stakeholders
Standard:	The evaluation study is of conceptual value in understanding what has happened and possibly in shaping future policy and practice
Comment and Analysis:	The evaluation was described as an "eye-opener" by the interviewed stakeholders. Both senior managers and the MEC were described as having gained new insight about the PTOG, such as what bus drivers go through. Also, because of some of the challenges raised in the report, the DTSL met some of the affected bus companies to "find a way forward".
Rating:	4: The evaluation study is of conceptual value in understanding what has happened and some interviewed stakeholders indicated the likelihood of it constructively shaping policy and practice

## References

---

Northern Cape Department of Transport, Safety and Liaison (DTSL). 2017. Terms of Reference for the Northern Cape Department of Transport, Safety and Liaison Evaluation Plan: Departmental Evaluation Working Group.

DTSL. n.d. Departmental evaluation plan 2016/2017 (proposal) (PowerPoint presentation). Compiled by Dr V. E. Sithole

DTSL. n.d. Re: Update on PTOGs Evaluation. Present by Dr V. E. Sithole. (PowerPoint presentation)

DTSL. n.d. Data collecting questions for Provincial Transport Operating Grants (PTOGs).

DTSL. n.d. Provincial Transport Operations Grants (PTOG) Evaluation. April 2015 - December 2016.

DTSL. 2017. Departmental Evaluation Plan. Prepared by Dr V. E. Sithole.

DTSL. n.d. Departmental evaluation plan 2016/2017 (proposal) (PowerPoint presentation). Compiled by Dr V. E. Sithole

DTSL. 2017. Submission for PTOGs Evaluation Data Gathering. Compiled by Dr V. E. Sithole.

## List of Interviewees

---

Ms Renee Williams, Senior Manager: Policy and Planning; and Dr Vusi Sithole, Acting Manager: Office of the Head of Department. Northern Cape Department of Transport, Safety and Liaison. Joint telephone interview, 23 January 2018.