



# planning, monitoring & evaluation

---

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

## Report on the Assessment of Government Evaluations

Evaluation Title:

Implementation Evaluation of the National Curriculum Statement Grade R to 12 - Focusing on the Curriculum and Assessment Policy Statements (CAPS)

Evaluation Number:

601

Start Date:

Friday, 04 March 2016

Completion Date:

Thursday, 25 May 2017

Period of Evaluation:

14 months

Submitted:

Tuesday, 16 January 2018 by Cara Hartley

Approved:

Tuesday, 16 January 2018 by Cara Hartley

## Evaluation Details

Evaluation Title:	Implementation Evaluation of the National Curriculum Statement Grade R to 12 - Focusing on the Curriculum and Assessment Policy Statements (CAPS)
Evaluation Number:	601
Evaluation Completion Date:	Thursday, 25 May 2017
Created:	Thursday, 02 November 2017 by Cara Hartley
Submitted:	Tuesday, 16 January 2018 by Cara Hartley
Approved:	Tuesday, 16 January 2018 by Cara Hartley
Period of Evaluation:	14 months
Known Cost:	R 2 214 582,00
Known Cost Type:	Referenced
Initiated By:	Department of Planning, Monitoring and Evaluation & Department of Basic Education
Initiated By Internal:	Yes
Service Providers:	JET Education Services
Undertaken By Internal:	No

## Assessors

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

## Assessment Documents

Document Name:	Document Type:	Added By:	Added On:
CAPS Implementation Evaluation - Final Full Report 26 May17 5.00.pdf	Evaluation report	Cara Hartley	Thursday, 02 November 2017
Revised CAPS Evaluation TORs 17 September 2015 10h00.pdf	Terms of Reference (ToR) for the evaluation	Cara Hartley	Tuesday, 12 December 2017
Literature Review - Implementation Evaluation of CAPS 15Jan17.docx	Any other relevant documentation pertaining to the evaluation process	Cara Hartley	Tuesday, 12 December 2017
JET DBE ESA Field Work Report 22_August_2016.docx	Progress reports	Cara Hartley	Tuesday, 12 December 2017
CAPS Evaluation - Inception Report 26 January 2016.docx	An inception report	Cara Hartley	Tuesday, 12 December 2017
CAPS Evaluation - Final JET Proposal 26 Jan2016 (2).docx	An inception report	Cara Hartley	Tuesday, 12 December 2017
CAPS Approved Summary Report 25 May17 Final 5.00.pdf	Evaluation report	Cara Hartley	Tuesday, 12 December 2017
CAPS Eval - District interviews - Blank.docx	Assessment tools	Cara Hartley	Tuesday, 12 December 2017
Signed SLA.pdf	A letter of engagement or contractual agreement with the service provider	Cara Hartley	Monday, 15 January 2018

Peer Reflection on CAPS Evaluation.docx	Any other relevant documentation pertaining to the evaluation process	Cara Hartley	Monday, 15 January 2018
Re APPROVED CAPS EVALUATION REPORT (FULL AND SUMMARY).msg	Any other relevant documentation pertaining to the evaluation process	Cara Hartley	Monday, 15 January 2018
Refilwe Theory of change workshop lessons 8 September 2015.docx	Any other relevant documentation pertaining to the evaluation process	Cara Hartley	Monday, 15 January 2018
DPME Comments on CAPS report 17 02 12 JB.docx	Any other relevant documentation pertaining to the evaluation process	Cara Hartley	Monday, 15 January 2018
Assessment Report at Moderation.pdf	Assessment Report at Moderation	Mike Leslie	Tuesday, 16 January 2018

## Quality Assessment Summary

This evaluation is of a good standard, scoring 3.73 out of 5 in the quality assessment. The evaluation purpose was to evaluate to what extent the National Curriculum Statement Grade R to 12 (focusing on the Curriculum Assessment Policy Statement, CAPS) has been implemented, and how implementation may be strengthened. The evaluation report presents valid findings drawn from multiple sources, to achieve this purpose. The evaluation scored quite well on nearly all the Overarching Considerations against which it was scored, as shown in the spider diagram above - in this sense it was a well-balanced evaluation process and report.

The evaluation's Planning and Design phase scored 3.81. The Terms of Reference were well-conceptualised and comprehensive, the team was well-resourced to conduct the evaluation, and the inception phase served to bring stakeholders to a common understanding about the evaluation plan. The time and budget allocated for the project were tight, but adequate.

The Implementation Phase scored the highest, 3.90. This phase benefited from a literature review that was comprehensive and particularly relevant to the evaluation, helping to inform the instruments and report in key ways. The evaluation also benefited from the fact that the school-level data collection instruments were piloted, and that challenges were overcome to ensure that data of the intended coverage and content was collected. Another strength is the capacity development element, which was guided by a capacity building structure and supported by brief written reports by the capacity building participants on what they were learning. The approach to capacity development in this evaluation is an example of good practice.

The Reporting Phase also scored well, at 3.82. The report is of a relatively good quality, written in accessible language, with only minor errors. The analysis that it presents, appears to have been done to a good standard. It also effectively integrates different sources of evidence to provide a solid foundation for its conclusions and answers to the evaluation questions. The recommendations went through several iterations and were ultimately adequate, though it was noted by DBE that some recommendations remained somewhat broad.

Finally, the Follow-up, Use and Learning Phase scored 3.33. This is a lower score than what the other phases were given, but is nevertheless more than adequate. The challenges with this phase include the fact that the evaluation went over budget and somewhat over time; and that some stakeholders undertook some written reflection but did not meet to reflect in a structured and participatory way on the lessons learned for evaluations going forward. Nevertheless, it became clear in this Phase that the evaluation has been of considerable symbolic as well as conceptual value. The evaluation findings have been quite widely shared with key stakeholders inside as well as outside government (notably, teachers' unions); and some of these stakeholders have participated in developing an improvement plan based on the recommendations. Because of this extensive engagement and the quality of the report, stakeholders are positive about the value of the report for shaping future policy and practice.

Considering the good participation of key stakeholders throughout the process and the good scores that the Planning, Implementation and Reporting Phases achieved, the evaluation report can be taken as a valid and credible description of the implementation of CAPS so far - noting the caveats and limitations which are clearly stated in the report. The clear judgement rendered regarding the theory of change, and the comprehensive literature review are also likely to be of value to evaluation users in South Africa and beyond.

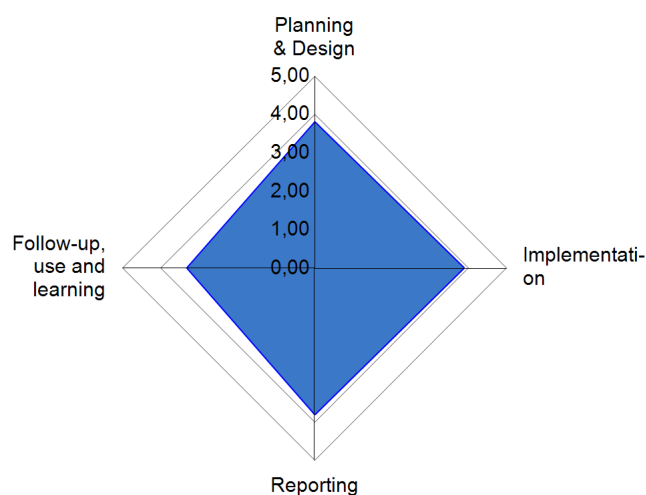
## Quality Assessment Scores

Phase of Evaluation	Score
Planning & Design	3,81
Implementation	3,90
Reporting	3,82
Follow-up, use and learning	3,33
Total	3,73

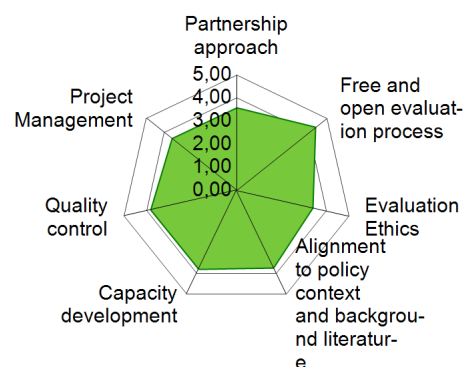
Overarching Consideration	Score
Partnership approach	3,56
Free and open evaluation process	4,39
Evaluation Ethics	3,40
Alignment to policy context and background literature	3,73
Capacity development	3,80

Quality control	3,82
Project Management	3,57
<b>Total</b>	<b>3,73</b>

Scores: Phases of Evaluation



Scores: Overarching Considerations



Phase of Evaluation	Area of Evaluation	Score
Planning & Design	Quality of the TOR	3,56
Planning & Design	Adequacy of resourcing	3,57
Planning & Design	Appropriateness of the evaluation design and methodology	4,27
Planning & Design	Project management (Planning phase)	4,00
Implementation	Evaluation ethics and independence	3,50
Implementation	Participation and M&E skills development	4,43
Implementation	Methodological integrity	3,79
Implementation	Project management (Implementation phase)	4,00
Reporting	Completeness of the evaluation report	3,50
Reporting	Accessibility of content	4,00
Reporting	Robustness of findings	4,00
Reporting	Strength of conclusions	4,33
Reporting	Suitability of recommendations	3,00
Reporting	Acknowledgement of ethical considerations	3,57
Follow-up, use and learning	Resource utilisation	2,00
Follow-up, use and learning	Evaluation use	3,67
<b>Total</b>	<b>Total</b>	<b>3,73</b>

## 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:	The TOR was well-structured and covered all of the following: background, purpose, evaluation questions, design & methodology, deliverables and timeframes, resource requirements (service provider qualifications and competencies), intended users and uses. In addition, it discussed the scope, milestones, and management arrangements. It quite clearly outlined the evaluation to be conducted.
Rating:	4: The evaluation was guided by a well-structured and complete TOR or internal evaluation proposal of a good standard
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:	<p>The evaluation purpose was "to evaluation whether the curriculum has been implemented as specified in the Curriculum and Assessment Policy Statements (CAPS) and how implementation can be strengthened"; and the evaluation type was specified in the TOR as an Implementation Evaluation. Thus the purpose and evaluation type were well matched. In terms of approach, like most DPME evaluations, the TOR called for a relatively participatory evaluation including a steering committee. Interviewees felt that this approach was suitable since it helped support buy-in by the Department of Basic Education units who have introduced the policy and will need to use the evaluation results.</p> <p>The National Curriculum Statement (NCS) Grades R to 12 (including CAPS) has been phased in since 2012, and in that first year it started in the Foundation Phase and Grade 10. Thus, this was where the curriculum has been implemented the longest. It was therefore appropriate that in terms of scope, the TOR specified that the evaluation should focus on these levels. Since the year 2015 was the fourth year of implementation in the Foundation Phase and Grade 10, it was also an appropriate time to move from analysing the design or introduction of the new curriculum, to its implementation. This focus on implementation was supported by the TOR in that it said the service provider need not focus on the policy document or protocol. The scope was further appropriately defined in terms of geographic coverage, to cover four quite different provinces.</p>
Rating:	4: The approach and type of the evaluation was well-suited to the purpose and scope of the evaluation TOR
Standard:	The TOR (or an internal evaluation proposal) identified the intended users of the evaluation and their information needs
Comment and Analysis:	The TOR listed the potential users of the evaluation and it specified how they would use it. For DPME's and DBE's potential use of the evaluation, the TOR indicated they would use it to understand specific aspects of the curriculum implementation; the TOR thereby stated (briefly) what aspects they needed information on. The intended uses were quite broad.
Rating:	3: The TOR (or an internal evaluation proposal) identified the intended users of the evaluation and their information needs

Standard:	Key stakeholders were involved in the scoping of the TOR and choosing the purpose of the evaluation
Comment and Analysis:	The DPME and DBE jointly developed the TOR. An interviewee mentioned specifically that two DBE officials were involved - one from the office of the Director-General and one from the Strategic Planning, Research and Coordination (RCME) unit. In this way there was DBE representation from a transversal as well as M&E-specific perspective in the process.
Rating:	3: Key stakeholders were involved in the scoping of the TOR and choosing the purpose of the evaluation

## Adequacy of resourcing

Standard:	The evaluation was adequately resourced in terms of time and budget allocated
Comment and Analysis:	<p>The Service Level Agreement indicates that 10 and a half months were allocated for the evaluation (inception report in mid-January; final presentation in November 2016). This is adequate for an evaluation that needed to go through the sequential steps of literature review, development of a theory of change (DBE had not developed one yet) and analytical framework, instrument development and qualitative fieldwork in four provinces.</p> <p>The budget constraints (see discussion later on) were not necessarily foreseeable from the outset and are therefore not taken into account in this standard, which is about planning.</p>
Rating:	3: The evaluation was adequately resourced in terms of time and budget allocated

Standard:	The team conducting the evaluation was adequately resourced in terms of staffing and skills sets
Comment and Analysis:	<p>An implementation evaluation of the NCS (focusing on CAPS) requires extensive education sector expertise, research expertise and evaluation expertise. The evaluation TOR also specified that the team leader should have at extensive experience working with government at a high level. The proposal shows that the JET team fulfilled this criteria; and the interviewed stakeholders affirmed this view. The team leader is an established expert in the evaluation sector who has among others, headed the National Education Evaluation and Development Unit. Furthermore, the project manager had experience in National Evaluation Plan evaluations, and could help to ensure the evaluation satisfied DPME's requirements. Two interviewees agreed that the original project plan had perhaps allocated too limited a role to her (i.e. original resourcing was somewhat less than ideal in this regard), but when the need for it arose, she usefully supplemented the team leader's limited experience with National Evaluation Plan evaluations. The evaluation also had an international expert, a professor who attained her PhD in the United Kingdom, has held consultancies and visiting fellowships in the United States, and has specifically conducted research on education in other African countries.</p>
Rating:	4: The evaluation was well resourced in terms of staffing and skills sets

## 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:	At the TOR stage, one key evaluation question (KEQ) was specifically focused on the theory of change - whether it is working as expected, and based on this, whether one can expect the outcomes as per the theory of change. There was also reference to the theory of change in the delineation of the evaluation scope. The final revised proposal, which accompanied the inception report, had a preliminary theory of change, with a visual representation and narrative discussion. The theory of change was described as having seven key nodes / delivery mechanisms, and the document discussed how these nodes would inform the way that the KEQ would be answered. In this way, the theory of change was well integrated into the TOR and Inception Report.
Rating:	5: The intervention logic or theory of change of the evaluand was well integrated into the TOR and the Inception Report, including visual representations, and informed the design of the evaluation
Standard:	The planned methodology was appropriate to the questions being asked
Comment and Analysis:	<p>The seven evaluation questions are typical of an implementation evaluation. The methodology was primarily qualitative, with 24 case study schools where fieldworkers would collect primary data over two days. Interviews with other stakeholders in the education system were also planned, which spoke to higher-level implementation questions. The revised proposal which accompanied the inception report, already briefly discussed the existing research on each node in the evaluation theory of change, thereby clarifying how the evaluation methodology would contribute to what is already known to answer the evaluation questions. In this way the methodology considered the data already available.</p> <p>The decision to take a more qualitative than quantitative approach was appropriate because, as was noted in the revised proposal and also mentioned by the interviewees for this quality assessment, there is plenty of quantitative evidence on education outcomes. What was of interest here was why outcomes are so weak, and a qualitative methodology featuring 24 schools across four provinces could allow for sufficient insight into daily implementation of the curriculum to pick up on "softer" reasons possibly not observable from quantitative datasets. It was nevertheless noted that broader coverage of data on day to day curriculum implementation at schools (perhaps through a larger, quantitative survey) could have supported generalisation of some of the lessons from the 24 case studies.</p>
Rating:	4: The planned methodology was well suited to the questions being asked and considered the data available



Standard:	The sampling planned was appropriate and adequate given the focus and purpose of evaluation
Comment and Analysis:	<p>The agreed sampling approach was to do 24 case studies in 4 provinces. In each province, a matched pair of high schools would be selected, plus a third one giving a different perspective; and likewise a matched pair and a third primary school. All schools would be in quintiles 1 to 3. The match would be made based on similarities in socio-economic status, location, management and governance, history and culture - but with the key variable of learner performance (ANA and NSC scores) differentiating the pair. This sampling methodology would yield insight into why similar schools applying the same curriculum are producing better and worse learner results, with the possibility that it is linked to instructional leadership or pedagogy.</p> <p>The approach has reportedly been much discussed among stakeholders that have engaged with the results. The first critique is that NSC and ANA results are not necessarily good indicators of school or teacher effectiveness. The evaluation report confirms that this approach did not yield pairs of schools of clearly different education quality - but there are no obvious alternatives that could have been used. The second critique is that 24 schools cannot be statistically representative of all schools in quintiles 1 to 3. This critique was noted during the evaluation planning, but steering committee decided to prioritise rich, qualitative data ("subtleties of comportment and culture... that would not necessarily be apparent in a survey" as the final report puts it) over wide coverage, since the evaluation's purpose was to understand implementation on the ground at schools. Furthermore, the interviewees pointed out that despite the small sample size, the evaluation found few differences in the variables of interest (they were quite uniformly poor), suggesting a common set of challenges with the implementation of the curriculum. Furthermore, the evaluation findings were in line with several other small-N studies, suggesting that despite the small sample, the findings give a valid indication of conditions in quintile 1 to 3 schools. Even though it would have been easier to dispel critiques with a larger sample, the sampling was good to provide valid findings, which would serve the focus and purpose of the evaluation well.</p> <p>The sample of interviewees in national and provincial education departments was also deemed comprehensive and appropriate by all.</p>
Rating:	4: The sampling planned was good given the focus, purpose and context of the 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>According to interviewees, stakeholders made good use of the inception phase to reach common agreement about the way forward - including ironing out differences of nuance between the service provider, DBE and DPME. An interviewee noted however, that the staff from the curriculum unit were not so engaged during the inception phase. If they had been, they may have helped to narrow the evaluation's areas of emphasis, which may have resulted in a more specific set of findings that was more aligned to the focal areas of relevance to the department. Still, other DBE staff were actively involved, and the effect of less participation by the curriculum unit was not so severe that the inception phase should be considered inadequate in retrospect. The general view of the interviewed stakeholders was of a good inception phase, with all being in agreement regarding issues of approach, sampling, methodology etc.</p>
Rating:	4: The inception phase was used to good effect to achieve a common agreement and understanding of 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:** The main consideration was to ensure confidentiality of schools and individual participants. There is some evidence of an effort to ensure this: the fieldworker training manual stated that confidentiality is assured in this study; in the instructions regarding the teacher tests, the fieldworkers are instructed to remind teachers of the complete confidentiality of the study. Furthermore, no individuals' or schools' names are mentioned in the final evaluation report. Teachers were also given the choice whether or not to participate in the teacher tests. No informed consent forms were signed by teachers or interviewees at other levels of the system, but according to interviewees this is not deemed necessary for DBE-commissioned studies of public schools.

**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:** No interviewees noted any interference with the evaluation, or any efforts to amend the tone or substance of the findings. The service provider noted that DBE willingly shared the requested data and information.

**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:** The evaluation benefited from both a steering committee and a technical working group (see the first few pages of the evaluation report). Members were from DPME and DBE; and within DBE from several relevant units. Interviewees were all satisfied that this consisted of the key stakeholders with regards to the NCS and CAPS. As noted above, some staff from the curriculum unit were considered not so actively involved in the earlier stages of the evaluation, with repercussions for the relevance and specificity of the evaluation findings. Nevertheless, on the whole, the steering committee is considered to have been very engaged throughout the process and to have offered detailed and considered feedback on all deliverables. Furthermore, all interviewed stakeholders felt that it was appropriate to keep the committees' membership limited to national DBE and DPME (instead of including, for instance, provincial education department representatives, education interest groups or teacher unions). All felt that such wide representation on these committees could have slowed the process down more, and that these stakeholders could rather be consulted at the Improvement Plan stage (as they were - see below).

**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:	In this evaluation, it was agreed that DPME (evaluators / evaluation support) and DBE (department responsible for the evaluand) would nominate junior staff / interns to participate in the evaluation. One DBE intern and one DPME intern were nominated as per the plan. A brief capacity development plan was developed in discussion with the service provider (a list of activities and dates from them to participate in). DPME introduced a capacity building structure which allows the capacity building participant to document the hours spent, content and skills learnt in each project stage. It also includes a brief reporting template for review by the capacity building participant's supervisor. The interns participated using this structure to document their learnings; the DBE staff member reportedly worked quite closely with the service provider in several stages of the project and thereby gained considerable valuable experience. This evaluation quality assessment the templates, and one example of the lessons documented by the participants. It is reported that both the participants found contract work - one in an evaluation consulting firm and the other with DBE - following their internships, and the DPME interviewee relates this to the capacity building experience they gained in this evaluation among others. This can be regarded as an example of good capacity building practice.
Rating:	5: Well thought-through capacity building of evaluators and partners responsible for the evaluand occurred as per the plan which was supported by evidence of learning throughout the process

## Methodological integrity

Standard:	A literature review was developed which informed the analytical framework and findings of the evaluation
Comment and Analysis:	A detailed literature review was developed and it was structured according to the nodes in the theory of change. It was relevant to the evaluation as follows: (1) to describe three other South African studies that followed a similar sampling approach - and the strengths and weaknesses of the approach; (2) it identified theoretical constructs which helped to inform the evaluation's analytical framework; (3) it described the findings of existing literature on the each of the nodes of the theory of change, i.e. it started identifying hypotheses to be tested in, and findings to be integrated with, the evaluation. In these ways the literature review provided a good foundation for the evaluation both in terms of analytical framework, and to inform the findings.
Rating:	4: A good quality literature review was developed which was insightful in terms of the analytical framework and provided good context for the findings
Standard:	The methods employed in the process of the evaluation were consistent with those planned and implemented adequately
Comment and Analysis:	The methods employed in the process of the evaluation did not deviate significantly from what had been agreed in the inception phase. Sampling was done as agreed, with variations in the approach for only two schools, for legitimate reasons which are stated in the full evaluation report. Unfortunately the data collection period had to be extended by more than a month because by the time fieldwork started, some high schools were writing June exams. Nevertheless, the fieldwork report includes a table showing that the fieldworkers ultimately managed to collect 99% of the possible data through the intended methods at the schools, as planned. Interviews were conducted with national and district-level respondents as planned (although some were represented by colleagues / subordinates). Provincial curriculum managers were not reached in person as planned and instead, a written questionnaire was sent out to which "not all" responded (it is not clear how many did). In all, the evaluation faced some challenges, but these were largely overcome and the data collection phase ultimately yielded good data in line with the coverage and content that was planned, albeit that it took a bit longer than planned.
Rating:	4: The methods employed in the process of the evaluation were consistent with those planned and implemented well (in terms of time, coverage, and content)

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:	The school-based data collection instruments were subjected to a thorough piloting process. As documented in the standalone piloting report, two districts in Gauteng were randomly selected and the agreed sampling method was applied to select a matched pair of schools in each district. The full two-day data collection process was then implemented in each of the schools, including all interviews, classroom observations, reviews of learner workbooks etc. As is clear from the piloting report, the pilot resulted in numerous revisions to the instruments and approach. This is an example of good practice. Unfortunately a higher rating cannot be given on this standard as the instruments for semi-structured data collection with district, provincial and national officials were not piloted.
Rating:	3: A pilot of basic data collection instrumentation occurred prior to undertaking data collection and it was used to inform the research process
Standard:	Data was collected from key stakeholders (e.g. implementers, governance structures, indirectly affected stakeholders) as data sources
Comment and Analysis:	As planned, data was collected from 24 schools as well as district, provincial and national officials. This covers all the key stakeholder groupings as planned. As discussed above, the data collection achieved the intended sample well; the only small issues in coverage were the provincial officials' response rate to the email questionnaires and the few district officials who sent colleagues / subordinates in their place.
Rating:	4: Data was collected from the intended key stakeholder groupings in line with the envisioned range and type of stakeholders (approx. 80-89% of intended)
Standard:	The methodology included engaging beneficiaries appropriately as a key source of data and information
Comment and Analysis:	The ultimate beneficiaries of the curriculum, learners, were engaged in the primary data collection in that the fieldwork included classroom observation. Furthermore, learners' books were reviewed as a secondary source of information on their engagement with the curriculum. Data on learners' performance in NSC and ANA tests was also integrated into the study. Teachers, the intermediate beneficiaries of the curriculum, were also well covered through the school-based data collection - classroom observations, interviews and teacher tests.
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:	All interviewees indicated that the steering committee, technical working group and service provider worked together well to facilitate achievement of the evaluation objectives. It was noted that the DPME officials provided valuable guidance; DBE members generally engaged thoroughly with the deliverables and gave good inputs; and the service provider was flexible and responded constructively to the steering committee's feedback. As noted earlier, an even more useful evaluation product could have been produced if the DBE curriculum unit staff had been more closely involved from the beginning; this is why a score of 4 instead of 5 is assigned here.
Rating:	4: The steering committee, technical working group and service provider worked together in a flexible and constructive manner facilitating 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 evaluation secretariat role was played by DPME. By all accounts, DPME helped to ensure steering committee participation, served a valuable role as collator of steering committee inputs on deliverables, corresponded promptly, arranged meetings as well as possible given members' scheduling constraints, and generally supported a constructive evaluation process. While the evaluation was not completed within original time frames, this is attributable to other factors, not the support provided by the secretariat.
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 first draft evaluation report was described by the DPME and DBE interviewees as somewhat too rough, to the extent that in retrospect, it should probably not have been shared with stakeholders. It required more considerable revisions than had been originally anticipated by the service provider or the steering committee members. It ultimately went through 6 iterations, as shown on page iii.
Rating:	2: A first draft of the evaluation report was of a poor quality and required major changes

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 well-structured and complete in terms of the following: executive summary; context (brief "background" section, with literature review section providing more detail), evaluation purpose, questions (under section 2: "the brief"); methodology; findings (including analysis), conclusions and recommendations. Thus the only element not explicitly included in the report is that of scope, but this can be inferred from the description of the methodology, as well as the discussion of the literature review (where the report is explicit about areas where only existing sources were used vs. areas where the evaluation contributed new primary research). In all, the report structure is easy to follow. Some useful appendices support the main findings, including an updated log frame showing which assumptions proved to hold true, or not, in the evaluation.
Rating:	5: The final evaluation report structure is excellent, complete and makes exceptional use of the appendices to supplement the main report content structure

## 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 quite clearly and understandably written. Abbreviations are introduced before they are used. There are only small and isolated errors with grammar, layout or language; it is sufficiently edited in this sense to be published. The interviewees consider the report to be user-friendly and written in a clear and accessible style.
Rating:	4: The final report is well written, accessible to the common reader and ready for publication with only minor spelling, grammar or formatting mistakes

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:	The data is presented in quotes, boxes, figures and tables. Appropriate conventions are used in presenting them, so that it is clear, for instance, what the axes on graphs represent. Where correlation is discussed with regards to teacher test scores and learner workbook coverage, R-squared values are given. A footnote would have been useful here but the values will be readily discernible to a reader familiar with the conventions, and the interpretation of the results is clear in the text. Generally, the text explains the relevance and interpretation of the presented data (tables, figures etc.) well, producing a clear flow of narrative which is supported by the presented data.
Rating:	4: Figures, tables and conventions are well used for a variety of types of data presentations and supporting explanations make them accessible to readers

## Robustness of findings

Standard:	Data analysis appears to have been executed to an adequate standard
Comment and Analysis:	The data analysis appears to have been executed well for all datasets. There are no obvious errors in the way the numerical figures have been calculated nor in the way that the data in general have been interpreted.
Rating:	4: Data analysis appears to have been well executed for all datasets
Standard:	Findings are supported by evidence which is sufficiently and appropriately analysed to support the argument, integrating sources of data
Comment and Analysis:	The findings are all based clearly on the data that is presented, and as discussed above, data appears to have been analysed well for all datasets. Throughout the report, the evidence is used with a clear link to the question or issue at hand, so that a clear line of reasoning emerges. Notably, the findings from the substantial body of literature that was reviewed, is well integrated with the new analysis of secondary and primary data.
Rating:	4: The evidence gathered is well analysed, integrated and supports the argument in key sections of the report, without presenting data which are not used in the argument
Standard:	There is appropriate recognition and exploration of the possibility of alternative interpretations
Comment and Analysis:	The report frequently presents more than one alternative explanation to a piece of data and then weighs them up. For instance, where it presents the finding that high-performing schools are not distinguishable from low-performing schools in terms of reported instructional leadership behaviour, three possible explanations are advanced. Similarly, two reasons are suggested why schools would report shortage of learning and teaching support materials despite large budget allocations to this. In this way, alternative explanations are dealt with quite transparently.
Rating:	4: There is clear recognition of the possibility of alternative interpretations and these are concisely presented without detracting from other findings

Standard:	The report appears free of significant methodological and analytic flaws
Comment and Analysis:	As discussed above, the report is quite transparent about its methodology and analysis, states limitations clearly, and generally appears to present findings based on sound analysis. There are no apparent significant methodological analytical flaws. The report documents the fact that a steering committee oversaw the evaluation and approved the reports. It also notes that there was a technical evaluation working group. It also mentions that the school data collection instruments were piloted. The involvement of a peer reviewer was confirmed in the interviews but was not mentioned in the final evaluation report. His/her comments both on the evaluation instruments and an iteration of the report were reportedly helpful.
Rating:	4: The report documents some of the methodological and analytical processes used to ensure that it is free of methodological and analytic flaws

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:	<p>The evaluation is very explicit about limitations to its methodology in particular. Rather than a single section dedicated to limitations, there are several pieces in the report where specific limitations are discussed. They appear quite comprehensive, including the limitations / "risks" / "caveats" of, among others:</p> <ol style="list-style-type: none"> <li>1. The ordinal variables used to score teachers' competence on each indicator of pedagogic competence.</li> <li>2. The sampling method (especially the reliability of ANA and NSC scores)</li> <li>3. The data collection approach (limitations of validity and reliability)</li> <li>4. Generalisability of the short teacher competency tests</li> </ol> <p>There are also some recommendations for further research. Not all the discussions of limitations / "risks" are accompanied by a discussion of what was done to mitigate them. In some instances, the the reader can be left with the impression that the limitations render the data of limited value to the evaluation.</p>
Rating:	4: Limitations of all aspects of the methodology and findings are clearly articulated and distinguish between different kinds of limitations

## Strength of conclusions

Standard:	Conclusions are derived from evidence
Comment and Analysis:	The report concludes with two sections: answers to the evaluation questions; and conclusions organised by the evaluation criteria. The former is more detailed with more discussion of evidence; the latter a more concise discussion of what the implications of the findings are for the curriculum's achievement of the criteria. The sections are simply a synthesis of the evidence presented in the report; there are no unexpected claims or any points that are new to a reader who has read the report. The conclusions have a strong foundation of well-analysed and integrated data (as discussed above).
Rating:	4: Conclusions are derived from evidence and well supported by multiple sources of data that has been well analysed

Standard:	Conclusions address the original evaluation purpose and questions
Comment and Analysis:	The evaluation conclusions address each key evaluation question directly and in some detail. In doing so, and in presenting the conclusions in relation to the evaluation criteria, the evaluation covers the purpose of evaluating the implementation of the NSC (especially CAPS) and identifying areas where it can be strengthened, well.
Rating:	4: Conclusions address the original evaluation purpose and questions well



Standard:	Conclusions are drawn with explicit reference to the intervention logic or theory of change
Comment and Analysis:	Since there is a key evaluation question devoted to the theory of change, it is addressed explicitly in the section on answers to the key evaluation questions. The conclusions include a visual depiction of the findings against the theory of change, with colour codes indicating what the evaluation found about each box (e.g. a red box means it is a blockage in the system). In this way it is very clear what the evaluation found regarding the theory of change. The proposed revised theory of change further shows where additional support, policies, research and development should fit into the theory to unblock the blockages. These supports, policies, research and development are then included and fleshed out in the recommendations. It is also noted that the evaluation report goes further in the appendix, where it provides a judgment on each of the assumptions in the log frame based on the evaluation findings and literature.
Rating:	5: The conclusions are exceptional in the manner that they provide a judgement on the intervention logic or theory of change and are clearly linked to design recommendations

## 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 service provider and then went through some rounds of comments from the steering committee (government officials including the unit responsible for the curriculum).
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:	<p>As mentioned above, the recommendations went through several rounds of inputs and revisions. DPME assessed the recommendations against its internal criteria, and revisions were made where DPME identified that they initially fell short. Following the revisions, the steering committee approved the recommendations. The DPME interviewee felt that this demonstrates the usefulness of the recommendations. The DBE interviewees, although they participated in giving the abovementioned inputs and approving the recommendations, expressed a lingering degree of dissatisfaction with the final recommendations, saying some of the proposed actions were already underway in the DBE, without the need for a recommendation. Other recommendations they found too broad (insufficiently specific) and therefore hard to operationalise. The DBE interviewees attributed this to the broad nature of the evaluation and evaluation questions, and felt that more could have been done from DBE's side to delineate the focal areas earlier on in the report. In contrast to the DBE interviewees, the service provider interviewee felt that the recommendations had become too numerous and detailed over the course of the revisions.</p> <p>These different responses suggest that the steering committee and the service provider did not entirely "find each other" through the process of drafting the recommendations. Nevertheless, when reading the recommendations they are clearly relevant (related to the evaluation findings and conclusions), and mostly specific (though some are still quite broad, as the DBE interviewees said). Among others, the recommendations are clear and specific about the responsible role player(s) for each recommendation. In terms of feasibility, affordability and acceptability, one can point to the fact that an Improvement Plan was successfully developed as an indication that the recommendations were adequate in this regard.</p>
Rating:	3: Recommendations are useful- they are relevant, specific, feasible, affordable and acceptable to an extent

## 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:	The full report does not elaborate on this point, but it notes that "interviews were conducted on the understanding that the names of individuals and institutions would remain strictly confidential".
Rating:	3: The full report documents some procedures intended to ensure confidentiality and to secure informed consent where necessary

Standard:	There are no risks to participants or institutions in disseminating the evaluation report on a public website
Comment and Analysis:	Names have been carefully anonymised. There are no risks to participants or institutions apparent in the full report, nor were any raised by the interviewees.
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 ran a few months over schedule, with the final report approved in May 2017 instead of November 2016. Delays in fieldwork, and several rounds of edits on the draft reports, were the main drivers of the overrun. This was not a severe overrun and, since the contract was for the period February 2016 to May 2017, there was no need to extend in order to accommodate it. The evaluation did however require much more of the team's time than originally anticipated, causing the service provider to make a considerable loss on it.
Rating:	2: The evaluation was completed outside of the planned timeframes and over budget, but with approval of the commissioning organisation

### Evaluation use

Standard:	Results of the evaluation have been presented to relevant stakeholders
Comment and Analysis:	The evaluation results have been presented to numerous audiences, including an improvement plan workshop (attended by, among others, curriculum staff from all 9 provincial departments of education and teacher unions), senior management of the DBE, the Heads of Education Departments Committee (HEDCOM) and Council of Education Ministers, Umalusi, the social cluster working session, and Cabinet.
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:	There was no shared reflective process between the steering committee and service provider. The fact that the final round of revisions to the report, and approval of it, was done by round-robin email correspondence, meant that there was no close-out meeting where such a shared reflection exercise could take place. In the absence of a steering committee meeting, the following reflective activities are nevertheless noted: the peer reviewer submitted a detailed reflection in writing (saying what was learned and how future evaluations could be strengthened w.r.t. peer review); and brief reflective comments were exchanged between the service provider and DPME in their final emails. The score of 2 is awarded to acknowledge these efforts.
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:	According to the DBE interviewees, the evaluation has had considerable symbolic value; it is "signalling what's important to the programme". The process of developing the Improvement Plan, with strong participation from unions and officials responsible from curriculum from all nine provinces, plus another round of sharing and consultation with provincial representatives, was seen as an indication of the high symbolic value of evaluating the curriculum. Interviewees also indicated that the presentation in Cabinet has elicited considerable discussion and interest in the implementation of CAPS and has raised the profile of the Improvement Plan. Furthermore, it is reported that a key decision maker in a related department described the evaluation as a "game changer" that would spur his department on to collaborate more closely with DBE to facilitate effective implementation of the curriculum.
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:	Interviewed stakeholders indicated that the evaluation did not necessarily say anything new, but that it provided research and evidence to substantiate what some have been claiming. In so doing, two interviewees suggested that the evaluation has added impetus to ongoing initiatives to improve future practice (for instance, efforts to improve initial teacher training; and research into in-service training). The DPME official indicated that some stakeholders have interrogated the methodology more closely, precisely because the findings are being noted by key role players in the education system (the department and unions, among others).
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

---

- DPME (n.d.), DPME's Comments on the CAPS Report.
- Anonymous peer reviewer, 2017, CAPS Implementation evaluation: final report: Reviewer's reflection.
- Keikabile, R. (n.d.), Lessons from the theory of change workshop held 8-9 September 2018. (Example of a capacity building report.)
- DPME, 2016. Service Level Agreement between DPME and JET Education Services (SLA file reference CFP/15/24).
- JET Education Services, 2017. Report on the Implementation Evaluation of the National Curriculum Statement Grade R to 12, Focusing on the Curriculum and Assessment Policy Statements (CAPS): Policy Summary, Executive Summary and Summary Report. 25 May 2017.
- JET Education Services, 2017. Report on the Implementation Evaluation of the National Curriculum Statement Grade R to 12, Focusing on the Curriculum and Assessment Policy Statements (CAPS): Full Report. 25 May 2017.
- JET Education Services, 2016. CAPS Review: Fieldwork Guide.
- JET Education Services, 2015. Implementation Evaluation of the National Curriculum Statement Grade R to 12, Focusing on the Curriculum and Assessment Policy Statements (CAPS): Inception Report (including Appendix B: Revised Proposal). 17 December 2015.
- Department of Planning, Monitoring and Evaluation (2015). Terms of Reference: the Implementation Evaluation of the National Curriculum Statement Grade R to 12, Focusing on the Curriculum and Assessment Policy Statements (CAPS). 17 September 2015.
- Mathe, J. and Taylor, N., 2017, Re: Approved CAPS evaluation report (full and summary). (Email correspondence regarding the finalisation of the evaluation.)
- DPME Evaluation and Research Unit, 2016. Capacity development and skills transfer plan template for interns and emerging evaluators. 10 November 2016.
- Taylor, N. and Modiba, M., 2016. the Implementation Evaluation of the National Curriculum Statement Grade R to 12, Focusing on the Curriculum and Assessment Policy Statements (CAPS): Literature Review. 15 January 2017.
- JET Education Services, 2016. Implementation Evaluation of the National Curriculum Statement Grade R to 12, Focusing on the Curriculum and Assessment Policy Statements (CAPS): Fieldwork Report. 22 August 2016.

## List of Interviewees

---

- Mr Jabu Mathe, Evaluation Manager, Department of Planning, Monitoring and Evaluation. Telephonic interview. 23 November 2017.
- Ms Carol Nuga Deliwe (Chief Director) and Dr Janeli Kotzé (Assistant Director), Strategic Planning, Research and Coordination, Department of Basic Education. Telephonic interview. 29 November 2017.
- Dr Nick Taylor. Research Fellow, JET Education Services. Telephonic interview. 29 November 2017.